

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA

RICHARLYSON ALVES D'EMERY

**Aplicação de Análises Estatística e Neural para
Reconhecimento de Sinais de Odores**

Recife, Fevereiro de 2007

Richarlyson Alves D'Emery

Aplicação de Análises Estatística e Neural para Reconhecimento de Sinais de Odores

Dissertação apresentada à Pós-Graduação em Biometria do Departamento de Estatística e Informática da Universidade Federal Rural de Pernambuco como requisito parcial à obtenção do grau de Mestre em Biometria.

Orientador: Prof. Dr. Wilson Rosa de Oliveira Jr.

Recife

Universidade Federal Rural de Pernambuco

2007

Ficha catalográfica
Setor de Processos Técnicos da Biblioteca Central – UFRPE

D376a D'Emery, Richarlyson Alves
Aplicação de análises estatística e neural para reconhecimento de sinais de odores / Richarlyson Alves D'Emery. – 2007.
191 f. : il.

Orientador : Wilson Rosa de Oliveira Júnior
Dissertação (Mestrado em Biometria) - Universidade Federal Rural de Pernambuco. Departamento de Estatística e Informática.
Inclui apêndice e bibliografia

CDD 574.018 2

1. Nariz artificial
 2. Sensor de aroma
 3. Reconhecimento de padrão
 4. Diabete *Mellitus*
- I. Oliveira Júnior, Wilson Rosa de
II. Título

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA

APLICAÇÃO DE ANÁLISES ESTATÍSTICA E NEURAL PARA
RECONHECIMENTO DE SINAIS DE ODORES

RICHARLYSON ALVES D'EMERY

Dissertação julgada adequada para obtenção do título de mestre em Biometria, defendida e aprovada por unanimidade em 28/02/2007 pela banca examinadora.



Prof. Dr. Wilson Rosa de Oliveira Junior
Universidade Federal Rural de Pernambuco


Comissão Examinadora:



Profa. Dra. Tatijana Stosic
Universidade Federal Rural de Pernambuco



Profa. Dra. Laélia Pumilla Botelho Campos
Universidade Federal Rural de Pernambuco



Prof. Dr. Francisco Luiz dos Santos
Universidade Católica de Pernambuco

DEDICATÓRIA

Para minha família, amigos e professores.

AGRADECIMENTOS

Primeiramente gostaria agradecer aos meus pais que souberam me educar e colocar no rumo certo, alcançando este título de Mestre. Mais uma vez muito obrigado a minha mãe

Sueli A. D'Emery, e ao meu pai Emerson M. D'Emery.

Agradeço a Deus por ter conhecido minha esposa Paula Shelida a quem posso contar em todos os momentos, e por seu carinho e compreensão.

As minhas irmãs quem me proporcionam momentos de alegria e de perturbação. Ao meu padrasto Wilson pelos momentos de amizade e companheirismo.

Especialmente a professora Marizete S. Santos, por ter acreditado e confiado em mim desde o momento em que nos conhecemos, e hoje além de um grande mestre uma amiga em que posso contar a qualquer momento.

A Wilson Rosa, meu professor desde a graduação e hoje um orientador com quem pude contar em todos cruciais momentos.

A empresa CRIATRONICS que apoiou este trabalho e aos seus pesquisadores e colaboradores que me auxiliaram nos momentos mais críticos. Em especial a Francisco Santos um notável pesquisador e amigo. Agradeço também ao Laboratório de Análises Clínicas

Pasm pela parceria fundamental que proporcionou a realização do estudo.

Ao grande amigo Ítalo Belo, que me apoiou em todos os momentos, saboreando suco de limão, em momentos de reflexão, em discussões acadêmicas e em tantos outros.

Obrigado ao meu amigo Carlos Batista pelo inteligente investimento de descontração, muitos R\$ 2,30 de momentos de alegria, tristeza e comemorações. Em tais momentos o surgimento de idéias inovadoras e infundáveis discussões acadêmicas, pessoais e conversas fiadas. A você grande companheiro, salve a Bohemia.

Aos meus amigos de estudo Oscar Raposo e Luciano Souza, uma nova amizade que perdurará para sempre. A todos os demais estudantes do mestrado em Biometria que comigo caminharam nesses dois anos de esforço e dedicação.

A Ana Cláudia, pela parceria na pesquisa realizada desde o início até a finalizaçã..

A Universidade Federal Rural de Pernambuco, uma séria instituição de ensino que me proporciona mais esse degrau no rumo ao sucesso.

Todos vocês são muito importantes para mim.

*É melhor tentar e falhar, que preocupar-se e ver a vida passar;
é melhor tentar, ainda que em vão, que sentar-se fazendo nada
até o final. Eu prefiro na chuva caminhar, que em dias tristes
em casa me esconder. Prefiro ser feliz, embora louco, que em
conformidade viver ...".*

Martin Luther King

"A imaginação é mais importante que o conhecimento."

Albert Einstein

RESUMO

Esta dissertação investiga a utilização de um protótipo de nariz artificial no prognóstico da Diabetes Mellitus. O trabalho envolve cinco partes principais: (1) construção de uma base de dados de odores a partir de sensores de aroma; (2) avaliação da base de dados construída a partir de uma técnica estatística multivariada; (3) utilização das Redes Neurais Artificiais (RNs) *Multilayer Perceptron* (MLP) e de Função de Base Radial (RBF) no reconhecimento de odores; (4) avaliação das abordagens estatísticas e conexionistas e (5) o estudo de caso. A técnica estatística multivariada utilizada foi a Análise de Componentes Principais (PCA) como extração de característica, além da avaliação de desempenho por Teste de Hipótese. Também foi utilizado o classificador *K-Nearest-Neighbour* (KNN) para classificação dos sinais de odores. A abordagem conexionista envolve as RNs MLP e RBF treinadas a partir da base de dados de odor construída. Para as entradas das RNs foram utilizadas a Normalização como pré-processamento de dados e PCA para extração de características e redução da dimensionalidade dos dados. Foram analisados os sinais gerados por quatro sensores de um protótipo de nariz artificial exposto à urina de pacientes diabéticos e não-diabéticos, além dos sinais característicos dos sensores sem a presença de substâncias voláteis. A utilização da extração de característica por PCA melhorou consideravelmente o desempenho dos classificadores utilizados. Nos experimentos realizados, o classificador KNN classificou corretamente todo o conjunto de dados pertencente à base de dados dividida em treinamento e teste após a extração de características. As melhores RNs MLP obtiveram uma classificação de 77% para os dados Normalizados sem a utilização de PCA e de 100% após PCA. Já as melhores RNs RBF obtiveram uma classificação de 66% para os dados Normalizados e de 100% após PCA na classificação dos odores.

Palavras-chave: Narizes Artificiais, Sensores de Aroma, Reconhecimento de Padrões, Diabetes Mellitus.

ABSTRACT

This work investigates the use of an electronic nose prototype for prognosis of Diabetes Mellitus. The work involves five main parts: (1) building of odors database by aroma sensors; (2) an evaluation of the odors database through multivariate statistics techniques; (3) use of Multilayer Perceptron (MLP) and Radial Basis Function (RBF) Artificial Neural Networks (ANN) in the odors recognition; (4) an evaluation of statistics and connectionist approaches; and (5) a case study. The multivariate statistical method used was the Principal Components Analysis (PCA) as characteristic extraction, beyond of an evaluation of performance for Hypothesis Test. Also had been used the K-Nearest-Neighbour (KNN) classifier for classification of odors signs. The connectionist approach involves the MLP e RBF ANN trained with de odors database. For the ANN's inputs were used the Normalization as pre-processing of data and PCA for characteristic extraction and reduction of data dimensionality. The signals generated by the four sensors of an electronic nose prototype exposed to the urine from the diabetics and non-diabetics patients and the characteristics signals of the sensors without the presence of volatiles substances were analyzed. The use of characteristic extraction by PCA improved the performance of the classifiers used. In the experiments, the KNN classified correctly all the data set from the database divided in training and test after PCA. The better MLP ANN obtained a classification of 77% to the normalized data without PCA use and of 100% after PCA while the RBF ANN obtained a classification of 66% to the normalized data and the 100% after PCA in the odors classification.

Keywords : Electronic Noses, Aroma Sensors, Pattern Recognition, Diabetes Mellitus.

SUMÁRIO

| | | |
|----------|---|-----------|
| 1 | INTRODUÇÃO..... | 16 |
| 1.1 | Motivação..... | 16 |
| 1.2 | Objetivos da dissertação..... | 19 |
| 1.3 | Descrição do trabalho..... | 20 |
| 1.4 | Organização da dissertação..... | 22 |
| 2 | A TECNOLOGIA IMITANDO O CORPO HUMANO | 24 |
| 2.1 | Introdução..... | 24 |
| 2.2 | O Nariz Biológico | 27 |
| 2.2.1 | Anatomia e fisiologia do sistema olfativo..... | 28 |
| 2.2.1.1 | As vias olfativas | 30 |
| 2.2.2 | Teorias da olfação | 31 |
| 2.2.2.1 | Teoria dos odores primários..... | 32 |
| 2.2.2.2 | Teoria estereoquímica | 33 |
| 2.2.2.3 | Teoria vibracional | 35 |
| 2.3 | Narizes Artificiais | 36 |
| 2.3.1 | Sensores de Aroma | 39 |
| 2.4 | Nariz biológico versus nariz artificial..... | 41 |
| 2.5 | Resumo do capítulo | 44 |
| 3 | DIABETES | 45 |
| 3.1 | Introdução..... | 45 |
| 3.2 | Os diversos diabetes | 47 |
| 3.3 | Sintomas e diagnósticos..... | 50 |
| 3.4 | Tratamento..... | 53 |
| 3.5 | Resumo do capítulo | 55 |
| 4 | SISTEMA DE AQUISIÇÃO E PRÉ-PROCESSAMENTO | 56 |
| 4.1 | Introdução..... | 56 |
| 4.2 | O protótipo do sistema de aquisição..... | 57 |
| 4.2.1 | Descrição geral..... | 57 |
| 4.2.2 | Câmara de aquisição | 59 |
| 4.2.3 | Sensores utilizados..... | 60 |
| 4.2.4 | Leitor de características..... | 61 |
| 4.2.5 | A aquisição dos dados..... | 62 |
| 4.3 | Preparação dos dados..... | 66 |

| | | |
|----------|---|------------|
| 4.3.1 | Métodos locais | 68 |
| 4.3.2 | Métodos globais | 70 |
| 4.4 | Resumo do capítulo | 72 |
| 5 | ANÁLISE DE PADRÕES PARA NARIZES ARTIFICIAIS | 73 |
| 5.1 | Introdução..... | 73 |
| 5.2 | Análise de Componentes Principais | 74 |
| 5.2.1 | Componentes principais exatas extraídas da matriz de covariâncias | 76 |
| 5.2.2 | Estimação das componentes principais: matriz de covariâncias | 81 |
| 5.2.3 | Análise de Componentes Principais via matriz de correlação | 83 |
| 5.2.4 | Determinação do número k de componentes principais | 86 |
| 5.2.4.1 | Puramente matemáticos | 86 |
| 5.2.4.2 | Sob o ponto de vista prático..... | 89 |
| 5.3 | Métodos não-paramétricos | 89 |
| 5.3.1 | Estimador de densidade por Kernel..... | 91 |
| 5.3.2 | Nearest-neighbour..... | 92 |
| 5.4 | Redes neurais artificiais | 95 |
| 5.4.1 | Redes neurais multilayer perceptron..... | 101 |
| 5.4.2 | Redes de função de base radial | 104 |
| 5.4.3 | Comparação entre redes MLP e RBF..... | 108 |
| 5.5 | Resumo do capítulo | 109 |
| 6 | RESULTADOS DA ANÁLISE DOS SINAIS DE ODORES | 111 |
| 6.1 | Introdução..... | 111 |
| 6.2 | Análise de Componentes Principais | 112 |
| 6.2.1 | Análise dos sensores utilizados..... | 113 |
| 6.2.1.1 | PCA via matriz de covariâncias amostrais..... | 113 |
| 6.2.1.2 | PCA via matriz de correlação | 121 |
| 6.2.2 | Análise dos sensores expostos à urina de pacientes diabéticos e não-diabéticos..... | 125 |
| 6.2.2.1 | PCA via matriz de covariâncias amostrais..... | 126 |
| 6.2.2.2 | PCA via matriz de correlação | 129 |
| 6.3 | Classificação por <i>K-nearest-neighbour</i> | 132 |
| 6.4 | Experimento Computacional: Classificação de Padrões por RNs | 135 |
| 6.4.1 | Divisão da base de dados em conjuntos de treinamento, validação e teste | 136 |
| 6.4.2 | Pré-processamento e extração de características dos dados | 136 |
| 6.4.3 | Topologia utilizada..... | 137 |
| 6.4.4 | Metodologia e treinamento..... | 138 |
| 6.4.5 | Aspectos observados..... | 139 |
| 6.4.6 | Métodos de comparação entre as abordagens..... | 140 |
| 6.4.7 | Resultados obtidos para as RNs MLP..... | 142 |
| 6.4.8 | Resultados obtidos para as RNs RBF | 153 |
| 6.4.9 | Comparação entre as abordagens..... | 165 |
| 6.5 | Resumo do capítulo | 166 |

| | |
|--|------------|
| 7 CONCLUSÃO | 169 |
| 7.1 Considerações finais | 169 |
| 7.2 Contribuições deste trabalho..... | 171 |
| 7.3 Proposta para trabalhos futuros..... | 172 |
| APÊNDICE A | 173 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 179 |

LISTA DE FIGURAS

| | |
|---|-----|
| Figura 2.1 – Ilustração da cavidade nasal humana [EMCH, 2000]..... | 28 |
| Figura 2.2 – Localização das células no epitélio olfativo [Leffingwell & Associates, 2001]..... | 29 |
| Figura 2.3 – Ilustração do bulbo olfativo e células mitrais [Leffingwell & Associates, 2001]..... | 30 |
| Figura 2.4 – Área olfativa medial e lateral [Guyton, 1981]..... | 31 |
| Figura 2.5 – Quadro comparativo da olfação humana com o nariz artificial [Santos, 2000]..... | 41 |
| Figura 4.1 – Diagrama do protótipo de aquisição de dados [Santos, 2000]..... | 57 |
| Figura 4.2 – Fotografia do protótipo de aquisição de dados medindo os sinais de odores provenientes da reação dos sensores de aroma com o odor da urina de um paciente..... | 58 |
| Figura 4.3 – (a) Câmara de aquisição de sinais de odores provenientes de uma amostra e (b) de várias amostras..... | 59 |
| Figura 4.4 – Nomenclatura dos sensores utilizados, onde S é referente a sensor, 01 a classe tipo I, BK é referente à cor do substrato e reagente do sensor, nesse caso black (BB é black e blue, BR é black e red, e BG é black e green), 3 é número de camadas do sensor, e D3 é o tamanho do sensor nesse caso 3 cm x 3 cm. | 61 |
| Figura 4.5 – Curva temporal formada pelos valores dos sinais resistivos dos sensores sendo (a) na coleta de linha de base e (b) exposição de uma amostra de urina de um paciente diabético aos sensores de aroma..... | 65 |
| Figura 4.6 – Procedimentos da normalização para a resposta do sinal de um sensor: (a, d) dados originais, (b) auto-escalamento do sinal do sensor, (c) normalização do sinal do sensor e (e) normalização do vetor..... | 69 |
| Figura 5.1 – Interpretação geométrica da Análise de Componentes Principais , $p = 2$ | 81 |
| Figura 5.2 - O scree-plot mostra as componentes principais seguindo a tendência de estabilidade, aproximação de zero, dos autovalores $\hat{\lambda}_i$ para um vetor aleatório X..... | 88 |
| Figura 5.3 – (a) No classificador KNN, um novo ponto, representado pelo quadrado preto, é classificado pela classe formada pela maioria dos pontos mais próximos pertencentes ao conjunto de treinamento, neste caso $K = 3$. (b) Na classificação pela abordagem nearest-neighbour ($K=1$) o limite de decisão resultante é formado de hiperplanos que formam bissetores perpendiculares pelos pares de pontos das diferentes classes. | 94 |
| Figura 5.4 – Modelo não-linear de um neurônio. Nesse modelo o bias (b_k) é incluindo, sua presença aumenta ou diminui a entrada da função de ativação, dependendo se ele é positivo ou negativo, respectivamente..... | 96 |
| Figura 5.5 – Grafo arquitetural de um perceptron de múltiplas camadas com duas camadas ocultas..... | 102 |
| Figura 5.6 – Função Logística – Efeito da variação do parâmetro a | 103 |
| Figura 5.7 – Rede de função de base radial..... | 106 |
| Figura 6.1 – Variância explicada em cada uma das componentes principais..... | 116 |
| Figura 6.2 – Projeção dos sensores em três componentes principais..... | 117 |
| Figura 6.3 – Projeção dos sensores em três componentes principais..... | 119 |
| Figura 6.4 – Projeção dos sensores em duas componentes principais..... | 120 |
| Figura 6.5 – Projeção dos sensores amplificados em três componentes principais..... | 121 |
| Figura 6.6 – variância explicada em cada uma das componentes principais..... | 123 |
| Figura 6.7 – Projeção dos sensores utilizados em três componentes principais..... | 124 |
| Figura 6.8 – Distribuição das variâncias explicadas utilizando a estimação de componentes principais..... | 125 |
| Figura 6.9 – Variância explicada em cada uma das componentes principais para os sensores..... | 128 |
| Figura 6.10 – Projeção dos sensores expostos a urina dos pacientes em três componentes..... | 128 |
| Figura 6.11 – Variância explicada em cada uma das componentes principais para os sensores..... | 131 |
| Figura 6.12 – Projeção dos sensores expostos a urina dos pacientes em três componentes..... | 131 |
| Figura 6.13 – Distribuição das variâncias explicadas utilizando a estimação de componentes principais..... | 132 |
| Figura 6.14 – Função de grau para $c = 0$ | 139 |

| | |
|---|-----|
| Figura 6.15 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas com a função de ativação logística usando Normalização como pré-processamento: (a) MLP com $\mathbf{h} = 0,01$, (b) MLP com $\mathbf{h} = 0,05$ e (c) MLP com $\mathbf{h} = 0,10$ | 144 |
| Figura 6.16 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas com a função de ativação tangente hiperbólica usando Normalização como pré-processamento: (a) MLP com $\mathbf{h} = 0,01$, (b) MLP com $\mathbf{h} = 0,05$ e (c) MLP com $\mathbf{h} = 0,10$ | 145 |
| Figura 6.17 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas com a função de ativação logística após PCA: (a) MLP com $\mathbf{h} = 0,01$, (b) MLP com $\mathbf{h} = 0,05$ e (c) MLP com $\mathbf{h} = 0,10$ | 148 |
| Figura 6.18 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas a função de ativação tangente hiperbólica após PCA: (a) MLP com $\mathbf{h} = 0,01$, (b) MLP com $\mathbf{h} = 0,05$ e (c) MLP com $\mathbf{h} = 0,10$ | 149 |
| Figura 6.19 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\mathbf{a} = 5\%$ usando Normalização como pré-processamento: (a) RBF com $\mathbf{h} = 0,01$, (b) RBF com $\mathbf{h} = 0,05$ e (c) RBF com $\mathbf{h} = 0,10$ | 154 |
| Figura 6.20 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\mathbf{a} = 10\%$ usando Normalização como pré-processamento: (a) RBF com $\mathbf{h} = 0,01$, (b) RBF com $\mathbf{h} = 0,05$ e (c) RBF com $\mathbf{h} = 0,10$ | 155 |
| Figura 6.21 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\mathbf{a} = 20\%$ usando Normalização como pré-processamento: (a) RBF com $\mathbf{h} = 0,01$, (b) RBF com $\mathbf{h} = 0,05$ e (c) RBF com $\mathbf{h} = 0,10$ | 156 |
| Figura 6.22 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\mathbf{a} = 5\%$ após PCA: (a) RBF com $\mathbf{h} = 0,01$, (b) RBF com $\mathbf{h} = 0,05$ e (c) RBF com $\mathbf{h} = 0,10$ | 159 |
| Figura 6.23 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas a função de ativação gaussiana com $\mathbf{a} = 10\%$ após PCA: (a) RBF com $\mathbf{h} = 0,01$, (b) RBF com $\mathbf{h} = 0,05$ e (c) RBF com $\mathbf{h} = 0,10$ | 160 |
| Figura 6.24 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas a função de ativação gaussiana com $\mathbf{a} = 20\%$ após PCA: (a) RBF com $\mathbf{h} = 0,01$, (b) RBF com $\mathbf{h} = 0,05$ e (c) RBF com $\mathbf{h} = 0,10$ | 161 |
| Figura A.1 – Diagrama de dispersão com identificação de valores aberrantes bidimensionais..... | 175 |

LISTA DE TABELAS

| | |
|--|------------|
| <i>Tabela 3.1 – Nível de glicose no sangue para diabéticos.....</i> | <i>52</i> |
| <i>Tabela 4.1 – Descrição dos sensores utilizados no protótipo de nariz artificial.....</i> | <i>63</i> |
| <i>Tabela 4.2 – Exemplo de aquisição dos dados da coleta de linha de base dos sensores de aroma.....</i> | <i>64</i> |
| <i>Tabela 4.3 – Exemplo de aquisição dos dados da exposição de uma amostra de urina de um paciente diabético aos sensores de aroma.....</i> | <i>64</i> |
| <i>Tabela 6.1 – Estatísticas descritivas para a aquisição dos dados coletados de linha de base dos sensores de aroma.....</i> | <i>113</i> |
| <i>Tabela 6.2 – Dados relativos às médias dos 308 sensores de aroma utilizados para compor as componentes principais a serem estimadas.....</i> | <i>114</i> |
| <i>Tabela 6.3 – Correlação entre as três componentes principais escolhidas e os sensores utilizados.</i> | <i>117</i> |
| <i>Tabela 6.4 – Estatísticas descritivas para a aquisição dos dados da exposição das amostras de urinas dos pacientes diabéticos aos sensores de aroma.....</i> | <i>126</i> |
| <i>Tabela 6.5 – Estatísticas descritivas para a aquisição dos dados da exposição das amostras de urinas dos pacientes não-diabéticos aos sensores de aroma.....</i> | <i>126</i> |
| <i>Tabela 6.6 – Correlação entre as três componentes principais escolhidas e a resposta dos sensores.....</i> | <i>129</i> |
| <i>Tabela 6.7 – Matriz de confusão para as 77 amostras de urina num conjunto de 50% de treino e 50% de teste analisadas com o classificador KNN.....</i> | <i>133</i> |
| <i>Tabela 6.8 – Matriz de confusão para as 77 amostras de urina num conjunto de 70% de treino e 30% de teste analisadas com o classificador KNN.....</i> | <i>134</i> |
| <i>Tabela 6.9 – Matriz de confusão para as 77 amostras de urina após PCA num conjunto de 50% de treino e 50% de teste analisadas com o classificador KNN.....</i> | <i>134</i> |
| <i>Tabela 6.10 – Matriz de confusão para as 77 amostras de urina num conjunto de 70% de treino e 30% de teste analisadas com o classificador KNN.....</i> | <i>134</i> |
| <i>Tabela 6.11 – Desempenho geral da classificação pelo método KNN nas 77 amostras de urina num conjunto de 50% de treino e 50% de teste analisadas a partir da base de dados original e com PCA.....</i> | <i>135</i> |
| <i>Tabela 6.12 – Desempenho geral da classificação pelo método KNN nas 77 amostras de urina num conjunto de 70% de treino e 30% de teste analisadas a partir da base de dados original e com PCA.....</i> | <i>135</i> |
| <i>Tabela 6.13 – Ranking de desempenho das RNs MLP com a funções de ativação logística e tangente hiperbólica através do MEQMs para os dados normalizados.....</i> | <i>143</i> |
| <i>Tabela 6.14 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação logística usando Normalização como pré-processamento de dados.....</i> | <i>146</i> |
| <i>Tabela 6.15 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação tangente hiperbólica usando Normalização como pré-processamento de dados.</i> | <i>146</i> |
| <i>Tabela 6.16 – Ranking de desempenho das RNs MLP com as funções de ativação logística e tangente hiperbólica através do MEQMs para dados após PCA.....</i> | <i>147</i> |
| <i>Tabela 6.17 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação logística após PCA.....</i> | <i>150</i> |
| <i>Tabela 6.18 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação tangente hiperbólica após PCA.....</i> | <i>150</i> |
| <i>Tabela 6.19 – Classificação das respostas das RNs MLP com as funções de ativação logística (LOG) e tangente hiperbólica (TANHIP) para os dados antes (normalização dos dados) e após a extração de características.....</i> | <i>151</i> |
| <i>Tabela 6.20 – Estatística do teste de Wilcoxon para o desempenho das RNs MLP com normalização no pré-processamento e PCA como extração de características.....</i> | <i>152</i> |
| <i>Tabela 6.21 – Ranking de desempenho das RNs RBF com a função de ativação gaussiana (GAUSS) com nível de significância (α) de 5%, 10% e 20% através do MEQMs para dados normalizados.....</i> | <i>153</i> |
| <i>Tabela 6.22 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 5% usando Normalização como pré-processamento.....</i> | <i>157</i> |

| | |
|---|-----|
| <i>Tabela 6.23 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (a) de 10% usando Normalização como pré-processamento</i> | 157 |
| <i>Tabela 6.24 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (a) de 20% usando Normalização como pré-processamento</i> | 157 |
| <i>Tabela 6.25 – Ranking de desempenho das RNs RBF com a função de ativação gaussiana (GAUSS) com nível de significância (a) de 5%, 10% e 20% através do MEQMs para dados após PCA.</i> | 158 |
| <i>Tabela 6.26 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (a) de 5% após PCA.</i> | 162 |
| <i>Tabela 6.27 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (a) de 10% após PCA.</i> | 162 |
| <i>Tabela 6.28 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (a) de 20% após PCA.</i> | 162 |
| <i>Tabela 6.29 – Classificação das respostas das RNs RBF com função de ativação gaussiana para os dados antes (normalização dos dados) e após a extração de características.</i> | 164 |
| <i>Tabela 6.30 – Estatística do teste de Wilcoxon para o desempenho das RNs RBF com normalização no pré-processamento e PCA como extração de características.</i> | 165 |
| <i>Tabela 6.31 – Estatística do teste de Wilcoxon para o desempenho das RNs simuladas com PCA como extração de características.</i> | 166 |

Capítulo 1

Introdução

Neste capítulo é apresentada a motivação desta dissertação, analisando aspectos inerentes ao processo de reconhecimento de odores em narizes artificiais. Na seção 1.1 expõe-se brevemente o problema de reconhecimento de odores, o qual será aprofundado na dissertação. Na seção 1.2 demarcam-se os objetivos deste trabalho. A seção 1.3 contém uma sucinta descrição do trabalho realizado na dissertação. E na seção 1.4 é fornecida uma visão dos capítulos da dissertação.

1.1 Motivação

A busca por dispositivos artificiais que apresentam algum tipo de comportamento inteligente, similar ao exibido por muitos organismos biológicos, sempre fascinou os cientistas. Nas pesquisas por sistemas que possam exibir características biológicas, uma das mais intensas batalhas que o homem vem travando no progresso científico e tecnológico é tentar produzir aparatos eletrônicos aptos a ajudar, simular, aguçar ou estender os sentidos humanos além de seus limites naturais.

A partir de pesquisas realizadas com órgãos sensoriais humanos foi possível construir aparelhos que, embora limitados, reproduzem o comportamento de alguns sentidos como paladar (línguas eletrônicas), olfato (narizes eletrônicos), audição (dispositivos de surdez), tato (sensores de pressão), visão (filmadora e câmeras digitais), entre outros.

No entanto, dentre os órgãos dos sentidos, o paladar e o olfato ficaram durante muito tempo esquecidos, perante o avanço tecnológico obtido em pesquisas e desenvolvimentos científicos e comerciais dos demais órgãos sensoriais. O sistema olfativo, em especial, se deve pelos muitos aspectos inerentes ao processo de reconhecimento de odores em sistemas biológicos ainda não estarem bem esclarecido pela ciência, havendo muitas teorias na tentativa de explicá-los [dos Santos, 1999]. Este fato motivou o estudo dos processos envolvidos na olfação humana; e mesmo sem a completa compreensão da fisiologia olfativa, impulsionaram o estudo e o desenvolvimento de vários modelos de sistemas olfativos artificiais baseados em evidências biológicas [Li e Hopfield, 1989; Santos, 2000; Li e Hertz, 2000].

Atualmente denominam-se *Narizes Artificiais* ou *Narizes Eletrônicos*¹ os modelos e dispositivos olfativos artificiais inspirados em sistemas olfativos biológicos. Os narizes artificiais são sistemas capazes de detectar e classificar odores, vapores e gases automaticamente; dentre eles destacam-se os sensores de aroma responsáveis por efetuar a medição dos compostos odorantes no ambiente [D'Emery et al., 2005b; Santos, 2000; dos Santos, 1999; Gardner et al., 1994] e o sistema de reconhecimento de padrões, responsável pelo reconhecimento e classificação dos sinais de odores provindos dos narizes artificiais [D'Emery et al., 2005c; Santos, 2000; Gardner e Bartlett, 1992; Gardner, 1991].

Os narizes artificiais têm sido utilizados com grande sucesso em diversas áreas [Benedetti et al., 2004; Pavlou et al., 2002; Llecha, 2001], e tem despertado especial interesse na saúde, onde podem ser utilizados como dispositivos de prognósticos de determinadas doenças, dentre elas o Diabetes [Mohamed et al., 2002; Ping et al., 1997]. Existem muitas outras aplicações como nas indústrias alimentícias e petrolíferas. De forma geral narizes artificiais podem ser colocados em posições estratégicas, como em lugares de difícil acesso no combate

¹ Estes dispositivos também são referenciados comercialmente com *e-nose* do inglês Electronic Nose.

a doenças, detectando e identificando vazamentos químicos em estágios iniciais, antes que os problemas alcancem níveis perceptíveis e perigosos para os seres humanos e para natureza.

Devido à importância dos resultados produzidos por narizes artificiais, o sistema de reconhecimento de padrões tem recebido crescente atenção em pesquisas e desenvolvimentos. Igualmente importante para as mais diversas áreas, o reconhecimento de padrões é um campo em crescente expansão no domínio da inteligência computacional. Dentre as abordagens investigadas objetivando o reconhecimento de padrões, resultados promissores vêm sendo encontrados tanto nas abordagens estatísticas como computacionais [Zanchettin e Luder-mir, 2005, 2003; Benedetti et al., 2004; Llecha, 2001; Braga et al., 2000; Santos, 2000]. Os modelos estatísticos geralmente advêm de análises multivariada, enquanto que os computacionais são derivados da subárea inteligência artificial como as redes neurais.

Na pesquisa por sistemas que possam compartilhar ou combinar características, diversos pesquisadores têm constatado que técnicas de aprendizagem de dados são, em muitos aspectos, complementares. Assim, muitos esforços têm sido realizados no sentido de investigar possíveis formas de cooperação entre métodos inteligentes ou mesmo métodos de inteligência computacional e métodos estatísticos [Benedetti et al., 2004; Zanchettin et al., 2003; Pavlou et al., 2002; Llecha, 2001].

O objetivo deste trabalho foi o de investigar técnicas estatísticas e computacionais que possam contribuir para o aprimoramento do sistema de reconhecimento de padrões do protótipo de um nariz artificial [D'Emery et al., 2005b], a ser utilizado no prognóstico da doença Diabetes Mellitus a partir da urina de pacientes. São abordadas técnicas estatísticas multivariada já consagradas na literatura como a *Análise de Componentes Principais* (PCA, do inglês *Principal Component Analysis*) e o classificador *K-Nearest-Neighbour* (KNN) [Cover e Hart, 1967], além das Redes Neurais Artificiais (RNs), como as RNs *Multilayer Perceptron* (MLP) [Rumelhart et al., 1986] e RNs de *Função de Base Radial*, capazes de extrair caracte-

rísticas, reduzir a dimensionalidade dos dados, seleção dos atributos mais importantes e efetuar a classificação de padrões.

1.2 Objetivos da dissertação

Neste trabalho se estuda a tecnologia dos narizes artificiais, desde a aquisição dos dados até o sistema de reconhecimento de padrões, bem como as diversas abordagens já apresentadas na literatura para este fim, apresentando os procedimentos, vantagens e desvantagens.

Dentre os objetivos específicos deste trabalho, temos a obtenção dos sinais dos sensores de aroma, fundamentais na construção da base de dados a partir de um sistema de aquisição, descrito no Capítulo 4. Há também análises estatísticas e neurais dos dados, descrito no Capítulo 6. A primeira investigação é realizada a partir de uma técnica estatística de análise multivariada, que tem por objetivo a interpretação, visualização e redução da dimensionalidade dos dados, além de uma técnica para classificação dos mesmos. A segunda investigação utiliza técnicas de redes neurais artificiais visando o treinamento e classificação dos padrões do conjunto de dados.

Nesta dissertação se propõe a avaliação de um Sistema de Reconhecimento de Padrões para o prognóstico da doença Diabetes Mellitus, que futuramente poderá servir de base para comparação com outros modelos.

Vale ressaltar que o principal objetivo dessa pesquisa, é a busca por sistemas que sejam eficientes, robustos e fáceis de operar, e capazes de fornecer prognósticos de qualidade para problemas complexos.

1.3 Descrição do trabalho

Este trabalho consistiu das seguintes etapas: (1) construção da base de dados de odores a partir de sensores de aroma; (2) avaliação da base de dados construída a partir de uma técnica estatística multivariada; (3) utilização das Redes Neurais Artificiais (RNs) Multi-layer Perceptron (MLP) e de Função de Base Radial (RBF) no reconhecimento de odores; (4) avaliação das abordagens estatísticas e conexionistas e (5) o estudo de caso, com a aplicação dos modelos investigados no prognóstico da doença Diabetes Mellitus.

A fase de desenvolvimento de sensores de aroma compreendeu a sua utilização em uma aplicação específica, visando estudar o seu comportamento. A construção da base de dados está intimamente relacionada aos sensores, uma vez que é formada pela respostas desses sensores perante a exposição aos vapores das substâncias sob teste.

A avaliação da base de dados compreendeu uma análise a partir de uma técnica multivariada, a qual tem por objetivo a interpretação, visualização e redução da dimensionalidade dos dados. Esta técnica, conhecida como Análise de Componentes Principais [Jackson, 1991], foi utilizada como método de visualização e interpretação dos sinais de odores, com o objetivo de verificar características e particularidades dos padrões que estão sendo investigados, além de otimizar o desempenho das redes neurais. Também se utilizou um classificador estatístico, o K-Nearest-Neighbour visando à classificação dos padrões dos dados.

A seleção dos modelos neurais, para os sistema de reconhecimento de padrões do protótipo de nariz artificial, englobou o estudo das técnicas utilizadas na literatura para classificação de odores. Verificou-se que a maioria das soluções conexionistas faz uso das RNs MLP, empregando o algoritmo *Backpropagation* [Rumelhart et al., 1986] para o aprendizado da rede. Também se utilizou as RNs RBF utilizando o mesmo algoritmo na aprendizagem.

Observou-se também, na literatura, a utilização de diversas funções de ativação nas RNs. Neste trabalho foram utilizadas as funções Sigmóide Logística e Tangente Hiperbólica para as RNs MLP, enquanto que para as RNs RBF a função utilizada foi a Gaussiana.

Com o intuito de melhorar o desempenho de classificação das técnicas acima referenciadas, no reconhecimento dos sinais de odores, além de técnicas de pré-processamento como a Normalização, também foram realizados testes com técnicas de redução de dimensionalidade, sendo esta a Análise de Componentes Principais.

Todos os métodos investigados foram implementados e testados com o programa de simulação *MATLAB*², utilizando os pacotes estatísticos disponíveis no programa: “*Statistics ToolBox*”. Para as simulações das redes neurais foi utilizado o software *R*³, utilizando o pacote de redes neurais “*neural: Neural Network*”. Foram utilizadas ainda as linguagens de programação *C* e *QBASIC* na fase de aquisição dos dados e o *Microsoft Excel*, por vezes utilizado na manipulação dos padrões. Vale ressaltar que as demais abordagens pesquisadas foram implementadas a partir de especificações obtidas na revisão bibliográfica.

No estudo de caso, foram tratados os padrões de odores, obtidos a partir das exposições dos sensores de um protótipo de nariz artificial, cedidos pela empresa CRIATRONICS⁴, exposto a urina de pacientes diabéticos e não-diabéticos, fornecida pelo Laboratório de Análises Clínica Pasm.

² MATLAB, marca registrada da empresa The Math Works, Inc.

³ R, marca registrada pelo grupo The R Development Core Team

⁴ CRIATRONICS - Centro de Robótica, Inteligência e Automação LTDA. Disponível em <<http://www.criatronics.com.br>>. Acesso em 2 jan. 2007.

1.4 Organização da dissertação

No Capítulo 1 é apresentada a motivação que inspirou a realização desse trabalho, assim como os objetivos gerais e específicos da dissertação e um breve descrição sobre os capítulos apresentados.

No Capítulo 2 é apresentada uma introdução ao sistema olfativo humano e aos narizes artificiais. Para o sistema olfativo humano são descritas suas partes constituintes e algumas teorias sobre como é realizado o reconhecimento de odores. Para os narizes artificiais, são descritos seus componentes, como os sensores de aroma, além de suas diversas aplicações e desenvolvimentos. Também é realizada uma comparação entre os modelos.

No Capítulo 3 é realizada uma revisão da literatura científica sobre a doença Diabetes Mellitus, destacando suas classificações, principais sintomas, formas de diagnóstico além de como é realizado o tratamento na prevenção e controle da doença. O reconhecimento da urina de pessoas com e sem Diabetes Mellitus é a principal aplicação no objeto de estudo desta dissertação.

No Capítulo 4 são descritas as partes pertencentes ao protótipo de um nariz artificial, ou seja, o sistema de aquisição de dados, destacando os sensores de aroma. Neste capítulo, também se descreve a metodologia utilizada nos experimentos de aquisição de dados, além de uma revisão da literatura científica sobre alguns métodos de pré-processamento de dados.

A análise dos padrões de odores utilizada nesta dissertação é descrita no Capítulo 5. É descrita uma técnica estatística multivariada, a Análise de Componentes Principais, fornecendo um conhecimento prévio sobre o problema abordado, além de um classificador estatístico, o K-Nearest-Neighbour. Também são descritas as técnicas conexionistas: a Rede Neural *Multilayer Perceptron* e a de Função de Base Radial. O funcionamento de cada uma

destas redes é detalhadamente descrito, de forma a dar uma visão geral sobre cada uma das abordagens.

No Capítulo 6 são apresentados os experimentos realizados, a partir da base de dados construída, em cada uma das técnicas investigadas e descritas nos capítulos anteriores. É descrita a metodologia utilizada para cada abordagem e são destacados os resultados de cada abordagem, além de uma comparação entre algumas delas.

As considerações finais sobre o trabalho apresentado nesta dissertação são tecidas no Capítulo 7. Além das considerações finais, este capítulo apresenta uma descrição das contribuições desta dissertação e algumas propostas para trabalhos futuros.

Ao final desta dissertação é fornecido um apêndice que serve de complemento evidenciando algumas técnicas fundamentais utilizadas em sistemas de padrões que envolvem os narizes artificiais.

Capítulo 2

A tecnologia imitando o corpo humano

Neste capítulo é apresentada uma breve explanação sobre a estrutura do sistema olfativo humano e sobre os narizes artificiais, dos quais são analisadas suas partes constituintes, bem como as diversas abordagens existentes sobre seu funcionamento. Na seção 2.1 é apresentado um resumo sobre as diversas pesquisas que visam entender o organismo humano e a tentativa de imitá-lo. Na seção 2.2 é descrito o sistema olfativo humano. Na seção 2.3 são descritas as tecnologias do nariz artificial. Na seção 2.4 é feita uma comparação entre o nariz biológico e o artificial. A seção 2.5 é um resumo do capítulo.

2.1 Introdução

O que será mais importante: sentir ou perceber? Mais importante do que sentir é perceber aquilo que sentimos. De nada nos serviria captar todos os estímulos do meio externo se não fôssemos capazes de nos consciencializar pelo menos um. O ato de perceber consiste em tornar consciente a informação que nos chega ao cérebro pelos neurônios sensitivos.

Quando um homem vislumbra a face de uma mulher, cheira ou saboreia um delicioso aperitivo, ou até mesmo ao ouvir a voz de um grande amor, o reconhecimento é imediato. Numa fração de milésimos de segundos, após a estimulação da pele, dos olhos, nariz, língua ou ouvidos, cada pessoa reconhece se um estímulo é familiar, desagradável ou até perigoso. Neste processo, vários sentidos estão envolvidos.

Apesar da grande quantidade de pesquisas na área, o sentido do olfato ainda é o menos conhecido, existindo diversas teorias explicativas para o processo de reconhecimento de odores. Pesquisadores em diversas áreas vêm desvendando o princípio do funcionamento do olfato e propondo modelos artificiais, com diversas finalidades como, por exemplo, o reconhecimento de determinadas substâncias, como vinho, frutas, doenças contagiosas, na indústria alimentícia e outros [D'Emery et al., 2005c; Batista e D'Emery, 2005; Benedetti et al., 2004; UB, 2002; Llecha, 2001; Santos, 2000; dos Santos, 1999].

Além do estudo da olfação um outro sentido também estudado é o da visão, onde propriedades como o modelo de cores, distância dos corpos e processo de reconhecimento já são bem difundidos. Na audição, equipamentos auxiliam pessoas que não mais possuem boa receptividade de som. Para o paladar, sistemas como línguas eletrônicas visam reconhecer amostras através do contato de sensores com substâncias [Lavigne et al., 1998]. Neste trabalho o sentido a ser estudado é a olfação.

Embora os mistérios que cercam o cérebro e as dificuldades em se estudar o sistema olfativo ainda sejam problemas contemporâneos, muitos pesquisadores têm dedicado esforços nesse século e no final do século passado para elucidá-los. Estes estudos deram inclusive a pesquisadores, o prêmio Nobel de medicina [Levy, 2004]. A detecção do odor pelo nariz humano é feita com base na reação físico-química entre as células olfativas e as moléculas de odor, sendo o processamento realizado pelas diversas células dispersas numa pequena região do sistema olfativo. Portanto, existe um sistema altamente acoplado nestas várias etapas e até hoje se procura encontrar respostas mais conclusivas para todas as fases deste processo de detecção, identificação e reconhecimento dos odores feitos pelo cérebro.

A grande quantidade de pesquisas nesta área de olfação levou pesquisadores à construção de narizes artificiais que poderá certamente ajudar o homem em muitas aplicações úteis, a exemplo do controle ambiental através do monitoramento da qualidade do ar, detecção

de vazamento de óleo, tráfico de drogas, entre outras. As primeiras teses, na América Latina, de desenvolvimento completo de um nariz artificial ou eletrônico capaz de detectar e classificar odores, vapores e gases utilizando a tecnologia de polímeros condutores e redes neurais podem ser encontradas em [Santos, 2000; Souza, 2000; dos Santos, 1999]. As aplicações nestes trabalhos foram com a identificação e o reconhecimento de alguns gases como o butano, propano, etano, metano, vinhos branco, tinto e rose e safras de vinho merlot, além de uísque e detecção de mistura em gasolina.

Várias outras aplicações têm sido sugeridas nesta linha de detecção e classificação de odores, a exemplo de um estudo para aplicações na indústria petrolífera, onde estes dispositivos podem ser usados como dispositivos de segurança na detecção e reconhecimento de gases tóxicos, antes que sua concentração alcance níveis perceptíveis e perigosos para os seres humanos e para a natureza [Zanchettin et al., 2003]. Outra área de aplicação onde o odor pode ser de grande utilidade é na detecção de enfermidades, embora, no Brasil, não tenha sido historicamente muito explorado [Batista e D'Emery, 2006; Carneiro et al., 2006].

O ser humano é acometido de diversas enfermidades e nem sempre essas enfermidades são detectadas em um período ideal para se fazer estes diagnósticos, por razões diversas. O exame clínico pode ser realizado pelo uso do odor para detectar determinados tipos de enfermidades. Um nariz treinado pode identificar, por exemplo, se uma pessoa tem câncer de bexiga com base no odor da urina do paciente. Acredita-se que substâncias tumorais voláteis presentes na urina de pacientes com câncer poderiam ser detectadas pelo olfato humano, uma vez que o cão pode identificar [Beers, 1999].

Vejamos como ocorre o funcionamento da olfação, através de um breve estudo sobre o nariz biológico, e conhecer um pouco sobre as propostas de narizes artificiais.

2.2 O Nariz Biológico

O sistema olfativo, juntamente com os outros sentidos tem função relevante para a sobrevivência dos seres vivos, pois permite que estes se relacionem com o mundo ao seu redor, podendo perceber e identificar uma grande quantidade de odores através das moléculas liberadas pelas substâncias odoríferas. Auxiliar a busca ao alimento, ajudar na percepção do sabor, identificar situações de perigo e participar da reprodução são algumas das funções exercidas através do olfato.

O olfato é o sentido mais primitivo, com seu aparecimento a cerca de 3.500 milhões de anos, quando as primeiras células desenvolveram um sentido químico para perceber as informações ao seu redor. É também o primeiro a aparecer na escala evolutiva, o que pode ser visto, por exemplo, com a sua evolução nos peixes e nos anfíbios, onde o olfato se separa anatomicamente do paladar a 400 milhões de anos. Também as plantas utilizam o odor de suas flores para atrair os insetos e disseminar seu pólen e os insetos sociais, como as formigas que se reconhecem através do odor [Miret, 2004].

O olfato humano é considerado rudimentar quando comparado ao de outros mamíferos. O epitélio olfativo humano contém cerca de 20 milhões de células sensoriais, cada uma com cerca de seis pêlos sensoriais, enquanto que um cachorro, por exemplo, tem mais de 100 milhões destas células e cada uma contendo pelo menos 100 pêlos sensoriais. O olfato de um cachorro é bem mais apurado do que o dos humanos, ele é capaz de sentir o cheiro de uma pessoa que esteve num determinado local e tenha saído há horas [Miret, 2004].

Foi demonstrado em numerosos estudos que o olfato humano passa uma fase de aprendizagem até os 20 anos, e tende a manter-se constante até os 40 anos e decair a partir dos 50. E também que as mulheres possuem um melhor olfato e que ao envelhecer perdem em menor grau do que os homens [Miret, 2004].

2.2.1 Anatomia e fisiologia do sistema olfativo

A anatomia da *cavidade nasal* (Figura 2.1) tem início nas janelas do nariz, que está situada debaixo da caixa craniana e acima da boca, e está dividida pelo septo nasal em cavidade direita e esquerda [Vilela, 2006].

Toda a cavidade nasal é revestida por um epitélio secretor de muco e apresenta na parte inferior da cavidade uma mucosa vermelha, essa coloração por ser rica em vasos sanguíneos, apresentando também glândulas secretoras de muco que mantêm a região úmida. Na porção superior está presente a mucosa amarela, onde estão localizadas as células receptoras olfativas, e por isso também chamada de mucosa olfativa. Esta difere da mucosa respiratória circunjacente pela presença das glândulas tubulares de Bowman, pela ausência de batimentos ciliares rítmicos que são característicos da mucosa respiratória e por um pigmento marron-amarelado distintivo. Está localizada no corneto superior adjacente ao septo nasal [Bullock, 1998].

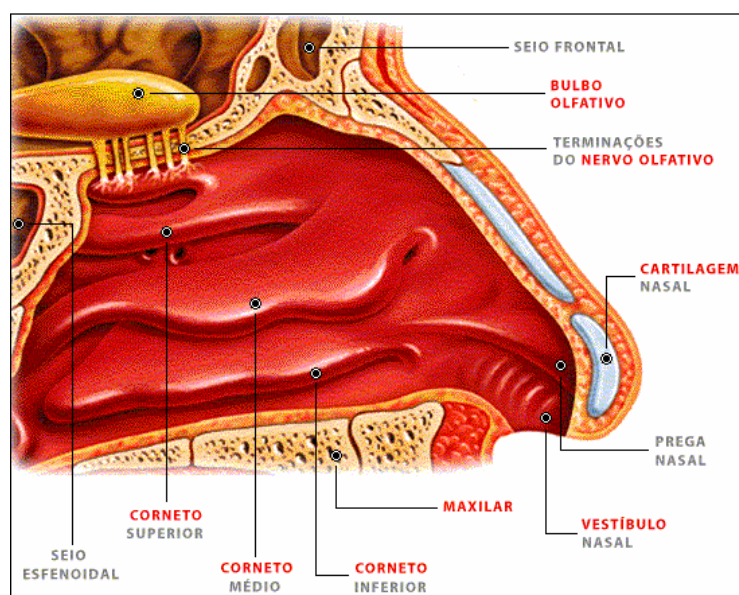


Figura 2.1 – Ilustração da cavidade nasal humana [EMCH, 2000]

A parede lateral de cada cavidade, apresenta três saliências que são chamadas de *cornetos*, sendo superior, médio e inferior. Todos se projetando para o interior da cavidade.

Em cada cavidade nasal, a mucosa olfativa tem uma área de superfície de aproximadamente 2 cm [Andrew et al., 2001].

Os receptores olfativos, em sua maioria, estão localizados na mucosa olfativa que fica localizada junto aos cornetos superiores e na parte superior da mucosa do septo nasal, estendendo-se em parte para o corneto médio. Com relação à ossatura do crânio, a mucosa olfativa está localizada imediatamente abaixo da lâmina crivada do etmóide, que a separa do bulbo olfativo, na cavidade craniana.

Na Figura 2.2, o epitélio da mucosa nasal é composto por três tipos de células: as células de sustentação ou suporte, as células receptoras ou células olfativas, as células Basais [Berne e Levy, 2004; Bear, 2002; Aires, 1999]

Entre as células olfativas são encontradas as glândulas de Bowman, que secretam o muco na superfície da membrana olfativa. O muco flui constantemente e é substituído a cada 10 minutos, e consiste de mucopolissacarídeos, proteínas, anticorpos e enzimas, além de sais dissolvidos em água [Guyton, 1997].

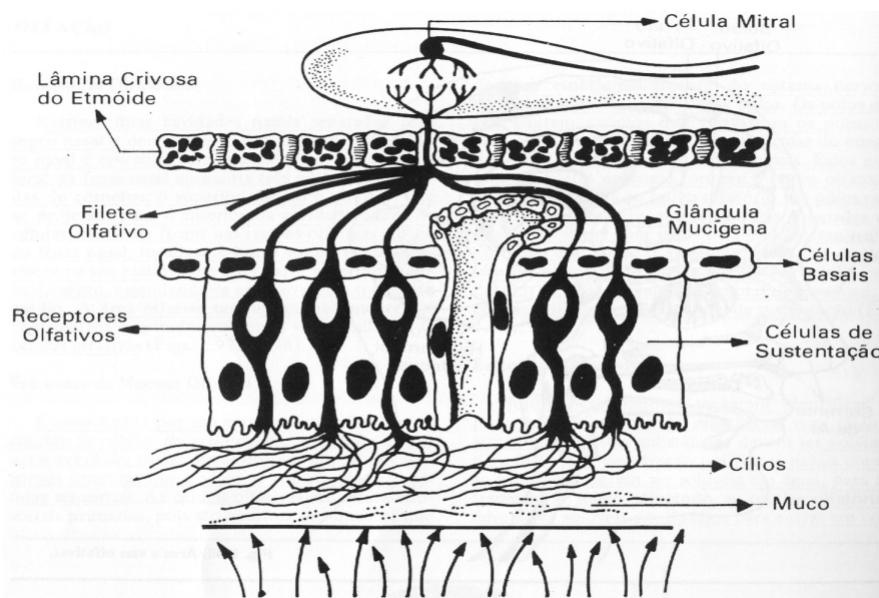


Figura 2.2 – Localização das células no epitélio olfativo [Leffingwell & Associates, 2001]

2.2.1.1 As vias olfativas

As vias olfativas são constituídas pelo bulbo olfativo, o trato olfativo e o córtex olfativo. Na Figura 2.3 pode ser visualizado o bulbo olfativo onde as células olfativas emitem axônios que após atravessarem a lâmina crivada do etmóide, se projetam para os dois bulbos olfativos. Os bulbos olfativos são expansões do cérebro, estrutura cortical, e são cheios de circuitos dendríticos, sinapses recíprocas incomuns e altos níveis de neurotransmissores [Be-ar, 2002; Aires, 1999].

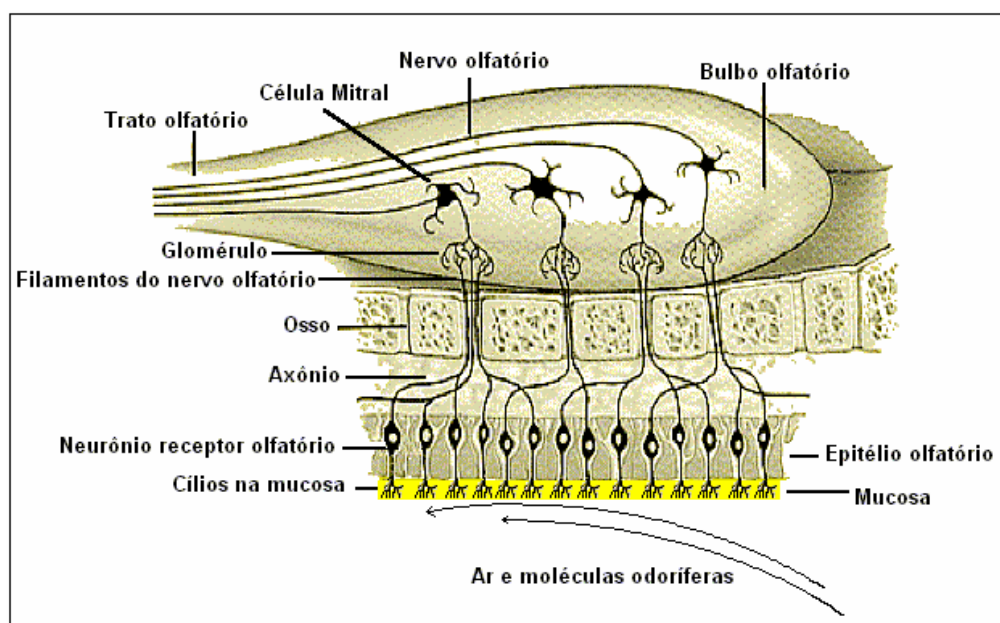


Figura 2.3 – Ilustração do bulbo olfativo e células mitrais [Leffingwell & Associates, 2001]

O trato olfativo, assim como o bulbo é uma extensão do cérebro. Os axônios das células mitrais e em tufo deixam o bulbo olfativo e entram no trato olfativo, cursando pela parte ventral do lobo frontal. No interior do trato olfativo existe um núcleo chamado de núcleo olfativo anterior, onde os neurônios recebem conexões sinápticas dos neurônios do bulbo e projetam-se através da comissura anterior para o bulbo olfativo contralateral.

À medida que o trato olfativo aproxima-se da estrutura perfurada anterior, que fica na base do cérebro, ele se divide nas estrias olfativas lateral e medial (Figura 2.4) que delimitam o trígono olfativo.

A área olfativa medial consiste num grupo de núcleos localizados nas regiões medionasais do cérebro, anterior e superior ao hipotálamo. Os núcleos mais conspícuos são os núcleos septais, localizados na linha média e que enviam sinais para o hipotálamo e outras partes do sistema límbico. A área olfativa lateral é composta principalmente pelo córtex pré-piriforme e piriforme, mais a porção cortical dos núcleos amigdaloídes. A partir dessas áreas os sinais seguem para quase todas as partes do sistema límbico, principalmente para o hipocampo [Guyton, 1997].

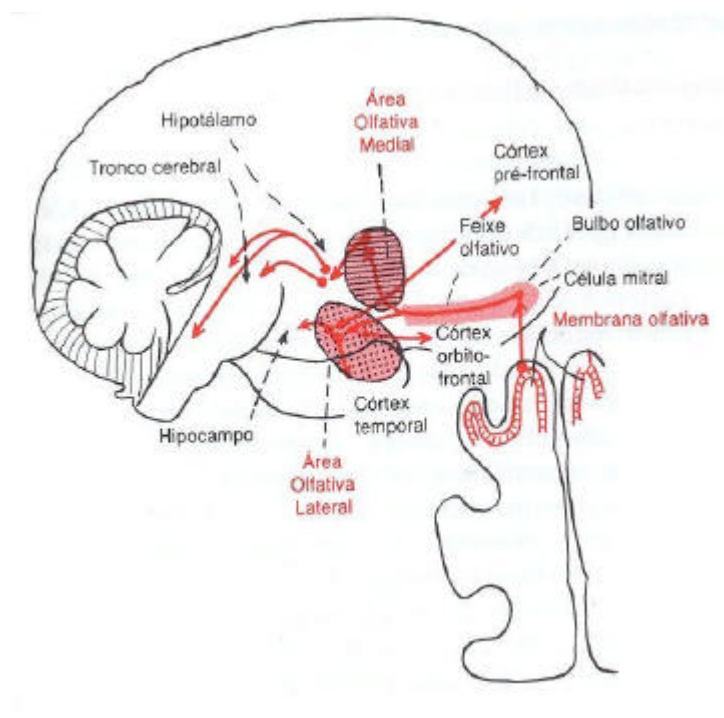


Figura 2.4 – Área olfativa medial e lateral [Guyton, 1981]

2.2.2 Teorias da olfação

Quando inalamos, as moléculas de odor que estão fluando no ar entram em contato com o epitélio olfativo e a partir deste contato as informações são passadas ao cérebro. A

sensação será tanto mais intensa quanto maior for a quantidade de receptores estimulados, o que dependerá da concentração da substância odorífera presente no ar.

Existe uma grande dificuldade em estudar o sistema olfativo devido a sua grande complexidade, e saber como a detecção, identificação e reconhecimento de um odor é feito pelo cérebro. Diante desta dificuldade várias teorias foram elaboradas e reformuladas com o passar do tempo na tentativa de fornecer respostas mais conclusivas.

Dentre as muitas, atualmente é aceito que as interações específicas entre moléculas de odores e proteínas receptoras na superfície dos cílios das células olfativas geram um potencial de ação elétrico ao nível de receptor, e esse potencial elétrico gerado é monofásico, negativo, único, dura alguns segundos e tem uma voltagem que se aproxima de 10 milivolts. Porém não existe um consenso como a partícula induz essa descarga elétrica ao nível de receptor.

Em 1959, Davies e Taylor sugeriram que a molécula olfativa se difunde através da membrana da célula do receptor e seu poder odorante estaria relacionado com sua capacidade de penetração mecânica na membrana do receptor. Os limiares de percepção dos odores seriam tanto mais baixos quanto maior o poder de rigidez e o peso das moléculas. Fortunato e Nicoline sugeriram a teoria energética ou óxido redutor, na qual admite que as moléculas de odores em contato com o muco realizam através de seu poder redutor, uma oxidação favorecida pela abundância de oxidase e peroxidase contida no interior deste muco [Davies e Taylor, 1959]. A seguir serão abordadas algumas das teorias formuladas, dentre elas a teoria dos odores primários, a estereoquímica e a vibracional.

2.2.2.1 Teoria dos odores primários

Na teoria dos odores primários são analisados diferentes odores como combinações de um pequeno número de odores primários. Com o tempo pesquisadores criaram várias

classificações de odores primários, a maioria dos fisiologistas estava convencida de que as numerosas sensações olfativas eram geradas por algumas poucas sensações primárias, assim como a visão e a gustação são geradas por algumas sensações selecionadas.

Em [Mueller, 1966] podem ser vistas várias classificações sugeridas para os odores primários. A mais antiga como a de um botânico Lineu, apresentou a classificação composta por uma lista com oito odores que foram denominados básico ou primários. Os odores sugeridos por Lineu foram: o aromático, fragrante, ambrosíaco, aliáceo, hircino, fétido, nauseabundo e repulsivo. Henning, em 1924, propõe uma lista com apenas seis odores primários onde ele aceita apenas três odores propostos anteriormente e acrescenta mais três. Os odores sugeridos por ele foram: aromático, fragrante, pútrido, etéreo, resinoso e queimado. Já [Crocker e Henderson, 1927] propõem apenas quatro odores, aceitando apenas dois dos odores proposto por Henning, e acrescentando mais dois, os quais foram: o fragrante, ácido, queimado e coprílico [Ross e Harriman, 1949].

Com base em estudos psicológicos, na tentativa de classificar estas sensações, eram descritos os seguintes odores: cancoforado, almiscarado, floral, mentolado, Etéreo, pungente e pútrido. Essa classificação tem sido utilizada com base para outras propostas como a do inglês John Amoore [Amoore, 1971] que incorporou os conceitos destes odores primários e sugeriu a *hipótese estereoquímica*.

2.2.2.2 Teoria estereoquímica

Em meados dos anos 50 os químicos observaram que as cadeias C4-C8 de determinados aldeídos/álcoois tinham odores fortes e que tinham uma mudança no odor com o aumento do comprimento da cadeia. O anel de benzeno 6-C alterava extremamente seu cheiro de acordo com o local que suas correntes laterais estavam situadas e foi visto também que

seus anéis maiores (átomos 14-19C) poderiam ser rearranjados consideravelmente sem alterar seu odor. Então a partir dessas observações foi proposta uma teoria que estava baseada na forma molecular dos odores. Em 1951, Moncrieff formulou sua própria teoria com a hipótese de “chave e fechadura” onde ele fez uma relação com o princípio utilizado na enzimologia de “chave e fechadura”, nesse princípio os cientistas acreditam que a forma das moléculas enzimáticas encaixe-se perfeitamente à forma dos substratos, como uma chave encaixa-se na fechadura correspondente. Moncrieff criou a hipótese de chave e fechadura adaptada ao odor, onde afirmava que os odores primários tinham locais ou sítios distintos no receptor olfativo. Em sua teoria, Moncrieff também afirma que uma substância para ser um odor deve possuir duas características: ter volatilidade e possuir uma configuração molecular que seja complementar a determinados locais no sistema do receptor [Jacob, 2006; Moncrieff, 1951].

Em 1952, Amoore propôs a teoria estereoquímica do odor baseada no modelo chave fechadura de Moncrieff, analisando mais de 600 substâncias e incorporando o conceito dos odores primários, propôs que existiam sete odores primários devido a sua alta frequência entre os 600 compostos orgânicos estudados. Segundo sua teoria as células olfativas distinguem cinco odores primários baseados na forma que são: cancoforado, almiscarado, etéreo, floral e menta e dois com base nas cargas elétricas: picante e pútrido. Todas as substâncias dentro de uma dessas sete categorias teriam uma configuração molecular comum [Ferreira e Acar, 1975].

Amoore também afirmou que os receptores olfativos seriam dotados de cavidade cuja forma corresponderia ao modelo dos diversos tipos de moléculas odoríferas. Os odores que são percebidos seria uma mistura dos odores primários. Como por exemplo, o odor de amêndoa seria a resultante de um odor cancoforado mais um odor floral mais um odor de menta, o odor de alho seria a resultante de um odor etéreo mais um odor picante mais um pútrido [Jacob, 2006; Ferreira e Acar, 1975].

2.2.2.3 Teoria vibracional

A teoria vibracional foi proposta pela primeira por [Dyson, 1938], ele sugeriu que a frequência de muitos odores está no infravermelho, e essa ressonância estaria associada ao seu cheiro, pois as diferentes frequências de infravermelho podiam causar cheiros diferentes. Um dos exemplos para explicar a teoria vibracional, foi observando que traças masculinas são atraídas às chamas de velas porque a emissão do infravermelho da chama de uma vela é idêntica a do ferormônio, que é um odor característico das traças fêmeas para atrair os machos. Dyson acreditava que os receptores olfativos entravam seletivamente em ressonância com as moléculas de odor, as quais o cérebro interpretava como odor. Ele sugeriu que o cérebro seria capaz de interpretar padrões de ativação de detectores sintonizados vibracionalmente como odores.

Mais tarde, em 1954, esta teoria foi defendida também por Wright, o qual supôs que a característica de uma molécula responsável por seu odor poderia ser seu espectro infravermelho, semelhante à visão, audição e olfação que dependem de vibrações eletrônicas de diferentes grandezas [Wright, 1977].

Em 1996, Luca Turin, um investigador francês da *University College of London*, publicou um artigo na Revista *Chemical Senses* que apresentava a hipótese “vibratória” para o olfato. Turin se baseou nas definições de Dyson e Wright, diferindo apenas por apresentar um mecanismo detalhado e plausível para a transdução biológica de vibrações moleculares onde o elétron construiria um túnel pelo receptor. Turin afirmou que assim como a visão e a audição, o olfato se comportaria como um espectroscópio e que os receptores olfativos responderiam a vibração das moléculas e não a sua forma molecular. Ele realizou uma série de experimentos nos quais afirmava que duas substâncias de forma química muito similar, porém com diferentes vibrações mostravam-se diferentes e que duas substâncias de forma química diferentes,

mas com vibrações similares mostravam-se iguais [Miret, 2004; Turin e Yashii, 2003; Gardner e Persaud, 2000; Turin, 1996].

Turin mostrou o mecanismo como os receptores olfativos detectam a molécula de odor e ativam a proteína-G. Quando a proteína olfativa do receptor se liga a um odorante, a construção do túnel do elétron pode ocorrer através de um receptor específico se a energia vibracional igualasse a abertura de energia entre níveis preenchidos e vazios do elétron. O elétron que construiu o túnel no receptor da célula olfativa, então ativaria a proteína-G, que por sua vez ativa uma cascata de eventos para transmissão de impulso. Os Receptores consequentemente “são ajustados” à frequência vibracional de odorantes particulares [Gardner e Persaud, 2000]. O esquema de Turin pode ser visto em [Turin, 1996]

2.3 Narizes Artificiais

Após um breve estudo sobre o nariz biológico, os narizes artificiais, envolvem diversas áreas como a Biologia, Química, Eletrônica, Computação, Estatística, entre outras, ou seja, um estudo multidisciplinar com várias especificidades a serem respeitadas. Porém para o desenvolvimento de tal dispositivo é necessário entender um problema constante e ainda em aberto que consiste no processo de aquisição e reconhecimento dos sinais de odores.

Os narizes artificiais têm uma história muito recente, ainda que suas origens estejam baseadas nos sistemas desenvolvidos nos anos sessenta [Wilkins e Hatman, 1964], porém o termo “nariz artificial” e os primeiros sistemas inteligentes da olfação artificial não apareceram até a segunda metade da década de oitenta. Pode-se afirmar, de todas as maneiras, que o interesse por esse tipo de sistema e sua investigação deu um salto espetacular na primeira metade da década de noventa, e partir daí passou-se a aceitar a definição de nariz artificial como sendo “um instrumento composto por uma matriz de sensores químicos sensíveis a certas

substâncias e algoritmos inteligentes capazes de reconhecer aromas simples e complexos” [Gardner e Bartlett, 1994].

Desde os anos noventa existem narizes artificiais comerciais que permitem realizar estudos sobre a utilização desta tecnologia em diferentes aplicações [Fígaro, 2006]. Estes equipamentos de caráter genérico já que não foram desenhados especificamente para uma determinada aplicação. Entretanto, devido ao estado atual dessa tecnologia, existem muitas aplicações em que a otimização desses equipamentos é de vital importância para um funcionamento eficaz. Diante disso existe uma outra classe de narizes artificiais, os dedicados a aplicações específicas, e devido a essa exclusividade passaram a se chamar “ASEN”, do inglês *Application Specific Eletronic Noses*, ou seja, narizes artificiais para aplicações específicas, e só recentemente estão aparecendo no mercado equipamentos que seguem essa filosofia como o “*Freshsense*”, sistema desenhado exclusivamente para o monitoramento do estado de conservação de pescado [Element Ltd, 2006].

Em poucos segundo somos capazes de detectar, no ar, a presença de substâncias em concentrações tão baixas, e que para um dispositivo artificial detectá-las teriam de ser extremamente sensíveis. Vivemos em um mundo rodeados de substâncias onde na atmosfera, na forma gasosa, sua maior parte é ou se tornam invisíveis. Porém, não menos perigosas, pois a maior parte dos gases inflamáveis, tóxicos e altamente explosivos são transparentes. O que dificulta a prevenção de acidentes é que muito desses gases também são inodoros. Testes realizados com sensores químico-resistivos de aroma demonstraram que narizes eletrônicos são capazes de fazer distinção entre substâncias que os humanos não têm condições de fazer. Por exemplo, um nariz artificial pode distinguir álcool etílico de metílico, no entanto, seres humanos não têm um olfato muito apurado para diferenciar um do outro, o que frequentemente provoca a morte de pessoas que preparam bebidas alcoólicas utilizando metanol, substância

tóxica que mata ou sega, em doses menores, enquanto o etanol produz apenas ressaca e quando tomado em excesso [Santos, 2000].

Na medicina, por mais interessante que seja, não é um assunto novo a detecção de enfermidades através do odor, tal detecção foi largamente praticada pelos chineses e gregos desde 2000 anos a.C. Embora esses praticantes não descobrissem nem definissem a patologia das bactérias causadoras da doença, perceberam claramente que o paciente infectado apresentava mudanças no odor dos excrementos produzidos pelo corpo como, por exemplo, na urina, no suor e na saliva. Os primeiros estudos na área de liberação de odor microbiais datam de meados de 1923, quando Omelianski percebeu que espécies patogênicas como, por exemplo, a “*Mycobacterium tuberculosis*” e a “*Pseudomonas aeruginosa*”, exalavam odores característicos [Pavlou e Turner, 2000; Omelianski, 1923]. Testes iniciais demonstram que, num futuro próximo, será possível utilizar narizes artificiais, não só para um rápido diagnóstico de doenças, bem como para o monitoramento contínuo da evolução das doenças e seus estágios. Atualmente os narizes artificiais estão sendo empregados em várias áreas do diagnóstico médico. Estão sendo realizados pesquisas e projetos [D’Emery et al., 2007; Batista e D’Emery, 2006; Aathithan et al., 2001; Guernion et al., 2001; Pavlou e Turner, 2000] capazes de reconhecer e diagnosticar doenças pulmonares, gastrintestinais e no tratamento de doenças provocadas por bactérias causadoras de infecções urinárias [Fend et al., 2005]. Acreditamos também que sensores de odor seja um importante recurso para sistemas móveis, em um futuro breve, podendo ser instalados em laboratórios de diagnóstico para locais de difícil acesso [D’Emery et al., 2005a,d]. Na agricultura, frutas e hortaliças também são testadas na utilização de narizes artificiais [D’Emery et al., 2005c; Llecha, 2001], uma vez que apresentam odores próprios provenientes de seus compostos, com mudança antes e depois da colheita. A intensidade dos odores depende do estado em que se encontram e em geral são classificadas de acordo com a intensidade de respiração de cada espécie, denominadas de climatéricas e não-climatéricas [Az-

zolin, 2003]. Diversas tecnologias de narizes artificiais e alguns até mesmo comerciais podem ser encontradas em [Pearce et al., 2003].

2.3.1 Sensores de Aroma

Os narizes artificiais fazem uso de sensores que ao serem estimulados por determinados compostos odorantes mudam suas características físico-químicas, e são essas mudanças depois de mensuradas as responsáveis pelo reconhecimento do odor em questão. São muitos os tipos de sensores existentes, cada um com características específicas, onde sua seletividade é vinculada a um determinado tipo de substância, nesta seção veremos alguns tipos de sensores utilizados em narizes artificiais.

Podemos destacar duas categorias de sensores: os sensores de absorção e os semicondutores quimicamente sensíveis como os citados acima. Os sensores de absorção são formados por uma camada de material absorvente quimicamente seletivo, aplicada em um dispositivo de base acústica ou óptica. Ao entrarem em contato com as moléculas de odor, esta camada sofre mudanças em suas propriedades ópticas, como índice de refração, mecânicas, como espessura, densidade e elasticidade, produzindo um sinal relacionado à concentração do odor em contato com a superfície [Santos, 2000].

Os quimiosensores são dispositivos capazes de converter uma quantidade química em um sinal elétrico dependendo da concentração específica das partículas, assim como os átomos, moléculas ou íons presentes em gases ou líquidos. Os tipos de quimiosensores que podem ser usados em narizes artificiais precisam reagir com as moléculas odorantes na fase gasosa, dos quais são tipicamente moléculas orgânicas voláteis com diferentes massas molares. Alguns já são explorados e até mesmo industrializados. Princípios elétricos, térmicos e ópticos podem constituir estes quimiosensores de acordo com sua classe, ou seja, semicondu-

tores de óxido de metal (MOS), polímeros orgânicos semicondutores (CP), quimiocondutores, transistores MOS de efeito de campo (MOSFET), micro balança de cristal de quartzo (QCM), superfícies de ondas acústicas, fluorescentes e muitos outros podem ser usados facilmente como sensores para um nariz artificial. Detalhes sobre os tipos de quimiosensores apresentados aqui podem ser encontrados em alguns trabalhos [dos Santos, 2003; Pearce et al., 2003; Toko, 2000; Göpel et al., 1999, Gardner, 1994; Hauptmann, 1993].

Os semicondutores de óxido metálico são formados por uma película de cerâmica eletricamente aquecida, sobre a qual se deposita um filme fino e poroso de óxido metálico, dopado com diversos metais. As moléculas de odor interagem com o oxigênio absorvido na superfície, afetando a condutividade e a sensibilidade do filme, controladas por temperatura. Também se podem modificar as características deste sensor através da variação do dopante utilizado [Gardner et al., 1994].

Quando os sensores alcançam temperaturas próximas de 400°C, diante da ausência de oxigênio, os elétrons livres fluem facilmente entre as fronteiras de dióxido de estanho, por exemplo. No ar puro, os oxigênios, que retém os elétrons livres devido a sua afinidade eletrônica, ficam absorvidos na superfície do dióxido de estanho criando uma barreira de potencial nas fronteiras. Estas barreiras dificultam a livre circulação de elétrons livre o que reduz a resistência do sensor.

Os polímeros condutores pertencem a uma família de materiais, os quais, divergindo dos plásticos comuns, conduzem corrente elétrica. A condutividade elétrica é possível pela estrutura de ligações químicas duplas e simples, distribuídas alternadamente, nas cadeias moleculares dos polímeros. As propriedades ópticas e elétricas destes plásticos podem ser modificadas pela variação da temperatura, pressão, interação com solventes, ou pela aplicação de um potencial elétrico [Santos, 2000].

Os sensores construídos com polímeros condutores apresentam algumas vantagens em relação aos demais compostos, entre as quais se destacam [Persaud e Travers, 1997]: rápida cinética de absorção de calor, baixo consumo de energia (*microwatts*), resistência a compostos que tornam inativos sensores inorgânicos de semicondutores, como compostos a base de enxofre e construído especificamente para determinada aplicação.

2.4 Nariz biológico versus nariz artificial

Aqui será feito um paralelo entre o nariz biológico e o artificial, tendo como objetivo o melhor entendimento da funcionalidade de tais sistemas. Mesmo sabendo da diferença entre narizes artificiais e o humano, quanto à estrutura sensorial, desde formato a características físico-químicas, ambos apresentam as mesmas etapas no processo de detecção e processamento dos sinais de odores. Na Figura 2.5, podemos identificar as etapas de reconhecimento dos sinais de odor do sistema olfativo biológico e as fases equivalentes no processo de detecção do odor através de um nariz artificial.

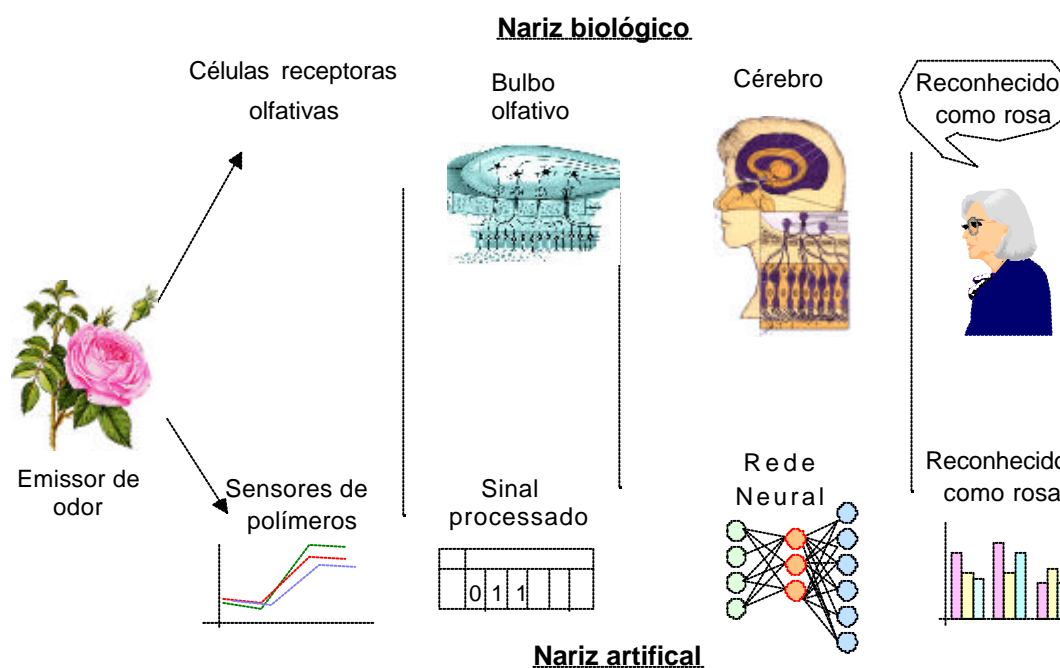


Figura 2.5 – Quadro comparativo da olfação humana com o nariz artificial [Santos, 2000]

A primeira fase no sistema olfativo humano é composta por um conjunto de células receptoras olfativas, situadas na parte superior do nariz, responsáveis pelo primeiro contato com as moléculas de odor que estão expostas no ar. A equivalente fase do nariz artificial é feita com o conjunto de sensores de odor, no exemplo da Figura 2.5 sensores à base de polímeros condutores sofrem uma mudança em suas resistividades elétricas quando em contato com as moléculas de odor.

Na tentativa de explicar como é feita a interação físico-química da primeira fase entre os gases que atingem o epitélio olfativo e os neurosensores, capítulo 2, seção 2.1, assim como muitos cientistas das mais diversas áreas que o estudo do nariz biológico envolve, [Figueiredo, 2007; Batista, 2006] a partir de uma teoria no reconhecimento do odor construíram um modelo neural que a simulasse. E nosso grupo de pesquisa, integrantes da empresa CRIATRONICS, também está estudando estas teorias no nariz artificial através de várias técnicas de construção de sensores e circuitos eletrônicos para aquisição de sinais de odores. No nosso caso, os sensores utilizados envolvem reações físico-químicas onde a reação dos sinais de odor com os sensores corresponde a valores de resistência elétrica que são armazenadas no computador ou valores da tensão elétrica presente no filme pigmentado. De forma geral, os sensores de um nariz artificial são como os neurônios receptores no epitélio olfativo biológico, responsável por converter o estímulo provocado pelos compostos odorantes frente aos seus sensores, em sinais perceptíveis e mensurados.

A segunda fase surge logo após a reação físico-química da primeira fase, onde começa o processamento ou etapa de identificação, também chamado reconhecimento, do sinal. Diante das diversas teorias, já citadas, do nariz biológico, não existe teoria universalmente aceita que explique como este processamento é feito. Mas nesta dissertação adotamos os trabalhos de [Batista, 2006; Li e Hartz, 2000, 1999; Li, 1995; Axel, 1995; Freeman, 1991;

Yao e Freeman, 1990], pesquisadores que estudam este tema há muito tempo e defendem ambas as abordagens.

O bulbo olfativo é o primeiro ponto de processamento e transmissão da via olfatória, segundo [Freeman, 1991], é uma região de convergência dos sinais onde diversas camadas de células realizam uma espécie de pré-tratamento dos dados de odor. A função destas células neste estágio é reduzir o ruído pela compressão dos sinais e amplificação da saída, isto aumenta a seletividade e a sensibilidade do sistema olfativo. O estágio correspondente no nariz artificial é o pré-processamento dos dados, tendo como objetivo eliminar ruídos e adequá-los para a entrada de dados de técnicas de reconhecimento de padrão. Existem alguns fatores que podem ser tratados nesta fase de pré-processamento, como deslocamento do sensor, escala ou faixa de valores.

Após a segunda fase, pré-processamento do sinal de odor, o próprio bulbo, através de suas células, pode fazer o reconhecimento do sinal de odor [Freeman, 1991] ou então os sinais são enviados ao córtex olfativo, região pertencente ao cérebro, onde o reconhecimento do sinal de odor é feito [Axel, 1995].

As duas últimas fases, terceira e última fase, o reconhecimento do sinal de odor no sistema olfativo pode ainda ser executada pelo cérebro após tratamento (terceira fase) e codificação (última fase) do sinal que é enviado para regiões do cérebro [Freeman, 1991].

Cientistas acreditam que o córtex piriforme no cérebro realiza funções na memória associativa onde a assinatura das respostas das células olfativas é associada a noções do cheiro, embora existam outras regiões do cérebro que também estão envolvidas, como o hipocampo. A última etapa no nariz eletrônico é caracterizada por um sistema de reconhecimento de padrões para identificar o odor, dos quais os sinais de odores podem ser pré-classificados por técnicas extratoras de características como PCA, e utilizadas redes neurais artificiais como técnica de reconhecimento do padrão de cheiro no nariz artificial.

2.5 Resumo do capítulo

Neste capítulo foram apresentadas as tecnologias que são inspiradas nos sentidos do ser humano, assim como apresentou uma descrição teórica sobre o sistema olfativo humano e sobre os narizes artificiais. Foram discutidas as principais etapas do funcionamento do nariz biológico além das teorias existentes sobre o reconhecimento de odor. No que diz respeito aos narizes artificiais foram citados alguns trabalhos científicos que utilizam tais tecnologias nas mais diversas áreas de pesquisa. Foi comparado o modelo do nariz biológico com o artificial nas etapas de seus funcionamentos.

No próximo capítulo será abordada a doença Diabetes Mellitus que é tema de aplicação do nariz artificial como estudo de caso desta dissertação.

Capítulo 3

Diabetes

Neste capítulo é apresentada uma explanação sobre a doença diabetes, da qual será objeto dos testes de sensibilidade e reconhecimento do nariz artificial usado nesta dissertação. Na seção 3.1 é explanada a doença. Na seção 3.2 apresenta os diversos tipos de diabetes. Na seção 3.3 são abordados os sintomas que os pacientes com diabetes apresentam e como é feito o diagnóstico da doença. Na seção 3.4 são descritos alguns tratamentos e prevenções para o Diabetes. Na seção 3.5 é apresentado um resumo do capítulo.

3.1 Introdução

Diabetes mellitus é um grupo de enfermidades metabólicas caracterizadas por hiperglicemia (aumento dos níveis de glicose no sangue), resultado de defeitos na secreção de insulina, em sua ação ou ambos. Trata-se de uma complexa doença na qual coexiste um transtorno global do metabolismo dos carboidratos, lipídios e proteínas, bem como pelo desenvolvimento tardio de complicações vasculares e neuropáticas [Price, 1987; Haberly, 1985].

O termo diabetes, em grego: *o d?af?t??*, foi inventado por Aretaeus de Cappadocia. É derivado da palavra, grega *d?afa??e??*, *diabaínein* que literalmente quer dizer “que passam completamente,” ou “sifão”, referência a um dos sintomas principais do diabetes, ou seja, a produção excessiva de urina. Em 1675, Thomas Willis adicionou o *mellitus*, do latim, que significa um gosto doce. Isto havia sido observado por muito tempo, em épocas antigas antes

dos gregos, chineses, egípcios ou pelos indianos. Em 1776, Matthew Dobson confirmou que o gosto doce era por causa de um tipo de açúcar na urina e no sangue das pessoas com diabetes [Dobson, 1776]

Os antigos indianos estudaram o diabetes observando se as formigas eram atraídas pela urina de uma pessoa, e chamaram a indisposição física de “doença de urina doce” (*Madhumeha*). Para os coreanos, chineses e japoneses também possui o mesmo significado. Enquanto o termo diabetes geralmente se refere a diabetes mellitus, existem outros tipos de diabetes. O mais comum destes é o *insipidus*, em que a urina não é doce.

Um dos mais importantes processos metabólicos do organismo humano é a conversão de alimentos em energia e calor. De forma geral, os alimentos são constituídos de três nutrientes principais: carboidratos, proteínas e gordura. No processo digestivo os carboidratos são transformados em glicose, ou seja, açúcar no sangue, as proteínas em aminoácidos e gorduras em ácidos graxos. Podemos retirar energia de qualquer uma das três categorias, mas os carboidratos são, em especial, importantes, pois são rapidamente convertidos em glicose quando precisamos rapidamente de energia. Entre as refeições, o fígado libera a glicose estocada à corrente sanguínea, mantendo normais os níveis de glicose no sangue [dos Santos, 2006].

O pâncreas é o órgão responsável pela produção do hormônio denominado insulina. Este hormônio é responsável pela regulação da glicemia⁵. Para que as células das diversas partes do corpo humano possam realizar o processo de respiração aeróbica, onde utiliza glicose como fonte de energia, é necessário que a glicose esteja presente na célula. Portanto, as células possuem receptores de insulina que, quando acionados “abrem” a membrana celular para a entrada da glicose presente na circulação sanguínea. Uma falha na produção de insulina

⁵ glicemia é o nível de glicose no sangue

resulta em altos níveis de glicose no sangue, já que esta última não é devidamente dirigida ao interior das células. Visando manter a glicemia constante, o pâncreas também produz outro hormônio antagônico à insulina, denominado *glucagon*. Ou seja, quando a glicemia cai o glucagon é secretado visando restabelecer o nível de glicose na circulação [dos Santos, 2006].

Sendo assim, para ajudar a penetração do suprimento de açúcar em cada célula do corpo, o pâncreas envia insulina para a corrente sanguínea, fazendo com que o hormônio chegue aos receptores de insulina na superfície destas células. Só quando a insulina se liga à superfície das células é que elas podem absorver a glicose da corrente sanguínea. Quando o nível de glicemia, açúcar no sangue, aumenta após as refeições, a quantidade de insulina (mas conhecida como insulina da hora da refeição) também aumenta, para que este excesso de glicose possa ser rapidamente absorvido pelas células. O fígado pára de secretar glicose e passa a estocar glicose no sangue para usá-la posteriormente e quando a insulina termina seu trabalho ela se degrada. Sendo assim, o corpo tem que renovar constantemente seu estoque de insulina.

Atualmente no Brasil estima-se que cerca de cinco milhões de pessoas sejam diabéticos, sendo que metade delas não sabe que possuem diabetes [Isaacson, 2001]. De acordo com a Organização Mundial de Saúde, em 2006 cerca de 171 milhões de pessoas estão doentes, e esse índice aumenta rapidamente, estimando-se que em 2030 esse número dobre.

3.2 Os diversos diabetes

Segundo a Organização Mundial de Saúde o Diabetes é classificada em quatro tipos: tipo I, tipo II, gestacional e insípida [WHO, 1999]. O termo universal *diabetes mellitus tipo I* substituiu diversos termos anteriores, incluindo o diabetes infantil, diabetes juvenil ou

diabetes insulino-dependente. O *diabetes mellitus tipo II* substituiu também diversos termos mais velhos, incluindo o diabetes da fase adulta, diabetes relacionado a obesidade ou o diabetes não-insulino-dependente. Além destes números, não há nenhum padrão, assim um tipo II que não tenha se tornado insulino-dependente tem sido chamado às vezes o tipo III, o mesmo termo também é usado para o diabetes gestacional em alguns casos [ADA, 2006]. A distinção entre o que se sabe hoje como o diabetes tipo I e o tipo II foi inicialmente feito claramente pelo Sir Harold Percival (Harry) Himsworth em 1935 e foi publicada em janeiro 1936 [Himsworth, 1936].

No diabetes mellitus tipo I as células β do pâncreas, que normalmente produzem insulina, foram destruídas devido a processos auto-imunes ou idiopáticos⁶. Quando pouca ou nenhuma insulina vem do pâncreas, o corpo não consegue absorver a glicose do sangue, ou seja, as células começam a “passar fome” e o nível de glicose no sangue fica constantemente alto. A solução é injetar insulina subcutânea⁷ para que possa ser absorvida pelo sangue. Ainda não se produziu uma forma de insulina que possa ser administrada oralmente já que a insulina é degradada pelo estômago em uma forma inativa.

Uma vez que o distúrbio se desenvolve, não existe maneira de “reviver” as células produtoras de insulina do pâncreas. O transplante de um pâncreas sadio, ou apenas, o transplante de células produtoras de insulina de um pâncreas sadio já foram tentados, mas ainda são considerados transplantes experimentais. Portanto, a dieta correta e o tratamento com insulina ainda são necessários por toda a vida de um diabético.

Não se sabe o quê causa a destruição das células produtoras de insulina do pâncreas ou o porquê do diabetes aparecer em certas pessoas e em outras não. Fatores hereditá-

⁶ *Idiopático* diz-se da doença que não está ligada a outra.

⁷ *Insulina subcutânea* é um tipo de injeção ministrada embaixo da pele.

rios parecem ter o seu papel, mas o distúrbio, praticamente, nunca é diretamente herdado. Os diabéticos, ou as pessoas com diabetes na família, não devem ter restrições quanto a ter filhos.

Embora não se saiba o que causa o diabetes mellitus tipo II, sabe-se que neste caso o fator hereditário tem uma importância bem maior do que no diabetes tipo I. Também existe uma conexão entre a obesidade e esse tipo de diabetes, e surge normalmente após os 40 anos, embora a obesidade não leve, necessariamente ao diabetes. O diabetes tipo II é um distúrbio comum, afetando de 2% a 10% da população e é a forma mais comum da doença, sendo encontrada em mais de 90% da população diabética [WHO, 1999].

Todos os diabéticos tipo II produzem insulina quando diagnosticados e, a maioria, continuará produzindo insulina pelo resto de suas vidas. O principal motivo que faz com que os níveis de glicose no sangue permaneçam altos está na incapacidade das células musculares e adiposas de usar toda a insulina secretada pelos pâncreas, anomalia também chamada de “resistência insulínica”. Assim, muito pouco da glicose presente no sangue é aproveitado por estas células.

O *diabetes gestacional* envolve também uma combinação da inadequada secreção e recepção de insulina, assemelhando-se ao diabetes tipo II em diversos aspectos, podendo persistir ou desaparecer depois do parto. Mesmo que possa ser passageiro, o diabetes gestacional pode prejudicar a saúde do feto ou da mãe, e aproximadamente cerca de 20% a 50% das mulheres com diabetes gestacional desenvolvem o diabetes tipo II diabetes e ocorre em aproximadamente de 2% a 5% de todas as grávidas. É provisório e inteiramente tratável, mas se não tratado, pode causar problemas com a gravidez, incluindo macrosomia (peso elevado do bebê), mal-formação fetal e doença de coração congênita, requer a supervisão médica cuidadosa durante a gravidez.

Existe ainda o *diabetes insípido*, porém este é um tipo menos comum, denota a incapacidade de o rim reter urina. É causado por uma deficiência do hormônio antidiurético, ou

por uma insensibilidade dos rins a esse hormônio. Urinar em excesso e sede excessiva (especialmente para a água fria) são típicos do diabetes insípido.

Outros tipos específicos de diabetes abrangem vários defeitos genéticos nas células beta, resistência à insulina genética, doenças do pâncreas, defeitos hormonais, ou causados por ingestão de fármacos ou drogas. A décima versão do *International Statistical Classification of Diseases* possui uma entidade de diagnóstico chamada “*mellitus malnutrition-relacionado do diabetes*”. E recomendou, em 1999, que fosse proposto uma nova taxonomia para novos tipos de diabetes [WHO, 1999].

3.3 Sintomas e diagnósticos

Em um paciente diabético tipo I, se não tratado, o metabolismo da glicose pelas células do corpo é reduzida e o excesso de glicose presente no sangue é eliminado pela urina. Esta condição produz alguns sintomas, sendo os mais frequentes: perda de peso, fadiga, poliúria (excesso de urina) e polidipsia (sede excessiva). Uma pessoa com estes sintomas pode ser facilmente diagnosticada como portadora do diabetes tipo I. Os sintomas do diabetes tipo II são menos pronunciados e esta é a razão para considerar este tipo de diabetes mais brando que o tipo I, logo os portadores do diabetes tipo II apresentam diversos níveis de gravidade e sintomas, e podem passar por um período, até de muitos anos, onde não se suspeitará do diabetes [Miller, 1976].

No diabetes insípido urinar em excesso pode ocorrer durante todo o dia e a noite. Nas crianças, podem interferir com o apetite, o ganho de peso e o crescimento também. Podem apresentar-se com febre, vômito ou diarreia. Os adultos podem permanecer saudáveis por

décadas contanto que seja ingerida bastante água para compensar as perdas urinárias, entretanto, há um risco contínuo da desidratação.

Pacientes diabéticos têm maior propensão em desenvolver hipertensão, arteriosclerose, doenças oculares, doenças renais, e devem principalmente estar atento a feridas não cicatrizantes em extremidades do corpo (pés, pernas, e mãos) que podem levar à amputação do membro.

Para diagnosticar apropriadamente o diabetes, o médico deve saber a quantidade exata de glicose presente no sangue do paciente. Esta quantidade de glicose é expressa em milimols por litro (mmol/L), referindo-se ao número de moléculas de açúcar por litro de sangue, sendo este valor expressado em miligramas de açúcar por decilitro (mg/dL).

Com relação a exames laboratoriais, o exame mais comum para medir o nível de glicose no sangue chama-se “glicemia de jejum”. É um teste feito através do sangue venoso. Em indivíduos não diabéticos, o nível normal de glicose no sangue é de aproximadamente 5 mmol/L (90 mg/dL), uma variação de 70 até 110 mg/dL. Logo após uma refeição, este nível aumenta aproximadamente para 7 mmol/L (126 mg/dL). O nível raramente cai abaixo de 3,5 mmol/L (63 mg/dL). Se o resultado ficar em torno de 110 a 125 mg/dL, não indica diagnóstico da doença, porém levanta suspeita [WHO, 1999]. Os laboratórios podem fazer um exame chamado hemoglobina glicosada, ou HbA1c, que indica os níveis de glicose no sangue durante os últimos 120 dias.

Ocorrendo um resultado igual ou acima de 126 mg/dL, em pelo menos dois exames consecutivos, fica então confirmado o diagnóstico do diabetes mellitus. Já com uma glicemia superior a 140 mg/dL, mesmo sendo recolhida a qualquer hora do dia, já se confirma o diagnóstico do diabetes. Em geral, não se encontra açúcar na urina se o nível de glicose no sangue for menor que 10 mmol/L (180 mg/dL).

Não é fácil definir exatamente o nível ideal de glicose para todos os diabéticos. A Tabela 3.1 fornece uma idéia aproximada do nível de glicose no sangue a ser atingido, podendo ser estabelecidos outros limites por recomendações médicas.

Tabela 3.1 – Nível de glicose no sangue para diabéticos.

| Situação | mmol/L | mg/dL |
|-----------------------------------|--------|-------|
| Em jejum e antes das refeições | 7,0 | 126 |
| Uma hora depois de uma refeição | 10,0 | 180 |
| Duas horas depois de uma refeição | 8,0 | 144 |
| Antes de dormir (mínimo) | 6,0 | 108 |

A presença de glicose na urina é um meio indireto de testar o excesso de glicose presente no sangue, porém ele não lhe dirá nada sobre o nível exato deste excesso de açúcar ou se o nível de açúcar está muito baixo. A maneira mais comum de testar a *glicosúria* (presença de açúcar na urina) é utilizar uma fita reagente específica. A fita muda de cor para indicar a presença do açúcar. Geralmente, esse teste é feito duas vezes por dia, de acordo com as necessidades que o médico detectar. Como no teste de urina, existem, também, medidores que podem “ler” a fita e dar uma medição precisa.

Outra forma de avaliação do diabetes é através dos níveis de cetona. As cetonas são produzidas quando o corpo começa a degradar o tecido adiposo numa tentativa de alimentar às células “famintas”. Se o diabetes, não é bem controlado, o seu corpo pode produzir quantidades excessivas de cetonas que podem causar uma condição grave conhecida como *cetoacidose*. Embora esta condição se desenvolva vagarosamente, o paciente com diabetes deve reduzir seu nível de glicose no sangue quando testes regulares de urina ou sangue mostrar níveis irregularmente altos. O nível de cetona na urina também pode ser medido por tabletes ou fitas de níveis, devendo sempre ser realizado quando estiver com febre, diarreia ou até mesmo estressado. Mesmo sabendo das diversas formas de controlar o diabetes, ainda pode-se

realizar testes na urina, sobretudo para controlar as cetonas, quando a glicemia é superior a 200 mg/dL [Neto, 2004].

3.4 Tratamento

O objetivo do tratamento é alcançar um bom controle metabólico, isto é, a utilização adequada de glicose. Desta forma evita-se grande parte das complicações que ao longo prazo pode alterar a qualidade de vida, tanto do diabético insulino-dependente como do não insulino-dependente.

Os diabéticos que raramente monitoram seus níveis de glicose no sangue podem estar controlando muito mal o distúrbio sem se darem conta disso, porque os sintomas não são sempre óbvios. Porém, muitos diabéticos se sentem melhor quando monitoram o nível de glicose no sangue. Existem indicações que um bom controle da glicose no sangue retardará ou prevenirá o desenvolvimento das posteriores “complicações do diabetes”, dentre elas estão: aumento dos riscos de um ataque cardíaco, circulação sanguínea deficiente, perda de sensibilidade nas pernas e nos pés, olho diabético e doenças renais.

Uma boa dieta pode prevenir o diabetes. Os alimentos pode, a grosso modo, ser dividido em duas categorias: os que contêm açúcares “rápidos” (carboidratos de absorção rápida) e os que contêm açúcares “lentos” (carboidratos de absorção lenta). Os alimentos com açúcares rápidos são, por exemplo, geléias, doces, balas, frutas, sucos e leite, estes produzem altos níveis de glicose no sangue (dependendo da quantidade consumida), porque o açúcar chega a corrente sanguínea em um curto período de tempo. Portanto, é melhor combina-los com açúcares lentos, encontrados em alimentos como batatas, vegetais e arroz. Os açúcares lentos são mais seguros para o diabético porque eles chegam à corrente sanguínea mais len-

tamente e dão ao corpo a chance de absorvê-los antes que eles se “acumulem” no sangue. As fibras dos alimentos retardam a absorção de açúcares. Os açúcares rápidos podem ser consumidos nos períodos onde o nível de glicose no sangue está muito baixo. Algumas regras gerais de uma boa alimentação:

- (i) Comer de 4 a 6 pequenas refeições por dia;
- (ii) Manter horários rígidos para as refeições, porém não pular-las;
- (iii) Comer apenas quantidades recomendadas pelo médico, nutricionista e/ou educador em diabetes;
- (iv) Comer pães de fibras ou de grãos inteiros, evitando o pão branco;
- (v) Comer verduras e legumes diariamente;
- (vi) Evitar gorduras, açúcares e álcool.

Os exercícios aumentam a sensibilidade do corpo à insulina e, portanto, tendem a diminuir o nível de glicose no sangue. Para o diabético, qualquer tipo de atividade física (trabalho em casa, caminhar ou correr) deve ser considerado como exercício. Exercícios regulares e programados são melhores porque impactos súbitos, de exercícios mais intensos, podem trazer problemas para o controle da glicose no sangue.

No tratamento com insulina, ela só pode ser administrada por injeção porque é destruída no estômago se administrada oralmente. Embora a insulina administrada subcutaneamente seja tão boa quanto à insulina produzida pelo pâncreas, ela é mais difícil de ser regulada. O pâncreas normal sente o aumento da glicose no sangue depois de uma refeição, e imediatamente, ajusta o suprimento de insulina. A insulina injetada, porém, é absorvida pelo sangue independentemente das quantidades de glicose presentes. Existem alguns aparelhos, como o tipo “caneta”, que fazem a administração de insulina ser mais fácil e conveniente.

3.5 Resumo do capítulo

Este capítulo apresentou a doença Diabetes, descrevendo os diversos tipos da doença e quais as principais características que levam a esta doença estar presente em todo o mundo. No decorrer do capítulo foram apresentados os principais sintomas quando um paciente está com a doença, assim como os mecanismos de detecção. Abordou alguns tratamentos utilizados para prevenção e controle da doença.

O próximo capítulo apresenta a tecnologia para aquisição dos dados que serão estudados nos demais capítulos.

Capítulo 4

Sistema de Aquisição e Pré-processamento

Neste capítulo é apresentado o sistema de aquisição do protótipo de nariz artificial utilizado, assim como os procedimentos de pré-processamento de sinais de odores. Na seção 4.1 são discutidos os narizes artificiais. Na seção 4.2 é apresentada a descrição do protótipo usado na aquisição dos sinais de odores. Na seção 4.3 descreve o pré-processamento dos sinais de odores. Na seção 4.4 é apresentado um resumo do capítulo.

4.1 Introdução

Como já comentado anteriormente no capítulo 2 seção 2.2, os primeiros sistemas de narizes eletrônicos experimentais apareceram nos anos oitenta. De fato, levaram várias décadas investigando e aperfeiçoando este tipo de instrumento, porém ainda estamos longe do perfeito sistema olfativo humano. Atualmente são projetados narizes artificiais genéricos que possa ajudar um operário em suas atividades em que seja necessário utilizar a olfação humana.

No desenvolvimento de um instrumento, como um nariz artificial, várias tecnologias são necessárias, em especial o tipo de sensor utilizado, que contém características próprias tanto de aplicações como principalmente suas características de construção.

Diante de tantas tecnologias existentes, algumas são de caráter genérico, não foram desenhadas especificamente para uma única aplicação. Nosso trabalho envolve a utiliza-

ção de um protótipo experimental destinado à aquisição de dados de substâncias voláteis, em especial, a urina humana visando o prognóstico de diabetes.

4.2 O protótipo do sistema de aquisição

4.2.1 Descrição geral

Um nariz eletrônico é um equipamento de medida formado por uma série de módulos que trabalham de forma conjunta, cujo objetivo é quantificar e/ou classificar amostras no estado gasoso. O protótipo de aquisição de dados construído neste trabalho pode ser distinguido em três módulos básicos, um de amostra, um de sensores e um sistema computacional, sendo este último dividido em aquisição e processamento. Este protótipo utilizado segue o modelo proposto por [Santos, 2000].

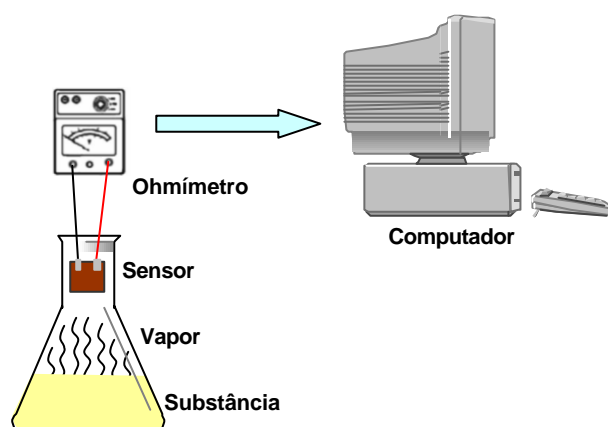


Figura 4.1 – Diagrama do protótipo de aquisição de dados [Santos, 2000]

Na Figura 4.1 é apresentado o esquema geral do protótipo proposto por [Santos, 2000]. Nele podemos ver os diferentes elementos que formam o equipamento de aquisição de dados, ou seja, uma câmara contendo a substância a ser analisada e o sensor de aroma, o leitor

de característica do sensor de aroma e o computador que armazena e processa os sinais elétricos que provém dos sensores de aroma.

Na Figura 4.2 pode ser vista uma fotografia do protótipo medindo os sinais de odores provindos da urina de um paciente analisado. O desenvolvimento do protótipo desse trabalho contou com o apoio da empresa CRIATRONICS, assim como a parceria do Laboratório de Análises Clínicas Pasm com a disponibilização do seu espaço físico e as amostras de urinas.

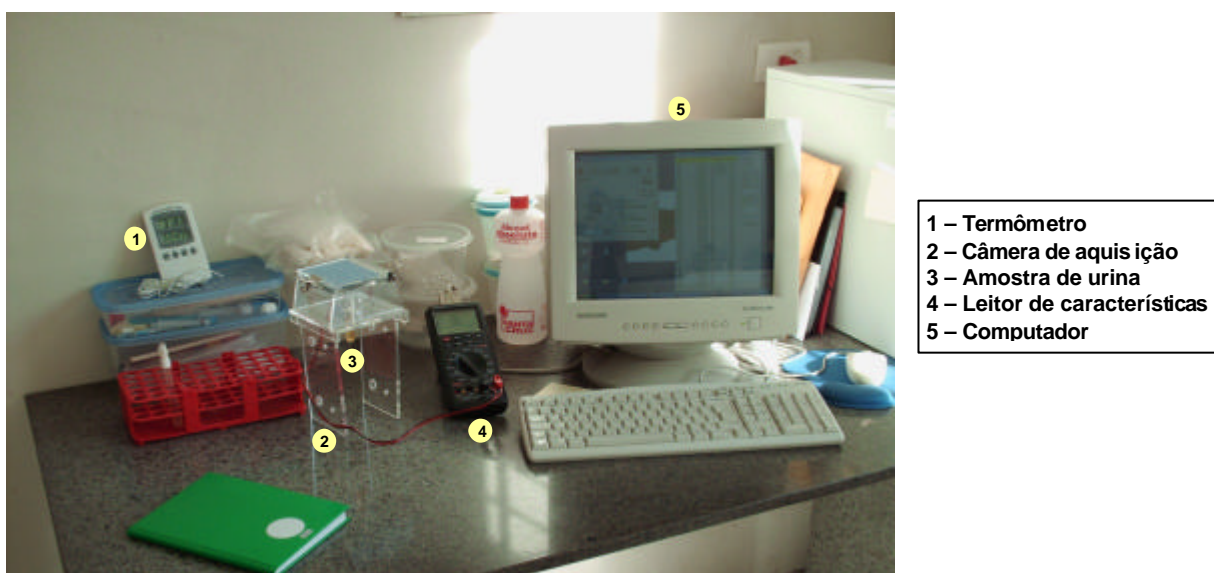


Figura 4.2 – Fotografia do protótipo de aquisição de dados medindo os sinais de odores provenientes da reação dos sensores de aroma com o odor da urina de um paciente.

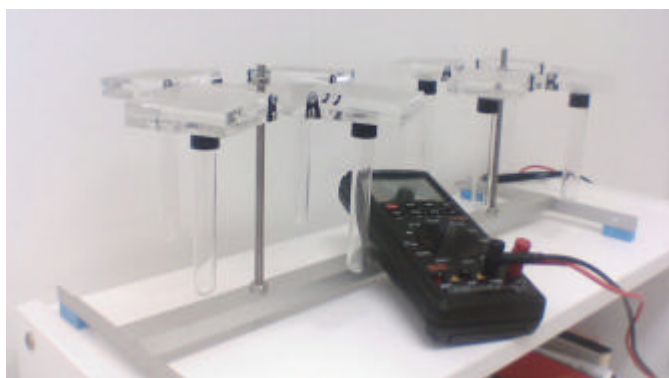
O processo de medida com o protótipo começa com a leitura do sensor de aroma durante um período de tempo para estabilizar o sensor, depois a introdução da amostra de urina no tubo de ensaio da câmara de aquisição, onde permanece um período de tempo necessário para a leitura da reação do sensor de aroma com o odor da urina. Após a leitura da reação lêem-se os sinais dos sensores de aroma referentes à dessorção das moléculas do odor da urina. Esse processo será discutido na seção 4.4.

4.2.2 Câmara de aquisição

Do ponto de vista construtivo, a câmara de aquisição é um recipiente hermético de acrílico, que proporciona a inserção de sensores de aroma e tubo de ensaio de tamanhos variados. Em sua parte frontal dispõe de uma tampa que permite a inserção dos sensores e os isola do contato com o ar. Dispõe também de um orifício que pode ser atravessado pelo tubo de ensaio e permite que as amostras no tubo de ensaio entrem em contato com o sensor. O tubo de ensaio é removível, permitindo sua troca após a análise de uma urina por outro contendo uma nova amostra. Os sensores são colocados sobre contatos que conduzem as variações das características elétricas do sensor para um aparato eletrônico responsável pela leitura dos sinais e que os envia ao computador. Como recurso auxiliar também é utilizado um medidor de temperatura do ambiente, evitando que as mudanças bruscas de temperatura interfiram no experimento. Na Figura 4.3 (a) pode ser vista uma fotografia da câmara de aquisição para uma amostra, enquanto na Figura 4.3 (b) pode ser vista uma câmara para várias amostras com uma matriz de sensores.



(a)



(b)

Figura 4.3 – (a) Câmara de aquisição de sinais de odores provenientes de uma amostra e (b) de várias amostras

4.2.3 Sensores utilizados

Os sensores de aroma utilizados foram produzidos pela empresa CRIATRONICS. As características elétricas da reação dos sensores de aroma com o odor da urina podem ser adquiridas através das características elétricas do sensor, ou seja, capacitância, corrente elétrica, tensão elétrica ou resistência elétrica, sendo esta última a adotada. Além do mais, não requerem uma complexa instrumentação de medição e podem ser utilizados em tempo real.

Em geral, sensores de aroma, depois de submetidos às substâncias voláteis, perdem a sensibilidade na mudança de características elétricas da reação com tais substâncias, também chamado de saturação. Para evitar a saturação é necessário um processo de purga, ou seja, a descontaminação dos sensores feitas através de um jato de ar, preferivelmente puro, assim como a utilização de uma camada de filtragem de impurezas denominada *septum*. [D'Emery et al. 2005a, 2005b]. Nos experimentos realizados, os sensores expostos ao odor da urina foram descartados em cada experimento.

Os sensores utilizados são semicondutores, pois seu funcionamento se baseia na variação de resistência que apresentam em sua capa ativa na presença de diferentes compostos voláteis. Na literatura este tipo de sensor é classificado como sensores resistivos. [Moseley e Prudenziati, 1994; Moseley e Tofield, 1987; Clifford e Tuma, 1983]

Foram utilizados quatro tipos de sensores, divididos em duas classes:

Classe 1: S01BK3D3

Classe 2: S02BB3D3, S02BR3D3, S02BG3D3

A nomenclatura dos sensores está relacionada às características dos sensores que podem ser vistas na Figura 4.4.

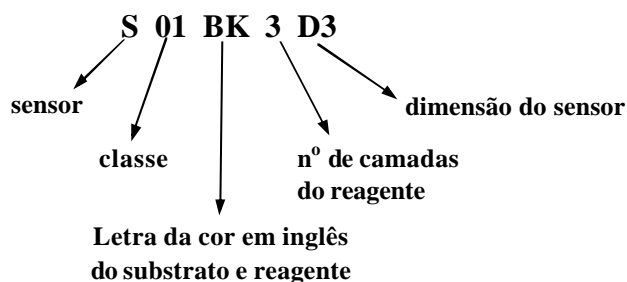


Figura 4.4 – Nomenclatura dos sensores utilizados, onde S é referente a sensor, 01 a classe tipo I, BK é referente à cor do substrato e reagente do sensor, nesse caso *black* (BB é *black* e *blue*, BR é *black* e *red*, e BG é *black* e *green*), 3 é número de camadas do sensor, e D3 é o tamanho do sensor nesse caso 3 cm x 3 cm.

4.2.4 Leitor de características

Como comentado na seção 4.2.1, o módulo de leitura do sinal é composto por uma interface de aquisição e um computador convenientemente equipado e controlado por dois programas de aquisição.

O primeiro programa é responsável pela aquisição dos dados enviados pelo instrumento de leitura da resistividade dos sensores, além de controlar a porta de comunicação de transferência de dados feitas a partir da interface RS-232 da porta serial do computador, realizando uma amostra periódica das características elétricas do sensor de aroma. Ao finalizar o processo de medição, o conjunto de dados é armazenado em um arquivo de dados no computador contendo a evolução temporal da característica elétrica medida.

Posteriormente, uma vez realizadas as medidas pertinentes, um segundo programa extrai as características importantes e aplica algoritmos de pré-processamento de sinais e reconhecimento de padrões. Os algoritmos de reconhecimento de padrões são: a Análise de Componentes Principais e as Redes Neurais Artificiais a serem descritas no capítulo 5.

4.2.5 A aquisição dos dados

O processo de aquisição dos dados foi realizado em parceria com o Laboratório de Análises Clínicas Pasm. Vários tipos de sensores podem ser utilizados no protótipo de aquisição de dados do nariz artificial, porém foram utilizados os sensores descritos na seção 4.2.3. Cada sensor utilizado no protótipo é responsável por uma resposta específica a cada odor analisado. Ou seja, toda substância apresentada ao protótipo produz nos sensores alterações físico-químicas que os caracterizam.

A base de dados de odores, utilizada nesta dissertação, foi obtida a partir da exposição dos sensores do protótipo do nariz artificial a urinas cedidas pelo Laboratório de Análises Clínicas Pasm de pacientes diabéticos e não-diabéticos. A partir dessa exposição foram armazenados os sinais característicos provenientes dos sensores submetidos ao odor da urina.

A aquisição dos padrões de odores no nariz artificial envolveu três etapas distintas:

- 1ª etapa:** Coleta da linha de base, ou seja, a aquisição da característica do sensor durante um período de 180 segundos, onde os sensores são expostos ao ambiente visando estabilização;
- 2ª etapa:** Coleta das variações das características elétricas (resistência) dos sensores de aroma, frente às urinas dos pacientes diabéticos e não-diabéticos durante um período de 600 segundos; e
- 3ª etapa:** Coleta da resistência elétrica do sensor após a exposição às urinas, ou seja, etapa de dessorção das moléculas de odores da urina nos sensores de aroma durante um período de 180 segundos.

O processo de aquisição é de aproximadamente 16 minutos para cada amostra de urina analisada. Para a análise foram utilizadas amostras com um volume de 3 mL de urina por paciente.

Conforme descrito na seção 4.3.2 os sensores utilizados na aquisição dos dados são denominados pelas suas características de fabricação. Na análise, os sensores utilizados estão apresentados na Tabela 4.1.

Tabela 4.1 – Descrição dos sensores utilizados no protótipo de nariz artificial.

| Sensor | Nome do Sensor |
|--------|----------------|
| 01 | S01B3D3 |
| 02 | S02BB3D3 |
| 03 | S02BG3D3 |
| 04 | S02BR3D3 |

A partir dos sensores apresentados na Tabela 4.1, os sinais de odores foram obtidos com a exposição dos quatro sensores a uma amostra de urina. Em todas as repetições, tanto para a leitura de característica do sensor quanto para os dois tipos de amostras de urina, provenientes de pacientes diabéticos e não-diabéticos, o valor da resistência elétrica dos sensores foi registrado pelo leitor de característica a cada 1 segundo. Com o processo da aquisição dos padrões de cada sensor estendeu-se por 18 minutos, cada sensor obteve o registro de 960 valores de resistência, para cada odor apresentado ao protótipo.

Porém, o estudo é realizado nas etapas distintas da aquisição, ou seja, a primeira etapa referente aos sinais característicos dos sensores (linha de base) e a segunda etapa, referente à exposição dos sensores as amostras de urina. A terceira etapa não foi avaliada neste trabalho.

O conjunto formado pelos quatro valores, de cada um dos sensores num instante de tempo, foi considerada uma instância da base de dados. Desta forma para a primeira etapa cada repetição contém 720 padrões (4 sensores com 180 segundos de aquisição) e para a segunda etapa 2.400 padrões (4 sensores com 600 segundo de aquisição).

Foram avaliados 77 pacientes, sendo 18 pacientes diabéticos e 59 não-diabéticos, com suas amostras de urinas submetidas aos quatro sensores. Logo a base de dados é formada por

um total 240.240 padrões, sendo 55.440 padrões da primeira etapa e 184.800 padrões da segunda etapa.

A Tabela 4.2 apresenta um exemplo de uma das repetições de aquisição dos dados para a primeira etapa de aquisição, e a Tabela 4.3 uma das repetições para a segunda etapa.

Tabela 4.2 – Exemplo de aquisição dos dados da coleta de linha de base dos sensores de aroma.

| Tempo (s) | Variação da resistência elétrica dos sensores | | | |
|--------------|---|----------|----------|----------|
| | S01B3D3 | S02BB3D3 | S02BR3D3 | S02BG3D3 |
| 0 | 19,22 | 34,00 | 31,40 | 32,80 |
| 1 | 19,24 | 34,00 | 31,40 | 33,00 |
| 2 | 19,24 | 34,00 | 31,40 | 33,00 |
| 3 | 19,24 | 34,00 | 31,40 | 32,80 |
| ... | ... | ... | ... | ... |
| 90 | 19,36 | 34,20 | 31,40 | 32,80 |
| 91 | 19,38 | 34,00 | 31,40 | 32,80 |
| 92 | 19,38 | 34,00 | 31,40 | 32,80 |
| 93 | 19,38 | 34,20 | 31,40 | 32,80 |
| ... | ... | ... | ... | ... |
| 177 | 19,60 | 34,20 | 31,40 | 32,60 |
| 178 | 19,60 | 34,20 | 31,60 | 32,80 |
| 179 | 19,60 | 34,20 | 31,40 | 32,60 |

Tabela 4.3 – Exemplo de aquisição dos dados da exposição de uma amostra de urina de um paciente diabético aos sensores de aroma.

| Tempo (s) | Variação da resistência elétrica dos sensores | | | |
|--------------|---|----------|----------|----------|
| | S01B3D3 | S02BB3D3 | S02BR3D3 | S02BG3D3 |
| 0 | 19,70 | 34,20 | 31,40 | 32,60 |
| 1 | 19,68 | 34,40 | 31,60 | 32,80 |
| 2 | 19,68 | 34,40 | 31,60 | 32,80 |
| 3 | 19,70 | 34,20 | 31,40 | 32,80 |
| ... | ... | ... | ... | ... |
| 300 | 20,28 | 35,80 | 32,60 | 34,00 |
| 301 | 20,28 | 35,80 | 32,60 | 34,00 |
| 302 | 20,28 | 35,80 | 32,60 | 34,00 |
| 303 | 20,28 | 35,80 | 32,60 | 34,00 |
| ... | ... | ... | ... | ... |
| 597 | 20,72 | 37,00 | 33,60 | 35,20 |
| 598 | 20,72 | 37,00 | 33,60 | 35,00 |
| 599 | 20,72 | 37,00 | 33,60 | 35,20 |

Como o processo de aquisição dos dados é realizado de forma discretizada no tempo, cada sensor gera uma curva temporal. A Figura 4.5 apresenta um exemplo da evolução temporal dos valores da resistência, para os quatro sensores, durante a fase de aquisição, sendo a Figura 4.5a para os valores dos sinais resistivos característicos dos sensores e Figura 4.5b para a reação com as amostras analisadas. Pode-se perceber a evolução temporal dos valores medidos no determinado espaço de tempo, porém, maiores análises seriam necessárias para verificar as características temporais desta série.

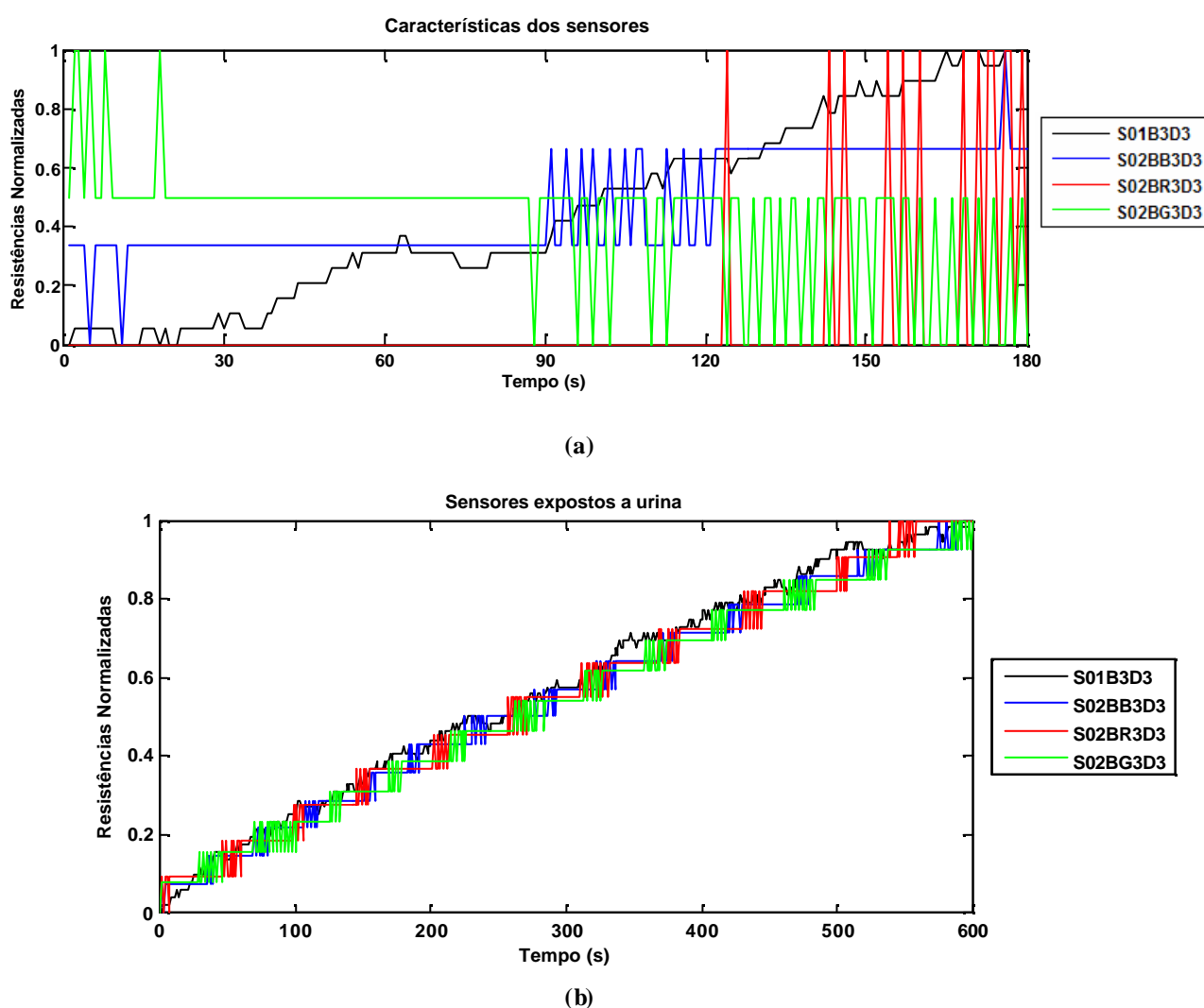


Figura 4.5 – Curva temporal formada pelos valores dos sinais resistivos dos sensores sendo (a) na coleta de linha de base e (b) exposição de uma amostra de urina de um paciente diabético aos sensores de aroma.

4.3 Preparação dos dados

Esta fase, conhecida como *pré-processamento de dados*, é um passo prévio das aplicações de técnicas de reconhecimento de padrões e têm como finalidade a análise dos dados, extração de descritores, normalização e remoção ou diminuição de erros na precisão das análises posteriores a que serão submetidos os sinais de odores. Esta etapa realiza operações sobre os sinais provenientes dos sensores de aroma, a fim de corrigir distorções existentes nos dados ou melhorar o desempenho de algoritmos de classificação que serão utilizados.

Em geral, os métodos de pré-processamento de dados se caracterizam pelo resultado obtido em sua aplicação. No entanto, um problema comum em análise e aquisição de dados é a presença de pontos aberrantes. Entende-se por ponto aberrante (*outlier*) a uma observação que apresenta um comportamento atípico em relação ao restante dos dados, não seguindo a distribuição da maioria deles. Quando se tem uma única variável, o valor aberrante caracteriza-se por assumir um valor mais alto ou mais baixo que os demais.

Sendo assim, na escolha de um método para o tratamento dos sinais de odores, algumas características inerentes ao processo de aquisição dos dados devem ser consideradas:

- (i) os sensores utilizados podem apresentar distúrbios ou ruídos, devido tanto a fabricação dos sensores quanto a erro de medidas;
- (ii) erro de transcrição ou digitação;
- (iii) o processo de aquisição dos dados tendem a ser instáveis, ou seja, os sensores apresentam variações durante o processo de aquisição dos dados, aos quais podem se transformar em *outliers* na base de dados;
- (iv) outros odores no ambiente podem causar interferências nos sinais dos sensores do sistema de aquisição de dados;

- (v) erro ao considerar uma unidade amostral que não pertence à população de interesse, por exemplo, na submissão da urina de pacientes aos sensores de aroma, observações de pacientes não diabéticos, erroneamente incluídos na base de dados de diabéticos ou por erro de diagnóstico, podem resultar em valores aberrantes;
- (vi) variabilidade natural dos dados

Com exceção do item (vi), todas as demais causas estão relacionadas a erros e, conseqüentemente, a identificação de valores aberrantes nessas situações exige sua retirada da amostra final. No apêndice A será abordado sobre valores aberrantes multidimensionais e métodos de identificação.

Outra questão a considerar no pré-processamento é quanto à normalização dos padrões, o ideal é que os dados estejam compreendidos entre intervalos bem definidos. A normalização se faz necessária para que valores esparsos não interfiram no processo de aprendizado e convergência do algoritmo utilizado no reconhecimento dos sinais de odores. As características do classificador utilizado também podem influenciar na escolha do método de pré-processamento. Em certos algoritmos não-paramétricos é assumido implicitamente que as distâncias em diferentes direções do espaço de entrada têm o mesmo peso, e variáveis com grandes valores numéricos podem dominar os efeitos de variáveis com valores menores mas de igual modo importantes na definição do modelo, vejamos alguns: algoritmos da vizinhança-mais-próxima, algoritmo de propagação retroativa, redes neurais, redes *Kohonen*, algoritmos de *clustering*.

Devidas às propriedades dos sensores, muitas vezes faz-se necessária a redução dos dados a uma mesma escala, abordada no capítulo 5, seção 5.2.3. As técnicas de normalização podem ser agrupadas em duas classes: métodos locais e globais. Os métodos locais operam através do conjunto de sensores sobre cada cheirada a fim de compensar as variações amostra por amostra causadas pela concentração da substância analisada e a reação com os

sensores, entre outras causas. Por outro lado, os métodos globais, operam através de toda a base de dados dos sinais de um único sensor, por exemplo, o histórico completo de cada sensor, e são geralmente empregados para compensar as diferenças na escala de cada sensor. A seguir, denotaremos por $x_s^{(p)}$ a resposta de cada sensor ‘s’ para o p -ésimo exemplo na base de dados. [Pearce et al., 2003]

4.3.1 Métodos locais

O método local mais utilizado é a normalização vetorial, em que a característica do vetor de cada cheirada individual é dividida pela sua norma [Pearce et al., 2003] e, como resultado, é forçado a encontrar-se sobre uma hipersfera de raio unitário, como mostrado na Figura 4.6 d, e.

$$Y_s^{(p)} = \frac{x_s^{(p)}}{\sqrt{\sum_s (x_s^{(p)})^2}} \quad (4.1)$$

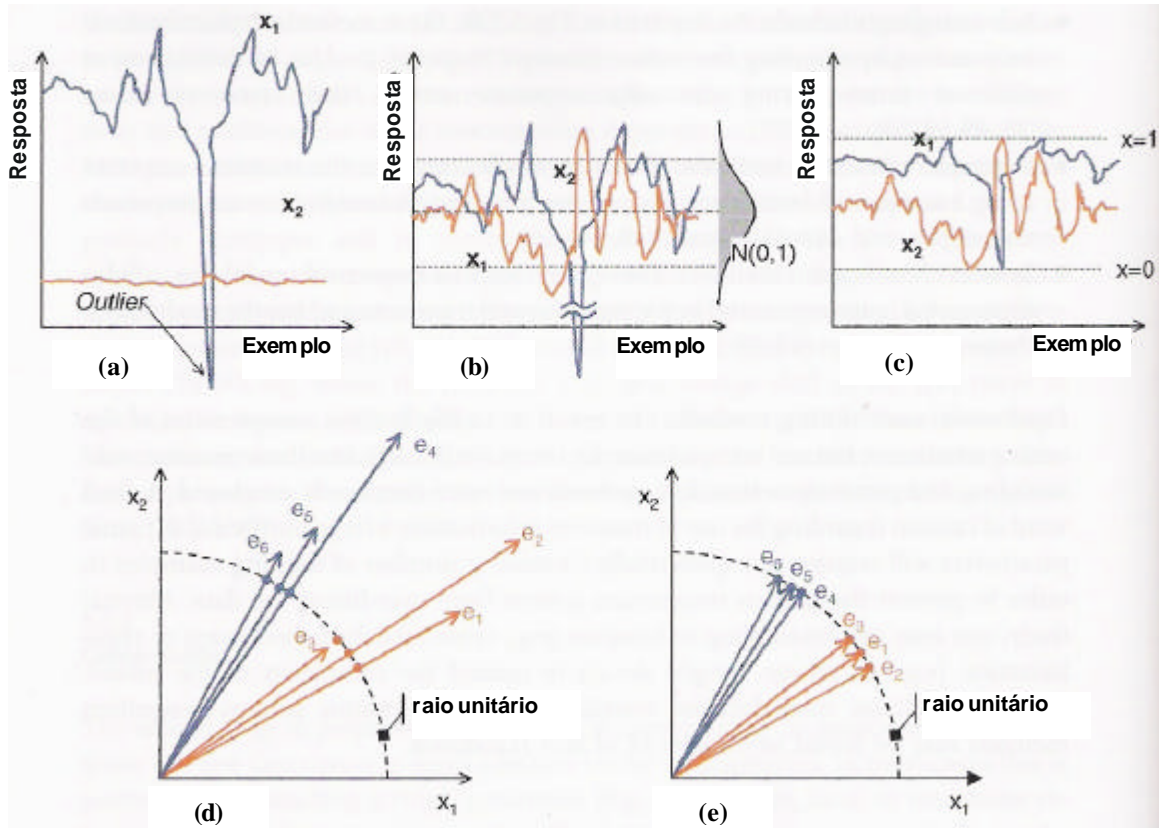


Figura 4.6 – Procedimentos da normalização para a resposta do sinal de um sensor: (a, d) dados originais, (b) auto-escalonamento do sinal do sensor, (c) normalização do sinal do sensor e (e) normalização do vetor.

Quando conhecidas as concentrações das amostras, o vetor de normalização pode ser empregado para compensar as diferenças entre tais concentrações. Assumindo a força da relação $x_{s,a}^{(p)} = \mathbf{a}_{s,a} [C_a^{(p)}]^b$ de sensores de aroma do tipo *metal-oxide chemoresistors* [Di Natale et al., 1995b], onde $x_{s,a}^{(p)}$ é a resposta de cada sensor 's' para a *p*-ésima amostra de odor 'a', $\mathbf{a}_{s,a}$ é a sensibilidade do sensor 's' ao odor 'a', e $[C_a^{(p)}]$ é a concentração da *p*-ésima amostra do odor 'a', então:

$$Y_{s,a}^{(p)} = \frac{x_{s,a}^{(p)}}{\sqrt{\sum_s (x_{s,a}^{(p)})^2}} = \frac{\mathbf{a}_{s,a} [C_a^{(p)}]^b}{\sqrt{\sum_s (\mathbf{a}_{s,a} [C_a^{(p)}]^b)^2}} = \frac{\mathbf{a}_{s,a}}{\sqrt{\sum_s (\mathbf{a}_{s,a})^2}} \quad (4.2)$$

Para a extensão que essas suposições simplificadas abrangem, o vetor normalização pode ser usado para compensar as variações de concentração de cada amostra. Neste contexto, o vetor normalização pode ser aplicado em situações onde cada odor tem uma única concentração, mas a discriminação deve ser aplicada sobre a qualidade da base de dados dos sinais de odor (por exemplo, a direção do vetor resposta $\vec{x}^{(p)}$) melhor que a intensidade do odor (por exemplo, a magnitude de $\vec{x}^{(p)}$). Entretanto este método não deveria ser usado quando o vetor amplitude carrega informações relevantes.

4.3.2 Métodos globais

Alguns procedimentos globais são comumente empregados nos dados de narizes artificiais:

- i. Sensor autoescalonável, em que a distribuição dos valores de cada sensor através de toda a base de dados é ajustado para ter média zero e desvio padrão um:

$$Y_s^p = \frac{x_s^p - \bar{x}_s}{\mathbf{s}_s} \quad (4.3)$$

- ii. Sensor normalizado, em que a escala dos valores de cada sensor 's' é ajustado para um novo intervalo do sensor normalizado s':

$$Y_s^p = \frac{x_s^p - \min_{\forall p} [x_s^p]}{\max_{\forall p} [x_s^p] - \min_{\forall p} [x_s^p]} * \left(\max_{\forall p} [x_{s'}^p] - \min_{\forall p} [x_{s'}^p] \right) + \min_{\forall p} [x_{s'}^p] \quad (4.4)$$

Se a normalização for realizada para o intervalo [0,1] a equação 4.4 será simplificada para:

$$Y_s^p = \frac{x_s^p - \min_{\forall p} [x_s^p]}{\max_{\forall p} [x_s^p] - \min_{\forall p} [x_s^p]} \quad (4.5)$$

- iii. Normalização para respostas de sensores não-lineares num intervalo $[1, 1]$ usando uma *função sigmoïdal*. A fórmula aplicada por este tipo de normalização é a seguinte:

$$Y_s^p = \frac{1 - e^{-\mathbf{b}}}{1 + e^{-\mathbf{b}}} \quad (4.6)$$

onde $\mathbf{b} = \frac{x_s^p - \bar{x}_s}{\mathbf{s}_s}$. A normalização sigmoïdal é apropriada quando se pretende:

incluir pontos isolados no conjunto de dados a analisar e, evitar que os valores mais comuns sejam comprimidos, sem contudo perder a habilidade de representar valores muito isolados.

Métodos globais são tipicamente usados para assegurar que a magnitude dos sensores é comparável, impedindo conseqüentemente procedimentos de reconhecimento de padrões de ser oprimido por sensores com valores arbitrariamente grandes. Por exemplo, procedimentos vizinhos-mais-próximos são extremamente sensíveis a características com pesos maiores, redes neurais do tipo *perceptrons* multicamadas podem saturar suas funções de ativação sigmoïdal por grandes entradas. Sensores normalizados fazem uso de entradas dinâmicas grandes, porém, como ilustrado na Figura 4.6 a, c, é muito sensível a *outliers* desde que a escala é determinada pelos *outliers* da base de dados. Por outro lado, auto-escalamento, não fornecem limites reduzidos. Entretanto, isto deve ser notado que ambas as técnicas podem amplificar ruídos de todos os sensores (particularmente aqueles que não carregam informação) que possui pesos relativamente proporcionais.

Métricas logarítmicas também podem ser usadas para compensar altos efeitos de concentração não-linear [Abe et al., 1987]. Isto também é conhecido como *transformação*

Box-Cox [Box e Cox, 1964], na qual poderia ser empregada para compensar não-linearidades, assim como comprimir as escalas dos sensores:

$$Y_s^p = \begin{cases} \frac{(x_s^p)^I - 1}{I}, & I \neq 0 \\ \ln(x_s^p), & I = 0 \end{cases} \quad (4.7)$$

Para o tratamento de sinais de odores em narizes artificiais, diversos modelos de normalização já foram propostos, dentre os quais outros métodos podem ser encontrados em [Gardner e Bartlett, 1994].

4.4 Resumo do capítulo

Este capítulo apresentou uma descrição do protótipo de nariz artificial para aquisição de dados utilizados para construção da base dados a partir de sua exposição à urina de alguns pacientes com a doença descrita no capítulo anterior, ou seja, o Diabetes. Foram discutidas as características dos sinais provenientes dos sensores utilizados no protótipo do nariz artificial antes e após expostos a urina dos pacientes. No decorrer do capítulo foram descritos os métodos de pré-processamento dos sinais de odores que serão utilizados nos sistemas de reconhecimento de padrões.

No próximo capítulo serão abordadas algumas técnicas multivariada estatísticas e neurais de reconhecimento de padrão dispostas na literatura, geralmente utilizadas no reconhecimento de padrões dos narizes artificiais.

Capítulo 5

Análise de padrões para Narizes Artificiais

Neste capítulo são apresentadas as técnicas utilizadas, nesta dissertação, como sistemas de reconhecimento de sinais de odores através da análise estatística e neural. Na seção 5.1 são apresentadas as técnicas utilizadas. Na seção 5.2 é apresentada a técnica de extração de característica Análise de Componentes Principais. Na seção 5.3 são apresentados classificadores não-paramétricos, como a técnica usada nesta dissertação: K-Nearest-Neighbour. Na seção 5.4 são apresentadas as abordagens conexionistas de reconhecimento de padrões, utilizadas, as técnicas são: Redes Neurais Multilayer Perceptron e Redes Neurais com Funções de Base Radial. Na seção 5.5 é apresentado um resumo do capítulo.

5.1 Introdução

Em geral, narizes eletrônicos em seus sistemas de reconhecimento de padrões fazem uso das mais diversas técnicas estatísticas clássicas, usando modelos de probabilidade, que foram inicialmente desenvolvidas e usadas no campo da matemática aplicada, chamada quimiometria na área de química. Outras abordagens envolvem a Inteligência computacional, dentre elas as Redes Neurais Artificiais (RNs). Nas próximas seções serão apresentados alguns métodos que podem ser aplicados para as análises multicomponentes de odor.

5.2 Análise de Componentes Principais

Introduzida por Karl Pearson em 1901, a técnica denominada de componentes principais, popularmente chamada de PCA, do inglês *principal components analysis*, está fundamentada no artigo de [Hotelling, 1933].

PCA é um método estatístico multivariado, não-paramétrico, não-supervisionado e linear, baseado na transformação *Karhunen-Loève*, usado para modelos de classificação que produzem resultados qualitativos em narizes artificiais no reconhecimento de padrão [Pearce, et al. 2003].

PCA tem como objetivo principal explicar a estrutura de variância e covariância de um vetor aleatório, composto de p -variáveis aleatórias, através da construção de combinações lineares das variáveis originais. As componentes principais (PCs, do inglês *principal component*) são estas combinações lineares e que não são correlacionadas entre si. Se tivermos p -variáveis originais que representam um dado então será possível obter p componentes principais. No entanto, o objetivo é obter a redução do número de variáveis a serem avaliadas e a interpretação das combinações lineares construídas, ou seja, a informação contida nas p -variáveis originais é substituída pela informação contida em k componentes principais não correlacionadas, sendo $k < p$. Sendo assim, a variabilidade do vetor aleatório composto das p -variáveis original é aproximado pela variabilidade do vetor aleatório que deverá conter as k componentes principais. Desta forma, a qualidade da aproximação dependerá do número de componentes mantidas no sistema e pode ser medida através da avaliação da proporção de variância total explicadas por essas componentes escolhidas.

PCA é uma forma de identificar padrões em dados, e de expressá-los de tal forma a destacar suas similaridades e diferenças. Devido à alta dimensionalidade dos dados, sua representação gráfica nem sempre é possível, daí PCA torna-se uma poderosa ferramenta para

analisar tais dados. Uma outra vantagem é que uma vez encontrados padrões nos dados, comprime-se os dados reduzindo o número de dimensões sem muita perda de informação.

Com o surgimento de novas tecnologias de aquisição de dados, dentre elas o nariz eletrônico, onde reduzir e extrair informações relevantes de um volumoso conjunto de dados obtidos e em que muitas vezes a informação essencial não depende de variáveis isoladas, mas sim da relação entre as mesmas. PCA aborda tais problemas e por isso é uma das técnicas mais utilizadas pelos cientistas da computação, químicos analíticos, por todos aqueles investigadores que trabalham com o nariz eletrônico já que permite reduzir, representar e extrair informação relevante ao mesmo tempo [Gardner, 1991].

PCA é um método que assume colinearidade entre as variáveis envolvidas. Em outras palavras, se trata de um algoritmo linear que pode funcionar incorretamente em processos não-lineares como podem ser as reações físico-químicas entre sensores e substâncias voláteis. No entanto, funcionam surpreendentemente bem em muitas aplicações com os narizes artificiais, em que as concentrações de voláteis não são muito elevadas e o comportamento dos sensores não é excessivamente não-linear.

A utilização em narizes artificiais limita-se a representar multidimensionalidade em conjunto de medidas para ver se podem determinar grupos (“*clusters*”) espontâneos entre as diferentes medidas realizadas previamente, ainda que existam aplicações em que o método possa ser supervisionado [Llecha, 2002].

Assumindo uma distribuição de probabilidade do vetor aleatório é normal p-variada, as componentes principais, além de não correlacionadas, são independentes [Mingoti, 2005]. Entretanto, a suposição de normalidade não é requisito para que a técnica possa ser utilizada, fator que viabiliza sua aplicação no estudo dos sinais de odores de narizes artificiais que nem sempre possuem seus dados seguindo uma distribuição normal.

A obtenção das componentes principais envolve a decomposição da matriz de covariâncias do vetor aleatório de interesse. Caso seja utilizada qualquer transformação do vetor aleatório, então as componentes serão determinadas pela matriz de covariância relativa ao vetor transformado. Algumas transformações podem ser vistas no capítulo 4, seção 4.3.

Uma vez determinadas as componentes principais, os seus valores numéricos, denominados de escores, podem ser calculados para cada elemento amostral. Deste modo, os valores de cada componente podem ser analisados, fazendo uso de técnicas estatísticas como análise de variância e análise de regressão, dentre outras.

Na literatura são muitos os exemplos que utilizam técnicas de PCA. A utilização de PCA em sensores e narizes artificiais pode ser encontrada em [Llecha, 2001; Llobet et al., 1999; Di Natale et al., 1995a; Gardner e Bartlett, 1992; e Gardner, 1991].

5.2.1 Componentes principais exatas extraídas da matriz de covariâncias

Seja $X = (X_1, X_2, \dots, X_p)'$ um vetor aleatório com vetor de médias $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p)'$ e matriz de covariância $\sum_{p \times p}$. Sejam $\mathbf{I}_1 \geq \mathbf{I}_2 \geq \dots \geq \mathbf{I}_p$ os autovalores da matriz $\sum_{p \times p}$, com os respectivos autovetores normalizados e_1, e_2, \dots, e_p , isto é, os autovetores e_i satisfazem as seguintes condições:

$$(vii) \quad e_i' e_j = 0 \text{ para todo } i \neq j;$$

$$(viii) \quad e_i' e_i = 1 \text{ para todo } i = 1, 2, \dots, p;$$

$$(ix) \quad \sum_{p \times p} e_i = \mathbf{I}_i e_i, \text{ para todo } i = 1, 2, \dots, p;$$

sendo o autovetor e_i denotado por $e_i = (e_{i1}, e_{i2}, \dots, e_{ip})'$. Considere o vetor aleatório $Y = O'X$, onde $O_{p \times p}$ é a matriz ortogonal de dimensão $p \times p$, constituída dos autovetores normalizados da matriz $\Sigma_{p \times p}$, isto é,

$$O_{p \times p} = \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{p1} \\ e_{21} & e_{22} & \cdots & e_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ e_{1p} & e_{2p} & \cdots & e_{pp} \end{bmatrix} = [e_1 \ e_2 \ \cdots \ e_p] \quad (5.1)$$

O vetor Y é composto de p combinações lineares das variáveis aleatórias do vetor X , tem vetor de médias igual $O'm$ e a matriz de covariância $\Lambda_{p \times p}$, que é uma matriz diagonal, cujos elementos são iguais a $a_{ii} = \lambda_i$, $i = 1, 2, \dots, p$ isto é,

$$\Lambda_{p \times p} = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \lambda_2 & \\ \mathbf{0} & & \lambda_p \end{bmatrix}$$

Portanto, as variáveis aleatórias que constituem o vetor Y são não correlacionadas entre si. Deste modo, surge a idéia de utilizar as combinações lineares em Y , como uma forma alternativa de se representar a estrutura de covariâncias do vetor X , tentando obter uma redução do espaço de variáveis, passando da dimensão p , para uma dimensão $k < p$. Portanto, ao invés de se utilizar o vetor aleatório original na análise de dados, utilizam-se as k combinações lineares principais. Os vetores aleatórios, X e Y , têm a mesma variância total e a mesma variância generalizada, sendo que o vetor Y tem a vantagem de ser composto por variáveis aleatórias não correlacionadas, facilitando na interpretação conjunta dessas variáveis.

Algumas definições para o método:

Definição 1: A j -ésima componente principal da matriz $\Sigma_{p \times p}$, $j = 1, 2, \dots, p$ é definida como:

$$Y_j = e_j' X = e_{j1}X_1 + e_{j2}X_2 + \dots + e_{jp}X_p \quad (5.2)$$

A esperança e a variância da componente Y_j são, respectivamente, iguais a:

$$E[Y_j] = e_j' \mathbf{m} = e_{j1} \mathbf{m}_1 + e_{j2} \mathbf{m}_2 + \dots + e_{jp} \mathbf{m}_p$$

$$\text{Var}[Y_j] = e_j' \sum_{pxp} e_j = \mathbf{I}_j$$

Sendo $\text{Cov}[Y_j, Y_k] = 0$, $j \neq k$. Cada autovalor \mathbf{I}_i representa a variância de uma componente principal Y_j . Como os autovalores estão ordenados em ordem decrescente, a primeira componente é a de maior variabilidade e a p -ésima é a de menor.

Definição 2: A proporção da variância total de X que é explicada pela j -ésima componente principal é definida por:

$$\frac{\text{Var}[Y_j]}{\text{VariânciaTotaldeX}} = \frac{\mathbf{I}_j}{\text{Traço}\left(\sum_{pxp}\right)} = \frac{\mathbf{I}_j}{\sum_{i=1}^p \mathbf{I}_i} \quad (5.3)$$

Pelo teorema de decomposição espectral [Mingoti, 2005], as variâncias total e generalizada do vetor aleatório X podem ser descritas através da variância total e variância generalizada do vetor aleatório Y uma vez que,

$$\text{traço}\left(\sum_{pxp}\right) = \sum_{i=1}^p \mathbf{s}_{ii} = \sum_{i=1}^p \mathbf{I}_i, \text{ onde } \mathbf{s}_{ii} = \text{Var}[X_i], i = 1, 2, \dots, p$$

$$\left|\sum_{pxp}\right| = \prod_{i=1}^p \mathbf{I}_i = |\Lambda_{pxp}|$$

Assim, em termos destas duas medidas globais de variação, os vetores X e Y são equivalentes. Em geral, a razão em (5.3) é multiplicada por 100, indicando o resultado em porcentagem. É evidente que a primeira componente principal tem a maior proporção de explicação da variância total de X .

Definição 3: A proporção da variância total que é explicada pela k primeiras componentes principais é definida como:

$$\frac{\sum_{j=1}^k \text{Var}[Y_j]}{\text{VariânciaTotaldeX}} = \frac{\sum_{j=1}^k \mathbf{I}_j}{\text{Traço}\left(\sum_{p \times p}\right)} = \frac{\sum_{j=1}^k \mathbf{I}_j}{\sum_{i=1}^p \mathbf{I}_i} \quad (5.4)$$

Se as k primeiras componentes principais explicam uma grande parte da variância total do vetor X , pode-se restringir o foco da atenção apenas ao vetor aleatório $(Y_1, Y_2, \dots, Y_k)'$. Desta forma, um conjunto de k -dimensional de variáveis aleatórias poderá ser examinado, ao invés de um conjunto p dimensional, sem que com isto se perca muita informação sobre a estrutura de variâncias e covariâncias original do vetor X . Pelo teorema da decomposição espectral, ao restringir-se o foco de atenção somente para as k primeiras componentes principais, a matriz de covariâncias $\sum_{p \times p}$ estará sendo aproximada pela fórmula em (5.5):

$$\sum_{p \times p} \approx \sum_{j=1}^k \mathbf{I}_j e_j e_j' \quad (5.5)$$

Cada parcela da soma em (5.5) envolve uma matriz de dimensão $p \times p$ correspondente apenas à informação da j -ésima componente principal, $j = 1, 2, \dots, k$. Sendo assim, a variabilidade original do vetor X estará sendo aproximada pela soma de k matrizes, cada uma representando o sistema de variabilidade relacionado a uma componente. Quando $k=p$, tem-se que a matriz de covariância $\sum_{p \times p}$ é reproduzida com exatidão pela soma de matrizes relacionadas às componentes principais, isto é,

$$\sum_{p \times p} = \sum_{j=1}^p \mathbf{I}_j e_j e_j'$$

Definição 4: Uma outra forma de se definir as componentes principais é apresentada a seguir. Considere o seguinte sistema de combinações lineares de X constituído de p equações do tipo:

$$Y_i = a_i' X = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p, \quad i = 1, 2, \dots, p$$

Deste modo tem-se que:

$$\text{Var}[Y_i] = a_i' \sum_{pxp} a_i$$

$$\text{Quando } \text{Cov}(Y_i, Y_j) = a_i' \sum_{pxp} a_j, \quad i \neq j, \quad i, j = 1, 2, \dots, p$$

Suponha que desejássemos encontrar os valores dos coeficientes a_{ij} tais que $a_i' a_i = 1$, de modo que as combinações lineares Y_1, Y_2, \dots, Y_p fossem não correlacionadas entre si e tivesse variância máxima. Então, pode ser demonstrado [Johnson e Wichern, 2002] que a variância máxima de $Y_i = a_i' X$, sob as restrições $a_i' a_i = 1$ e $\text{cov}(Y_i, Y_j) = 0 \quad j < i$, é obtida quando $a_i = e_i$, ou seja, o autovetor normalizado correspondente ao autovalor λ_i , sendo Y_i chamada de *i-ésima* componente principal. Deste modo, as p componentes principais são construídas, sendo únicas, exceto pela troca de sinal de todos os seus coeficientes. Senso assim, se Y_i é uma componente principal então $-Y_i$ também será uma componente principal de ordem i .

Devido a própria forma de construção, a primeira componente principal é sempre mais representativa em termos de variância total e a p -ésima é sempre de menor representatividade. A Figura 5.1 mostra graficamente as componentes principais no caso de duas variáveis, cada ponto no sistema de coordenadas X_1 e X_2 é projetado ortogonalmente no novo sistema de coordenadas Y_1 e Y_2 .

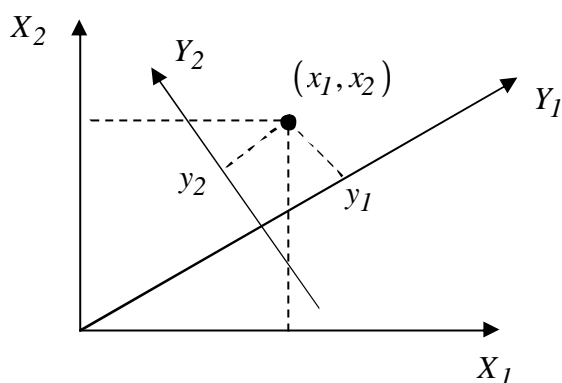


Figura 5.1 – Interpretação geométrica da Análise de Componentes Principais , $p = 2$

5.2.2 Estimação das componentes principais: matriz de covariâncias

A estimação das componentes principais está atrelada ao fato da matriz $\Sigma_{p \times p}$ ser desconhecida que será estimada através dos dados amostrais proveniente da coleta de dados, neste trabalho, provenientes de um experimento científico. A matriz $\Sigma_{p \times p}$ é estimada pela matriz de covariância amostral $S_{p \times p}$:

$$S_{p \times p} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix} \quad (5.6)$$

Sendo $S_{ij} \neq S_{ji}$, $j \neq i$ e S_{ii} definido por:

$$S_{ii} = \frac{\sum_{l=1}^n (X_{il} - \bar{X}_i)^2}{n-1} \quad (5.7)$$

que é a variância amostral da i -ésima variável,

$$S_{ij} = \frac{\sum_{l=1}^n (X_{il} - \bar{X}_i)(X_{jl} - \bar{X}_j)}{n-1} \quad (5.8)$$

sendo esta a covariância amostral entre a i -ésima e j -ésima variáveis.

Sejam $\hat{\mathbf{l}}_1, \hat{\mathbf{l}}_2, \dots, \hat{\mathbf{l}}_p$ e $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$ os autovalores matriz $S_{p \times p}$ e os correspondentes autovetores normalizados respectivamente, então a j -ésima componente principal amostral estimada é definida por:

$$\hat{Y}_j = \hat{e}_j' X = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + \dots + \hat{e}_{jp} X_p, j = 1, 2, \dots, p \quad (5.9)$$

Onde as componentes principais amostrais estimadas possuem as seguintes propriedades:

Propriedade 1: A variância estimada de \hat{Y}_j é igual a $\hat{\mathbf{l}}_j, j = 1, 2, \dots, p$

Propriedade 2: A covariância entre as componentes principais \hat{Y}_j e \hat{Y}_k é igual a zero, para todo $j \neq k$, ou seja, estas componentes são não correlacionadas.

Propriedade 3: A variância total explicada pela j -ésima componente amostral é dada por:

$$\frac{\text{Var}[\hat{Y}_j]}{\text{Variância Total Estimada de } X} = \frac{\hat{\mathbf{l}}_j}{\text{traço}(S_{p \times p})} = \frac{\hat{\mathbf{l}}_j}{\sum_{i=1}^p \hat{\mathbf{l}}_i}$$

Propriedade 4: A correlação estimada entre a j -ésima componente principal amostral e variável aleatória $X_i, i = 1, 2, \dots, p$ é dada por:

$$r_{\hat{Y}_j, X_i} = \frac{\hat{e}_{ji} \sqrt{\hat{\mathbf{l}}_j}}{\sqrt{s_{ii}}}$$

onde s_{ii} é a variância amostral da variável aleatória X_i .

Propriedade 5: Pelo teorema da decomposição espectral, a matriz de covariância S_{pxp} pode ser expressa como:

$$S_{pxp} = \sum_{j=1}^p \mathbf{1}_j \hat{e}_j \hat{e}_j'$$

Ou aproximada por (5.10), caso se use apenas as primeiras k componentes principais amostrais.

$$S_{pxp} \approx \sum_{j=1}^k \mathbf{1}_j \hat{e}_j \hat{e}_j' \quad (5.10)$$

Na prática, para fazer uso das k componentes principais amostrais consideradas mais relevantes na análise dos dados, é necessário calcular os seus valores números para cada elemento amostral, ou seja, os escores das componentes.

5.2.3 Análise de Componentes Principais via matriz de correlação

A estimação das componentes principais, obtidas a partir da matriz de covariâncias \sum_{pxp} são influenciadas pelas variáveis de maior variância, sendo de pouca utilidade nos casos em que existe uma discrepância muito acentuada entre essas variâncias. Tal discrepância pode ser causada pela diferença das unidades de medidas das variáveis, e que podem ser minimizadas se uma transformação for efetuada nos dados originais, de modo a equilibrar os valores de variância ou a colocar os dados na mesma escala de medida. Tomaremos nesta seção a transformação em que cada variável é padronizada pela sua média e desvio padrão, sendo aplicada à técnica de componentes principais aplicada à matriz de covariâncias das va-

riáveis padronizadas. Este procedimento equivale a obter as componentes principais através da matriz de correlação $P_{p \times p}$ das variáveis X_i originais, $i = 1, 2, \dots, p$.

Seja $Z_i = \frac{(X_i - \mathbf{m}_i)}{\mathbf{s}_i}$ onde $E(X_i) = \mathbf{m}_i$ e $Var(X_i) = \mathbf{s}_i^2$, $i = 1, 2, \dots, p$. A matriz

$P_{p \times p}$, é a matriz de covariância das variáveis Z_i . E se a técnica de componentes principais for aplicada à $P_{p \times p}$, as componentes principais construídas serão combinações lineares das variáveis X_i padronizadas.

Sejam os autovalores de $P_{p \times p}$, $\mathbf{I}_1 \geq \mathbf{I}_2 \geq \dots \geq \mathbf{I}_p$ e os correspondentes autovetores normalizados por $e_1 \geq e_2 \geq \dots \geq e_p$, onde $e_i = [e_{i1} \ e_{i2} \ \dots \ e_{ip}]'$, então a j -ésima componente principal de $P_{p \times p}$, $j = 1, 2, \dots, p$ é definida por:

$$Y_j = e_j' Z = e_{j1} Z_1 + e_{j2} Z_2 + \dots + e_{jp} Z_p \quad (5.11)$$

Onde as componentes principais possuem as seguintes propriedades:

Propriedade 1: A variância Y_j é igual a \mathbf{I}_j , $j = 1, 2, \dots, p$;

Propriedade 2: A covariância entre Y_j e Y_k é igual a zero, para qualquer $j \neq k$;

Propriedade 3: A correlação entre a componente Y_j e a variável padronizada Z_i é igual a:

$$r_{Y_j, Z_i} = r_{Y_j, X_i} = e_{ij} \sqrt{\mathbf{I}_j}$$

logo as variáveis Z_i com os maiores coeficientes na componente principal Y_j são mais correlacionadas com a componente;

Propriedade 4: A variância total do vetor aleatório $Z = (Z_1, Z_2, \dots, Z_p)'$ é igual ao traço de $P_{p \times p}$, que é igual ao valor de p , ou seja, o número de variáveis medidas em cada elemento

amostral. Sendo assim, a proporção da variável total explicada pela j -ésima componente principal é igual a $\frac{I_j}{p}$, $j = 1, 2, \dots, p$.

Os componentes principais obtidos da decomposição de P_{pxp} não são numericamente iguais aos coeficientes obtidos através da decomposição da matriz de covariâncias Σ_{pxp} . De forma geral, a explicação de variância total, quando se utiliza a matriz P_{pxp} , necessita-se de um número maior de componentes para explicar a mesma quantidade de variância total obtida utilizando-se a matriz Σ_{pxp} .

Na prática, a matriz P_{pxp} é estimada pela matriz de correlação amostral R_{pxp} do vetor aleatório X , sendo as componentes principais construídas utilizando-se R_{pxp} :

$$R_{pxp} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1p} \\ R_{21} & R_{22} & \cdots & R_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ R_{p1} & R_{p2} & \cdots & R_{pp} \end{bmatrix} \quad (5.12)$$

Sendo $R_{ii} = 1$ e $R_{ij} = R_{ji}$, $j \neq i$ definido por:

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \quad (5.13)$$

que é o coeficiente de correlação amostral entre a i -ésima e j -ésima variáveis, conhecido como coeficiente de correlação de Pearson [Triola, 2005].

A partir da propriedade 3, algumas conclusões podem ser extraídas da correlação entre duas variáveis. Sejam duas variáveis A e B , então:

- (i) Se $r_{A,B} > 0$, então A e B são positivamente correlacionados, quanto maior o valor de A mais correlacionada estará com B .
- (ii) Se $r_{A,B} = 0$, então A e B são independentes.

- (iii) Se $r_{A,B} < 0$, então A e B são negativamente correlacionados, quando os valores de uma variável crescem o da outra decresce.

5.2.4 Determinação do número k de componentes principais

Quando se deseja reduzir a dimensionalidade do espaço amostral, isto é, a sumariação da informação das p -variáveis originais em k componentes principais, onde $k < p$, faz-se necessário estabelecer critérios de escolha para o valor de componentes, que deverão ser mantidas, vejamos três procedimentos que podem determinar a escolha de k :

5.2.4.1 Puramente matemáticos

Procedimento 1: Análise de representatividade em relação à variância total

Deve-se manter no sistema um número de componentes k que conjuntamente representem uma porcentagem γ 100% da variância total, onde $0 < \gamma < 1$ é um valor pré-determinado. Na prática, busca-se o valor k tal que:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^p \lambda_j} = \gamma \quad (5.14)$$

Não existe um limite definido para o valor γ , e sua escolha deverá ser feita de acordo com a natureza do fenômeno investigado. A porcentagem de explicação de variância total pode ser alta, maior que 90%, com as duas primeiras componentes. O objetivo, ao se utilizar PCA é sumarizar a informação das p -variáveis originais de uma forma simples. A utilidade das componentes decresce com o crescimento do número de componentes necessárias

para se chegar ao valor escolhido de γ , ou seja, quanto maior k , maior será a dificuldade de interpretação. Em geral, quando as componentes principais são extraídas da matriz de correlação, necessita-se de um número maior de componentes para se alcançar o valor de γ , em comparação com o número requerido via matriz de covariâncias.

Quando a matriz de correlação é utilizada para extração das componentes principais, a variância total é igual ao número de variáveis originais p . Segundo o critério de [Kaiser, 1958] pode-se escolher o valor de k a fim de manter no sistema apenas as componentes relacionadas àqueles autovalores $\hat{\lambda}_i \geq 1$, ou seja, mantêm-se as combinações lineares que conseguem explicar pelo menos a quantidade de variância de uma variável original padronizada. Da mesma forma, quando a análise é feita com a matriz de covariâncias, podem-se manter no sistema as componentes relacionadas aos autovalores que são maiores ou iguais a $\hat{\lambda}_m$, definido por:

$$\hat{\lambda}_m = \frac{\sum_{j=1}^p \hat{\lambda}_j}{p} \quad (5.15)$$

e que representa a variância média das variáveis originais $X_j, j = 1, 2, \dots, p$.

O *scree-plot* [Cattell, 1966] é um gráfico (Figura 5.2) que auxilia na escolha do valor de k , que mostra os valores numéricos dos autovalores $\hat{\lambda}_i$ de acordo a respectiva ordem i . Em geral podemos observar o ponto em que os valores de $\hat{\lambda}_i$ tendem a se estabilizar, ou seja, o ponto do qual os autovalores $\hat{\lambda}_i$ se aproximam de zero.

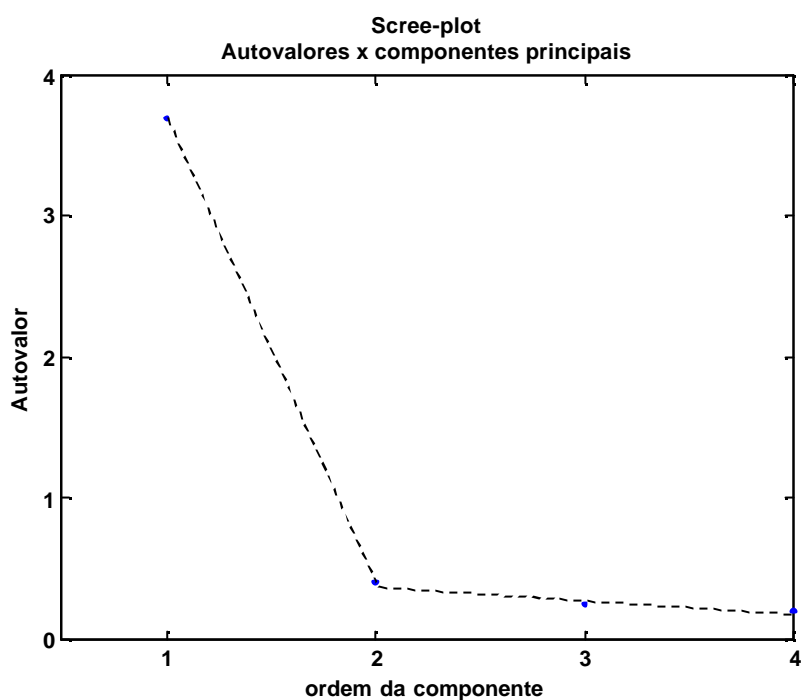


Figura 5.2 - O scree-plot mostra as componentes principais seguindo a tendência de estabilidade, aproximação de zero, dos autovalores \hat{I}_i para um vetor aleatório X .

Procedimento 2: Análise da qualidade de aproximação da matriz de covariâncias ou correlação.

A escolha do valor k de componentes determina uma aproximação para a matriz da qual foram extraídas. Na prática, quando as componentes são extraídas das matrizes de covariâncias ou de correlação amostrais, tem-se as seguintes aproximações para as matrizes

S_{pxp} e R_{pxp} , respectivamente:

$$S_{pxp} \approx \sum_{i=1}^k \hat{I}_i \hat{e}_i \hat{e}_i'$$

$$R_{pxp} \approx \sum_{i=1}^k \hat{I}_i \hat{e}_i \hat{e}_i'$$

onde em cada caso, $\left(\hat{I}_i, \hat{e}_i \right)$ representam os respectivos autovalores e autovetores normaliza-

dos de S_{pxp} e R_{pxp} .

Sendo assim, o valor de k poderia ser escolhido de modo a se ter uma aproximação razoável para as matrizes S_{pxp} e R_{pxp} . As parcelas que mais contribuem para a aproximação dessas matrizes são correspondentes aos autovalores significativamente maiores que zero. Logo, as componentes associadas a autovalores pequenos os próximos a zero poderiam ser eliminadas.

5.2.4.2 Sob o ponto de vista prático

Procedimento 3: Análise prática das componentes

Para que as componentes possam ser utilizadas adequadamente, a escolha das k componentes deve ser passível de interpretação. A escolha do valor k pode ser pautada pela natureza prática das componentes encontradas. Eventualmente, pode ocorrer que a componente mais útil sob o ponto de vista prático do pesquisador seja a de ordem 4, por exemplo. No entanto, para que este possa utilizá-la, será necessário manter pelo menos $k=4$ componentes principais no sistema, que será de difícil interpretação gráfica e a depender do tamanho das variáveis, um alto custo computacional. A situação ideal é aquela em que os componentes principais de maior interesse do pesquisador são as de maior explicação relativa à da variância total e levam a um valor pequeno de k .

5.3 Métodos não-paramétricos

Uma das dificuldades das abordagens não paramétricas é especificar distribuições de probabilidade tendo uma forma específica governada por um pequeno número de parâmetros dos quais são determinados pelo conjunto de dados. Inicialmente iremos supor que as

observações pertencem a uma desconhecida função densidade de probabilidade $p(x)$ num espaço D -dimensional e assumindo se euclidiano e que se deseja estimar o valor de $p(x)$. Para esta discussão iremos considerar uma pequena região R contendo x . A probabilidade associada a esta região é dada por:

$$P = \int_R p(x) dx \quad (5.16)$$

Supondo que um conjunto de dados seja de N observações de $p(x)$ e uma vez que cada ponto tenha uma probabilidade P de estar em R , o número total de pontos K que está em R será distribuído por uma distribuição de probabilidade binomial:

$$Bin(K|N, P) = \frac{N!}{K!(N-K)!} P^K (1-P)^{N-K} \quad (5.17)$$

A fração média de pontos que está na região é $E[K/N]=P$ e a variância é $Var[K/N]=P(1-P)/N$ [Bishop, 2006]. Para N grande, a distribuição estará em torno da média e então:

$$K \approx NP \quad (5.18)$$

Se, entretanto assumirmos que a região R é suficientemente pequena que a probabilidade $p(x)$ é também constante sobre a região, então:

$$P \approx p(x)V \quad (5.19)$$

onde V é o volume de R . Combinando (5.18) e (5.19), teremos a densidade estimada na sob a forma:

$$p(x) = \frac{K}{NV} \quad (5.20)$$

Vale ressaltar que a validade de (5.20) depende de duas contraditórias suposições, ou seja, que a região R é suficientemente pequena que a densidade é aproximadamente constante sobre a região e ainda é suficientemente grande (em relação ao valor da densidade) que

o número K de pontos na região é suficiente para uma distribuição binomial ser categoricamente nítida.

O resultado de (5.20) pode ser explorado de duas diferentes formas. Podemos fixar K e determinar o valor de V dos dados, no qual consiste na técnica K -vizinhos-mais-próximos, (KNN, do inglês *K-nearest-neighbour*), ou pode-se fixar V e determinar K a partir dos dados, seguindo a abordagem *kernel*⁸. Pode ser mostrado que ambos que tanto KNN e *kernel* estimador de densidade converge para a densidade de probabilidade real no limite $N \rightarrow \infty$ desde que V é inversamente proporcional a N , e K diretamente proporcional a N . [Duda e Hart, 2001, 1973]. Serão abordados dois métodos tanto a estimação por *kernel* quanto por *nearest-neighbour*, sendo este último, em especial, o utilizado neste trabalho.

5.3.1 Estimador de densidade por Kernel

Para o método *kernel* assumiremos uma região R sendo um pequeno hipercubo centrado sobre o ponto x em que se deseja estimar a densidade de probabilidade. A fim de saber o número K de ponto dentro desta região, é conveniente definir a seguinte função

$$k(u) = \begin{cases} 1, & |u_i| \leq 1/2, i = 1, \dots, D \\ 0, & \text{otherwise} \end{cases} \quad (5.21)$$

no qual representa um cubo unitário centrado na origem. A função $k(u)$ é um exemplo de uma *função kernel*, também chamada de Parzen. A partir de (5.21) $k((x - x_n)/h)$ será um se o ponto x_n estiver num cubo de tamanho h centrado em x , e zero caso contrário. O número total de pontos do cubo será dado por

⁸ Também conhecido como núcleo

$$K = \sum_{n=1}^N k\left(\frac{x-x_n}{h}\right) \quad (5.22)$$

Substituindo (5.22) em (5.20) então teremos o seguinte resultado para a densidade estimada de x

$$p(x) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h^D} k\left(\frac{x-x_n}{h}\right) \quad (5.23)$$

onde usamos $V = h^D$ para o volume do hipercubo de tamanho h em D dimensões. Usando a simetria da função $k(u)$, pode-se reinterpretar esta equação, não como um cubo unitário centrado em x , mas como a soma dos N cubos centrados sobre os N pontos x_n .

Também se pode escolher qualquer outra função kernel $k(u)$ em (5.23) sujeitas as condições

$$k(u) \geq 0, \quad (5.24)$$

$$\int k(u) du = 1 \quad (5.25)$$

onde a distribuição de probabilidade resultante é não-negativa em toda região e integrável a um. Esta classe de modelo de densidade dada por (5.23) é *kernel* estimador de densidade, ou *Parzen* estimador. Isto possui um grande mérito, pois não envolve nenhuma computação na fase de treinamento, pois requer apenas do conjunto de treinamento. Entretanto, esta é uma das maiores fraquezas uma vez que custo computacional de avaliar a densidade cresce linearmente com o tamanho do conjunto dos dados.

5.3.2 Nearest-neighbour

Uma das dificuldades da abordagem kernel para estimar densidade é que o parâmetro h envolve o tamanho do kernel é fixo para todos os kernels. Em regiões de alta densi-

dade, um grande valor de h pode conduzir a um *over-smoothing*⁹. Entretanto, reduzindo h pode levar a falsas estimativas no espaço dos dados onde a densidade é menor. Logo a melhor escolha de h pode depender na localização no espaço dos dados. Este tipo de problema é dedicado aos métodos vizinhos-mais-próximos para estimar a densidade.

Assumindo o resultado geral (5.20) para estimação de densidade local, em vez de fixar V e determinar o valor de K a partir dos dados, será considerado um valor fixo de K e usar os dados para achar um apropriado valor para V . Para isto pode-se tomar uma pequena esfera centrada sobre o ponto x ao qual se pode estimar $p(x)$ permitindo seu raio crescer até conter precisamente K pontos. A estimada densidade $p(x)$ então dada por (5.20) com V ajustado ao volume da resultante esfera. Esta técnica é conhecida como *K-Nearest-Neighbours*.

Para problemas de classificação pode-se aplicar tal técnica em cada classe separadamente fazendo uso do Teorema de Bayes [Bishop, 2006], porém neste trabalho será utilizado o extrator de características descrito na seção 5.2, ou seja, a Análise de Componentes Principais.

Para classificar o novo ponto, nós identificamos os K vizinhos do conjunto de treinamento e então se atribuí o novo ponto a classe que tiver o maior número de representações dentre os conjuntos. Empates podem ser quebrados aleatoriamente. O caso particular em que $K = 1$ é chamado *regra nearest-neighbour*, por que o ponto a ser classificado é simplesmente atribuído à mesma classe do seu vizinho mais próximo (ponto mais próximo) de conjunto de treinamento. A Figura 5.3 ilustra o classificador KNN.

⁹ Operação estatística para substituir os termos de uma série irregular de dados experimentais pelos valores de uma função de um deles, a fim de se chegar a uma série regular ou mais regular que a anterior e, eventualmente, contínua.

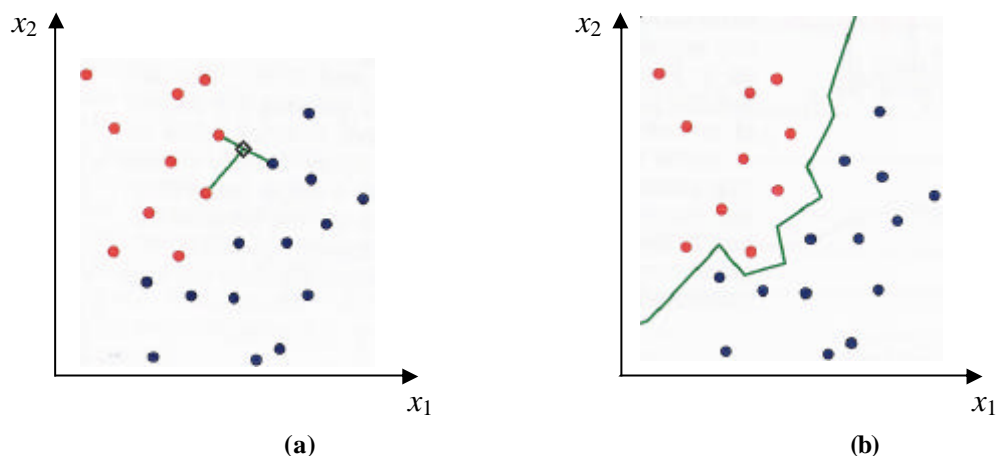


Figura 5.3 – (a) No classificador KNN, um novo ponto, representado pelo quadrado preto, é classificado pela classe formada pela maioria dos pontos mais próximos pertencentes ao conjunto de treinamento, neste caso $K = 3$. (b) Na classificação pela abordagem *nearest-neighbour* ($K=1$) o limite de decisão resultante é formado de hiperplanos que formam bissetores perpendiculares pelos pares de pontos das diferentes classes.

Uma propriedade interessante do classificador KNN quando $K=1$ é que, no limite $N \rightarrow \infty$, o erro nunca é maior que duas vezes o erro mínimo realizado para um classificador ótimo, ou seja, um classificador que utilize a verdadeira distribuição dos dados [Cover e Hart, 1967].

A técnica de *nearest-neighbour* é fácil de ser utilizada em classificação de padrões, não é necessário conhecer a lei de probabilidade que rege cada classe. Entretanto existe um problema, no que diz respeito à memória requerida pelo algoritmo, já que o tamanho da memória será da ordem do número de padrões usados como referência. KNN é um classificador onde o aprendizado é baseado na analogia. O conjunto de treinamento é formado por vetores n -dimensional e cada elemento deste conjunto representa um ponto no espaço n -dimensional.

O problema é resolvido após instanciar um valor para K (o número de vizinhos mais próximos), ou seja, tem-se que um padrão será classificado na classe que contiver um elemento, cuja distância entre ele e o padrão será a menor dentre todos os outros elementos e o padrão. Um problema inerente à técnica é, então, descobrir uma medida confiável para classificar os padrões. Essa técnica faz o uso de uma medida de distância. As técnicas de medidas que podem ser utilizadas na técnica de *nearest-neighbour* são a distância de Hamming, distância Square e a distância Euclidiana, sendo esta última a utilizada.

Seja $X = (x_1, x_2, \dots, x_p)$ e $Y = (y_1, y_2, \dots, y_p)$ dois pontos do \mathfrak{R}^p , logo:

$$d(x, y) = \left(|x_1 - y_1|^q + |x_2 - y_2|^q + \dots + |x_p - y_p|^q \right)^{\frac{1}{q}}, \quad (5.26)$$

onde $q \in N$, é conhecida como distância Minkowski, sendo a distância Euclidiana um caso particular desta quando $q = 2$, ou seja:

$$d(x, y) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \dots + |x_p - y_p|^2}, \quad (5.27)$$

KNN é um classificador que possui apenas um parâmetro livre (o número de K -vizinhos) que é controlado pelo pesquisador com o objetivo de obter uma melhor classificação. Este processo de classificação pode ser computacionalmente exaustivo se considerado um conjunto com muitos dados. Para determinadas aplicações, no entanto, o processo é bem aceitável.

5.4 Redes neurais artificiais

Ao final dos anos 80 ressurgiu a área das Redes Neurais Artificiais (RNs, com o adjetivo artificial implicado), também é conhecida como connexionismo ou sistemas de processamento paralelo e distribuído. É uma forma de computação não-algorítmica caracterizada por sistemas que tentam imitar a o funcionamento cerebral humano [Braga et al., 2000].

O estudo das RNs iniciou-se com a motivação do reconhecimento de que o cérebro humano é capaz de processar diferentemente do computador digital convencional. O cérebro é um computador (sistema de processamento de informação) altamente complexo, não-linear e paralelo. Organiza seus neurônios, de forma a realizar certos processamentos, a exemplo de reconhecimento de padrões, percepção e controle motor, muito mais rápido que qualquer computador digital existente. E o desenvolvimento das RNs têm sido fortemente

influenciado a partir do modelo original de um neurônio como um dispositivo binário [McCulloch e Pitts, 1943].

De forma geral, uma RN é uma máquina que é projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou função de interesse; a rede é normalmente implementada utilizando-se componentes eletrônicos ou é simulada por um programa num computador digital. Porém em sua construção existe o processo de aprendizagem, e segundo [Haykin, 2001]:

“Uma RN é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples, que tem a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos:

- 1 - O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem,
- 2 - Forças de conexões entre neurônios, conhecidas como pesos sinápticos, são utilizados para armazenar o conhecimento adquirido.”

Um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma RN. A Figura 5.4 ilustra o modelo de um neurônio, que forma a base para o projeto de uma RN.

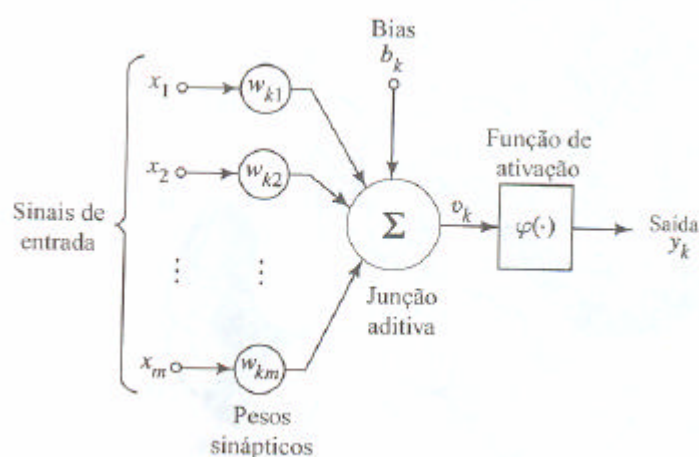


Figura 5.4 – Modelo não-linear de um neurônio. Nesse modelo o bias (b_k) é incluído, sua presença aumenta ou diminui a entrada da função de ativação, dependendo se ele é positivo ou negativo, respectivamente.

São identificados três elementos básicos do modelo neuronal:

1. Um conjunto de sinapses, cada uma caracterizada por um peso. Um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico. Vale ressaltar o primeiro índice do peso sináptico w_{kj} se refere ao neurônio em questão e o segundo ao terminal de entrada da sinapse a qual o peso se refere.
2. Um *somador* para somar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio; as operações consistem num *combinador linear*.
3. Uma *função de ativação* para restringir a amplitude da saída de um neurônio. Também é conhecida como *função restritiva* já que limita o intervalo permissível de amplitude do sinal de saída a um valor finito. Tipicamente o intervalo normalizado de amplitude de saída de um neurônio é escrito como o intervalo unitário fechado $[0, 1]$ ou alternativamente $[-1, 1]$.

Em termos matemáticos, podemos escrever um neurônio k como o seguinte par de equações:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (5.28)$$

e

$$y_k = \mathbf{j}(u_k + b_k) \quad (5.29)$$

onde x_1, x_2, \dots, x_m são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio k ; u_k é a saída do combinador linear devido aos sinais de entrada; b_k é o bias; $\mathbf{j}(\cdot)$ é a função de ativação; e y_k é o sinal de saída do neurônio.

Quando RNs são usadas para análise de dados, ressalta-se a diferença entre modelos de RNs e algoritmos de RNs, além de que muitos modelos RNs são similares ou idênticos a algumas populares técnicas estatísticas, a exemplo modelos lineares generalizados, regressão polinomial, regressão não-paramétrica, análise de discriminante, análise de componentes

principais e clusters, especialmente quando a ênfase está na predição de complicados fenômenos melhor que na explanação. Na literatura existem poucos modelos de RNs, que não tem equivalências estatísticas precisas, que podem ser usados para análise de dados, a exemplo de mapas auto-organizáveis, aprendizagem de vetores de quantização e *counterpropagation*. Mas, por outro lado, equivalências são evidenciadas, onde RNs são uma larga classe de regressão não-linear e modelos de discriminante, modelos de redução de dados e sistemas dinâmicos não-lineares, a exemplo das RNs *multilayer perceptron* (MLP) que nada mais são do que regressão não-linear e modelos de discriminante e que podem ser implementadas em softwares estatísticos [Warren, 1994].

Ainda segundo [Warren, 1994] existem consideráveis superposições entre RNs e a Estatística. A Estatística relaciona-se a análise de dados, em RNs estatística significa aprender a generalizar dados com ruído. Algumas RNs não estão relacionadas à análise de dados, a exemplo da modelagem de sistemas biológicos, algumas não aprendem, por exemplo, rede Hopfield [Li e Hopfield, 1989], e então têm pouco haver com estatística. Algumas RNs podem aprender satisfatoriamente a partir de dados livre de ruídos, por exemplo, ART ou *perceptron* [Braga et al., 2000], entretanto não poderiam ser consideradas métodos estatísticos. Mas, em sua maioria, as RNs podem aprender a generalizar efetivamente a partir de dados com ruídos e são semelhantes ou idênticas a métodos estatísticos.

Segundo [Ripley, 1993], existe uma relação direta entre algumas redes neurais e métodos estatísticos, vejamos algumas:

- 1 - RN feedforward sem camada escondida (inclusive funcional - ligação RNs e RNs de alta-ordem) são basicamente modelos lineares generalizados.
- 2 - RN *feedforward* com uma camada escondida é relacionado com *projection pursuit regression*.
- 3 - RNs probabilísticas são idênticas a *kernel discriminant analysis*.

4 – RNs de regressão geral são idênticas a *Nadaraya-Watson kernel regression*

5 – Rede Kohonen para *adaptive vector quantization* é muito similar à análise de clusters de k-médias.

6 – Aprendizado Hebbian é relacionado à Análise de Componentes Principais

E há algumas áreas de RNs que parecem não ter equivalência estatística na literatura, a exemplo:

1 – Mapas auto-organizáveis de Kohonen

2 – Aprendizado por reforço, entretanto este é tratado em pesquisas literárias como processo de decisão de Markov.

Existem outras evidências, tais como as RNs *feedforward* que são um subconjunto de classes de regressão não-linear e modelos de discriminante. Estatísticos têm estudado as propriedades desta classe geral, mas não tem considerado o específico caso de RNs *feedforward* antes que de tais fossem popularizadas no campo das RNs. Ainda, muitos resultados da teoria estatística de modelos não-lineares aplicados diretamente a RNs *feedforward*, e os métodos que são comumente usados para modelos adaptativos não-lineares, como vários algoritmos, a exemplo de *Levenberg-Marquardt* e *conjugate gradient algorithms*, podem ser usados no treinamento de RNs *feedforward*. O algoritmos *Levenberg-Marquardt* utiliza uma aproximação pelo método de Newton [Hagan e Menhaj, 1994].

Enquanto RNs são freqüentemente definidas em termos de seus algoritmos ou implementações, métodos estatísticos são usualmente definidos em termos de seus resultados. Porém, às vezes, ocorrem indagações onde RNs, ao contrário de modelos estatísticos, não requerem suposições de distribuição. De fato, RNs envolvem exatamente a mesma sorte das suposições de distribuição como os modelos estatísticos, mas estatísticos estudam a consequência e a importância dessas suposições enquanto os estudiosos de RNs, em sua maioria, os ignoram. Por exemplo, o método de treinamento mínimo-quadrados são largamente usados

por estatísticos e estudiosos de RNs, os estatísticos realizam o treinamento mínimo-quadrado envolvendo suposições implícitas de distribuição em que as estimativas dos mínimo-quadrado têm determinadas propriedades para otimizar o ruído que são distribuídos normalmente com variação igual para todos os casos do treinamento e que é independente entre casos diferentes. Estas propriedades de otimização são conseqüências do fato que a estimação dos mínimo-quadrado é a máxima verossimilhança sob aquelas circunstâncias [Warren,1994]

Mas é bom esclarecer que RNs e estatística não são metodologias de competição para análise de dados. Existem consideráveis correspondências entre os dois campos. RNs incluem muitos modelos, assim como MLPs, que são utilizadas para aplicações estatísticas. A metodologia estatística é diretamente aplicada em RNs de várias formas, incluindo critérios de estimação, otimização algorítmica, intervalo de confiança, diagnósticos, e métodos gráficos. E ambas as áreas serão beneficiadas se houver uma boa comunicação e inter-relação entre elas.

De forma geral estatísticos podem ver com interesse a disseminação explosiva dos estudos das redes neurais. Nas palavras de [Anderson et al., 1990]:

“Redes neurais é estatística para amadores. Uma rede corretamente projetada, ao aprender e ao responder, representa uma boa inferência estatística, baseada sobre o que ela viu quando aprendeu e o que ela vê quando responde. Muitas redes escondem a estatística do usuário”.

Na frase acima, “amadores” tem um sentido pejorativo, e é deixado para o leitor julgar se redes neurais é um amadorismo na tentativa de fazer estatística através de computadores. Neste trabalho, além da análise estatística, duas RNs são utilizadas a MLP e a rede de função de base radial (RBF, do inglês *radial-basis function*).

5.4.1 Redes neurais multilayer perceptron

Dentre os diversos modelos de RNs existentes, a rede MLP [Rumelhart et al., 1986] é a mais difundida, devido a sua facilidade de implementação e entendimento, tendo sido aplicada com sucesso em diversos problemas difíceis, através do seu treinamento de forma supervisionada com o algoritmo de retropropagação de erro, mais conhecido como *error back-propagation*. Este algoritmo é baseado na regra de aprendizagem por correção de erros. Dentre os diversos trabalhos que a utilizaram destacam-se [Zanchettin, 2003; Santos, 2000; Di Natale et al., 1995a; Gardner et al., 1994; Gardner et al., 1990].

Basicamente, a aprendizagem por retropropagação de erro consiste em dois passos através das diferentes camadas da rede: um passo para frente à propagação, e um passo para trás, a retropropagação. No passo para frente, um padrão de entrada é aplicado aos nós sensoriais da rede e seu efeito se propaga através da rede, camada por camada. Assim, um conjunto de saídas é produzido como resposta real da rede. Uma vez definido os pesos sinápticos da rede, no passo de propagação são todos fixos, e no passo para trás de acordo com uma regra de correção de erro. De forma geral, a resposta da rede é subtraída de uma resposta desejada para produzir um *erro*. Este erro é então propagado para trás através da rede, contra as conexões sinápticas, daí o nome *back-propagation*. O ajuste dos pesos tenta fazer com que a resposta da rede se mova para mais perto da resposta desejada, em um sentido estatístico. O algoritmo pode ser encontrado detalhadamente em [Haykin, 2001].

A Figura 5.5 mostra o grafo arquitetural de uma rede MLP com duas camadas ocultas e uma camada de saída.

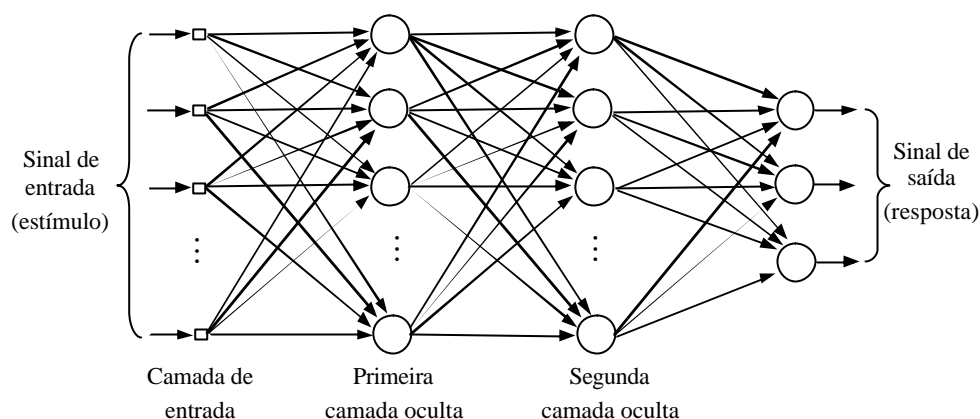


Figura 5.5 – Grafo arquitetural de um perceptron de múltiplas camadas com duas camadas ocultas

Um perceptron de múltiplas camadas tem três características distintas:

1. O modelo de cada neurônio da rede inclui uma *função de ativação não-linear*. Vale destacar que a não-linearidade é suave, ou seja, diferenciável em qualquer ponto. Uma forma normalmente utilizada de não-linearidade que satisfaz esta exigência é uma *não-linearidade sigmóide*, geralmente definida pela função logística:

$$y_j = \mathbf{j}(v_j, a) = \frac{1}{1 + \exp(-av_j)} \quad (5.30)$$

onde v_j é o campo local induzido, ou seja, a soma ponderada de todas as entradas sinápticas acrescidas do bias do neurônio j , e y_j é a saída do neurônio. Esta função oferece uma importante flexibilidade dada por sua inclinação na origem, $\mathbf{j}(0, a) = a/4$, ser variável com o parâmetro a , conforme ilustra a Figura 5.6. A escolha da função logística tem motivação biológica, pois procura levar em conta a fase refratária de neurônios reais, além disso, apresenta propriedades de simetria e de completa diferenciabilidade, ou seja, pertence à classe de funções C^∞ . Para os nossos experimentos $a = 1$.

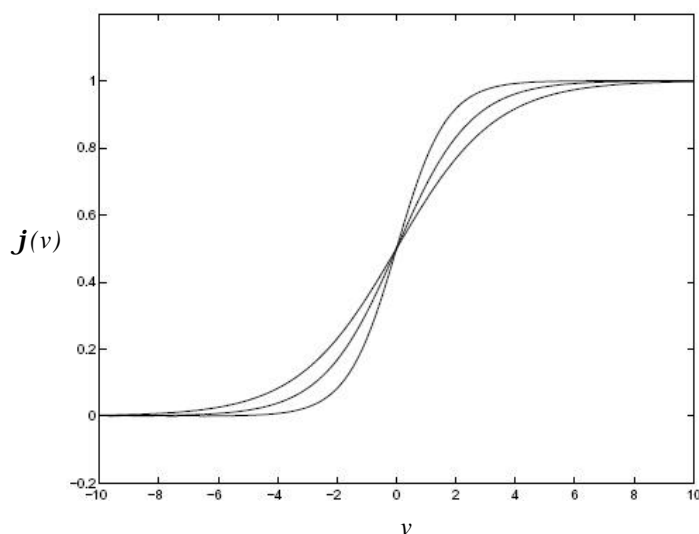


Figura 5.6 – Função Logística – Efeito da variação do parâmetro a .

2. A rede contém uma ou mais camadas de neurônios ocultos, que não são parte da entrada ou da saída da rede. Estes neurônios ocultos capacitam a rede a aprender tarefas complexas extraindo progressivamente as características mais significativas dos padrões de entrada.
3. A rede exibe um alto grau de *conectividade*, determinado pelas sinapses da rede. Todas as camadas são conectadas, porém não ocorrem ligações entre camadas não adjacentes.

Uma rede MLP treinada com o algoritmo *back-propagation* pode, em geral, aprender mais rápido (em termos do número de interações de treinamento necessárias) quando a função de ativação sigmóide incorporada no modelo do neurônio da rede for antissimétrica do que quando ela for não-simétrica (i. e., função ímpar de seu argumento) se:

$$\mathbf{j}(-v) = -\mathbf{j}(v) \quad (5.31)$$

Um exemplo de uma função de ativação anti-simétrica é uma não-linearidade sigmóide na forma de uma *tangente hiperbólica*, definida por:

$$\mathbf{j}(v) = a \tanh(bv) \quad (5.32)$$

onde a e b são constantes. Segundo [LeCun, 1993], valores adequados para essas constantes são $a = 1,7159$ e $b = 2/3$.

A função tangente hiperbólica assim definida tem as seguintes propriedades úteis:

- i. $\mathbf{j}(1) = 1$ e $\mathbf{j}(-1) = -1$
- ii. Na origem, a inclinação (i.e., o ganho efetivo) da função de ativação fica próxima de 1.
- iii. A segunda derivada de $\mathbf{j}(v)$ atinge seu valor máximo em $v = 1$.

Outro exemplo de uma função de ativação tangente hiperbólica é definido por:

$$y_j = \mathbf{j}(v_j, a) = \frac{1 - \exp(-2av_j)}{1 + \exp(-2av_j)} \quad (5.33)$$

onde v_j é o campo local induzido, ou seja, a soma ponderada de todas as entradas sinápticas acrescidas do bias do neurônio j , e y_j é a saída do neurônio.

A função tangente hiperbólica também muito utilizada como função de ativação, corresponde unicamente a um deslocamento da função logística. Para os nossos experimentos $a = 0$.

5.4.2 Redes de função de base radial

Adicionalmente às RNs com múltiplas camadas, um tipo específico de rede que tem tido recente aplicação e também será utilizada neste trabalho é a RN de função de base radial (RBF), onde os neurônios geralmente são *funções radiais multivariadas*, com vários argumentos de entrada, e com comportamento radial em relação a cada argumento, ou seja, as funções radiais decrescem (ou crescem) monotonicamente à medida que, se distanciam do ponto central.

O projeto de uma RN supervisionada pode ser executado de várias formas. O algoritmo de retropropagação utilizado para o projeto de uma rede MLP (sob supervisão) pode ser visto como a aplicação de uma técnica recursiva conhecida na estatística como *aproximação estocástica*. Nesta seção veremos um enfoque diferente do anterior pois se trata de uma RN com um problema de ajuste de curva (*aproximação*) em um espaço de alta dimensionali-

dade. Neste ponto de vista aprender é equivalente a encontrar uma superfície, em um espaço multidimensional, que forneça o melhor ajuste para o conjunto de treinamento, com o critério de “melhor ajuste” sendo medido em um sentido estatístico. Senso assim, generalizar é equivalente ao uso desta superfície multidimensional para interpolar os dados de teste. Tal ponto de vista é a motivação que se encontra por trás do método das funções de base radial, no sentido que isto o aproxima dos trabalhos científicos em interpolação tradicional em um espaço multidimensional.

No contexto de uma RN, as unidades ocultas fornecem o conjunto de funções que constituem uma “base” arbitrária de padrões de entrada, quando eles são expandidos sobre o espaço oculto, daí estas funções são chamadas de funções de base radial. As funções de base radial foram primeiramente introduzidas na solução do problema de interpolação multivariada real [Light, 1992; Powell, 1985].

A construção de uma RN RBF básica envolve três camadas com papéis distintos. A camada de entrada é constituída por nós de fonte (unidades sensoriais) que conectam a rede ao seu ambiente. A segunda camada, a *única* camada oculta da rede diferentemente das RNs MLP, aplica uma transformação não linear do espaço de entrada para o espaço oculto, geralmente de alta dimensionalidade. A camada de saída é linear, fornecendo a resposta da rede ao padrão de ativação aplicado à camada de entrada. A Figura 5.7 ilustra uma RN RBF. Em [Cover, 1965], verifica-se uma justificativa matemática para a estratégia de uma transformação não-linear seguida de uma linear. De acordo com esse artigo, um problema de classificação de padrões disposto em um espaço de alta dimensionalidade tem maior probabilidade de ser linearmente separável do que em um espaço de baixa dimensionalidade, daí a razão de fazer com que a dimensão do espaço oculto está adiante relacionada à capacidade da rede aproximar um mapeamento de entrada-saída suave [Mhaskar, 1996; Niyogi e Girosi, 1996]; quanto mais alta for a dimensão do espaço oculto, mais precisa será a aproximação.

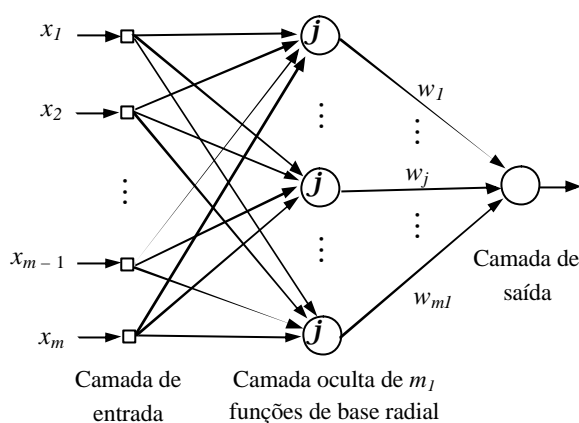


Figura 5.7 – Rede de função de base radial.

Ainda segundo [Cover, 1965], considerando uma classe natural de mapeamentos obtidos utilizando-se uma combinação linear de produtos de r coordenadas vetoriais do padrão, então as superfícies de separação correspondentes a estes mapeamentos são referidas como *variedades racionais de ordem r* . uma variedade racional de ordem r em um espaço de dimensão m_0 é descrita por uma equação heterogênea de grau r envolvendo as coordenadas do vetor de entrada x , como mostrado por

$$\sum_{0 \leq i_1 \leq i_2 \leq \dots \leq i_r \leq m_0} a_{i_1, i_2, \dots, i_r} x_{i_1} x_{i_2} \dots x_{i_r} = 0 \quad (5.34)$$

onde x_{i_r} é a i -ésima componente do vetor de entrada x , e x_0 é fixo em uma unidade para expressar a equação em uma forma homogênea. Um produto de ordem r das componentes x_i de x , isto é, $x_{i_1} x_{i_2} \dots x_{i_r}$, é chamado um *monômio*. Para um espaço de entrada de dimensionalidade m_0 , existem

$$\frac{(m_0 - r)!}{m_0! r!} \quad (5.35)$$

monômios na Eq. (5.34). Exemplos dos tipos de superfícies de separação descritas pela Eq. (5.34) são hiperplanos, quadráticas e hipersferas. Estes exemplos são ilustrados na Figura 5.8 para uma configuração de cinco pontos em um espaço de entrada bidimensional. Em geral, a separabilidade linear implica a separabilidade esférica que implica a separabilidade quadrática; porém, o inverso não é necessariamente verdadeiro.

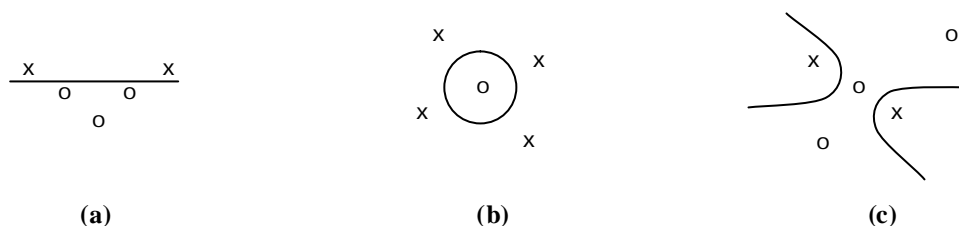


Figura 5.8 – Três exemplos de dicotomias separáveis por uma função oculta de diferentes conjuntos de cinco pontos em duas dimensões: (a) dicotomia linearmente separável; (b) dicotomia esfericamente separável; (c) dicotomia quadraticamente separável.

Em [Haykin, 2001] é apresentado o teorema de [Cover, 1965] sobre a separabilidade de padrões.

Existe um grande número de funções radiais que podem ser utilizadas para este fim. Sendo as mais comuns às funções do tipo:

$$h(x) = \mathbf{f}\left((x - c)^T R^{-1}(x - c)\right) \quad (5.36)$$

Onde \mathbf{f} é o tipo de função usada, c é a posição em que a função está centrada e R é a métrica utilizada. O termo $h(x) = \mathbf{f}\left((x - c)^T R^{-1}(x - c)\right)$ define a distância entre o vetor de entrada x e o vetor de centros e na métrica definida por R (cada entrada corresponde a uma RBF, que por sua vez possui um centro e um raio em relação a cada entrada). Os tipos mais comuns de RBFs utilizadas são:

$$\text{Gaussiana : } \mathbf{f}(x) = e^{-x^2}$$

$$\text{Multiquadrática : } \mathbf{f}(x) = \frac{1}{(1 + x^2)^2}$$

$$\text{Multiquadrática inversa : } \mathbf{f}(x) = \frac{1}{(1 + x^2)^2}$$

$$\text{Cauchy : } \mathbf{f}(x) = (1 + x^2)^{-1}$$

Neste trabalho usaremos a função Gaussiana.

5.4.3 Comparação entre redes MLP e RBF

As RNs MLP e RBF são exemplos de redes em camadas alimentadas adiante, não-lineares. Ambas são aproximadores universais. Portanto, observa-se que sempre existe uma RBF capaz de imitar precisamente uma MLP específica, ou vice-versa. Entretanto, estas duas redes diferem entre si em vários aspectos:

- i. Uma rede RBF (na sua forma mais básica) tem uma única camada oculta, enquanto que um MLP pode ter uma ou mais camadas ocultas.
- ii. Tipicamente, os nós computacionais de um MLP, localizados em uma camada oculta ou em uma camada de saída, compartilham um modelo neuronal comum. Por outro lado, os nós computacionais na camada oculta de uma rede RBF são bastante diferentes e servem a um propósito diferente daqueles da camada de saída da rede.
- iii. A camada oculta de uma rede RBF é não-linear, enquanto que a camada de saída é linear. Entretanto, as camadas ocultas e de saída de um MLP usado como classificador de padrões são normalmente todas não-lineares.
- iv. O argumento da função de ativação de cada unidade oculta em uma rede RBF calcula a *norma* (distância) euclidiana entre o vetor de entrada e o centro daquela unidade. Enquanto isso, a função de ativação de cada unidade oculta em um MLP calcula *produto interno* do vetor de entrada pelo vetor de peso sináptico daquela unidade.
- v. Os MLPs constroem aproximações globais de um mapeamento de entrada-saída não-linear. Por outro lado, as redes RBF utilizando não-linearidades localizadas com decaimento exponencial (por exemplo, funções gaussianas) constroem aproximações locais para mapeamento de entrada-saída não-lineares.

Sendo assim, que a para a aproximação de um mapeamento de entrada-saída não-linear, o MLP requer um número menor de parâmetros que a rede RBF para o mesmo grau de

precisão. A rede RBF difere do perceptron pela capacidade de implementar transformações lineares arbitrárias do espaço de entrada. Isto é bem ilustrado no *problema do XOR*¹⁰, que não pode ser resolvido por nenhum perceptron linear, mas que pode ser resolvido por uma rede RBF.

5.5 Resumo do capítulo

Este capítulo objetivou apresentar as técnicas utilizadas nesta dissertação no reconhecimento dos padrões de odores. Inicialmente descreveu-se a técnica estatística multivariada Análise de Componentes Principais utilizado como extrator de características a fim de verificar algumas características do problema abordado, uma vez que os dados do problema possuem uma alta dimensionalidade e são de difíceis interpretações. Na construção dos escores das principais componentes foram utilizadas duas abordagens, sendo a primeira via matriz de covariâncias amostral e a segunda via matriz de correlação. Apresentou-se também técnicas não-paramétricas como classificador de padrões a ser utilizado na classificação dos dados.

Foi apresentada algumas abordagens conexionistas de reconhecimento de padrões, as Redes Neurais Multilayer Perceptron (RNs MLP) e Redes Neurais com Funções de Base Radial (RNs RBF). O uso de redes neurais artificiais como sistemas de reconhecimento de padrões em narizes artificiais tem sido muito investigado, e a grande maioria destas abordagens faz uso da conhecida rede MLP.

A vantagem da utilização de redes neurais artificiais no reconhecimento de padrões é a não necessidade de conhecimento sobre os dados, devido a grande capacidade de

¹⁰ No problema XOR, existem entradas de padrões, por exemplo (1,1), (0,1), (0,0) e (1,0) em um espaço bidimensional, onde a saída binária 0 é resposta ao padrão de entrada (1,1) ou (0,0), e a saída binária 1 em resposta ao padrão de entrada (0,1) ou (1,0).

generalização da rede a partir de dados não-lineares. Sua desvantagem é que a definição de arquitetura ideal para a rede é baseada em um processo empírico, simples, porém exaustivo.

No próximo capítulo são apresentados os estudos dos sinais de odores utilizando as técnicas descritas neste capítulo.

Capítulo 6

Resultados da análise dos sinais de odores

Neste capítulo é apresentada a análise da base de padrões dos sinais de odores pelas técnicas descritas no capítulo anterior. Na seção 6.1 é feita uma introdução sobre as técnicas utilizadas. Na seção 6.2 é apresentada a análise da base de dados por Análise de Componentes Principais para visualização e interpretação dos sinais de odores. A classificação da base de dados pelo método K-Nearest-Neighbour é descrita na seção 6.3. Na seção 6.4 são apresentados os experimentos computacionais de classificação pelas Redes Neurais descritas no capítulo anterior. Na seção 6.5 é feito um resumo do capítulo.

6.1 Introdução

Um sistema de aquisição de dados de um nariz artificial, como o protótipo utilizado, produz dados com alta dimensionalidade. Ou seja, um grande número de variáveis que caracterizam ou descrevem cada observação, e que a visualização gráfica dos dados, com mais de três dimensões, é muito difícil de ser analisada. Sendo assim, métodos cujos objetivos é a redução da dimensionalidade de dados multivariados são largamente utilizados, pois possibilitam a interpretação de uma base de dados com grande magnitude e complexidade.

A estatística multivariada consiste em um conjunto de métodos estatísticos utilizados em situações nas quais várias variáveis são analisadas simultaneamente, em cada ele-

mento amostral. Em geral, as variáveis são correlacionadas entre si e quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns de estatística univariada.

Em linhas gerais, os métodos de estatística multivariada são utilizados com o propósito de simplificar ou facilitar a interpretação do fenômeno que está sendo estudado através da construção de índices ou variáveis alternativas que sintetizem a informação original dos dados; construir grupos de elementos amostrais que apresentam similaridade entre si, possibilitando a segmentação do conjunto de dados original; investigar as relações de dependência entre as variáveis respostas associadas ao fenômeno e outros fatores (variáveis explicativas), muitas vezes, com objetivos de predição; comparar populações ou validar suposições através de testes de hipóteses.

6.2 Análise de Componentes Principais

Dentre os vários métodos da estatística multivariada, neste estudo utilizamos a Análise de Componentes Principais (PCA) descrita no capítulo 5. Esta técnica pertence ao grupo da análise multivariada que consiste em técnicas exploratória de sintetização (ou simplificação da estrutura de variabilidade dos dados). O PCA pode ser utilizado para compressão, redução da dimensionalidade, extração de características e projeção de dados multivariados.

A seguir são apresentados os resultados da aplicação PCA na base de dados de padrões de odores.

O procedimento para melhorar a análise dos dados, de modo a reduzir a dimensão das amostras estudadas, é a escolha do número de componentes principais estimadas ($\hat{Y}_j, j =$

1, 2, ..., p) utilizadas na descrição do sistema. Esta escolha é realizada considerando a porcentagem da variância descrita por \hat{Y}_j e a variância residual dos dados.

6.2.1 Análise dos sensores utilizados

Inicialmente iremos analisar os sensores de aroma utilizados no protótipo de aquisição dos dados. Na seção 4.4, capítulo 4, observamos o comportamento dos dados e algumas estatísticas descritivas destes dados. Na Tabela 6.1. podemos observar algumas estatísticas descritivas das repetições de aquisição dos dados para a primeira etapa de aquisição, ou seja, a fase de coleta da linha de base.

Tabela 6.1 – Estatísticas descritivas para a aquisição dos dados coletados de linha de base dos sensores de aroma.

| Sensor | Estatística descritiva | | | | | |
|-----------------|------------------------|--------|---------|---------------|--------|--------|
| | n | Média | Mediana | Desvio Padrão | Mínimo | Máximo |
| S01B3D3 | 77 | 19,819 | 19,62 | 0,03836 | 16,52 | 24,24 |
| S02BB3D3 | 77 | 22,173 | 21,92 | 0,05823 | 17,90 | 35,60 |
| S02BR3D3 | 77 | 20,903 | 20,32 | 0,06296 | 17,68 | 32,60 |
| S02BG3D3 | 77 | 21,768 | 21,48 | 0,05658 | 17,60 | 33,60 |

Faremos um estudo entre os quatro tipos de sensores utilizados por PCA via matriz de covariâncias e matriz de correlação.

6.2.1.1 PCA via matriz de covariâncias amostrais

Para calcular as componentes principais iremos seguir a seção 5.2.2, capítulo 5. Logo a partir da matriz de covariâncias, a escolha das componentes principais estimadas é

realizada considerando a porcentagem de variância descrita pelas componentes e a variância residual dos dados.

A Tabela 6.2 apresenta dados relativos aos quatro tipos de sensores, um total de 308 sensores, utilizando a média dos sensores como variáveis para formar combinação linear de cada componente principal estimada:

$$\hat{Y}_j = e_{j1} \hat{X}_1 + e_{j2} \hat{X}_2 + e_{j3} \hat{X}_3 + e_{j4} \hat{X}_4, \quad j = 1, 2, \dots, 4.$$

onde X_1 é média de S01B3D3, X_2 a média de S02BB3D3, X_3 média de S02BR3D3 e X_4 a média de S02BG3D3. A escolha das componentes estará relacionada à porcentagem de representatividade da variância total dos dados, onde juntas devem representar mais que 90% dos dados originais, conforme a equação 5.14 e as condições citadas na seção 5.2.4.2, do capítulo 5.

Tabela 6.2 – Dados relativos às médias dos 308 sensores de aroma utilizados para compor as componentes principais a serem estimadas.

| Amostra | Média dos sensores | | | |
|---------|--------------------|----------|----------|----------|
| | S01B3D3 | S02BB3D3 | S02BR3D3 | S02BG3D3 |
| A1 | 19,393 | 34,077 | 31,414 | 32,771 |
| A2 | 22,429 | 25,267 | 26,007 | 26,051 |
| A3 | 18,205 | 20,518 | 18,310 | 20,775 |
| ... | ... | ... | ... | ... |
| A75 | 17,463 | 22,413 | 18,040 | 19,434 |
| A76 | 19,997 | 22,205 | 19,132 | 17,699 |
| A77 | 19,108 | 20,340 | 18,451 | 22,537 |

A matriz de covariância amostral das variáveis é dada por:

$$S_{4 \times 4} = \begin{bmatrix} 3,0656 & 0,5229 & 1,0311 & 0,6472 \\ 0,5229 & 8,3344 & 5,8512 & 5,4547 \\ 1,0311 & 5,8512 & 7,1328 & 5,0047 \\ 0,6472 & 5,4547 & 5,0047 & 6,4166 \end{bmatrix}$$

Os autovalores da matriz de covariâncias amostral são: $\hat{I}_1 = 18,3511$, $\hat{I}_2 = 3,0976$, $\hat{I}_3 = 1,8160$ e $\hat{I}_4 = 1,6845$. A variância total é igual a $\hat{I}_1 + \hat{I}_2 + \hat{I}_3 + \hat{I}_4 = 24,9493$ e os correspondentes autovetores normalizados são:

$$\hat{e}_1 = \begin{bmatrix} 0,0821 \\ 0,6245 \\ 0,5691 \\ 0,5285 \end{bmatrix} \quad \hat{e}_2 = \begin{bmatrix} -0,9464 \\ 0,2577 \\ -0,1891 \\ 0,0461 \end{bmatrix} \quad \hat{e}_3 = \begin{bmatrix} 0,1483 \\ 0,6484 \\ -0,0404 \\ 0,7456 \end{bmatrix} \quad \hat{e}_4 = \begin{bmatrix} 0,2749 \\ 0,3509 \\ -0,7992 \\ 0,4031 \end{bmatrix}$$

Portanto as três componentes principais para este estudo são:

$$\hat{Y}_1 = 0,0821X_1 + 0,6245X_2 + 0,5691X_3 + 0,5285X_4$$

$$\hat{Y}_2 = -0,9464X_1 + 0,2577X_2 - 0,1891X_3 + 0,0461X_4$$

$$\hat{Y}_3 = 0,1483X_1 + 0,6484X_2 - 0,0404X_3 + 0,7456X_4$$

$$\hat{Y}_4 = 0,2749X_1 + 0,3509X_2 - 0,7992X_3 + 0,4031X_4$$

As porcentagens de variância total explicadas pelas componentes são, respectivamente 73,55% para \hat{Y}_1 ; 12,42% para \hat{Y}_2 ; 7,28% para \hat{Y}_3 e 6,75% para \hat{Y}_4 . Senso assim, a primeira componente descreve somente 73,55% da variância total dos dados, o que é insuficiente para descrever o conjunto total dos dados. Isto se deve pelo fato da variância residual ainda ser muito alta; desta forma alguma informação poderá estar sendo perdida. Juntas, as três componentes principais \hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3 representam aproximadamente 93,25% da variância total.

Apesar da quantidade de variância descrita por \hat{Y}_3 (apenas 7,28%) ser pequena, sua inclusão entre as componentes analisadas pode ser interessante, satisfazendo a condição de relevância dos dados amostrais originais como estipulado.

A relação entre a variância explicada em cada uma das componentes principais pode ser facilmente visualizada na Figura 6.1.

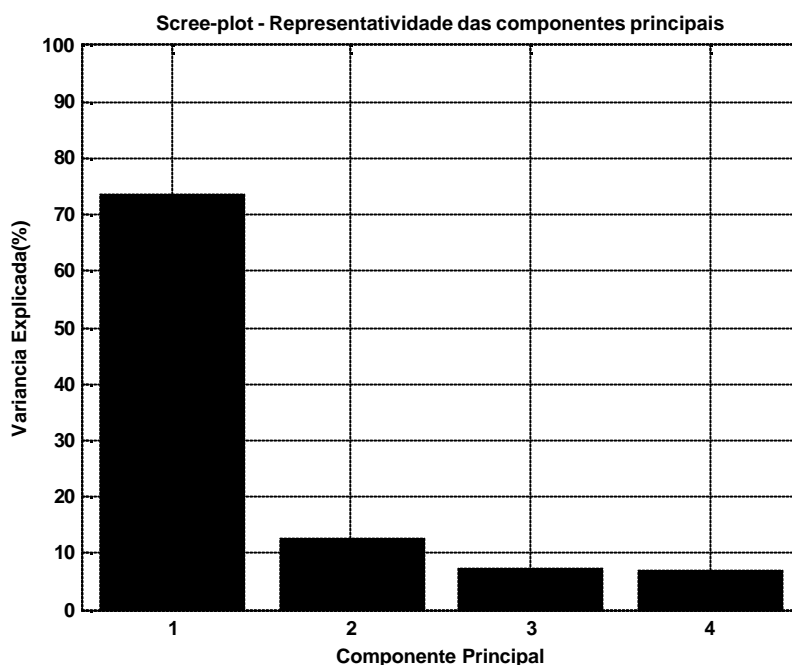


Figura 6.1 – Variância explicada em cada uma das componentes principais para os sensores de aroma via matriz de covariâncias

A projeção de cada padrão de sensor neste novo sistema de eixos fornece os escores; cada sensor possui um valor de escore para cada um dos novos eixos, ou seja, as componentes principais. A projeção destes valores, nas três primeiras componentes principais pode ser vista na Figura 6.2. A partir desse gráfico observamos que os sensores podem constituir a representação de uma substância analisada, uma vez que a coleta de linha de base está submetida ao ambiente, ou seja, o ar. Logo ao submeter os quatro sensores a substâncias diferentes pode ser possivelmente distinguíveis pela combinação linear dos sensores utilizados.

Na Figura 6.2 também podemos observar que alguns sensores aparecem afastados de seu agrupamento. Esta separação pode caracterizar a presença de *outliers* ou ruídos nos dados, os quais podem dificultar a fase de classificação dos odores dependendo do tipo de classificador utilizado.

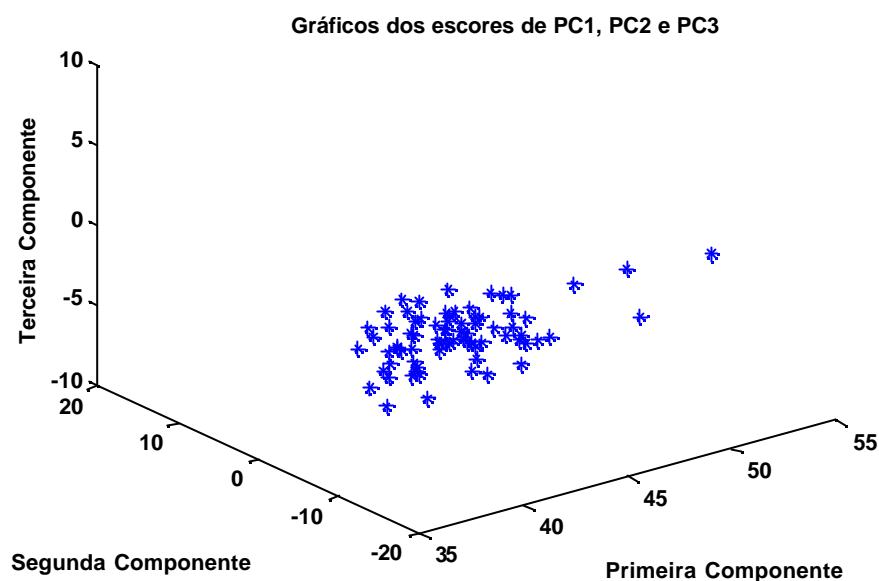


Figura 6.2 – Projeção dos sensores em três componentes principais construídas via matriz de covariâncias

A Tabela 6.3 apresenta a correlação das três componentes principais escolhidas para representar a variação dos dados com os respectivos sensores utilizados.

Tabela 6.3 – Correlação entre as três componentes principais escolhidas e os sensores utilizados.

| Sensor | Componentes principais | | |
|-----------------|------------------------|---------------|---------------|
| | \hat{Y}_1 | \hat{Y}_2 | \hat{Y}_3 |
| S01B3D3 | 0,2009 | -0,9514 | 0,1141 |
| S02BB3D3 | 0,9267 | 0,1571 | 0,3027 |
| S02BR3D3 | 0,9128 | -0,1246 | -0,0204 |
| S02BG3D3 | 0,8938 | 0,0321 | -0,3967 |

De um modo geral, as três componentes podem ser interpretadas da seguinte forma:

1ª componente: é basicamente um índice de desempenho global dos sensores em relação ao ambiente durante a coleta de linha de base. O coeficiente de maior grandeza numérica desta componente é relativo à S02BB3D3, variável de segunda maior variância amostral (ver Tabela 6.2), e o que menor representa é o S01B3D3, menor variância amostral (ver Tabela 6.2).

Quanto maior os valores dos sensores, maior será o valor numérico da componente. Pela Tabela 6.5, constata-se que essa componente está correlacionada com os quatro sensores, principalmente com S02BB3D3 e S02BR3D3, maiores variância amostral (ver Tabela 6.2).

2ª componente: A Tabela 6.2 mostra mais uma vez que o sensor S02BB3D3, melhor representa a coleta de linha de base dos sensores, possuindo maior correlação. Se considerarmos apenas as duas componentes principais teríamos os sensores S02BB3D3 e S02BG3D3 que melhor representaria o ambiente durante a coleta de linha de base.

3ª componente: Na terceira componente, maior uma vez, S02BB3D3 melhor representa a coleta de linha de base dos sensores, isto quer dizer que o sensor S02BB3D3 melhor representa o ambiente no momento da coleta de linha de base.

Os sensores também podem ser expostos a substâncias voláteis independentemente, ou seja, a partir de um dos sensores avaliar a presença de tal substância. Logo tomaremos os dados citados na Tabela 4.2 que apresenta dados relativos aos quatro tipos de sensores durante a coleta de linha de base, um total de 308 sensores, utilizando todo o sinal dos sensores (180s) como variáveis para formar a combinação linear de cada componente principal estimada:

$$\hat{Y}_j = e_{j1} \hat{X}_1 + e_{j2} \hat{X}_2 + \dots + e_{j180} \hat{X}_{180}, j = 1, 2, \dots, 180$$

onde \hat{X}_1 é a resistência elétrica no tempo t(1), \hat{X}_2 resistência elétrica no tempo t(2), e assim sucessivamente.

Para obter as componentes utilizamos PCA via matriz de covariâncias, cujas porcentagens de variância total explicadas pelas componentes são respectivamente 99,90% para \hat{Y}_1 ; 0,085% para \hat{Y}_2 e 0,005% para \hat{Y}_3 . Juntas, as três componentes principais \hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3 representam aproximadamente 99,99% da variância total do vetor original padronizado.

Na Figura 6.3 visualiza-se a projeção destes valores, nas três primeiras componentes principais, e também podemos observar que alguns sensores aparecem afastados de seu

agrupamento. Esta separação pode caracterizar a presença de *outliers* ou ruídos nos dados como já comentado anteriormente.

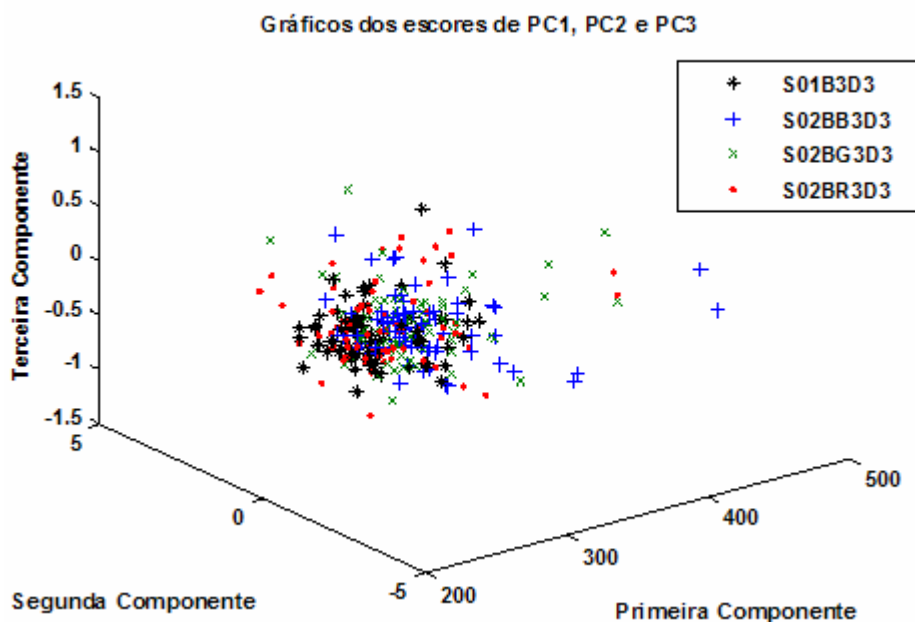


Figura 6.3 – Projeção dos sensores em três componentes principais construídas via matriz de covariâncias

Porém como a terceira componente só representa 0,005% da variância dos sensores utilizados, logo podemos utilizar apenas as duas primeiras componentes principais. A projeção destes valores, nas duas primeiras componentes principais pode ser vista na Figura 6.4

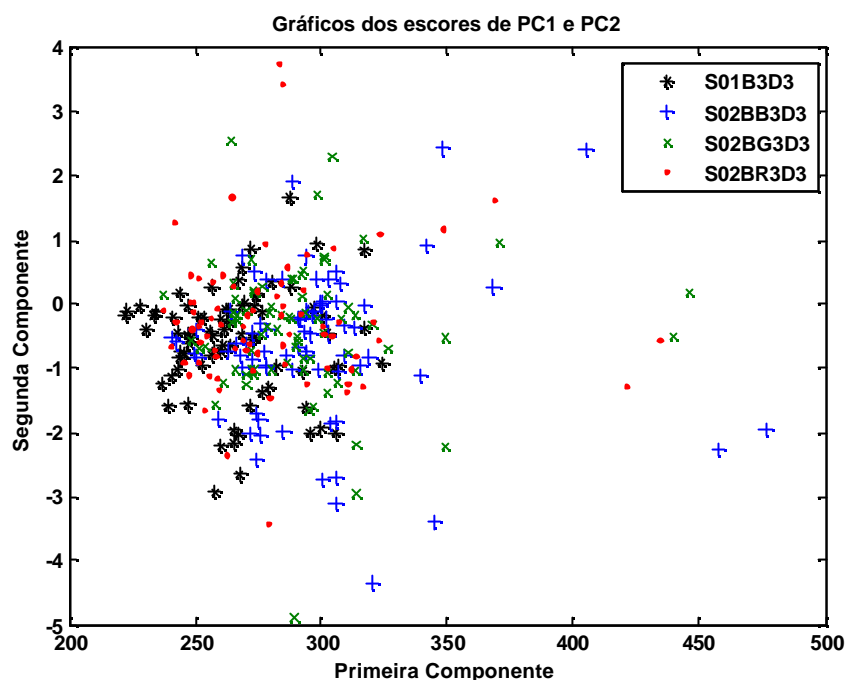


Figura 6.4 – Projeção dos sensores em duas componentes principais construídas via matriz de covariâncias

As formações dos grupos dos sensores dependem de suas propriedades físico-químicas, porém podemos amplificar os sinais da resistência elétrica dos sensores eletronicamente, logo consideraremos uma constante A_i , $i = 1, 2, \dots, p$ de amplificação. Considerando uma amplificação para cada tipo de sensor, então a combinação linear de cada componente principal será estimada por:

$$\hat{Y}_j = e_{j1} \hat{X}_1 A_i + e_{j2} \hat{X}_2 A_2 + \dots + e_{j180} \hat{X}_{180} A_i, \quad i=1, 2, 3, 4 \text{ e } j = 1, 2, \dots, 180$$

onde X_1 é a resistência elétrica no tempo $t(1)$, X_2 resistência elétrica no tempo $t(2)$, e assim sucessivamente. Através de PCA via matriz de covariâncias cujas porcentagens de variância total explicadas pelas componentes, respectivamente é de aproximadamente 99,99% para \hat{Y}_1 ; 0,0027% para \hat{Y}_2 e 0,00019% para \hat{Y}_3 . Juntas, as três componentes principais \hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3 representam aproximadamente 100% da variância total do vetor original padronizado.

Na Figura 6.5 é apresentada a projeção destes valores, nas três primeiras componentes principais, e também podemos observar os novos agrupamentos dos sensores utilizados.

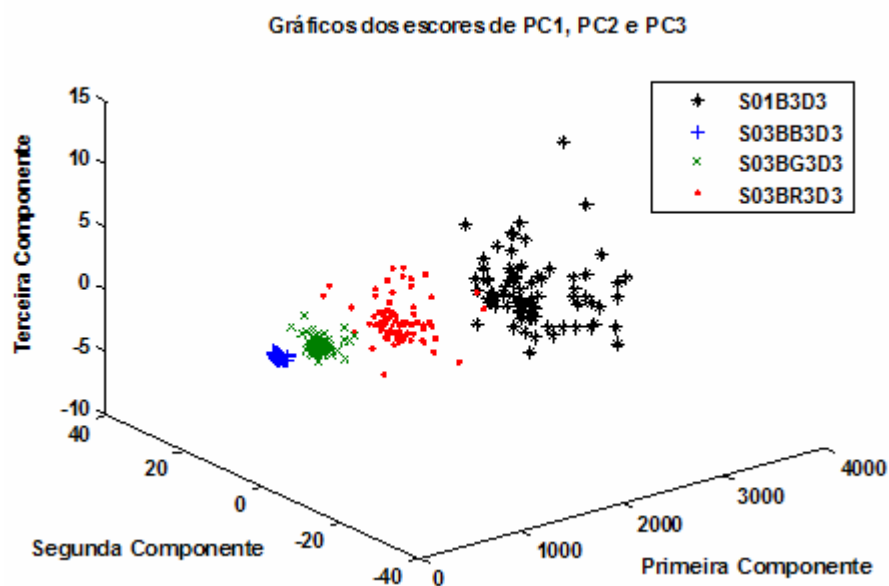


Figura 6.5 – Projeção dos sensores amplificados em três componentes principais construídas via matriz de covariâncias

6.2.1.2 PCA via matriz de correlação

Nesta seção estudaremos os sensores da Tabela 6.2 por PCA, porém utilizaremos a matriz de correlação para estimar as componentes principais. Para o conjunto de dados a matriz de correlação amostral é dada por:

$$R_{4 \times 4} = \begin{bmatrix} 1,0000 & 0,1034 & 0,2205 & 0,1459 \\ 0,1034 & 1,0000 & 0,7589 & 0,7459 \\ 0,2205 & 0,7589 & 1,0000 & 0,7398 \\ 0,1459 & 0,7459 & 0,7398 & 1,0000 \end{bmatrix}$$

Os autovalores da matriz de correlação amostral são: $\hat{I}_1 = 2,5441$, $\hat{I}_2 = 0,9619$, $\hat{I}_3 = 0,2624$ e $\hat{I}_4 = 0,2316$. A variância total é igual a $\hat{I}_1 + \hat{I}_2 + \hat{I}_3 + \hat{I}_4 = 4,00$ e os correspondentes autovetores normalizados são:

$$\hat{e}_1 = \begin{bmatrix} -0,1733 \\ -0,5667 \\ -0,5740 \\ -0,5651 \end{bmatrix} \quad \hat{e}_2 = \begin{bmatrix} 0,9783 \\ -0,1728 \\ -0,0129 \\ -0,1137 \end{bmatrix} \quad \hat{e}_3 = \begin{bmatrix} -0,0253 \\ 0,3589 \\ 0,4551 \\ -0,8145 \end{bmatrix} \quad \hat{e}_4 = \begin{bmatrix} 0,1107 \\ 0,7212 \\ -0,6806 \\ -0,0659 \end{bmatrix}$$

Portanto as três componentes principais para este estudo são:

$$\hat{Y}_1 = -0,1733Z_1 - 0,5667Z_2 - 0,5740Z_3 - 0,5651Z_4$$

$$\hat{Y}_2 = 0,9783Z_1 - 0,1728Z_2 - 0,0129Z_3 - 0,1137Z_4$$

$$\hat{Y}_3 = -0,0253Z_1 + 0,3589Z_2 + 0,4551Z_3 - 0,8145Z_4$$

$$\hat{Y}_4 = 0,1107Z_1 + 0,7212Z_2 - 0,6806Z_3 - 0,0659Z_4$$

Sendo Z_1 , Z_2 , Z_3 e Z_4 definidos respectivamente por:

$$Z_1 = \frac{X_1 - 19,819}{0,03836}$$

$$Z_2 = \frac{X_2 - 22,173}{0,05823}$$

$$Z_3 = \frac{X_3 - 20,903}{0,06296}$$

$$Z_4 = \frac{X_4 - 21,768}{0,05658}$$

As porcentagens de variância total explicadas pelas componentes são, respectiva-

mente 63,60% para \hat{Y}_1 ; 24,05% para \hat{Y}_2 ; 6,56% para \hat{Y}_3 e 5,79% para \hat{Y}_4 . Juntas, as três

componentes principais \hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3 representam aproximadamente 94,21% da variância total do vetor original padronizado.

A relação entre a variância explicada em cada uma das componentes principais pode ser facilmente visualizada na Figura 6.6.

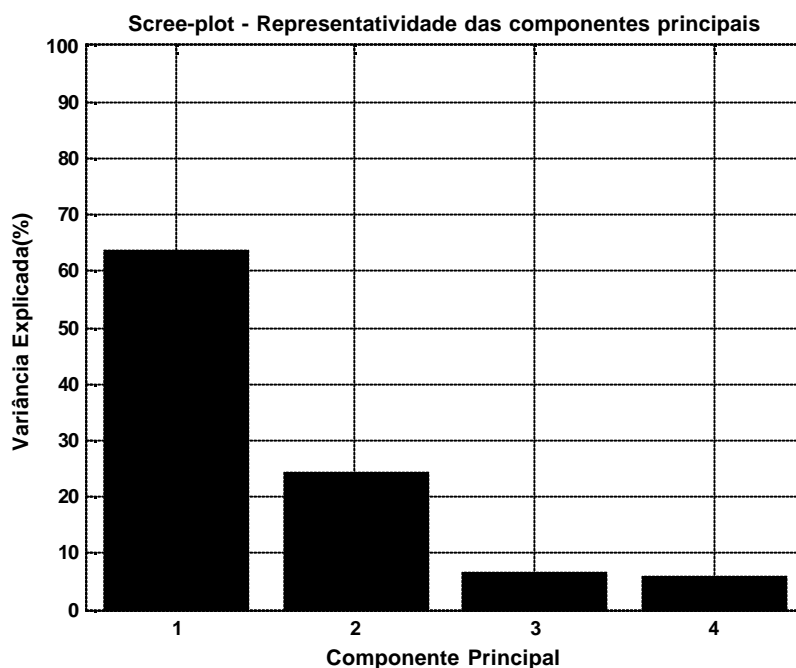


Figura 6.6 – variância explicada em cada uma das componentes principais para os sensores via matriz de correlação

A projeção destes valores, nas três primeiras componentes principais pode ser vistas na Figura 6.7. Também se observa a presença de *outliers* nos dados.

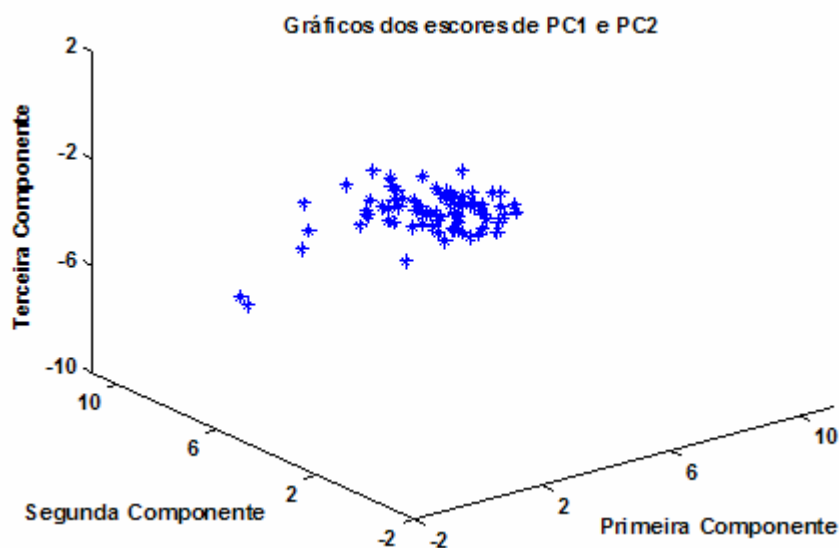


Figura 6.7 – Projeção dos sensores utilizados em três componentes principais construídas via matriz de correlação

A Figura 6.8 mostra o gráfico da distribuição da porcentagem da variância total explicada pelas três componentes para os sensores de aroma segundo as análises pelas matrizes de covariâncias e de correlação. A porcentagem total de explicação pela matriz de correlação (94,21%) é maior que a matriz de covariâncias (93,25%). Entretanto a porcentagem de explicação pela matriz de covariâncias é mais concentrada na primeira componente, ao contrário do que ocorre com a matriz de correlação, na qual a concentração é menor.

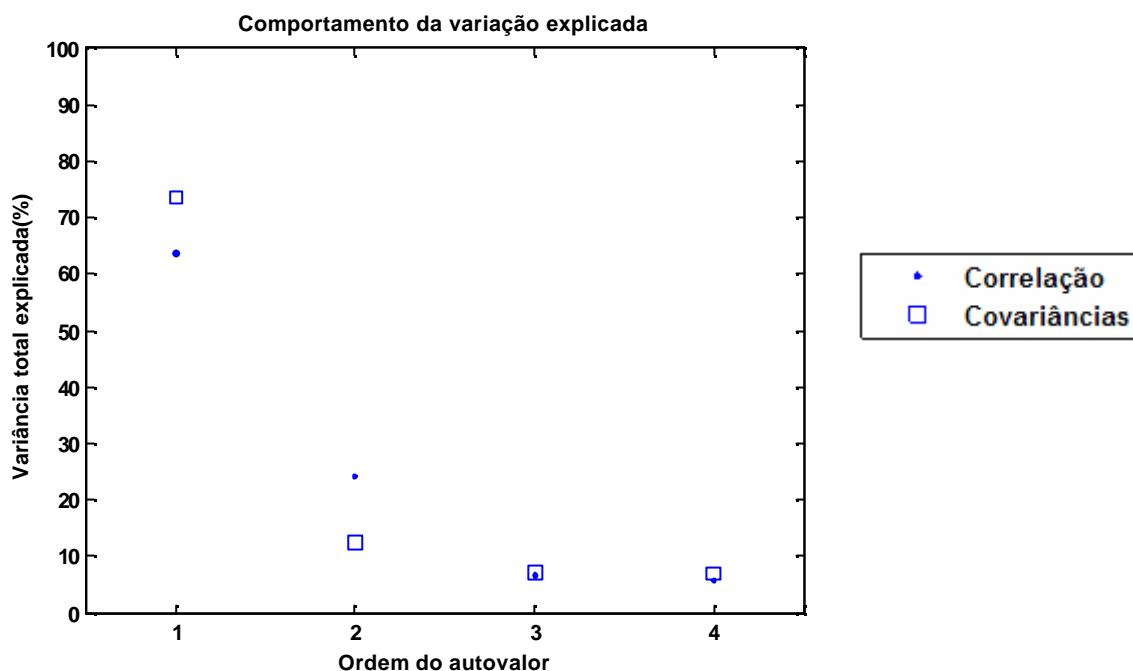


Figura 6.8 – Distribuição das variâncias explicadas utilizando a estimação de componentes principais via matriz de covariâncias e correlação.

6.2.2 Análise dos sensores expostos à urina de pacientes diabéticos e não-diabéticos.

Na seção anterior avaliamos a primeira etapa de aquisição referente à coleta de linha de base dos sensores por PCA via matriz de covariâncias e correlação. Agora iremos analisar os sinais de odores dos sensores submetidos ao odor da urina de pacientes diabéticos e não diabéticos e analisar se os sensores são capazes de distinguir a doença dos pacientes. Serão analisadas amostras de urinas de 77 pacientes sendo 18 diabéticos e 59 não diabéticos, todos expostos aos quatro tipos de sensores durante um período de 600s.

Na seção 4.3.5, do capítulo quatro observamos a curva característica dos sensores expostos à urina de um paciente não-diabético. Na Tabela 6.4 e Tabela 6.5 podemos observar algumas estatísticas descritivas para repetições de aquisição dos dados referentes à segunda

etapa de aquisição, sendo a primeira para um paciente diabético e a segunda de um não-diabético, respectivamente.

Tabela 6.4 – Estatísticas descritivas para a aquisição dos dados da exposição das amostras de urinas dos pacientes diabéticos aos sensores de aroma.

| Sensor | Estatística descritiva | | | | | |
|-----------------|------------------------|---------|---------|---------------|--------|--------|
| | n | Média | Mediana | Desvio Padrão | Mínimo | Máximo |
| S01B3D3 | 18 | 20,5019 | 20,325 | 0,0997 | 17,06 | 23,84 |
| S02BB3D3 | 18 | 22,2575 | 21,280 | 0,1623 | 18,02 | 37,00 |
| S02BR3D3 | 18 | 21,5967 | 20,850 | 0,1366 | 18,06 | 33,60 |
| S02BG3D3 | 18 | 22,7844 | 21,685 | 0,1728 | 20,30 | 35,20 |

Tabela 6.5 – Estatísticas descritivas para a aquisição dos dados da exposição das amostras de urinas dos pacientes não-diabéticos aos sensores de aroma.

| Sensor | Estatística descritiva | | | | | |
|-----------------|------------------------|---------|---------|---------------|--------|--------|
| | n | Média | Mediana | Desvio Padrão | Mínimo | Máximo |
| S01B3D3 | 59 | 20,5821 | 20,42 | 0,1431 | 16,62 | 26,06 |
| S02BB3D3 | 59 | 23,4250 | 23,20 | 0,1545 | 19,40 | 37,40 |
| S02BR3D3 | 59 | 21,5559 | 20,70 | 0,1172 | 18,06 | 34,00 |
| S02BG3D3 | 59 | 22,3689 | 22,16 | 0,1308 | 17,70 | 34,60 |

Faremos o estudo utilizando PCA via matriz de covariâncias e matriz de correlação e ao final compararemos a as técnicas.

6.2.2.1 PCA via matriz de covariâncias amostrais

Para calcular as componentes principais, as amostras de urina foram submetidas aos quatro tipos de sensores e iremos considerar os dados originais estão sendo representados pela média de cada um dos quatro sensores, de um total de 308 sensores analisados. Logo a partir da matriz de covariâncias e utilizando à média dos sensores como variáveis para formar combinação linear de cada componente principal estimada:

$$\hat{Y}_j = e_j \hat{X} = e_{j1} \hat{X}_1 + e_{j2} \hat{X}_2 + e_{j3} \hat{X}_3 + e_{j4} \hat{X}_4, \quad j = 1, 2, 3, 4.$$

onde X_1 é média de S01B3D3, X_2 a média de S02BB3D3, X_3 média de S02BR3D3 e X_4 a média de S02BG3D3. A escolha das componentes estará relacionada à porcentagem de representatividade da variância total dos dados, onde juntas devem representar mais que 90% dos dados originais, conforme a equação 5.14 e as condições citadas na seção 5.2.4.2, do capítulo 5.

Os autovalores da matriz de covariâncias amostral são: $\hat{I}_1 = 20,463$, $\hat{I}_2 = 3,3007$, $\hat{I}_3 = 2,2204$ e $\hat{I}_4 = 2,0707$. A variância total é igual a $\hat{I}_1 + \hat{I}_2 + \hat{I}_3 + \hat{I}_4 = 28,0544$ e os correspondentes autovetores normalizados são:

$$\hat{e}_1 = \begin{bmatrix} 0,0670 \\ 0,6289 \\ 0,5748 \\ 0,5191 \end{bmatrix} \quad \hat{e}_2 = \begin{bmatrix} -0,9674 \\ 0,2083 \\ -0,1410 \\ 0,0287 \end{bmatrix} \quad \hat{e}_3 = \begin{bmatrix} 0,0328 \\ -0,2765 \\ -0,4618 \\ 0,8421 \end{bmatrix} \quad \hat{e}_4 = \begin{bmatrix} -0,2419 \\ -0,6960 \\ 0,6606 \\ 0,1431 \end{bmatrix}$$

Portanto as três componentes principais para este estudo são:

$$\hat{Y}_1 = 0,0670X_1 + 0,6289X_2 + 0,5748X_3 + 0,5191X_4$$

$$\hat{Y}_2 = -0,9674X_1 + 0,2083X_2 - 0,1410X_3 + 0,0287X_4$$

$$\hat{Y}_3 = 0,0328X_1 - 0,2765X_2 - 0,4618X_3 + 0,8421X_4$$

As porcentagens de variância total explicadas pelas componentes são respectivamente 72,94% para \hat{Y}_1 ; 11,76% para \hat{Y}_2 e 7,91% para \hat{Y}_3 . Juntas, as três componentes principais \hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3 representam aproximadamente 92,62% da variância total.

A relação entre a variância explicada em cada uma das componentes principais pode ser facilmente visualizada na Figura 6.9.

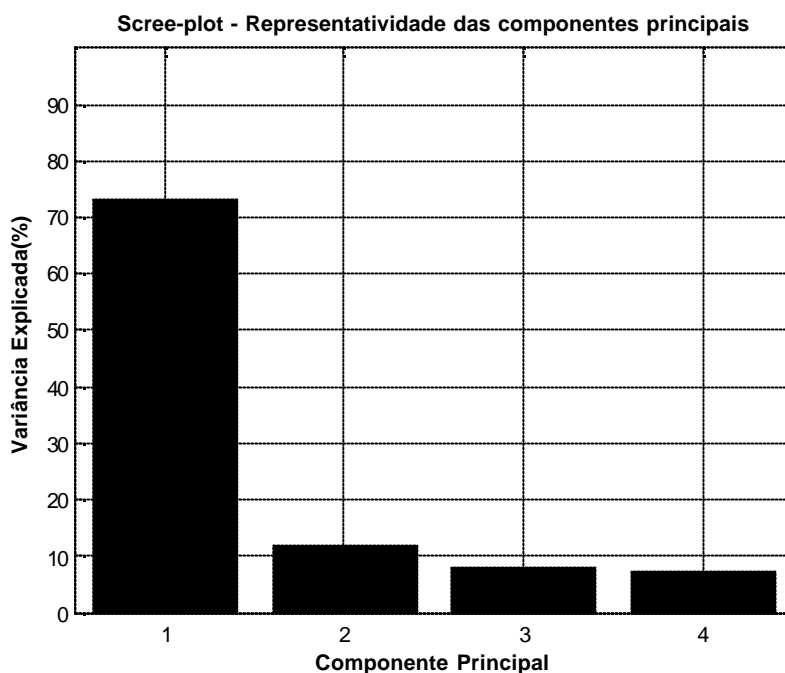


Figura 6.9 – Variância explicada em cada uma das componentes principais para os sensores expostos a urina dos pacientes via matriz de covariâncias

A projeção destes valores, nas três primeiras componentes principais pode ser vistas na Figura 6.10. A partir desse gráfico observamos os clusters formados entre os pacientes diabéticos e não-diabéticos.

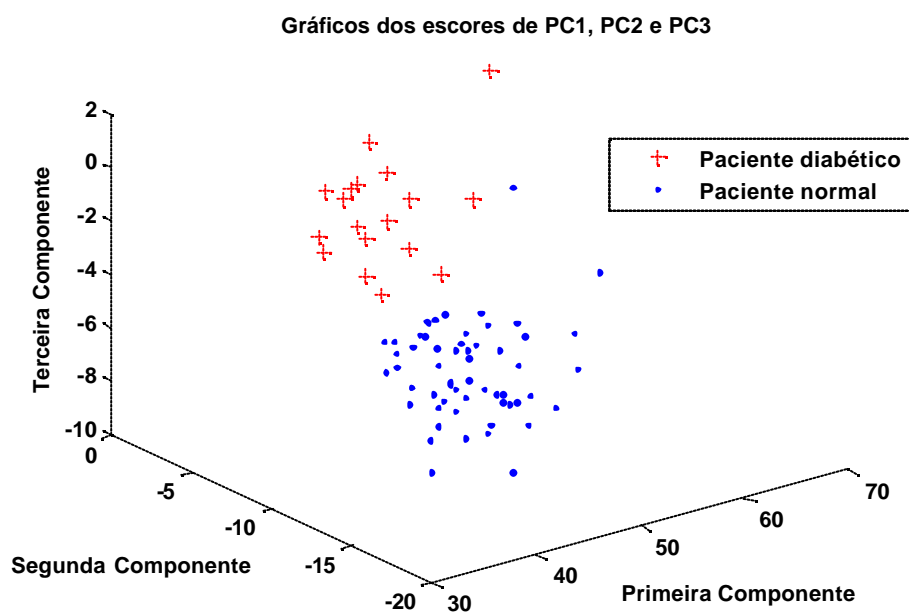


Figura 6.10 – Projeção dos sensores expostos a urina dos pacientes em três componentes principais construídas via matriz de covariâncias

A Tabela 6.6 apresenta a correlação das três componentes principais escolhidas para representar a variação dos dados com os respectivos sensores utilizados.

Tabela 6.6 – Correlação entre as três componentes principais escolhidas e a resposta dos sensores a urina dos pacientes

| Sensor | Componentes principais | | |
|-----------------|------------------------|---------------|---------------|
| | \hat{Y}_1 | \hat{Y}_2 | \hat{Y}_3 |
| S01B3D3 | 0,2040 | -0,8724 | 0,0677 |
| S02BB3D3 | 0,8826 | 0,0760 | 0,0362 |
| S02BR3D3 | 0,8815 | -0,0296 | 0,0856 |
| S02BG3D3 | 0,8294 | 0,1128 | -0,1397 |

6.2.2.2 PCA via matriz de correlação

Nesta seção estudaremos os sinais dos sensores expostos à urina dos pacientes via matriz de correlação. Também serão tomadas as médias das respostas dos sensores. Para o conjunto de dados a matriz de correlação amostral é dada por:

$$R_{4 \times 4} = \begin{bmatrix} 1,0000 & 0,0943 & 0,1679 & 0,1256 \\ 0,0943 & 1,0000 & 0,7548 & 0,7295 \\ 0,1679 & 0,7548 & 1,0000 & 0,7091 \\ 0,1256 & 0,7295 & 0,7091 & 1,0000 \end{bmatrix}$$

Os autovalores da matriz de correlação amostral são: $\hat{I}_1 = 2,496$, $\hat{I}_2 = 0,9703$, $\hat{I}_3 = 0,2948$ e $\hat{I}_4 = 0,2388$. A variância total é igual a $\hat{I}_1 + \hat{I}_2 + \hat{I}_3 + \hat{I}_4 = 4,00$ e os correspondentes autovetores normalizados são:

$$\hat{e}_1 = \begin{bmatrix} -0,1480 \\ -0,5743 \\ -0,5740 \\ -0,5646 \end{bmatrix} \quad \hat{e}_2 = \begin{bmatrix} 0,9863 \\ -0,1348 \\ -0,0323 \\ -0,0886 \end{bmatrix} \quad \hat{e}_3 = \begin{bmatrix} -0,0203 \\ 0,2469 \\ 0,5356 \\ -0,8005 \end{bmatrix} \quad \hat{e}_4 = \begin{bmatrix} 0,0689 \\ 0,7688 \\ -0,6097 \\ -0,1802 \end{bmatrix}$$

Portanto as três componentes principais para este estudo são:

$$\hat{Y}_1 = -0,1480Z_1 - 0,5743Z_2 - 0,5740Z_3 - 0,5646Z_4$$

$$\hat{Y}_2 = 0,9863Z_1 - 0,1348Z_2 - 0,0323Z_3 - 0,0886Z_4$$

$$\hat{Y}_3 = -0,0203Z_1 + 0,2469Z_2 + 0,5356Z_3 - 0,8005Z_4$$

Sendo Z_1 , Z_2 , Z_3 e Z_4 definidos respectivamente por:

$$Z_1 = \frac{X_1 - 19,819}{0,03836}$$

$$Z_2 = \frac{X_2 - 22,173}{0,05823}$$

$$Z_3 = \frac{X_3 - 20,903}{0,06296}$$

$$Z_4 = \frac{X_4 - 21,768}{0,05658}$$

As porcentagens de variância total explicadas pelas componentes são respectivamente 62,40% para \hat{Y}_1 ; 24,25% para \hat{Y}_2 e 7,37% para \hat{Y}_3 . Juntas, as três componentes principais \hat{Y}_1 , \hat{Y}_2 e \hat{Y}_3 representam aproximadamente 94,03% da variância total do vetor original padronizado. Logo se faz necessário utilizar a quarta componente para satisfazer a condição de relevância dos dados amostrais originais estipulados, ou seja, 90% de representatividade dos dados originais.

A relação entre a variância explicada em cada uma das componentes principais pode ser facilmente visualizada na Figura 6.11.

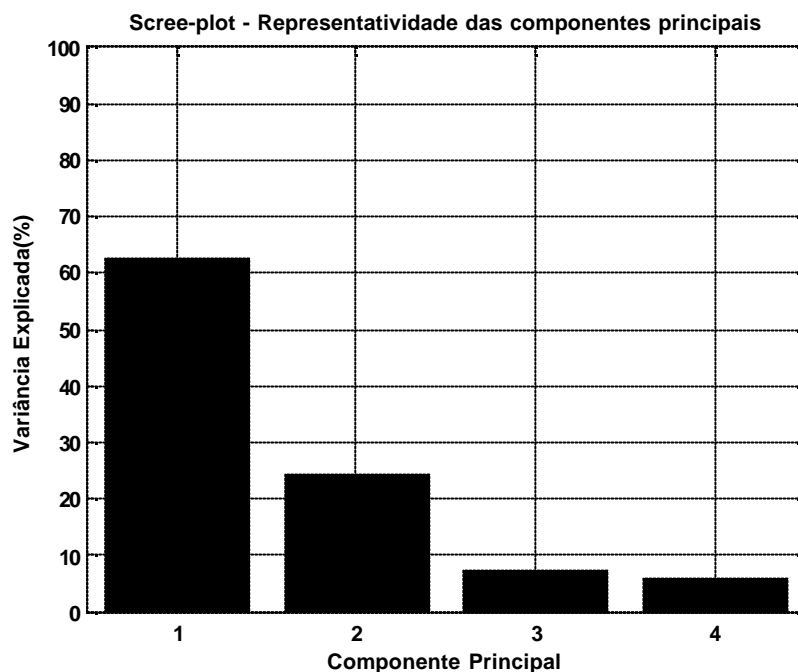


Figura 6.11 – Variância explicada em cada uma das componentes principais para os sensores expostos a urina dos pacientes via matriz de correlação

A projeção destes valores, nas três primeiras componentes principais pode ser vistas na Figura 6.12, uma vez que representar graficamente as quatro componentes seria de difícil interpretação.

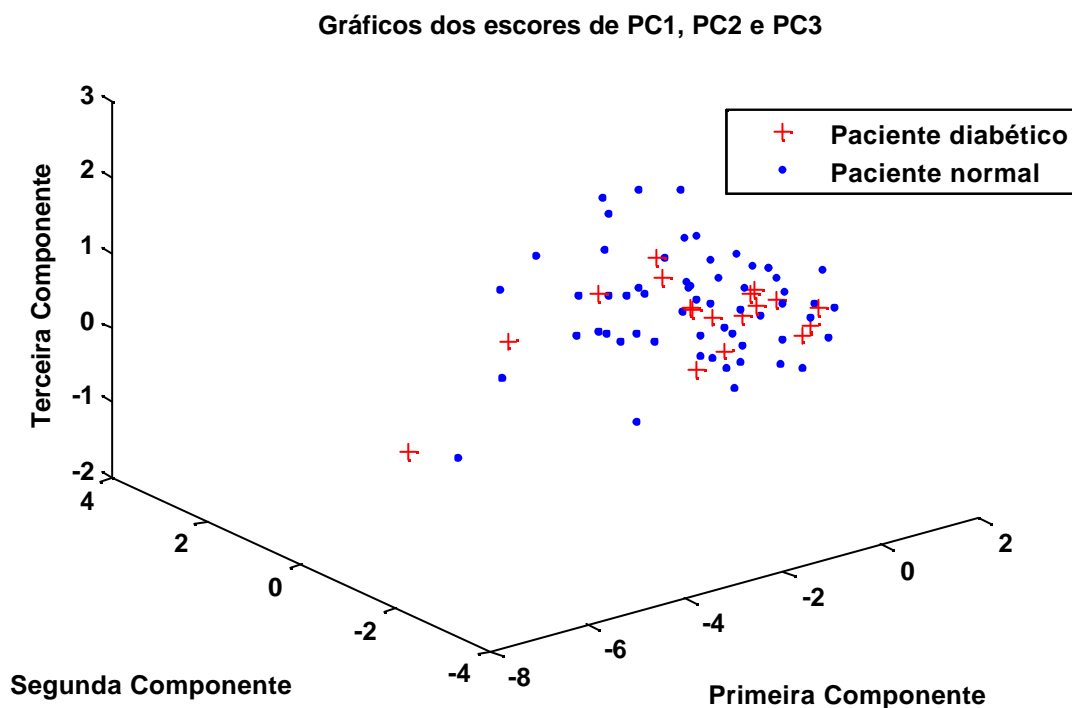


Figura 6.12 – Projeção dos sensores expostos a urina dos pacientes em três componentes principais construídas via matriz de correlação

A Figura 6.13 mostra o gráfico da distribuição da porcentagem da variância total explicada pelas três componentes para os sensores de aroma expostos a urina dos pacientes segundo as análises pelas matrizes de covariâncias e de correlação. A porcentagem total de explicação pela matriz de correlação (94,03%) é maior que a matriz de covariâncias (92,62%). A porcentagem de explicação pela matriz de covariâncias é mais concentrada na primeira componente, ao contrário do que ocorre com a matriz de correlação, na qual a concentração é menor.

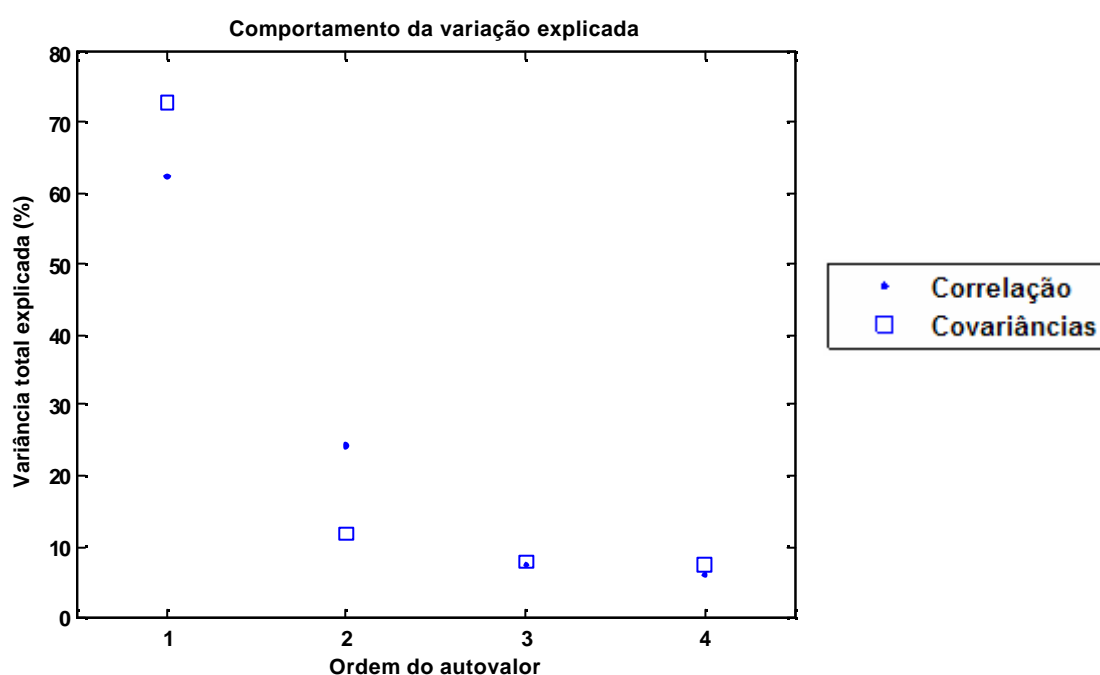


Figura 6.13 – Distribuição das variâncias explicadas utilizando a estimação de componentes principais através da matriz de covariâncias e correlação.

6.3 Classificação por *K-nearest-neighbour*

Dentre os vários métodos de classificação existentes, neste estudo utilizaremos o método de classificação *K-Nearest-Neighbour* (KNN) descrito no capítulo 5, seção 5.3. Como vimos KNN é um algoritmo de aprendizagem supervisionado onde a classe resultante de classificação se baseia na maioria dos pontos mais próximos, ou seja, os vizinhos mais próximos.

A finalidade deste algoritmo é classificar um novo objeto baseado em atributos das amostras do treinamento. Os classificadores não usam nenhum modelo ajustado, mas somente no conjunto de dados de treinamento, ou seja, na memória. Dado um ponto, encontra-se o número de K pontos (do treinamento) mais próximos. A classificação será baseada usando o voto de maioria entre a classificação dos K vizinhos. Todos os empates podem ser quebrados aleatoriamente, porém quando acontecer será tomada uma nova vizinhança $K+1$.

O classificador KNN foi aplicado à base de dados em duas aplicações distintas, a primeira levando em consideração apenas as médias das respostas dos sensores expostos à urina como descrito na Tabela 6.2, e a segunda após extração de características pela técnica PCA via matriz de covariância utilizando as três primeiras PCs, descrita no capítulo 6, seção 6.2. Como descrito anteriormente para realizar a classificação faz-se necessária dividir a base de dados em duas partes sendo um conjunto para treino e um para teste. Foram realizadas duas divisões, a primeira 50% para treino e 50% para teste e a segunda 70% para treino e 30% para teste.

Os resultados do classificador KNN podem ser visualizados na matriz de confusão. A matriz de confusão mostra a classificação desejada nas linhas e a classificação predita nas colunas. A situação ideal é quando se tem todas as respostas na linha diagonal referentes a cada classificação. As Tabelas 6.7 e 6.8 apresentam os resultados obtidos para o classificador KNN no conjunto de dados originais.

Tabela 6.7 – Matriz de confusão para as 77 amostras de urina num conjunto de 50% de treino e 50% de teste analisadas com o classificador KNN.

| Paciente | Saída do Classificador | | | |
|-----------|------------------------|--------|-----------|--------|
| | K = 1 | | K=3 | |
| | Diabético | Normal | Diabético | Normal |
| Diabético | 3 | 16 | 4 | 8 |
| Normal | 6 | 14 | 5 | 22 |

Tabela 6.8 – Matriz de confusão para as 77 amostras de urina num conjunto de 70% de treino e 30% de teste analisadas com o classificador KNN.

| Paciente | Saída do Classificador | | | |
|-----------|------------------------|--------|-----------|--------|
| | K = 1 | | K=3 | |
| | Diabético | Normal | Diabético | Normal |
| Diabético | 0 | 3 | 0 | 0 |
| Normal | 8 | 12 | 8 | 15 |

As Tabelas 6.9 e 6.10 apresentam os resultados obtidos para o classificador KNN na base de dados utilizando PCA como extrator de características.

Tabela 6.9 – Matriz de confusão para as 77 amostras de urina após PCA num conjunto de 50% de treino e 50% de teste analisadas com o classificador KNN.

| Paciente | Saída do Classificador | | | |
|-----------|------------------------|--------|-----------|--------|
| | K = 1 | | K=3 | |
| | Diabético | Normal | Diabético | Normal |
| Diabético | 8 | 0 | 8 | 0 |
| Normal | 0 | 15 | 0 | 15 |

Tabela 6.10 – Matriz de confusão para as 77 amostras de urina num conjunto de 70% de treino e 30% de teste analisadas com o classificador KNN.

| Paciente | Saída do Classificador | | | |
|-----------|------------------------|--------|-----------|--------|
| | K = 1 | | K=3 | |
| | Diabético | Normal | Diabético | Normal |
| Diabético | 8 | 0 | 8 | 0 |
| Normal | 0 | 15 | 0 | 15 |

A Tabela 6.11 mostra o desempenho geral do classificador KNN à base de odor para K=1 e K=3 fazendo uso da base de dados originais e utilizando PCA como extrator de características, sendo 50% da base de dados para treinamento e 50% para teste. E a Tabela 6.12 utilizando 70% do conjunto para treinamento e 30% para teste.

Tabela 6.11 – Desempenho geral da classificação pelo método KNN nas 77 amostras de urina num conjunto de 50% de treino e 50% de teste analisadas a partir da base de dados original e com PCA

| Paciente | Classificação | Taxa de classificação (%) | | | |
|-----------|---------------|---------------------------|-------|---------|-----|
| | | Sem PCA | | Com PCA | |
| | | K=1 | K=3 | K=1 | K=3 |
| Diabético | Acerto | 33,33 | 44,44 | 100 | 100 |
| | Erro | 66,67 | 55,56 | 0 | 0 |
| Normal | Acerto | 46,67 | 73,33 | 100 | 100 |
| | Erro | 53,33 | 26,67 | 0 | 0 |

Tabela 6.12 – Desempenho geral da classificação pelo método KNN nas 77 amostras de urina num conjunto de 70% de treino e 30% de teste analisadas a partir da base de dados original e com PCA.

| Paciente | Classificação | Taxa de classificação (%) | | | |
|-----------|---------------|---------------------------|-----|---------|-----|
| | | Sem PCA | | Com PCA | |
| | | K=1 | K=3 | K=1 | K=3 |
| Diabético | Acerto | 0 | 0 | 100 | 100 |
| | Erro | 100 | 100 | 0 | 0 |
| Normal | Acerto | 80 | 100 | 100 | 100 |
| | Erro | 20 | 0 | 0 | 0 |

O desempenho obtido para o classificador KNN usando PCA como extrator de características também obteve igual classificação utilizando as duas primeiras componentes principais em invés das três primeiras, ou seja, uma taxa de acerto de 100% na classificação dos pacientes diabéticos e não-diabéticos.

6.4 Experimento Computacional: Classificação de Padrões por RNs

Dentre os vários modelos de RNs existentes utilizamos os dois modelos de RNs descritos no capítulo anterior, as RNs MLP e RBF. Nesta seção serão apresentados os expe-

rimentos realizados com tais RNs no intuito de classificar a base de dados construída e descrita no capítulo 4.

6.4.1 Divisão da base de dados em conjuntos de treinamento, validação e teste.

A base de dados contendo 184.800 padrões foi reorganizada a fim de utilizar à média das respostas de cada sensor para cada paciente, ou seja, cada paciente será representado por um vetor de médias dos quatro sensores utilizados na construção dos padrões como mostrado na Tabela 6.2. Portanto, a base foi dividida aleatoriamente em três conjuntos de dados distintos como sugerido pelo relatório Proben1 [Prechelt, 1994]: conjunto de treinamento para as RNs, correspondendo a 50% da quantidade total de padrões; conjunto de validação, com 25% do total de padrões; e o conjunto de teste, com os 25% restantes.

6.4.2 Pré-processamento e extração de características dos dados

Nas simulações realizadas foi utilizado um método de pré-processamento global não-linear, também chamada *normalização sigmoideal* (Eq. (4.6)), de modo que os dados foram normalizados e compreendidos no intervalo $[-1, 1]$ visando à eliminação de *outliers* como descrito no capítulo 4, seção 4.3. Para a extração de características dos dados utilizou-se o PCA visando obter a redução de dimensionalidade dos dados e expressá-los de forma a pro-

duzir resultados qualitativos destacando suas similaridades e diferenças, como descrito no capítulo anterior, seção 5.2.

6.4.3 Topologia utilizada

Foram usadas RNs MLP e RBF, contendo apenas uma camada intermediária. A camada de entrada para um tipo de experimento possui quatro unidades de processamento, as quais correspondem a média dos valores dos padrões de cada sensor experimental e outra topologia utilizando três unidades de processamento de entrada provenientes das três primeiras componentes principais após PCA. A camada de saída possui uma unidade de processamento, correspondente ao prognóstico de classificação da doença Diabetes.

Foram usados dois modelos de RNs MLP, onde as unidades de processamento das camadas intermediárias e de saída para uma RN possui função de ativação logística (Eq. (5.30)) e para a outra tangente hiperbólica (Eq. (5.33)). O modelo de RN RBF possui função de ativação gaussiana com nível de significância (α) igual a 5%, 10% e 20%. As redes contêm todas as possíveis conexões de propagação direta (feedforward) entre camadas adjacentes, sem conexões entre unidade de processamento de camadas não-adjacentes.

Foram realizadas 10 simulações com 1.350 RNs, das quais foram utilizadas treze topologias distintas (com 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 30, 40 e 50 unidades de processamento na camada intermediária) com taxa de aprendizado (η) do algoritmo de aprendizagem de 1%, 5% e 10%, das quais serão avaliados os resultados. As arquiteturas das RNs simuladas serão denotadas pela nomenclatura 4*50*1, nesse caso quer dizer, quatro unidades de processamento na camada de entrada, cinquenta neurônios na camada intermediária e um neurônio na camada de saída.

6.4.4 Metodologia e treinamento

Como descrito anteriormente o algoritmo de treinamento utilizado nas RNs é o *back-propagation*. Para cada topologia, foram realizadas 10 (dez) inicializações de valores diferentes e aleatórios para as conexões (pesos) entre as camadas de processamento. Em todos os casos, foram escolhidas 200, 400, 600, 800 e 1000 interações (i).

Como critério de parada utilizou-se a quantidade máxima de interações utilizadas, porém podendo ser interrompida quando o critério GL_5 do *Proben1* [Prechelt, 1994] fosse violado. O critério GL_5 fornece uma idéia básica sobre a perda de generalização durante o treinamento da rede, sendo muito útil para evitar *overfitting*. É definido como o aumento do erro de validação em relação ao erro de validação mínimo obtido. Suas demonstrações podem ser vistas em [Prechelt, 1994].

As RNs são treinadas (ajustes dos pesos sinápticos) a partir das entradas fornecidas visando uma saída desejada, no caso a saída será 1(um) quando as entradas pertencerem a um paciente diabético e 0 (zero) quando a um paciente normal. Entretanto depois de treinadas as RNs e ao submeter o conjunto de teste as RNs produzem respostas reais, sendo assim é utilizada uma *função arredondamento*¹¹, também chamada *função degrau*, após a resposta do neurônio de saída com intuito de obter uma saída pertencente as duas classes, ou seja, pacientes diabéticos e não-diabéticos. A função arredondamento é definida por:

$$\mathbf{j}(y) = \begin{cases} 1, & \text{se } y \geq c \\ 0, & \text{se } y < c \end{cases} \quad (6.1)$$

onde y é a resposta do neurônio de saída da RN e c é uma constante de interesse.

¹¹ Na literatura da engenharia é normalmente conhecida como *função de Heaviside*

A função arredondamento é considerada para uma constante de interesse. Se a constante é 0, a sua derivada é $\mathbf{j}'(y) = 0$ para $\forall y \neq 0$ e não é definida em $y = 0$. A descontinuidade na origem e o valor nulo na derivada nos demais pontos praticamente anulam a utilidade prática dessa função $\mathbf{j}(\cdot)$, seu comportamento pode ser visto na Figura 6.14. Entretanto, a função terá fator fundamento visando à classificação dos padrões em duas classes distintas. Para os nossos experimentos será utilizado $c = 0,5$.

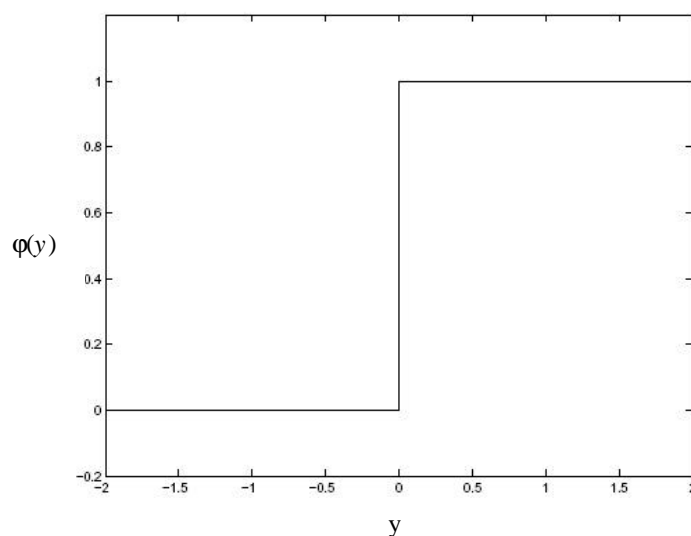


Figura 6.14 – Função degrau para $c = 0$.

Para os modelos e arquiteturas apresentadas foram realizadas um total de 1.950 simulações, utilizando como entrada das RNs os dados normalizados e os escores após PCA, dos quais 780 simulações para as RNs MLP e 1.170 simulações para RBF. Alguns dos resultados podem ser vistos em [D’Emery et al., 2007].

6.4.5 Aspectos observados

Como critério de avaliação das respostas das RNs em relação a predição do Diabetes foi utilizado o mínimo erro quadrático (EQMs), definido por:

$$EQMs = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (6.2)$$

onde y_i é o valor da variável de classificação da respectiva classe e $f(x_i)$ a resposta da RN antes da função arredondamento.

O principal parâmetro de comparação entre as abordagens investigadas foi o *Erro de Classificação* (EC) do conjunto de teste, que corresponde à razão da quantidade de padrões classificados de forma incorreta pela quantidade total de padrões após a função arredondamento.

Outro fator é o custo computacional, neste caso será feito um estudo de *Caso Escolhido* para os melhores resultados, que leva em conta o tempo médio de execução das topologias simuladas. Este fator relaciona-se diretamente à arquitetura e ao número de interações escolhidos.

Desta forma, os aspectos observados ao final do treinamento e que definiram o desempenho das abordagens estudadas foram os seguintes:

- i. Média do mínimo erro quadrático (MEQMs);
- ii. Erro de classificação do conjunto de teste (EC);
- iii. Custo computacional

6.4.6 Métodos de comparação entre as abordagens

Na maioria das áreas de pesquisa há uma busca contínua por novos métodos que melhorem em algum sentido o desempenho das abordagens existentes. De forma geral é preciso comparar as técnicas usuais com os métodos alternativos propostos. Esta comparação é uma questão importante que surge frequentemente no trabalho de pesquisa. Para se chegar a

qualquer conclusão é necessário coletar informações e fazer inferências a partir de evidências experimentais ou observacionais.

A partir das observações realizadas, para uma resposta coerente, é necessário generalizar os resultados obtidos no grupo estudado para a população de interesse. Esta tarefa exige cuidados especiais no planejamento dos experimentos, sendo que fatores que afetem a resposta avaliada devem ser controlados e considerados. O procedimento utilizado para determinar se o resultado de uma análise aponta para um caminho ou para outro é chamado Teste de Hipótese. Este procedimento é amplamente utilizado nas áreas de pesquisa onde as variáveis envolvidas estão sujeitas à grande variabilidade. A diferença entre procedimentos nos testes de hipótese são frequentemente elaboradas a partir das diferenças nas suposições que o investigador está disposto a fazer sobre a amostra dos dados. O teste de hipóteses utilizado na comparação dos classificadores investigados neste trabalho é o teste de Wilcoxon [Johnson e Wichern, 2002].

O teste de Wilcoxon pode ser usado na comparação de dados pareados, medidos nas escalas ordinal, intervalar ou de razão. Não há a exigência de que as amostras tenham distribuição normal. Tem 95% do poder do *teste t pareado* e, sua indicação, restringe-se às situações em que este último não pode ser utilizado. O princípio do teste consiste em avaliar se ocorreram modificações significativas nos dois conjuntos de dados. Quando as modificações ou diferenças são muito pequenas, elas podem ser devidas ao acaso, porém, quando são expressivas, é pouco provável que se devam ao acaso, sendo fruto de um fator causal. Em cada problema considerado, a pergunta de interesse é simplificada em duas reivindicações (hipóteses) competitivas, nas quais temos uma escolha: a hipótese nula (H_0) de encontro à hipótese alternativa (H_1).

O teste de hipóteses será utilizado para comparar os resultados obtidos por cada um dos classificadores. A partir destes testes será possível afirmar qual dos classificadores

avaliados possui o melhor desempenho na classificação dos pacientes do estudo. Os resultados foram considerados estatisticamente significativos quando p -valor $< 0,05$, ou seja, há evidências para rejeitar a hipótese nula em favor da hipótese alternativa com 95% de fator de confiança. O p -valor representa a probabilidade de se cometer o erro¹² tipo I, isto é, rejeitar a hipótese nula quando ela é verdadeira. Se a amostra é grande então p -valor é calculado usando uma aproximação normal e é fornecido o valor da estatística normal Z [Hollander e Wolfe, 1973].

6.4.7 Resultados obtidos para as RNs MLP

Como já descrito foram realizados dois experimentos distintos: um experimento onde o método de pré-processamento utilizado é somente a Normalização dos padrões captados pelos sensores de aroma do sistema de aquisição do protótipo de nariz artificial, de modo que estes valores fiquem compreendidos no intervalo $[-1, 1]$; e um experimento que utiliza o PCA como método de extração de característica reduzindo o número de entradas das RNs de quatro para três unidades de processamento.

Para cada um dos experimentos apresenta-se uma tabela contendo uma análise de melhor caso, caso médio e pior caso, escolhidos pelos erros das respostas das RNs, antes de submetidas a função arredondamento (Eq. (6.1)), bem como as arquiteturas (Arq) das RNs, valores das médias (MEQMs) (Eq. (6.2)) e desvio-padrão (s) dos erros das RNs antes da função arredondamento, o número de interações no treinamento (i), a taxa de aprendizado (h) e as respectivas funções de ativação (FA). Em seguida são tecidos comentários sobre o desempenho das RNs na solução do problema abordado.

¹² Quando se estabelece hipóteses se pode cometer dois tipos de erro estatísticos: erro tipo I (rejeita-se H_0 , quan-

Na Tabela 6.13 são apresentados os resultados obtidos no treinamento da RNs MLP com as funções de ativação logística (LOG) e tangente hiperbólica (TANHIP) fazendo uso da Normalização como método de pré-processamento

Tabela 6.13 – Ranking de desempenho das RNs MLP com a funções de ativação logística e tangente hiperbólica através do MEQMs para os dados normalizados.

| Ranking | Descrição de desempenho | | | | | |
|---------|-------------------------|------|------|--------|--------|--------|
| | Arq | h | i | MEQMs | s | FA |
| 1 | 4*50*1 | 0,01 | 200 | 4,9002 | 0,7075 | TANHIP |
| 2 | 4*20*1 | 0,01 | 800 | 4,9806 | 0,4316 | TANHIP |
| 3 | 4*10*1 | 0,05 | 200 | 4,9828 | 0,6877 | TANHIP |
| 4 | 4*50*1 | 0,10 | 200 | 4,9844 | 0,1807 | LOG |
| 5 | 4*50*1 | 0,01 | 1000 | 5,0023 | 0,5659 | TANHIP |
| ... | ... | ... | ... | ... | ... | ... |
| 193 | 4*16*1 | 0,05 | 200 | 5,4931 | 0,1431 | LOG |
| 194 | 4*14*1 | 0,01 | 1000 | 5,4957 | 0,1080 | LOG |
| 195 | 4*40*1 | 0,10 | 1000 | 5,4970 | 0,1502 | LOG |
| 196 | 4*12*1 | 0,01 | 400 | 5,5033 | 0,3763 | TANHIP |
| 197 | 4*4*1 | 0,05 | 400 | 5,5065 | 0,1737 | LOG |
| ... | ... | ... | ... | ... | ... | ... |
| 386 | 4*30*1 | 0,05 | 1000 | 7,6379 | 1,1885 | TANHIP |
| 387 | 4*40*1 | 0,10 | 1000 | 7,6384 | 1,7677 | TANHIP |
| 388 | 4*50*1 | 0,10 | 1000 | 8,0443 | 2,5460 | TANHIP |
| 389 | 4*30*1 | 0,10 | 1000 | 8,2169 | 2,1923 | TANHIP |
| 390 | 4*50*1 | 0,10 | 800 | 8,5522 | 1,6734 | TANHIP |

O desempenho das respostas das diversas arquiteturas MLP também pode ser avaliado nas Figuras 6.15 e 6.16, levando em consideração as funções de ativação logística e tangente hiperbólica, respectivamente.

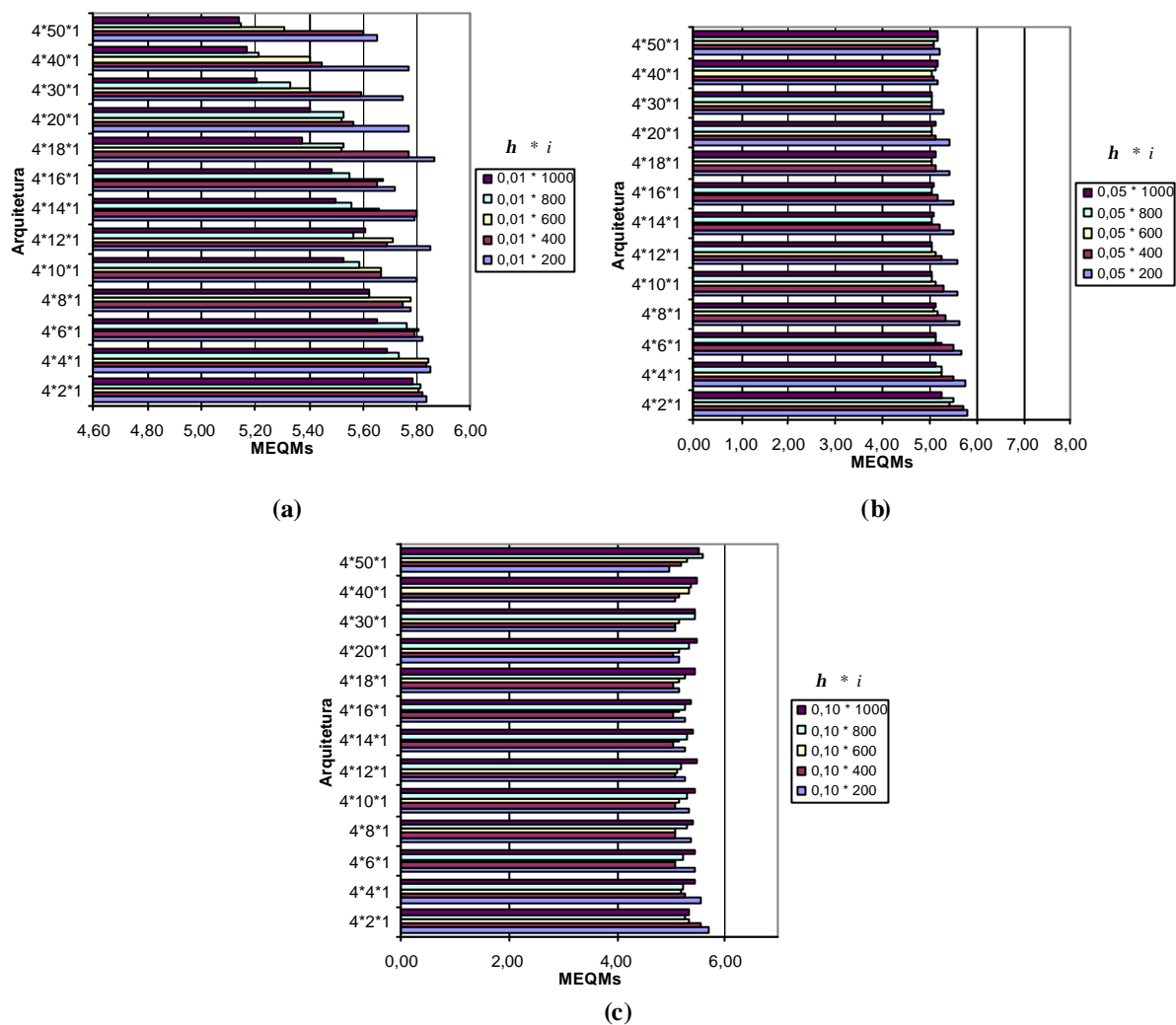


Figura 6.15 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas com a função de ativação logística usando Normalização como pré-processamento: (a) MLP com $h = 0,01$, (b) MLP com $h = 0,05$ e (c) MLP com $h = 0,10$

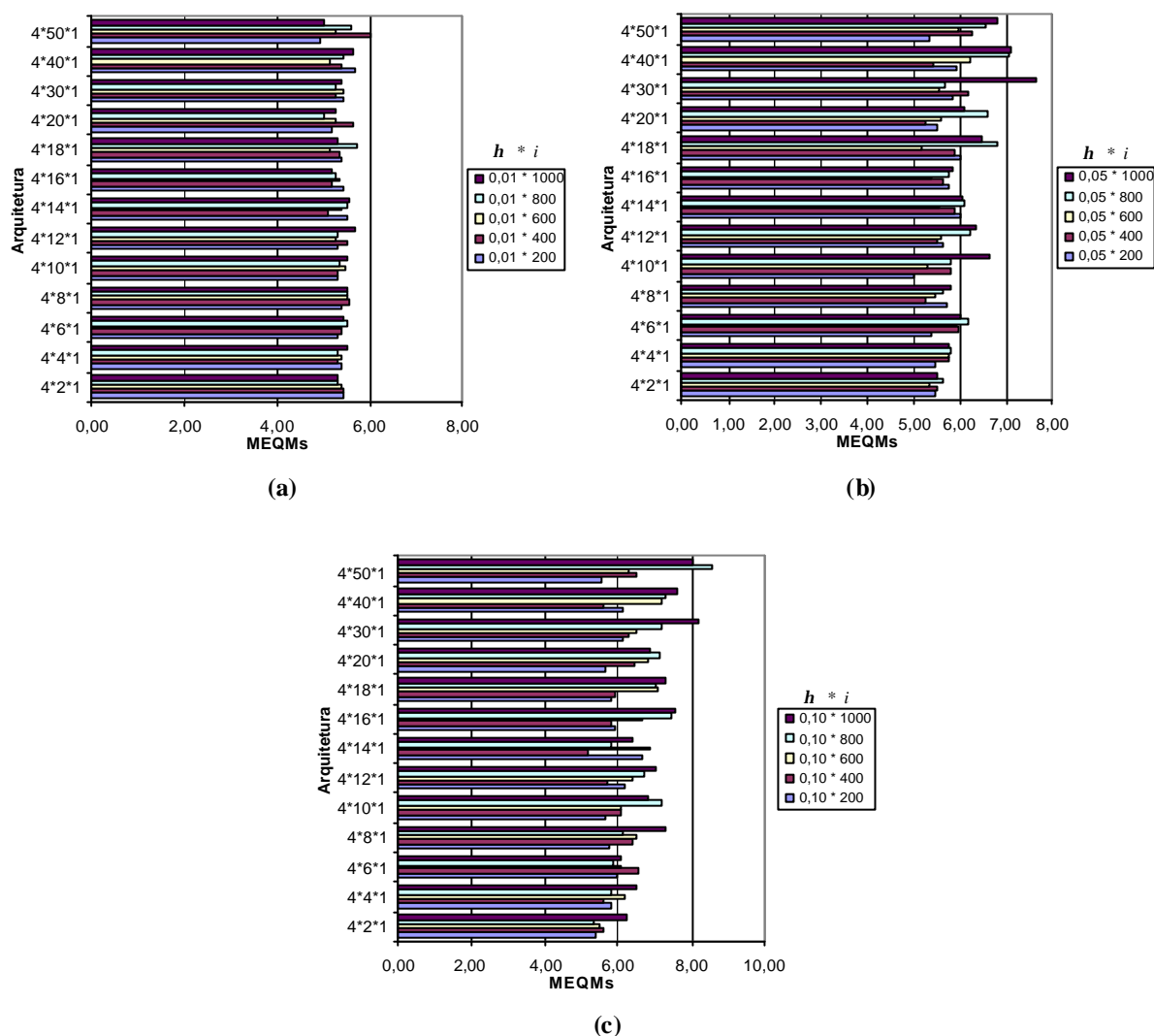


Figura 6.16 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas com a função de ativação tangente hiperbólica usando Normalização como pré-processamento: (a) MLP com $h = 0,01$, (b) MLP com $h = 0,05$ e (c) MLP com $h = 0,10$

Para estes experimentos, o menor MEQMs obtido dentre todas as topologias testadas foi de 4,9002 antes da função arredondamento, obtido pela topologia $4*50*1$, com a função de ativação tangente hiperbólica, seu erro de classificação é de 33,91%, sendo este o erro da análise de melhor caso.

Nas Tabelas 6.14 e 6.15 são apresentados os desempenhos das topologias analisadas no estudo de melhor caso, caso médio e pior caso para as RNs MLP utilizando somente a Normalização como método de pré-processamento.

Tabela 6.14 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação logística usando Normalização como pré-processamento de dados.

| Análise | Descrição de desempenho | | | | | | | |
|-------------|-------------------------|--------|------|-----|--------|--------|--------|-----|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA |
| Melhor caso | 1 | 4*50*1 | 0,10 | 200 | 4,9844 | 0,1807 | 0,3260 | LOG |
| Caso Médio | 97 | 4*30*1 | 0,01 | 800 | 5,3250 | 0,1071 | 0,3478 | LOG |
| Pior caso | 195 | 4*18*1 | 0,01 | 200 | 5,8667 | 0,1560 | 0,3478 | LOG |

Tabela 6.15 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação tangente hiperbólica usando Normalização como pré-processamento de dados.

| Análise | Descrição de desempenho | | | | | | | |
|-------------|-------------------------|--------|------|-----|--------|--------|--------|--------|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA |
| Melhor caso | 1 | 4*50*1 | 0,01 | 200 | 4,9002 | 0,7075 | 0,3391 | TANHIP |
| Caso Médio | 97 | 4*30*1 | 0,05 | 800 | 5,6674 | 0,9707 | 0,3608 | TANHIP |
| Pior caso | 195 | 4*50*1 | 0,10 | 800 | 8,5522 | 1,6734 | 0,3608 | TANHIP |

A partir dos resultados apresentados nas Tabelas 6.14 e 6.15, se percebe que de forma geral a utilização da Normalização como pré-processamento não contribuiu a fim de obter diferentes erros de classificação do conjunto de teste dos padrões apresentados. Mesmo observando que o menor MEQMs foi de 4,9002 para a arquitetura 4*50*1 com taxa de aprendizado de 1% e 200 interações fazendo uso da função de ativação tangente hiperbólica não difere significativamente das demais topologias com maiores MEQMs porém erros de classificação muito próximos.

Agora será abordado os experimentos das RNs MLP utilizando PCA como extração de características dos padrões da base de dados. Foras utilizadas as três primeiras componentes principais que juntas representam 92,62% da variância total da base dos dados, conforme visto na seção 6.2.2.1. Em virtude de termos o vetor de entrada formado por três valores a topologia das RNs passarão a ter 3 unidades de processamento na camada de entrada. Na Tabelas 6.16 são apresentados os resultados gerais obtidos.

Tabela 6.16 – Ranking de desempenho das RNs MLP com as funções de ativação logística e tangente hiperbólica através do MEQMs para dados após PCA.

| Ranking | Descrição de desempenho | | | | | |
|---------|-------------------------|------|------|---------|---------|--------|
| | Arq | h | i | MEQMs | s | FA |
| 1 | 3*30*1 | 0,05 | 800 | 0,0002 | 0,0002 | TANHIP |
| 2 | 3*50*1 | 0,01 | 1000 | 0,0002 | 0,0002 | TANHIP |
| 3 | 3*16*1 | 0,05 | 800 | 0,0002 | 0,0003 | TANHIP |
| 4 | 3*20*1 | 0,05 | 800 | 0,0002 | 0,0002 | TANHIP |
| 5 | 3*14*1 | 0,05 | 1000 | 0,0002 | 0,0002 | TANHIP |
| ... | ... | ... | ... | ... | ... | ... |
| 193 | 3*12*1 | 0,01 | 200 | 0,5845 | 1,8262 | TANHIP |
| 194 | 3*12*1 | 0,01 | 600 | 0,5864 | 1,8424 | TANHIP |
| 195 | 3*10*1 | 0,01 | 400 | 0,5947 | 1,8655 | TANHIP |
| 196 | 3*10*1 | 0,10 | 400 | 0,5968 | 1,8406 | LOG |
| 197 | 3*6*1 | 0,10 | 600 | 0,5975 | 1,8386 | LOG |
| ... | ... | ... | ... | ... | ... | ... |
| 386 | 3*2*1 | 0,10 | 1000 | 6,5333 | 1,0161 | TANHIP |
| 387 | 3*50*1 | 0,10 | 600 | 6,7065 | 4,4620 | TANHIP |
| 388 | 3*50*1 | 0,10 | 1000 | 7,1808 | 4,2183 | TANHIP |
| 389 | 3*50*1 | 0,10 | 200 | 8,7410 | 3,6895 | TANHIP |
| 390 | 3*50*1 | 0,10 | 400 | 13,8795 | 12,5107 | TANHIP |

O desempenho das respostas das diversas arquiteturas MLP também pode ser avaliado nas Figuras 6.17 e 6.18 levando em consideração as funções de ativação logística e tangente hiperbólica, respectivamente.

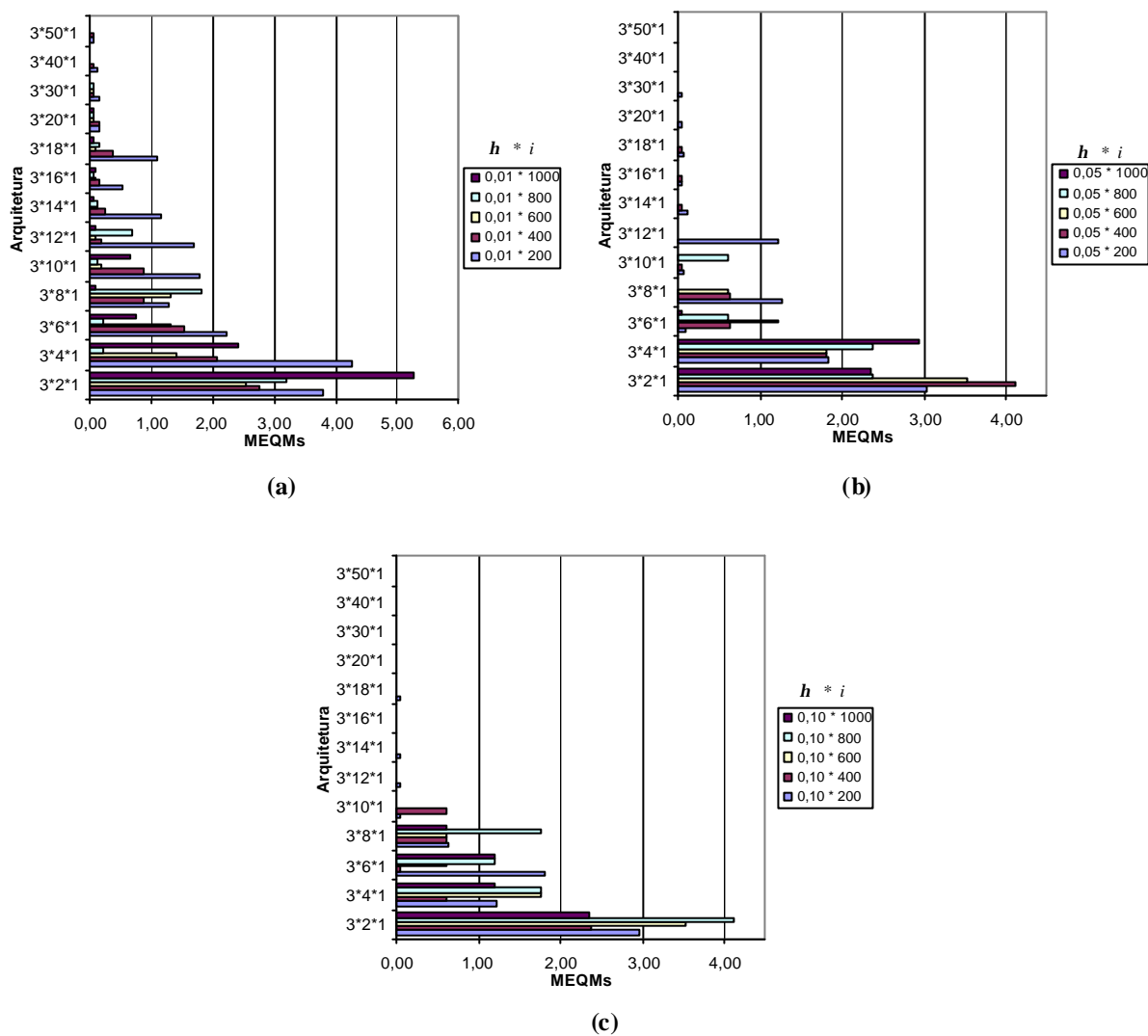


Figura 6.17 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas com a função de ativação logística após PCA: (a) MLP com $h = 0,01$, (b) MLP com $h = 0,05$ e (c) MLP com $h = 0,10$

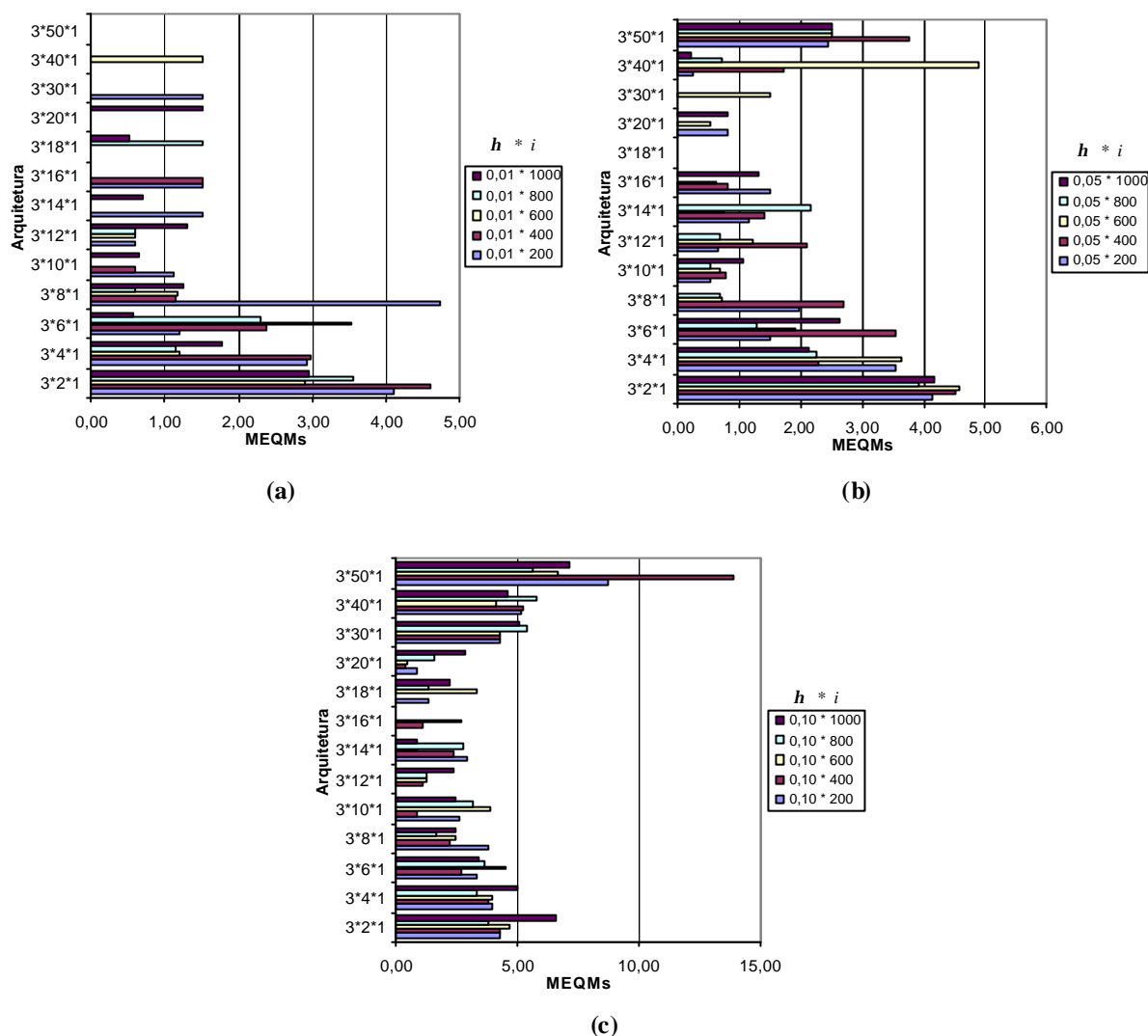


Figura 6.18 – Média dos erros das respostas das 10 simulações para as topologias MLP simuladas a função de ativação tangente hiperbólica após PCA: (a) MLP com $h = 0,01$, (b) MLP com $h = 0,05$ e (c) MLP com $h = 0,10$

Para estes experimentos usando a extração de características, o menor MEQMs obtido dentre todas as topologias testadas foi de 0,0002 antes da função arredondamento, obtido pela topologia 3*30*1 com taxa de aprendizado de 5% e 800 interações fazendo uso da função de ativação tangente hiperbólica. A rede classificou todo o conjunto de teste, ou seja, o erro de classificação foi de 0% , sendo este o erro da análise de melhor caso.

As Tabelas 6.17 e 6.18 são apresentados os desempenhos das topologias analisadas no estudo de melhor caso, caso médio e pior caso para as RNs MLP utilizando PCA como

método de extração de características. Também será utilizada uma análise de “caso escolhido” como descrito na seção 6.4.5.

Tabela 6.17 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação logística após PCA

| Análise | Descrição de desempenho | | | | | | | |
|----------------|-------------------------|--------|------|------|--------|--------|--------|-----|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA |
| Melhor caso | 1 | 3*50*1 | 0,10 | 1000 | 0,0011 | 0,0004 | 0 | LOG |
| Caso Médio | 97 | 3*10*1 | 0,10 | 200 | 0,0425 | 0,0265 | 0 | LOG |
| Pior caso | 195 | 3*2*1 | 0,01 | 1000 | 5,2631 | 1,7840 | 0,3130 | LOG |
| Caso Escolhido | 144 | 3*6*1 | 0,05 | 200 | 0,0909 | 0,0471 | 0 | LOG |

Tabela 6.18 – Análise de casos através do ranking de desempenho das RNs MLP com a função de ativação tangente hiperbólica após PCA

| Análise | Descrição de desempenho | | | | | | | |
|----------------|-------------------------|--------------|-------------|------------|---------------|---------------|---------------|---------------|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA |
| Melhor caso | 1 | 3*30*1 | 0,05 | 800 | 0,0002 | 0,0002 | 0 | TANHIP |
| Caso Médio | 97 | 3*12*1 | 0,01 | 1000 | 1,2927 | 2,7221 | 0,0695 | TANHIP |
| Pior caso | 195 | 3*50*1 | 0,10 | 400 | 13,8795 | 12,5107 | 0,2956 | TANHIP |
| Caso Escolhido | 92 | 3*6*1 | 0,01 | 200 | 1,2093 | 2,5307 | 0,0695 | TANHIP |

A partir dos resultados apresentados nas Tabelas 6.17 e 6.18, pode-se afirmar que de forma geral a utilização da técnica PCA como extração de característica teve um importante papel na classificação do conjunto de teste dos padrões apresentados por algumas arquiteturas. Mesmo tendo como menor MEQMs de 0,002 a arquitetura 4*30*1 com taxa de aprendizado de 5% e 800 interações fazendo uso da função de ativação tangente hiperbólica, esta rede apresenta um custo computacional elevado pelo grande número de interações, sendo assim dentre os casos escolhidos destaca-se a arquitetura 3*6*1 com 200 interações e taxa de aprendizado de 5% para a função de ativação logística, com MEQMs de 0,0909, classificando praticamente todo o conjunto de teste dos padrões apresentados.

A Tabela 6.19 apresenta um resumo da geral das RNs RMLP utilizando a Normalização como técnica de pré-processamento e PCA como extração de características enfatizando o percentual de classificação do conjunto de teste dos padrões das diversas arquiteturas nas análises de caso apresentados do estudo abordado.

Tabela 6.19 – Classificação das respostas das RNs MLP com as funções de ativação logística (LOG) e tangente hiperbólica (TANHIP) para os dados antes (normalização dos dados) e após a extração de características.

| Análise | Paciente | Classificação (%) | Técnica de classificação | | | |
|-----------------------|------------------|-------------------|--------------------------|--------|----------------|--------|
| | | | RN MLP sem PCA | | RN MLP com PCA | |
| | | | LOG | TANHIP | LOG | TANHIP |
| Melhor caso | Diabético | Acerto | 6,25 | 16,25 | 100 | 100 |
| | | Erro | 93,75 | 83,75 | 0 | 0 |
| | Normal | Acerto | 100 | 92,67 | 100 | 100 |
| | | Erro | 0 | 7,33 | 0 | 0 |
| Caso Médio | Diabético | Acerto | 0 | 13,75 | 100 | 80 |
| | | Erro | 100 | 86,25 | 0 | 20 |
| | Normal | Acerto | 100 | 90,67 | 100 | 100 |
| | | Erro | 0 | 9,33 | 0 | 0 |
| Pior caso | Diabético | Acerto | 0 | 8,75 | 10 | 90 |
| | | Erro | 100 | 91,25 | 90 | 10 |
| | Normal | Acerto | 100 | 81,33 | 100 | 60 |
| | | Erro | 0 | 18,67 | 0 | 40 |
| Caso Escolhido | Diabético | Acerto | - | - | 100 | 80 |
| | | Erro | - | - | 0 | 20 |
| | Normal | Acerto | - | - | 100 | 100 |
| | | Erro | - | - | 0 | 0 |

Na utilização de PCA, todas as arquiteturas das RNs MLP apresentam uma redução no MEQMs, conseqüentemente aumentando a taxa de classificação. Inicialmente as arquiteturas com menor custo computacional (menor número de neurônios na camada intermediária e de interações) possuem um elevado MEQMs e a medida que este custo computacional

aumenta as RNs encontraram um MEQMs ótimo, porém voltando a aumentar a medida que aumenta o custo computacional.

Apesar da melhora perceptível apresentada pelo segundo experimento com as RNs RMLP (com extração de característica) realizamos o teste de hipótese para verificar se esta melhora estatisticamente se confirma. Na Tabela 6.20 é apresentado o teste de hipóteses para os experimentos acima descritos. O teste é realizado a partir dos melhores resultados médios, ou seja, os resultados do “melhor caso”, obtidos por cada um dos experimentos (normalização como pré-processamento dos dados e PCA como extração de características). Neste teste verificamos se o desempenho da RN MLP com extração de características é superior ao desempenho da mesma com a Normalização como pré-processamento.

Tabela 6.20 – Estatística do teste de Wilcoxon para o desempenho das RNs MLP com normalização no pré-processamento e PCA como extração de características.

| $H_0 : \mathbf{m}_1 = \mathbf{m}_2$ | | |
|-------------------------------------|---------|---------------|
| $H_1 : \mathbf{m}_1 > \mathbf{m}_2$ | | |
| <i>p</i> -valor | z-valor | Decisão |
| 0,006 | 4,9851 | Rejeita H_0 |

A Tabela 6.210 apresenta os resultados do teste para dados pareados com nível de significância de 5%. Nesta tabela, μ_1 e μ_2 são respectivamente as médias dos erros dos experimentos com as RNs MLP com Normalização no pré-processamento e as médias dos erros dos experimentos utilizando PCA como extração de características. A partir do teste de hipótese se pode concluir que o desempenho médio das RNs MLP fazendo uso de PCA como extração de características é melhor que a das RNs MLP com Normalização no pré-processamento.

6.4.8 Resultados obtidos para as RNs RBF

Nos experimentos com as RNs RBF procura-se obter uma aproximação dos resultados visando se adequar ao problema de classificação de padrões no espaço da dimensionalidade dos dados, onde fará com que a dimensão do espaço oculto se relacione a capacidade da rede aproximar um mapeamento de entrada-saída desejado como discutido no capítulo anterior. Para isso será utilizada a função de base radial gaussiana.

Para cada um dos experimentos também serão apresentadas tabelas e gráficos contendo as mesmas análises realizadas nas simulações com as RNs MLP.

Na Tabela 6.21 são apresentados os resultados obtidos no treinamento das RNs RBF utilizando somente a Normalização como método de pré-processamento.

Tabela 6.21 – Ranking de desempenho das RNs RBF com a função de ativação gaussiana (GAUSS) com nível de significância (α) de 5%, 10% e 20% através do MEQMs para dados normalizados.

| Ranking | Descrição de desempenho | | | | | | |
|---------|-------------------------|------|------|--------|--------|-------|-----|
| | Arq | h | i | MEQMs | s | FA | a |
| 1 | 4*20*1 | 0,10 | 800 | 5,5543 | 0,2431 | GAUSS | 20% |
| 2 | 4*40*1 | 0,10 | 1000 | 5,5719 | 0,2333 | GAUSS | 20% |
| 3 | 4*6*1 | 0,10 | 800 | 5,6292 | 0,3347 | GAUSS | 5% |
| 4 | 4*18*1 | 0,10 | 800 | 5,6543 | 0,2640 | GAUSS | 5% |
| 5 | 4*50*1 | 0,10 | 200 | 5,6679 | 0,2654 | GAUSS | 20% |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 290 | 4*8*1 | 0,01 | 1000 | 5,8393 | 0,0354 | GAUSS | 10% |
| 291 | 4*30*1 | 0,01 | 1000 | 5,8394 | 0,0460 | GAUSS | 10% |
| 292 | 4*30*1 | 0,05 | 600 | 5,8399 | 0,3867 | GAUSS | 20% |
| 293 | 4*20*1 | 0,01 | 600 | 5,8400 | 0,0322 | GAUSS | 20% |
| 294 | 4*18*1 | 0,01 | 600 | 5,8402 | 0,0381 | GAUSS | 5% |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 581 | 4*4*1 | 0,10 | 400 | 6,2808 | 0,5242 | GAUSS | 5% |
| 582 | 4*30*1 | 0,10 | 1000 | 6,2841 | 0,6588 | GAUSS | 5% |
| 583 | 4*4*1 | 0,10 | 1000 | 6,4129 | 0,5912 | GAUSS | 5% |
| 584 | 4*16*1 | 0,10 | 200 | 6,4206 | 0,5608 | GAUSS | 20% |
| 585 | 4*4*1 | 0,10 | 600 | 6,6327 | 0,5649 | GAUSS | 10% |

O desempenho das respostas das diversas arquiteturas RBF também pode ser avaliado nas Figuras 6.19, 6.20 e 6.21 levando em consideração os níveis de significância utilizados na função de ativação gaussiana.

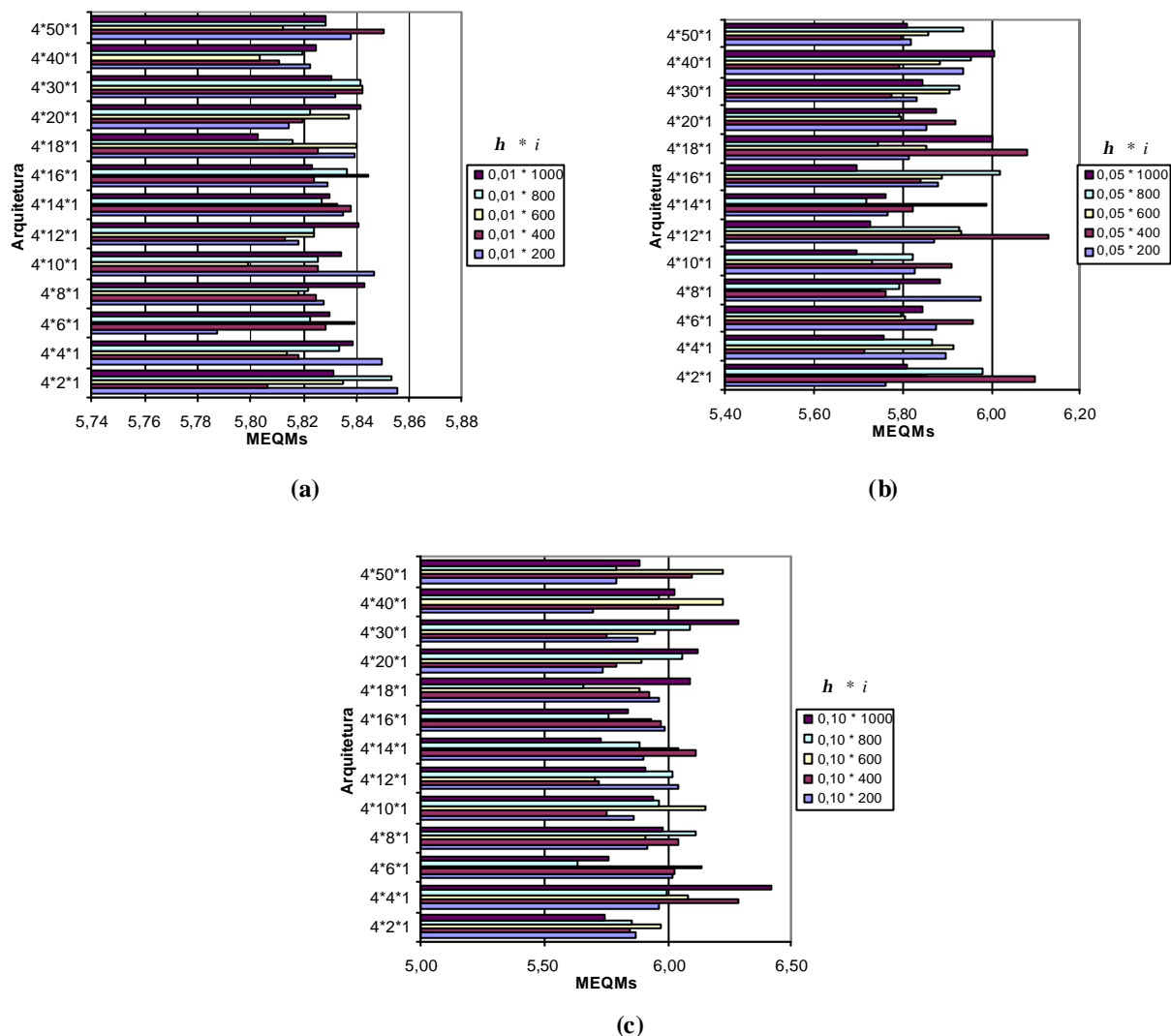


Figura 6.19 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\alpha = 5\%$ usando Normalização como pré-processamento: (a) RBF com $h = 0,01$, (b) RBF com $h = 0,05$ e (c) RBF com $h = 0,10$

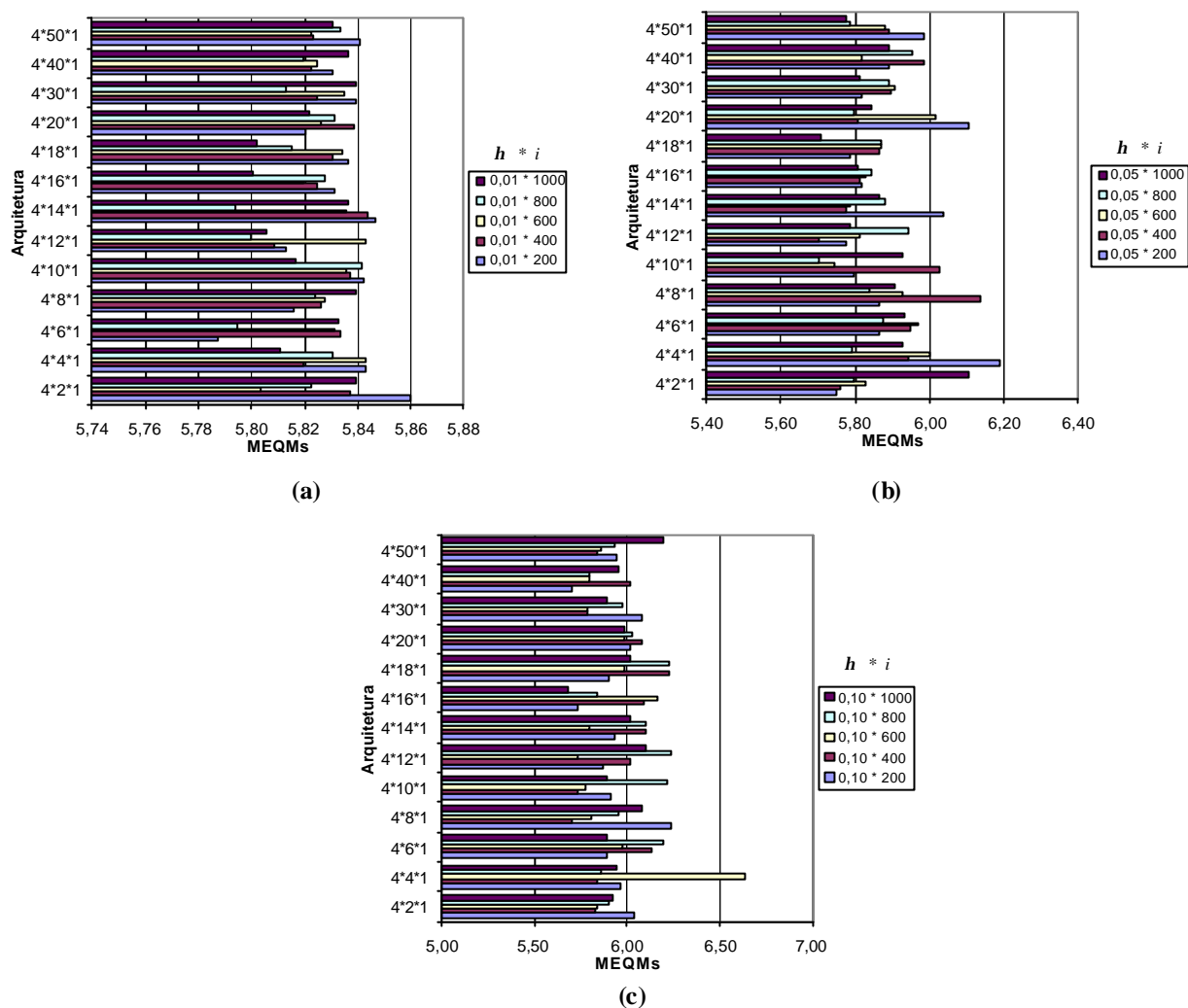


Figura 6.20 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\alpha = 10\%$ usando Normalização como pré-processamento: (a) RBF com $h = 0,01$, (b) RBF com $h = 0,05$ e (c) RBF com $h = 0,10$

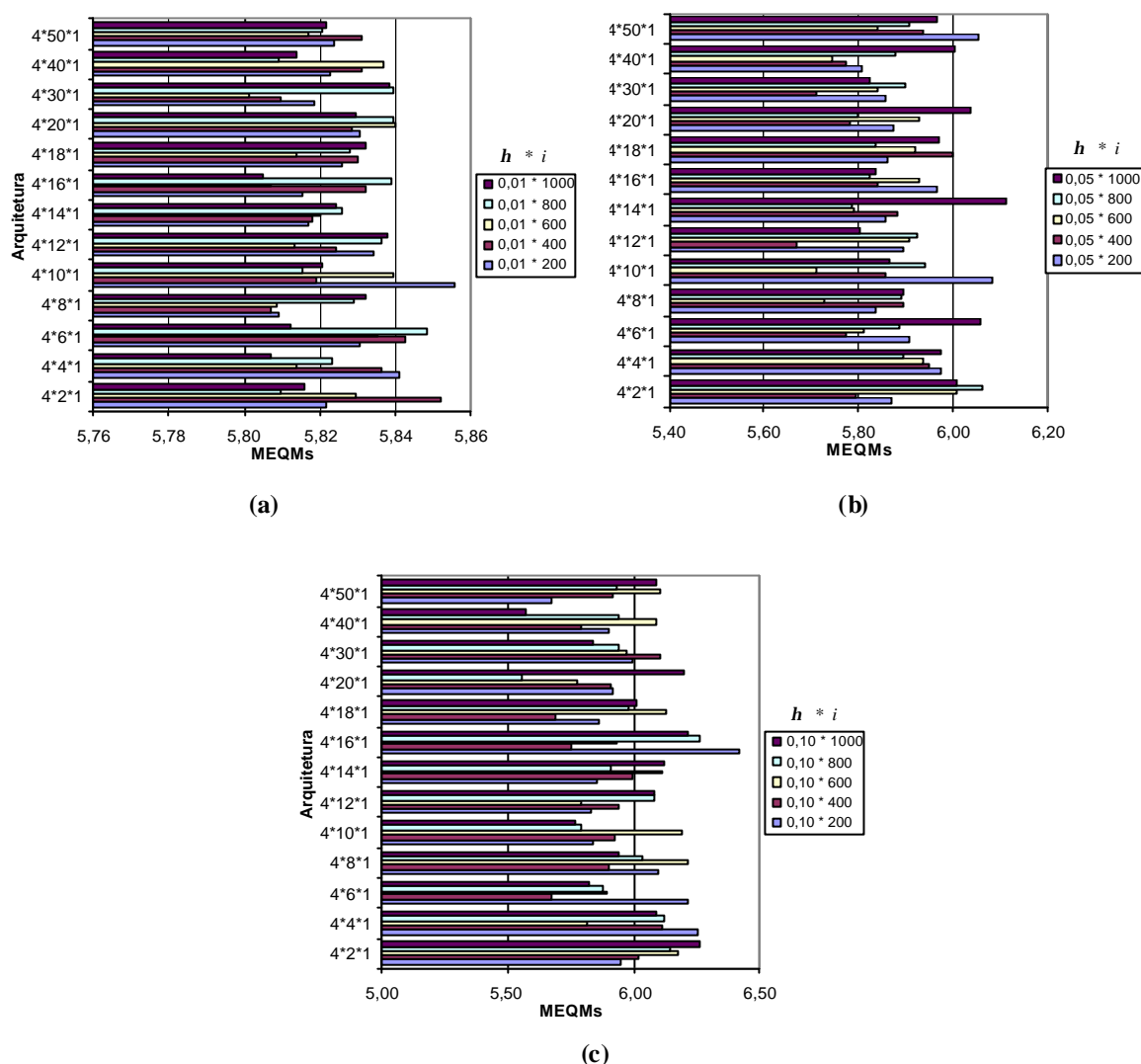


Figura 6.21 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\alpha = 20\%$ usando Normalização como pré-processamento: (a) RBF com $h = 0,01$, (b) RBF com $h = 0,05$ e (c) RBF com $h = 0,10$

Para estes experimentos, o menor MEQMs obtido dentre todas as topologias testadas foi de 5,55 antes da função arredondamento, obtido pela topologia 4*20*1, com $\alpha = 20\%$ da função de ativação gaussiana. O menor erro de classificação é de 34,78% , sendo este o erro da análise de melhor caso.

As Tabelas 6.22, 6.23 e 6.24 são apresentados os desempenhos das topologias analisadas no estudo de melhor caso, caso médio e pior caso para as RNs RBF utilizando somente a Normalização como método de pré-processamento.

Tabela 6.22 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 5% usando Normalização como pré-processamento

| Análise | Descrição de desempenho | | | | | | | | |
|-------------|-------------------------|--------|------|------|--------|--------|--------|-------|----|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA | a |
| Melhor caso | 1 | 4*6*1 | 0,10 | 800 | 5,6292 | 0,3347 | 0,3478 | GAUSS | 5% |
| Caso Médio | 97 | 4*18*1 | 0,01 | 200 | 5,8391 | 0,0419 | 0,3478 | GAUSS | 5% |
| Pior caso | 195 | 4*4*1 | 0,10 | 1000 | 6,4129 | 0,5912 | 0,3478 | GAUSS | 5% |

Tabela 6.23 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 10% usando Normalização como pré-processamento

| Análise | Descrição de desempenho | | | | | | | | |
|-------------|-------------------------|--------|------|------|--------|--------|--------|-------|-----|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA | a |
| Melhor caso | 1 | 4*16*1 | 0,10 | 1000 | 5,6810 | 0,5077 | 0,3478 | GAUSS | 10% |
| Caso Médio | 97 | 4*2*1 | 0,01 | 1000 | 5,8393 | 0,0545 | 0,3478 | GAUSS | 10% |
| Pior caso | 195 | 4*4*1 | 0,10 | 600 | 6,6327 | 0,5649 | 0,3478 | GAUSS | 10% |

Tabela 6.24 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 20% usando Normalização como pré-processamento

| Análise | Descrição de desempenho | | | | | | | | |
|-------------|-------------------------|--------|------|-----|--------|--------|--------|-------|-----|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA | a |
| Melhor caso | 1 | 4*20*1 | 0,10 | 800 | 5,5543 | 0,2431 | 0,3478 | GAUSS | 20% |
| Caso Médio | 97 | 4*50*1 | 0,05 | 600 | 5,8404 | 0,3080 | 0,3478 | GAUSS | 20% |
| Pior caso | 195 | 4*16*1 | 0,10 | 200 | 6,4206 | 0,5608 | 0,3478 | GAUSS | 20% |

A partir dos resultados apresentados nas Tabelas 6.22, 6.23 e 6.24, se percebe que de forma geral a utilização da Normalização como pré-processamento não contribuiu a fim de obter diferentes erros de classificação do conjunto de teste dos padrões apresentados. Mesmo observando que o menor MEQMs foi de 5,5543 para a arquitetura 4*20*1 com taxa de aprendizado de 10% e 800 interações fazendo uso da função de ativação gaussiana com 20% de nível de significância não difere significativamente das demais topologias com maiores MEQMs porém iguais erros de classificação.

Agora serão abordados os experimentos das RNs RBF utilizando PCA como extração de características dos padrões da base de dados. Para este treinamentos foram utilizados os mesmos dados e conjuntos dos experimentos com as RNs MLP Na Tabela 6.25 são apresentados os resultados gerais obtidos.

Tabela 6.25 – Ranking de desempenho das RNs RBF com a função de ativação gaussiana (GAUSS) com nível de significância (a) de 5%, 10% e 20% através do MEQMs para dados após PCA.

| Ranking | Descrição de desempenho | | | | | | |
|---------|-------------------------|------|------|--------|--------|-------|-----|
| | Arq | h | i | MEQMs | s | FA | a |
| 1 | 3*2*1 | 0,10 | 1000 | 0,1183 | 0,0131 | GAUSS | 5% |
| 2 | 3*6*1 | 0,10 | 1000 | 0,0794 | 0,0056 | GAUSS | 20% |
| 3 | 3*2*1 | 0,10 | 1000 | 0,1239 | 0,0236 | GAUSS | 10% |
| 4 | 3*8*1 | 0,10 | 1000 | 0,1299 | 0,0110 | GAUSS | 20% |
| 5 | 3*2*1 | 0,05 | 1000 | 0,1310 | 0,0281 | GAUSS | 20% |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 290 | 3*18*1 | 0,05 | 600 | 4,0450 | 0,1753 | GAUSS | 10% |
| 291 | 3*50*1 | 0,10 | 600 | 4,0461 | 0,3235 | GAUSS | 20% |
| 292 | 3*20*1 | 0,05 | 800 | 4,0487 | 0,1936 | GAUSS | 5% |
| 293 | 3*10*1 | 0,10 | 200 | 4,0634 | 0,4580 | GAUSS | 5% |
| 294 | 3*6*1 | 0,05 | 200 | 4,0665 | 0,2815 | GAUSS | 10% |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 581 | 3*12*1 | 0,01 | 200 | 5,7773 | 0,2694 | GAUSS | 5% |
| 582 | 3*50*1 | 0,01 | 200 | 5,8066 | 0,1207 | GAUSS | 20% |
| 583 | 3*50*1 | 0,01 | 200 | 5,8163 | 0,0590 | GSAUS | 5% |
| 584 | 3*16*1 | 0,01 | 200 | 5,8291 | 0,2079 | GAUSS | 5% |
| 585 | 3*50*1 | 0,10 | 200 | 5,8668 | 0,6348 | GAUSS | 5% |

O desempenho das respostas das diversas arquiteturas RBF também pode ser avaliado nas Figuras 6.22, 6.23 e 6.24 levando em consideração os níveis de significância utilizados na função de ativação gaussiana.

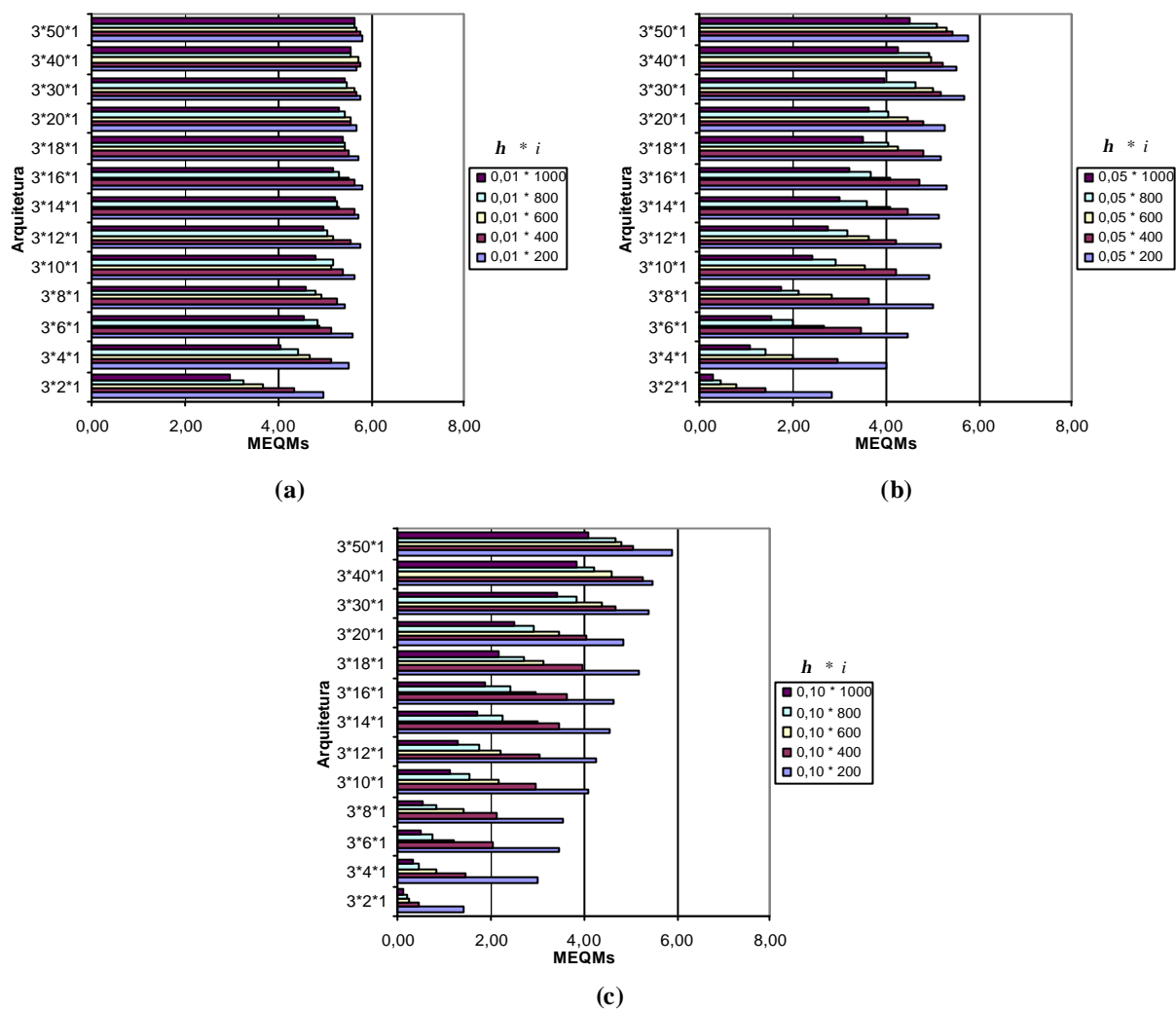


Figura 6.22 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas com a função de ativação gaussiana com $\alpha = 5\%$ após PCA: (a) RBF com $h = 0,01$, (b) RBF com $h = 0,05$ e (c) RBF com $h = 0,10$

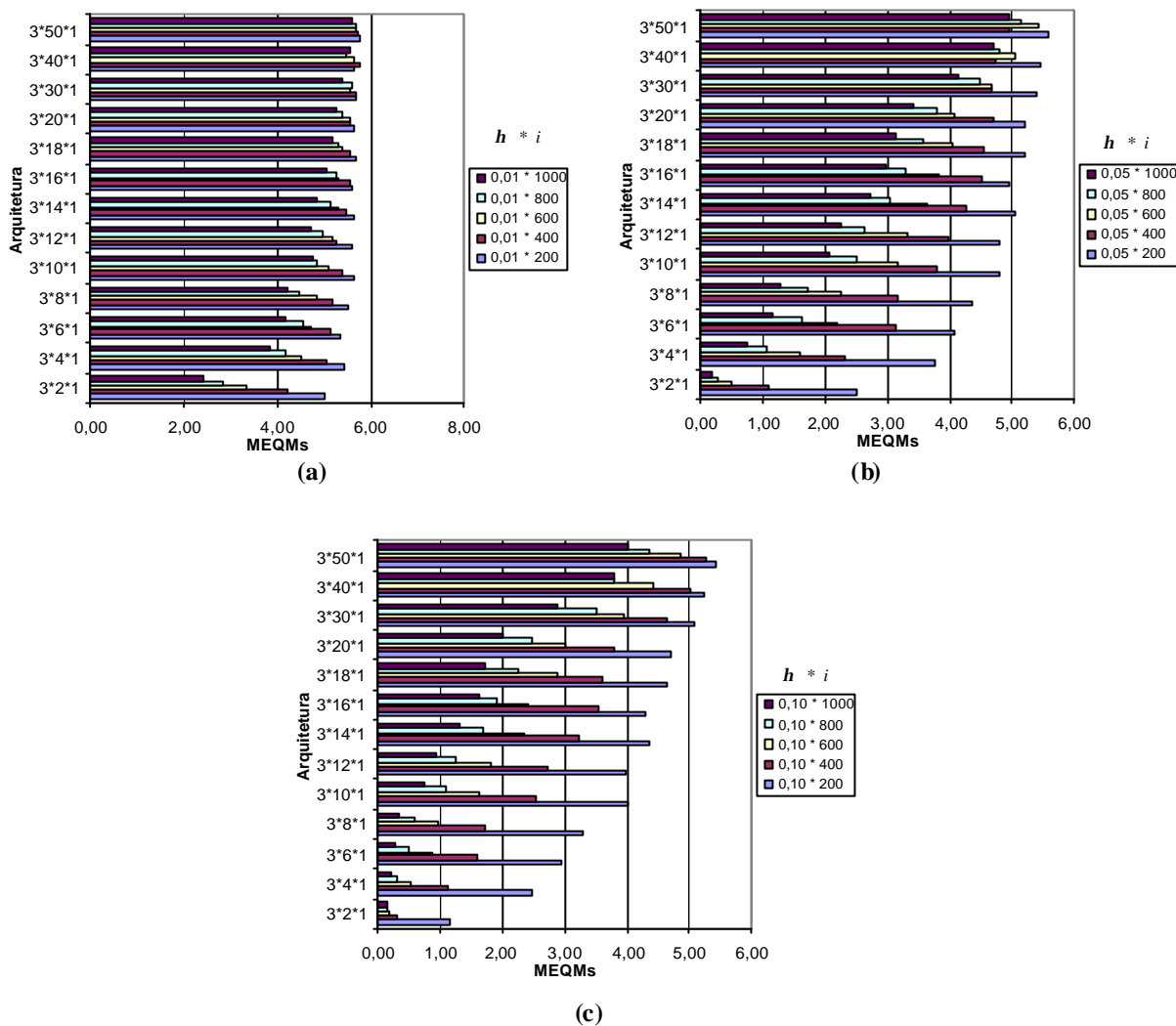


Figura 6.23 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas a função de ativação gaussiana com $\alpha = 10\%$ após PCA: (a) RBF com $h = 0,01$, (b) RBF com $h = 0,05$ e (c) RBF com $h = 0,10$

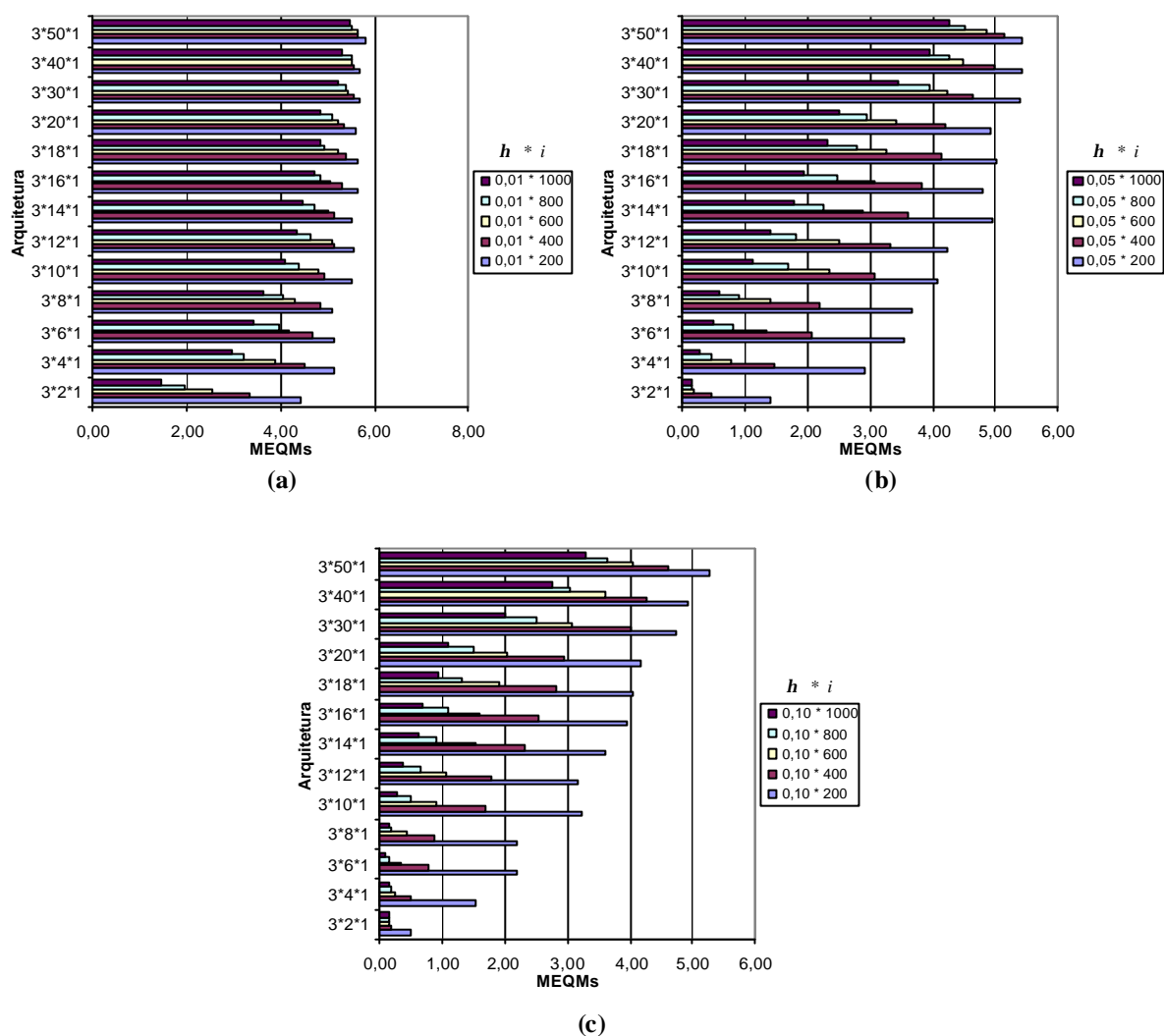


Figura 6.24 – Média dos erros das respostas das 10 simulações para as topologias RBF simuladas a função de ativação gaussiana com $\alpha = 20\%$ após PCA: (a) RBF com $h = 0,01$, (b) RBF com $h = 0,05$ e (c) RBF com $h = 0,10$

Para estes experimentos usando a extração de características, o menor MEQMs obtido dentre todas as topologias testadas foi de 0,1183 antes da função arredondamento, obtido pela topologia 2*2*1, com $\alpha = 5\%$ da função de ativação gaussiana. A rede classificou todo o conjunto de teste, ou seja, o erro de classificação foi de 0% , sendo este o erro de análise de melhor caso.

Nas Tabelas 6.26, 6.27 e 6.28 são apresentados os desempenhos das topologias analisadas no estudo de melhor caso, caso médio e pior caso para as RNs RBF utilizando PCA

como método de extração de características. Também será utilizada uma análise de “caso escolhido” como descrito na seção 6.4.5.

Tabela 6.26 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 5% após PCA

| Análise | Descrição de desempenho | | | | | | | | |
|----------------|-------------------------|--------------|-------------|------------|---------------|---------------|----------|--------------|-----------|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA | a |
| Melhor caso | 1 | 3*2*1 | 0,10 | 1000 | 0,1183 | 0,0131 | 0 | GAUSS | 5% |
| Caso Médio | 97 | 3*4*1 | 0,01 | 800 | 4,4312 | 0,2025 | 0,3478 | GAUSS | 5% |
| Pior caso | 195 | 3*50*1 | 0,10 | 200 | 5,8668 | 0,6348 | 0,3478 | GAUSS | 5% |
| Caso Escolhido | 19 | 3*2*1 | 0,10 | 200 | 1,3986 | 0,0993 | 0 | GAUSS | 5% |

Tabela 6.27 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 10% após PCA

| Análise | Descrição de desempenho | | | | | | | | |
|----------------|-------------------------|--------------|-------------|------------|---------------|---------------|----------|--------------|------------|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA | a |
| Melhor caso | 1 | 3*2*1 | 0,10 | 1000 | 0,1239 | 0,0236 | 0 | GAUSS | 10% |
| Caso Médio | 97 | 3*30*1 | 0,05 | 1000 | 4,1468 | 0,1567 | 0,3478 | GAUSS | 10% |
| Pior caso | 195 | 3*50*1 | 0,01 | 200 | 5,7631 | 0,0948 | 0,3478 | GAUSS | 10% |
| Caso Escolhido | 24 | 3*2*1 | 0,10 | 200 | 1,1387 | 0,1881 | 0 | GAUSS | 10% |

Tabela 6.28 – Análise de casos através do ranking de desempenho das RNs RBF com a função de ativação gaussiana com nível de significância (α) de 20% após PCA

| Análise | Descrição de desempenho | | | | | | | | |
|----------------|-------------------------|--------------|-------------|------------|---------------|---------------|----------|--------------|------------|
| | Ranking | Arq | h | i | MEQMs | s | EC | FA | a |
| Melhor caso | 1 | 3*6*1 | 0,10 | 1000 | 0,0794 | 0,0056 | 0 | GAUSS | 20% |
| Caso Médio | 97 | 3*6*1 | 0,01 | 1000 | 3,4203 | 0,2513 | 0,3478 | GAUSS | 20% |
| Pior caso | 195 | 3*50*1 | 0,01 | 200 | 5,8066 | 0,1207 | 0,3478 | GAUSS | 20% |
| Caso Escolhido | 25 | 3*2*1 | 0,10 | 200 | 0,4948 | 0,0690 | 0 | GAUSS | 20% |

A partir dos resultados apresentados nas Tabelas 6.26, 6.27 e 6.28, se percebe que de forma geral a utilização da técnica PCA como extração de característica teve um importan-

te papel na classificação do conjunto de teste dos padrões apresentados por algumas arquiteturas. Mesmo observando que o menor MEQMs foi de 0,11833 para a arquitetura, $3*2*1$ com taxa de aprendizado de 10% e 1000 interações fazendo uso da função de ativação gaussiana com 5% de nível de significância, porém apresenta um custo computacional elevado pelo grande número de interações, sendo assim dentre os casos escolhidos destacam-se a arquitetura $3*2*1$ com taxa de aprendizado de 10% e 200 interações fazendo uso da função de ativação gaussiana com 5%, 10% e 20% de nível de significância, com MEQMs de 1,3986, 1,1387, 0,4948, respectivamente, das quais classificam todo o conjunto de teste dos padrões apresentados.

Na utilização de PCA, todas as arquiteturas das RNs RBF apresentam um aumento no MEQMs a medida que o número de unidades de processamento da camada intermediária aumenta. Este aumento pode estar relacionado à perda de generalização da rede.

A Tabela 6.29 apresenta um resumo da geral das RNs RBF utilizando a Normalização como técnica de pré-processamento e PCA como extração de características enfatizando o percentual de classificação do conjunto de teste dos padrões das diversas arquiteturas nas análises de caso apresentados do estudo abordado.

Tabela 6.29 – Classificação das respostas das RNs RBF com função de ativação gaussiana para os dados antes (normalização dos dados) e após a extração de características.

| Análise | Paciente | Classificação (%) | Técnica de classificação | | | | | |
|-----------------------|------------------|-------------------|--------------------------|--------|--------|----------------|--------|--------|
| | | | RN RBF sem PCA | | | RN RBF com PCA | | |
| | | | a=0,05 | a=0,10 | a=0,20 | a=0,05 | a=0,10 | a=0,20 |
| Melhor caso | Diabético | Acerto | 0 | 0 | 0 | 100 | 100 | 100 |
| | | Erro | 100 | 100 | 100 | 0 | 0 | 0 |
| | Normal | Acerto | 100 | 100 | 100 | 100 | 100 | 100 |
| | | Erro | 0 | 0 | 0 | 0 | 0 | 0 |
| Caso Médio | Diabético | Acerto | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Erro | 100 | 100 | 100 | 100 | 100 | 100 |
| | Normal | Acerto | 100 | 100 | 100 | 100 | 100 | 100 |
| | | Erro | 0 | 0 | 0 | 0 | 0 | 0 |
| Pior caso | Diabético | Acerto | 0 | 0 | 0 | 0 | 0 | 0 |
| | | Erro | 100 | 100 | 100 | 100 | 100 | 100 |
| | Normal | Acerto | 100 | 100 | 100 | 100 | 100 | 100 |
| | | Erro | 0 | 0 | 0 | 0 | 0 | 0 |
| Caso Escolhido | Diabético | Acerto | - | - | - | 100 | 100 | 100 |
| | | Erro | - | - | - | 0 | 0 | 0 |
| | Normal | Acerto | - | - | - | 100 | 100 | 100 |
| | | Erro | - | - | - | 0 | 0 | 0 |

Apesar da melhora perceptível apresentada pelo segundo experimento com as RNs RBF (com extração de característica) realizamos o teste de hipótese para verificar se esta melhora estatisticamente se confirma. Na Tabela 6.30 é apresentado o teste de hipóteses para os experimentos acima descritos. O teste é realizado a partir dos melhores resultados médios, ou seja, os resultados do “melhor caso”, obtidos por cada um dos experimentos (normalização como pré-processamento dos dados e PCA como extração de características). Neste teste verificamos se o desempenho da RN RBF com extração de características é superior ao desempenho da mesma com a Normalização como pré-processamento.

Tabela 6.30 – Estatística do teste de Wilcoxon para o desempenho das RNs RBF com normalização no pré-processamento e PCA como extração de características.

| $H_0 : \underline{m}_1 = \underline{m}_2$ | | |
|---|---------|---------------|
| $H_1 : \underline{m}_1 > \underline{m}_2$ | | |
| <i>p</i> -valor | z-valor | Decisão |
| 0,00 | 9,1412 | Rejeita H_0 |

A Tabela 6.30 apresenta os resultados do teste de Wilcoxon com nível de significância de 5%. Nesta tabela, μ_1 e μ_2 são respectivamente as médias dos erros dos experimentos com as RNs RBF com Normalização no pré-processamento e as médias dos erros dos experimentos utilizando PCA como extração de características. A partir do teste de hipótese se pode concluir que o desempenho médio das RNs RBF fazendo uso do PCA como extração de características é melhor que a das RNs RBF com Normalização no pré-processamento.

6.4.9 Comparação entre as abordagens

Nas seções anteriores vimos a comparação entre as RNs de mesmo tipo. O objetivo nesta seção é comparar o desempenho dos melhores classificadores utilizados, ou seja, os casos escolhidos. Vale ressaltar que, para uma comparação idônea entre os classificadores, os experimentos foram realizados com a mesma base de dados e a mesma divisão dos conjuntos de treinamento e teste. O parâmetro de comparação é o erro de classificação antes da função arredondamento (MEQMs). Na Tabela 6.31 é apresentada uma comparação estatística, baseada no teste de hipóteses de Wilcoxon, entre os classificadores utilizados, onde \underline{m}_1 e \underline{m}_2 são as médias dos erros de classificação antes da função arredondamento para os classificadores da análise dos casos escolhidos.

Tabela 6.31 – Estatística do teste de Wilcoxon para o desempenho das RNs simuladas com PCA como extração de características.

| $H_0 : \mathbf{m}_1 = \mathbf{m}_2$ | | | | |
|-------------------------------------|-------------------------------|----------------------|---------|-------------------|
| $H_1 : \mathbf{m}_1 > \mathbf{m}_2$ | | | | |
| Modelo (m_1) | Modelo (m_2) | <i>p</i> -valor | z-valor | Decisão |
| MLP – LOG – PCA | MLP – TANHIP – PCA | $5,86 \cdot 10^{-5}$ | 5,8206 | Rejeita H_0 |
| | RBF ($\alpha = 0,05$) – PCA | 0,3946 | -0,8512 | Não Rejeita H_0 |
| | RBF ($\alpha = 0,10$) – PCA | 0,8909 | 0,1371 | Não Rejeita H_0 |
| | RBF ($\alpha = 0,20$) – PCA | 0,4991 | -0,6759 | Não Rejeita H_0 |
| MLP – TANHIP – PCA | RBF ($\alpha = 0,05$) – PCA | 0,5004 | -0,6738 | Não Rejeita H_0 |
| | RBF ($\alpha = 0,10$) – PCA | 0,8699 | 0,1638 | Não Rejeita H_0 |
| | RBF ($\alpha = 0,20$) – PCA | 0,0548 | -1,9203 | Não Rejeita H_0 |
| RBF ($\alpha = 0,05$) – PCA | RBF ($\alpha = 0,10$) – PCA | 0,9438 | 0,0705 | Não Rejeita H_0 |
| | RBF ($\alpha = 0,20$) – PCA | 0,8589 | 0,1778 | Não Rejeita H_0 |
| RBF ($\alpha = 0,10$) – PCA | RBF ($\alpha = 0,20$) – PCA | 0,9310 | 0,0866 | Não Rejeita H_0 |

6.5 Resumo do capítulo

Neste capítulo foram apresentados os resultados dos experimentos realizados com as técnicas de abordagem estatística e neural investigadas nesta dissertação. O objetivo de tais experimentos foi verificar qual das abordagens melhor obtém melhor desempenho na classificação no prognóstico da doença Diabetes através dos sinais dos sensores de aroma expostos a urina dos pacientes estudados.

Foi apresentada uma análise dos padrões de odores a partir da técnica PCA. O interesse de sua utilização reside no fato desta técnica permitir visualizar as semelhanças entre os sensores utilizados e estabelecer associações entre os odores e os sensores de acordo com suas propriedades de distribuição estatística, evidenciadas por este método de manipulação estatística. PCA também é importante por ser um método que proporciona a visualização da distribuição espacial dos dados, mesmo que de forma simplificada. Isto porque os componentes principais são aproximações com menores dimensões, das variáveis originais.

A partir dos resultados e da distribuição dos padrões nas três componentes principais, onde sua escolha está relacionada à capacidade de representação da variação total de 92,62% dos dados via matriz de covariâncias, pode-se verificar que os dois tipos de odores apresentados possuem uma separabilidade espacial. Vale ressaltar que na transposição dos dados para o plano das componentes principais houve uma simplificação da distribuição dos mesmos, fato pelo qual aparecem tão espacialmente disjuntos.

Além da análise por PCA foi testado o classificador KNN que aumentou a classificação significativamente entre os dois conjuntos de teste, sendo de 40% (apenas utilizando a Normalização como pré-processamento) para 100% com $K=1$ e de 35,56% para 100% com $K=3$ na divisão 50% da base para treino e 50% para teste. Já na divisão 70% para treino e 30% para teste o classificador após PCA aumentou de 40% (dados normalizados) para 100% com $K=1$ e de 50% para 100% com $K=3$.

Também se apresentou como um eficiente extrator de característica melhorando nos desempenhos de classificação das RNs simuladas. Foram simuladas diversas arquiteturas para as RNs variando o número de interações e a taxa de aprendizado, para as MLP foram utilizadas dois tipos de função de ativação: a logística e a tangente hiperbólica, já para as RBF utilizou-se a gaussiana variando o nível de significância entre 5%, 10% e 20%.

Verificou-se que nas RNs MLP usando a função de ativação tangente hiperbólica possui o maior poder de generalização em relação as que usam a função de ativação logística antes e após PCA. Na análise de melhor caso para os dados normalizados obteve um erro médio de classificação de 32,60% com a função de ativação logística e de 33,91% com a tangente hiperbólica, um erro relativamente alto se relacionada com a mesma análise fazendo uso da extração de característica por PCA onde conseguiu classificar todo o conjunto de teste apresentado fazendo uso das duas funções de ativação.

As RNs RBF também apresentaram altos erros de classificação quando utilizada a Normalização como pré-processamento, na análise de melhor caso obteve um erro médio de classificação de 34,78% para a função de ativação gaussiana com os níveis de significância de 5%, 10% e 20%. Assim como as RNs MLP também melhoraram significativamente a taxa de classificação após PCA, classificando todo o conjunto de teste na mesma análise de caso.

Nas RNs RBF se observou, particularmente, que existe um conjunto de arquiteturas ótimas quando relacionadas ao número de interações, e com o aumento significativo deste número a rede perde o poder de generalização.

Observou-se também que nos melhores casos as RNs possuíam um alto custo computacional em virtude do número de interações, sendo assim foram apresentadas outras redes de menor custo com excelentes taxas de classificação. Na comparação entre as RNs verificou-se que as redes MLP e RBF possuem erros médios de classificação estatisticamente iguais.

Portanto, PCA mostrou-se uma importante técnica no que diz respeito a sistemas de reconhecimento de padrões, já a Normalização como pré-processamento neste conjunto de dados de odor específico não obteve bons resultados.

Capítulo 7

Conclusão

Neste capítulo apresentam-se as considerações finais sobre o trabalho desenvolvido nesta dissertação. Na seção 7.1 são apresentadas algumas considerações sobre os resultados obtidos nos experimentos. Na seção 7.2 são descritas as contribuições desta dissertação e na seção 7.3 algumas propostas para trabalhos futuros.

7.1 Considerações finais

O desenvolvimento de dispositivos que simulem o comportamento do sistema olfativo biológico é uma área em grande expansão, principalmente na indústria. Muitas pesquisas estão sendo despendidas com a finalidade de simular a percepção humana. A maioria dos esforços visa refletir nos dispositivos eletrônicos a capacidade humana de perceber o mundo na forma de cinco sentidos ou até mesmo estende além das nossas potencialidades sensoriais.

O presente trabalho está contextualizado na construção de narizes artificiais, dispositivos eletrônicos cuja necessidade em diversas aplicações é indiscutível. O foco principal desta dissertação foi o estudo de sensores de aroma que poderão ser acoplados em tais dispositivos, além da construção de sistemas de reconhecimentos de padrões, buscando técnicas que apresentam resultados significativos na classificação de sinais de odores.

Foi apresentado o sistema de aquisição de dados desenvolvido, parte constituinte de um nariz artificial, além dos sensores de aroma utilizados na construção da base de dados.

Também foram apresentadas duas abordagens, uma estatística e uma neural nos sistemas de reconhecimento de padrões, como a Normalização na fase de pré-processamento, Análise por Componente Principais fazendo uso do classificador *K-Nearest-Neighbour* e as Redes Neurais Artificiais *Multilayer Perceptron* e de Função de Base Radial fazendo uso de várias topologias, das quais foram exploradas algumas funções de ativação e números de neurônios da camada oculta.

Na abordagem estatística, o classificador KNN mostrou-se eficiente quando combinado com PCA, aumentando a taxa de classificação para 100% do conjunto de teste apresentado, fato que não decorreu quando utilizado a Normalização na base de dados.

Dentre as abordagens conexionistas, as RNs MLP são as mais utilizadas na literatura para classificação de sinais de odores. As RNs RBF, por sua vez, têm apresentado resultados promissores na classificação de odores, como modelo alternativo ao MLP. Para as entradas das RNs foram utilizados os dados normalizados e os dados após PCA.

Nos resultados obtidos com as arquiteturas conexionistas, as RNs fazendo uso de PCA na fase de extração de características apresentaram resultados muito melhores que as redes fazendo uso de normalização de dados. Quando aliadas à escolha de determinadas funções de ativação podem ter desempenhos distintos, nas simulações a RN MLP com a função tangente hiperbólica obtiveram uma média de erro de classificação estatisticamente menor que as redes com a função logística, fato evidenciado no teste de hipótese realizado, já as RNs RBF a utilização da função de ativação gaussiana com diferentes níveis de significância obtiveram a média do erros de classificação estatisticamente iguais, evidenciadas a partir do teste de hipótese realizado.

Em geral, os classificadores estatísticos e neurais apresentaram-se eficientes na classificação de odores para o prognóstico do Diabetes Mellitus, onde ambos podem ser concomitantemente utilizados como ferramentas em sistemas de reconhecimento de padrões. Na

maioria das abordagens utiliza-se apenas o erro de classificação como critério de avaliação, entretanto neste trabalho foram utilizados, além desse, outros fatores como o custo computacional.

Considerando-se métodos estatísticos de comparação, as melhores abordagens para classificação de odores, com os dados testados nesta dissertação, foram as Redes Neurais fazendo uso de PCA como extração de características, destacando-se as RNs RBF.

A utilização de um sistema de reconhecimento de padrões na classificação de sinais de odores em narizes artificiais permite a construção de equipamentos ao mesmo tempo versáteis (a depender da tecnologia empregada), e também altamente específicos. Os narizes artificiais podem, por exemplo, ser configurados para identificar um odor característico para uma doença específica, mas faz-se necessário o treinamento do mesmo para identificá-los.

Para tarefas mais simples de reconhecimentos de odores já existem tecnologias suficientes para a detecção de odores, futuras pesquisas podem vir a considerar outras funcionalidades na construção dos narizes artificiais, como a aquisição automática de conhecimento (sistemas inteligentes), configuração automática dos dispositivos (visando a adequação para tipos de odores específicos), a praticidade e versatilidades dos dispositivos (miniaturização e comercialização dos dispositivos).

7.2 Contribuições deste trabalho

Além dos resultados obtidos neste trabalho servirem de base para futuras comparações com novos sistemas propostos para reconhecimento de odores, outras contribuições podem ser destacadas como:

- i. Construção de uma base de odores da doença Diabetes Mellitus;

- ii. Desenvolvimento de uma metodologia de aquisição de dados;
- iii. Validação de sensores de aroma para aplicações práticas;
- iv. Validação dos resultados obtidos com as RNs com Normalização como pré-processamento;
- v. Validação dos resultados obtidos com as RNs com PCA como extrator de características;
- vi. Comparação entre as abordagens conexionistas estudadas;
- vii. Apresentação das potencialidades das abordagens deste estudo aplicadas no funcionamento dos narizes artificiais;

7.3 Proposta para trabalhos futuros

Como continuidade deste trabalho, novos sensores de aroma poderiam ser utilizados, visando o enriquecimento e o aumento da base de dados atual. Utilização de outros métodos de pré-processamento visando otimizar a rede, assim como a escolha de outras funções de ativações e arquiteturas.

Na abordagem estatística, outras técnicas de classificação, como a Análise de Discriminantes. Outros métodos poderiam ser aplicados, dos quais possuem equivalências nas abordagens estatísticas e neurais e avaliar o desempenho dessas abordagens.

Novos dispositivos poderiam ser utilizados no processo de aquisição, assim como o melhoramento da metodologia de aquisição.

Apêndice A

Outliers Multivariados

Um problema comum em análise de dados é a presença de outliers, também chamado *pontos aberrantes*. Entende-se por um outlier a uma observação que apresenta um comportamento atípico em relação ao restante dos dados. Quando se tem uma única variável, o valor aberrante caracteriza-se por assumir um valor muito mais alto ou muito mais baixo que os demais, muito comum em experimentos envolvendo odor, uma vez que estão susceptíveis as variações do ambiente. Há várias causas possíveis para a ocorrência de valores aberrantes, dentre eles alguns foram citados no capítulo 4, seção 4.3.

Neste apêndice falamos sobre outliers multidimensionais e métodos de identificação.

Outliers unidimensionais

Conforme já mencionado, um outlier unidimensional caracteriza-se por situa-se longe da massa de dados. Há vários métodos de identificação de valores aberrantes. Serão abordados métodos que facilitarão a compreensão sobre a identificação de valores aberrantes multidimensionais.

Seja X uma variável aleatória com média \mathbf{m} um outlier deve se localizar distante da média. Dentre as várias maneiras de identificá-los, pode-se calcular as distâncias de todas as observações em relação a \mathbf{m} caso uma delas seja muito maior do que as demais, estaremos

diante de um possível outlier. Pode-se utilizar a distância euclidiana, como vimos na Eq. 5.27, ou a distância euclidiana ao quadrado:

$$d^2 = (X - \mathbf{m})^2 \quad (\text{A.1})$$

Seja $X \sim N(\mathbf{m}, \mathbf{S}^2)$, temos então que

$$d_M^2 = \frac{(X - \mathbf{m})^2}{\mathbf{S}} \sim \chi_1^2 \quad (\text{A.2})$$

Podemos utilizar o conhecimento sobre a distribuição de d_M^2 para identificar possíveis outliers. Nesse caso, para cada elemento amostral, calcula-se para cada observação x_i , $i = 1, \dots, n$ a probabilidade

$$P(d_M^2 \geq x_i)$$

se essa probabilidade for muito pequena, haverá indícios de que o ponto é um possível outlier, note que não se trata de um teste de hipóteses no sentido usual, é apenas uma regra de decisão. Como ponto de corte pode-se adotar a utilização do valor 0,001 [Hair et al., 1998]

Outliers bidimensionais

Admita uma situação onde se tem duas variáveis que serão estudadas conjuntamente. Nesse caso, em que situações um ponto é atípico?

A Figura A.1 apresenta um diagrama de dispersão no qual foram identificados pontos atípicos em relação à grande massa de dados, representada pelo cluster de pontos do canto inferior esquerdo. Note que todos os pontos posicionam-se num local inesperado dado o cluster de pontos, mas nem todos estão distantes dele.

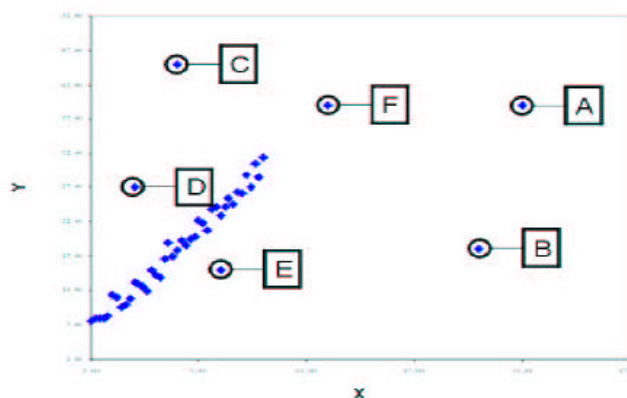


Figura A.1 – Diagrama de dispersão com identificação de valores aberrantes bidimensionais.

Caracterizando cada ponto tem-se:

- i. Ponto **A**: trata-se de um outlier unidimensional tanto para a variável X como para Y .
- ii. Pontos **B** e **C**: esses pontos diferenciam-se do anterior por não serem valores aberrantes unidimensionais para as duas variáveis. Note que **B** é aberrante para X , mas não para Y ; o inverso ocorre **C**.
- iii. Pontos **D** e **E**: esses pontos não seriam identificados numa análise unidimensional, uma vez que não se configuram como outliers unidimensionais. No entanto, eles ocupam uma região do plano na qual não esperaríamos encontrar valores, dada a nuvem de pontos.
- iv. Ponto **F**: esse ponto, assim como o **A** seria identificado como outlier unidimensional tanto para X como para Y , no entanto, apesar de localizar-se distante da massa de dados (o que caracterizaria como aberrante bidimensional), ele encontra-se no prolongamento do cluster de pontos.

Todos os pontos citados são outliers bidimensionais, no entanto nem todos causaríamos danos a todas as possíveis análises. por exemplo, o ponto **F** teria uma influência muito pequena sobre o coeficiente de correlação existente entre X e Y , no entanto, afetaria bastante as médias e as variáveis de X e Y . os pontos **D** e **E**, ao contrário, praticamente não afetariam as médias e variâncias, mas se ajustado um modelo de regressão, afastaria a reta da massa de

dados. A identificação do tipo de outlier, é portanto, importante ao se decidir sobre a retirada ou não da amostra.

Para a identificação de outliers bidimensionais seguirá como feito no caso unidimensional. Seja $x = (X, Y)^T$ um vetor aleatório com $E(x) = \mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2)^T$, temos então que a distância euclidiana ao quadrado entre x e \mathbf{m} é dada por

$$d^2 = (x - \mathbf{m})^T (x - \mathbf{m}) = (X - \mathbf{m}_1)^2 + (Y - \mathbf{m}_2)^2 \quad (\text{A.3})$$

Para identificação de outliers, pode-se determinar d^2 para todas as observações da amostra e verificar a existência de uma distância muito maior que as demais. Em caso afirmativo, a respectiva observação seria um possível outlier. Os pontos **A**, **B**, **C** e **F** poderiam ser identificados por essa estratégia. O mesmo não acontece com os pontos **D** e **E**.

Distância de Mahalanobis

Para identificar pontos semelhantes a **D** e **E**, definiremos uma nova distância, denominada de *distância de Mahalanobis*. Admita que $\text{Cov}(x) = \hat{\mathbf{a}}$. Defini-se a distância de Mahalanobis ao quadrado entre os pontos x e \mathbf{m} como

$$d_M^2(x, \mathbf{m}) = (x - \mathbf{m})^T \hat{\mathbf{a}}^{-1} (x - \mathbf{m}) \quad (\text{A.4})$$

A distância de Mahalanobis leva em consideração a estrutura de correlação existente nos dados, dessa forma pontos que não obedecem à estrutura geral apresentarão altas distâncias em relação à média.

Uma regra de identificação de outliers para dados bidimensionais pode ser construída se considerarmos que x segue uma distribuição normal bivariada. Sendo assim

$$d_M^2(x, \mathbf{m}) \sim c_2^2$$

Então como no caso unidimensional, pode-se calcular para cada observação x_i , $i = 1, \dots, n$ a probabilidade

$$P\left(d_M^2(x, \mathbf{m}) \geq x_i\right)$$

Valores pequenos dessa probabilidade (0,001) identificariam prováveis outliers. Essa regra poderia levar à identificação dos pontos **A**, **B**, **C**, **D** e **E** da Figura A.1. Apenas o ponto **F** seria de difícil identificação por estar posicionado no prolongamento do cluster de pontos.

Outliers multidimensionais

No caso multidimensional perde-se a capacidade de representar os pontos graficamente em eixos ordenados. Nesse caso, uma alternativa para identificação de possíveis outliers, são os métodos numéricos. Como nos casos anteriores, unidimensional e bidimensional, utilizaremos as abordagens: distância euclidiana para a identificação de pontos distantes do cluster dos dados e distância Mahalanobis para a identificação de pontos fora da tendência dos dados.

Seja $x = (X_1, \dots, X_p)^T$ um vetor aleatório com $E(x) = \mathbf{m} = (\mathbf{m}_1, \dots, \mathbf{m}_p)^T$ e $\text{Cov}(x) = \hat{\mathbf{a}}$. Então a distância euclidiana ao quadrado entre uma observação x e o vetor de média \mathbf{m} é dada por

$$d^2 = (x - \mathbf{m})^T (x - \mathbf{m}) = (X - \mathbf{m}_1)^2 + (X - \mathbf{m}_p)^2 \quad (\text{A.5})$$

Por sua vez a distância de Mahalanobis é dada por

$$d_M^2(x, \mathbf{m}) = (x - \mathbf{m})^T \hat{\mathbf{a}}^{-1} (x - \mathbf{m}) \quad (\text{A.6})$$

Se x seguir uma distribuição normal p -variada, temos que

$$d_M^2(x, \mathbf{m}) \sim \mathbf{c}_p^2$$

As duas distâncias podem ser utilizadas como proposto no caso bidimensional.

Comentários de ordem prática

Os resultados apresentados pressupõem o conhecimento sobre \mathbf{m} e $\hat{\mathbf{a}}$. Na prática, não temos esse conhecimento. Nesse caso, sugere-se substituir essas quantidades por seus estimadores usuais \bar{x} e \mathbf{S} . Os resultados passarão a ser aproximados e terão um melhor desempenho para grandes amostras.

Os outliers podem trazer influência sobre os valores de \bar{x} e $\hat{\mathbf{a}}$. Isso pode prejudicar sua identificação, visto que as referências de média e covariâncias podem estar distorcidas. Para minimizar esse efeito, sugere-se que ao avaliar se um determinado ponto é aberrante, deve-se excluir a observação no cálculo de \bar{x} e $\hat{\mathbf{a}}$.

O cálculo de probabilidade para a distância de Mahalanobis só é válido sob a hipótese de normalidade. Assim, pode ser necessária a utilização de transformações para viabilizar os resultados.

Referências Bibliográficas

- [Aathithan, 2001] AATHITHAN, S, PLANT, J. C, CHAUDRY, A. N e FRENCH, G. L. (2001) **Diagnosis of Bacteriuria by Detection of Volatile Organic Compounds in Urine Using an Automated Headspace Analyzer with Multiple Conducting Polymer Sensors**. In: Journal of Clinical Microbiology, vol. 39, n. 07, p. 2590–2593.
- [Abe et al., 1987] ABE, H., YOSHIMURA, T., KANAYA, S., TAKAHASHI, Y., MIYASHITA, Y.I. e SASAKI, S. (1987). **Automated Odor-Sensing System Based on Plural Semiconductor Gas Sensors and Computerized Pattern Recognition Techniques**. In: Analytica Chimica Acta, vol. 194, p.1-9
- [ADA, 2006] AMERICAN DIABETES ASSOCIATION. **Other "Types" of Diabetes**. Disponível em <<http://www.diabetes.org/home.jsp>>. Acesso em 18 dez. 2006.
- [Aires, 1999] AIRES, M. M. (1999) **Fisiologia**. 2 ed. Guanabara Koogan, Rio de Janeiro.
- [Anderson et al., 1990] ANDERSON, J. A., PELLIONISZ, A. e ROSENFELD, E. (Eds) (1990). **Neurocomputing 2: Directions for Research**. MIT Press, Cambridge, Massachusetts.
- [Axel, 95] AXEL, R. (1995). **The Molecular Logic of Smell**. In: Scientific American, p 130-137.
- [Amoore, 1971] AMOORE, J. E. (1971) **Olfactory genetics and anosmia**. In: BEIDLER, L.M. (Ed.). Handbook of Sensory Physiology. Springer-Verlag, Berlin, vol. 4, p. 245-256.
- [Andrew et al., 2001] ANDREW, D., BLAKELEY, A. G. H. e KIDD, C. (2001). **Fisiologia Humana**. 1 ed. Artmed.
- [Azzolini, 2003] AZZOLINI, M. (2003). **Fisiologia Pós-Colheita de Goiabas 'Pedro Sato': Estados de Maturação e Padrão Respiratório**. Dissertação de Mestrado. Escola Superior de agricultura Luiz de Queiroz, USP, São Paulo.
- [Batista, 2006] BATISTA, A. C. S. (2006). **Estudo e Desenvolvimento de um Modelo de Olfacção Biológica e Utilização de um Nariz**

Artificial na Detecção de Diabetes. Monografia - Departamento de Biologia, UFRPE, Recife.

- [Batista e D'Emery, 2006] BATISTA, A. C. S., D'EMERY, R.A. e SANTOS, M. S. (2006). **Uso de um Nariz Artificial na Detecção de Diabetes.** In: VI JEPEX – VII SIMPÓS.
- [Batista e D'Emery, 2005] BATISTA, A. C. S., SANTOS, M. S. e D'EMERY, R.A. (2005). **Detecção de Doenças através do Odor.** In: V JEPEX – VI SIMPÓS.
- [Bear, 2002] BEAR, M. F. (2002) **Neurociências – Desvendando o Sistema Nervoso.** 2 ed. Artmed.
- [Beers, 1999] BEERS, M. H. (1999) **Manual Merck.** Disponível em <<http://www.merck.com/mmpe/index.html>>. Acesso em 20 dez. 2006.
- [Benedetti et al., 2004] BENEDETTI, S., MANNINO, S., SABATINI, A. G. e MARCAZZAN, G. L. (2004). **Electronic Nose and Neural Network Use for the Classification of Honey.** In: *Apidologie*, vol. 35, p. 397- 402.
- [Berne e Levy, 2004] BERNE, R. M.e LEVY, M. N. (2004). **Fisiologia.** 5 ed., Elsevier.
- [Bishop, 2006] BISHOP, C. M. (2006). **Pattern Recognition and Machine Learning.** Springer.
- [Box e Cox, 1964] BOX, G. E. P. e COX, D. R. (1964). **An Analysis of Transformations.** In: *Journal of the Royal Statistical Society: Series B*, vol. 26, p. 211-252.
- [Braga et al., 2000] BRAGA, A. P., CARVALHO, A. P. L. e LUDEMIR, T. B. (2000). **Redes Neurais Artificiais: Teoria e Aplicações.** Rio de Janeiro: Livros Técnicos e Científicos.
- [Bullock, 1998] BULLOCK, J. (1998). **Fisiologia.** 3 Ed. Guanabara Koogan, Rio de Janeiro.

- [Carneiro et al., 2006] CARNEIRO, P. J. S., SANTOS, M. S. e FERNANDES, M. G. (2006). **Estudo e Detecção de Enfermidades Através do Odor.** In: 10^a Jornada de Iniciação Científica - PIBIC/FACEPE/CNPQ, Recife.
- [Cattell, 1966] CATTELL, R. B. **The Screen Test for the Number of Factors.** In: *Multivariate Behavioral Research*, vol. 1, p. 140-161, 1966.
- [Clifford e Tuma, 1983] CLIFFORD, P. K. e TUMA D.T. (1983). **Characteristics of Semiconductor Gas Sensor I, Steady-state gas response.** In: *Sensors and Actuators*, vol. 3, p. 233-254.
- [Cover e Hart, 1967] COVER. T. M. e HART, P. E.(1967). **Nearest Neighbour Pattern Classification.** In: *IEEE Transactions on Information Theory*, vol. IT-11, p. 21-27.
- [Cover, 1965] COVER, T. M. (1965). **Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition.** In: *IEEE Transactions on Electronic Computers*, vol. EC-14, p. 326-334.
- [Crocker e Henderson, 1927] CROCKER, E. C. e HENDERSON, L. F. (1927) **Analysis and classification of odors: an effort to develop a workable method.** In: *American Perfumer & Essential Oil Review*, vol. 22, p. 325–356.
- [D’Emery et al., 2007] D’EMERY, R.A., SANTOS, M. S e OLIVEIRA, W. R. (2007). **Electronic Nose for Non-Invasive Prognosis of Diabetes Mellitus.** In: *IEEE Computational Intelligence Society, 2007 International Joint Conference on neural Networks*, Orlando, Flórida, USA. (submitted).
- [D’Emery et al., 2005a] D’EMERY, R.A., SANTOS, M. S e DOS SANTOS, F.L. (2005). **Projeto e Implementação de um Nariz Eletrônico para um Robô Móvel.** In: *IX Jornada de Iniciação Científica*, Recife, p. 291-292.
- [D’Emery et al., 2005b] D’EMERY, R.A., SANTOS, M. S e DE OLIVEIRA, W. R. (2005). **Desenvolvimento de um Dispositivo Artificial para Detecção de Odor.** In: *V JEPEX – VI SIMPÓS.*

- [D'Emery et al., 2005c] D'EMERY, R.A., NASCIMENTO, R. J. e SANTOS, M. S (2005). **Uso de um Nariz Artificial para Monitoramento do Amadurecimento de Frutas**. In: V JEPEX – VI SIMPÓS.
- [D'Emery et al., 2005d] D'EMERY, R.A., SANTOS, M. S e dos SANTOS, F. L., (2005). **Projeto e Implementação de um Robô Móvel a Serviço do Homem** In: 57ª SBPC, Fortaleza.
- [Davies e Taylor, 1959] DAVIES, J. T. e TAYLOR, F. H. (1959) **The Role of Adsorption and Molecular Morphology in Olfaction: the Calculation of Olfactory Thresholds**. In: The Biological Bulletin, vol. 117, p. 222-238.
- [Di Natale et al., 1995a] DI NATALE, C., DAVIDE, F.A.M., D'AMICO, A., SBERVEGLIERI, G., NELLI, P., FAGLIA, G. e PEREGO, C. (1995). **Complex Chemical Pattern Recognition with Sensor Array: the Discrimination of Vintage years of Wine**. In: Sensors and Actuators B, vol. 24-25, p. 801-804.
- [Di Natale et al., 1995b] DI NATALE, C., DAVIDE, F. e D'AMICO, A. (1995). **Pattern Recognition in Gas Sensing: well-States Techniques and Advances**. In: Sensors and Actuators B, vol. 23, p. 111-118.
- [Dobson, 1776] DOBSON, M. (1776). **Nature of the urine in diabetes**. In: Med Obs Inqu, vol. 5, p. 298–310.
- [dos Santos, 2006] dos SANTOS, W. P. **Diabetes**. Disponível em <<http://www.maringasaude.com.br/drwesley/diabetes.shtml>>. Acesso em 20 out. 2006.
- [dos Santos, 2003] dos SANTOS, F. L., SOUZA, J. E. G., NETO, B., SANTOS, C. G., SANTOS, M. V. B. e MELO, C. P. (2003). **Free-grown Polypyrrole thin Films as Aroma Sensors**. In: Sensors and Actuators B - Chemical, vol. 88, p. 246-259.
- [dos Santos, 1999] dos SANTOS, F. L. (1999). **Caracterização Elétrica de Polímeros Condutores e de Sensores de Aroma**. Tese (Doutorado em Física), Curso de Pós-Graduação em Física, Departamento de Física, UFPE, Recife, 257p.

- [Duda e Hart, 2001] DUDA, R. O., HART, P. E. e STORK, D. G. (2001). **Pattern Classification**. 2nd Ed. Wiley.
- [Duda e Hart, 1973] DUDA, R. O. e HART, P. E. (1973). **Pattern Classification and Scene Analysis**. Wiley.
- [Dyson, 1938] DYSON, G. M. (1938). **The scientific basis of odour**. In: Chemistry and Industry, vol. 57, p. 647-651.
- [Element Ltd, 2006] Element Ltd. (2006). Disponível em <<http://www.element.is/>>. Acesso em 5 set. 2006.
- [EMCH, 2000] Enciclopédia Multimídia do corpo humano. (2000) Planeta DeAgostini, Planeta do Brasil LTDA, São Paulo.
- [Fend et al., 2005] FEND, R., GEDDES, R., LESELLIER, S., VORDERMEIER, H. M., CORNER, A. L., GORMLEY, E., COSTELLO, E., HEWINSON, R. G., MARLIN, D. J., WOODMAN, A. C. e CHAMBERS, M. A. (2005). **Use of an Electronic Nose to Diagnose Mycobacterium Bovis Infection in Badgers and Cattle**. In: Journal of Clinical Microbiology, vol 43(4), p. 1745–1751.
- [Ferreira e Acar, 1975] FERREIRA, A. G. e ACAR, A. (1975) **Olfato - Atualização (2.^a Parte)**. In: Brazilian Journal of Otorhinolaryngology, vol. 41, p. 276 – 284.
- [Figaro, 2006] FIGARO Engineering Inc. (2006). **Figaro Gas Sensor Technical Reference**. Disponível em <<http://www.figarosensor.com/>>. Acesso em 10 nov. 2006.
- [Figueiredo, 2007] FIGUEIREDO, A. V. P. (2007). **Comparação Estatística de Performance de Métodos de Redes Neurais para Sistema de Olfacção Biológica**. Dissertação (Mestrado em Biometria), Curso de Pós-Graduação em Biometria, Departamento de Estatística e Informática, UFRPE, Recife.
- [Freeman, 1991] FREEMAN, W. J. (1991). **The Physiology of Perception**, In: Scientific American, vol. 264, p. 78-85.
- [Gardner, 1994] GARDNER, J. W. (1994). **Microsensors – Principles and Applications**. John Wiley & Sons, p. 224-246.

- [Gardner et al., 1994] GARDNER, J. W., PEARCE, T. C., FRIEL, S., BARTLETT, P. N. e BLAIR, N. (1994). **A Multisensor System for Beer Flavour Monitoring Array using Pattern Recognition**. In: Sensors and Actuators B, vol. 18, p. 240-243.
- [Gardner e Bartlett, 1994] GARDNER, J. W. e BARTLETT, P. N. (1994). **A Brief History of Electronic Nose**. In: Sensors and Actuators B, vol. 18-19, p. 211-220.
- [Gardner e Bartlett, 1992] GARDNER, J. W. e BARTLETT, P. N.. (1992). **Sensors and Sensory Systems for an Electronic Nose**. In: NATO ASI Series E: Applied Sciences, Kluwer, Dordrecht, vol. 212, p. 346-351.
- [Gardner e Persaud, 2000] GARDNER, J. W. e PERSAUD, K. C. (2000). **Electronic Noses and Olfaction 2000**. Bristol: Institute of Physics Publishing. Disponível em <<http://www.iop.org/EJ/abstract/0957-0233/12/11/702>> Acesso em 21 dez. 2006.
- [Gardner, 1991] GARDNER, J. W. (1991). **Detection of Vapours and Odours from a Multisensor Array Using Pattern Recognition**. Part 1: principal components and cluster analysis. In: Sensors and Actuators B, vol. 4, p. 108-116.
- [Gardner et al., 1990] GARDNER, J. W. HINES, E. L. e WILKINSON, M. (1990). **Application of Artificial Neural Networks to an Electronic Olfactory System**. In: Measurement Science Technology, vol. 1, p. 446-451.
- [Göpel, 1997] GÖPEL, W., HESSE, J. e ZEMEL, J. N. (Eds) (1997). **Sensors: A Comprehensive Survey**. Germany: Book series and Sensors Update book volume series.
- [Guernion et al., 2001] GUERNION, N., RATCLIFFE, N. M. e SPENCER-PHILLIPS, P. T. N. (2001). **Identifying Bacteria in Human Urine: Current Practice and the Potential for Rapid, Near-Patient Diagnosis by Sensing Volatile Organic Compounds**. In: Clinical Chemistry and Laboratory Medicine. vol. 39, n. 10, p. 893-906.
- [Guyton, 1997] GUYTON, A. C. (1997). **Tratado de Fisiologia Médica**. 10 ed. Guanabara Koogan, Rio de Janeiro.

- [Guyton, 1981] GUYTON, A. C. (1981). **Fisiologia Humana**. 5 ed., Interamericana, Rio de Janeiro.
- [Haberly, 1985] HABERLY, L.B. (1985). **Neuronal Circuitry in Olfactory Cortex: Anatomy and Functional Implications**. In: Chemical Senses, vol. 10, p. 219-238.
- [Hagan e Menhaj, 1994] HAGAN, M. T. e MENHAJ, M. B. (1994). **Training Feed-forward Networks with the Marquardt Algorithm**. In: IEE Transaction on Neural Networks, vol. 5, p. 989 – 993.
- [Hair et al., 1998] HAIR, F. F. Jr, ANDERSON, R. E., TATHAM, R. L. e BLACK, W. C. (1998). **Multivariate Data Analysis**. 5th Ed, Upper Saddle River: Prentice Hall.
- [Hauptmann, 1993] HAUPTMANN, P. (1993). **Sensors – Principles and Applications**. Carl Hanser Verlag & Prentice hall, p. 115-153.
- [Haykin, 2001] HAYKIN, S. (2001). **Redes Neurais: Princípios e Práticas**. 2.ed. Porto Alegre: Bookman.
- [Himsworth, 1936] HIMSWORTH, H. P. (1936). **Diabetes Mellitus: its Differentiation into Insulin-Sensitive and Insulin-Insensitive Types**. Lancet, p. 127–130.
- [Hollander e Wolfe, 1973] HOLLANDER, M. e WOLFE, D. A. (1973). **Nonparametric Statistical Methods**, Wiley.
- [Hotelling, 1933] HOTELLING, H. (1933). **Analysis of a complex of statistical variables into principal components**. In: Journal of Educational Psychology, vol. 24, p. 417–441, 498–520.
- [Isaacson, 2001] ISAACSON, J. S. (2001) **Mechanisms Governity Dendritic g - Aminobutyric Acid (GABA) Release in the Rat Olfactory Bulb**. In: Proceedings of National academy of Sciences of USA, vol. 98, p. 337-342.
- [Jacob, 2006] JACOB, T. (2006). **Theories of Olfaction**. Disponível em: <<http://www.cf.ac.uk/biosi/staff/jacob/teaching/sensory/olfact1.html#Theories>> Acesso em 20 dez. 2006. Cardiff University, UK.

- [Jackson, 1991] JACKSON, J. E. (1991). **A User's Guide to Principal Components**. John Wiley and Sonsm Inc., p.1-25
- [Johnson e Wichern, 2002] JOHNSON, R. A. e WICHERN, D. W. (2002). **Applied multivariate statistical analysis**. New Jersey: Prentice Hall.
- [Kaiser, 1958] KAISER, H.F. (1958). **The Varimax Criterion for Analytic Rotation in Factor Analysis**. In: Psychometrika, vol. 23, p. 187-200.
- [Lavigne et al., 1998] LAVIGNE, J. V, SAVOY, S., CLEVINGER, M. B, RITCHIE, J. E., MCDONIEL, B., YOO, S., ANSLYN, E. V., MCDEVITT, J. T., SHEAR, J. B. e NEIKIRK, D. (1998). **Solution-Based Analysis of Multiple Analytes by a Sensor Array: Toward the Development of an "Electronic Tongue"**. In: Journal of the American Chemical Society, vol 120 (5), p. 6429-6430
- [Leffingwell & Associates, 2001] LEFFINGWELL & ASSOCIATES. (2001). **La Olfacción**. Disponível em <http://www.genaltruista.com/notas/00000262.htm>. Acesso em 19 dez. 2006.
- [Levy, 2004] LEVY, I. (2004). **Descoberto Mistério do Olfato**. In: Ciência Hoje On-line, 4 de outubro. Disponível em <http://cienciahoje.uol.com.br/materia/view/1364>. Acesso em 4 de out. de 2006.
- [Li, 1995] LI, Z. (1995). **Modeling the Sensoring Computations of the Olfactory Bulb**. In: Models of Neural Networks, vol. 2, J. L. van Hemmen, E. Domany, and K. Schulten, Springer-Verlag (Eds), New York.
- [Li e Hertz, 2000] LI, Z. e HERTZ, J. (2000). **Odour Recognition and Segmentation by a Model Olfactory Bulb and Cortex**. In: Computing Neural Systems, vol. 11, p. 83-102.
- [Li e Hertz, 1999] LI, Z. e HERTZ, J. (1999). **Odour Recognition and Segmentation by Coupled Olfactory Bulb and Cortical Networks**. In: Neurocomputing, vol. 26-27, p. 789-794.
- [Li e Hopfield, 1989] LI, Z. e HOPFIELD, J. (1989). **A Model of the Olfactory Bulb and its Oscillatory Processing**. In: Biological Cybernetics, vol. 61, p. 379-392.

- [Light, 1992] LIGHT, W. (1992). **Ridge Functions, Sigmoidal Functions and Neural Networks**. CHENEY, E.W., CHUI, C. K., SCHUMAKER, L. L. (Eds), Approximation Theory VII, Boston: Academic Press, p. 163-206.
- [Llecha, 2001] LLECHA, J. B. (2001). **Diseño de una Nariz Electrónica para la Determinación no Destructiva del Grado de Maduración de la Fruta**. Tesis – Departament de Teoria del senyal i comunicacions, Universitat Politècnica de Catalunya, UPC, Catalunya.
- [Llobert et al., 1999] LLOBERT, E., HINES, E.L., GARDNER, J.W. e FRANCO, S. (1999), **Non-destructive Banana Ripeness Determination using a Neural Network-based Electronic Nose**. In: Measurement Science and Technology, vol. 10, p. 538-548.
- [McCulloch e Pitts, 1943] MCCULLOCH, W.S. e PITTS, W. (1943). **A Logical Calculus of Idea Immanent in Neural Activity**. In: Bulletin Math. Biophysics, vol. 5, p. 115-133.
- [Mhaskar, 1996] MHASKAR, H. N. (1996). **Neural networks for Optimal Approximation of Smooth and Analytic Functions**. In: Neural Computation, vol. 8, p. 1731-1742.
- [Miller, 1976] MILLER, B. F. (org.). (1976). **O Livro da Saúde - Enciclopédia Médica Familiar**. In: Seleções do Reader's Digest, Lisboa. p. 554, 739-740, 749, 810-811.
- [Mingoti, 2005] MINGOTI, S. A. (2005). **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**. Belo Horizonte: UFMG.
- [Miret, 2004] MIRET, J. M. I. (2004). **El Olfato sus Receptores. La Historia de um Nobel, Unitat de Rinologia**. In: Servei D'otorino-Laringologia. ICEMEQ. Hospital Clínic de Barcelona. IDIBAPS. Acta Otorrinolaringol, vol. 55, p. 452-456.
- [Mohamed et al., 2002] MOHAMED, E. I., LINDER, R., PERRIELLO, G., DI DANIELE, N., POPPL, S. J. e DE LORENZO, A. (2002). **Predicting Type 2 Diabetes using an Electronic Nose-based Artificial Neural Network Analysis**. In: Diabetes, Nutrition & Metabolism, vol. 15(4), p. 215-221.

- [Moncrieff, 1951] MONCRIEFF, R. W. (1951). **The Chemical Senses**, 2 ed, Leonard Hill, London.
- [Monseley e Prudenziati, 1994] MOSELEY, P. T. e PRUDENZIATI, M. (1994). **Thick Film Sensors**. Elsevier Science. Amsterdam, p. 261-288
- [Monseley e Tofield, 1987] MOSELEY, P. T. e TOFIEL, B. C.(1987). **Solid-State Gas Sensors**. Adam Hilger, Bristol. p. 1-16
- [Mueller, 1966] MUELLER, C. G. (1966). **Psicologia Sensorial**, Zahar Editores, Rio de Janeiro.
- [Neto, 2004] NETO, M. R. L. (2004) **Conferência de Desenvolvimento de Consenso sobre Medicamentos Antipsicóticos, Obesidade e Diabetes**. In: Diabetes Care, vol. 27, p. 596-601
- [Niyogi e Girosi, 1996] NIYOGI, P. e GIROSI, F. (1996). **On the Relationship between generalizaTion Error, Hypothesis Complexity, and Sample Complexity for Radial Basis Functions**. In: Neural Computation, vol. 8, p. 819-842.
- [Omelianski, 1923] OMELIANSKI, V. L. (1923). **Aroma Producing Microorganisms**. In: The Journal of Bacteriology, vol. 8, p. 393-419.
- [Pavlou et al., 2002] PAVLOU, A. K., MAGAN, N., McNULTY, C., JONES, J. M., SHARP, D., BROWN, J. e TURNER, A. P. F. (2002) **Use of an Electronic Nose systEm for Diagnoses of Urinary Tract Infections**. In: Biosensors & Bioelectronics, vol. 17, p. 893-899.
- [Pavlou e Turner, 2000] PAVLOU, A. K. e TURNER, A. P. F. (2000) **Sniffing out the Truth: Clinical Diagnosis Using the Electronic Nose**. In: Clinical Chemistry and Laboratory Medicine, vol. 38, n. 2, p. 99-112.
- [Pearce et al., 2003] PEARCE, T. C., SCHIFFMAN, S. S., NAGLE, H. T. e GARDNER, J. W. (2003). **Handbook of Machine Olfaction: Electronic Nose Technology**. Wiley-VCH.
- [Persaud e Travers, 1997] PERSAUD, K. C. e TRAVERS, P. J. (1997). **Arrays of Broad Specificity Films for Sensig Volatile Chemicals**. In: Handbook of Conducting Polymers, vol. 1, Ed. by T. A. Skotheim, Marcel Dekker, p. 45.

- [Ping et al., 1997] PING, W., YIT, T., HAIBAO, X. e FARONG., S.(1997). **A Novel Method for Diabetes Diagnosis Based on Electronic Nose**. In: Biosensors & Bioelectronics, vol. 1, n. 9-10, p. 1031-1036.
- [Powell, 1985] POWELL, M. J. D. (1985). **Radial Basis Function for Multivariate interpolation: A Review**. In: IMA Conference on Algorithms for the Approximation of Functions and Data, Shrivenham, England, p. 143-167.
- [Prechelt, 1994] PRECHELT, L. (1994). **Proben1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules**. Technical Report 21/94, Fakultät für Informatik, Universität Karlsruhe, Germany.
- [Price, 1987] PRICE, J. L. (1987). **The Central Olfactory and Accessory Olfactory Systems**. In: T. E. Finger e W. L. Silver (Eds.), Neurobiology of Taste and Smell, p. 179-203.
- [Ripley, 2003] RIPLEY, B. D. (1996). **Pattern Recognition and Neural Networks**, Cambridge University Press.
- [Ross e Harriman, 1949] ROSS, S. e HARRIMAN, A. E. (1949). **A Preliminary Study of the Crocker-Henderson Odor-Classification System**. In: The American Journal of Psychology, vol. 62, n. 3, p. 399-404
- [Rumelhart et al., 1986] RUMELHART, D. E., HINTON, G. E. e WILLIAMS, R. J. (1986). **Learning Representations by Backpropagation Errors**. In: Nature, vol. 323, p. 533-536.
- [Santos, 2000] SANTOS, M. S. (2000) **Construção de um Nariz Artificial usando Redes Neurais**. Tese (Doutorado em Ciências da Computação), Curso de Pós-Graduação em Ciência da Computação, Centro de Informática, UFPE, Recife, 120p.
- [Souza, 2000] SOUZA, J. E. G. (2000) **Síntese do Polipirrol e sua Utilização como Sensor de Voláteis**. Dissertação (Mestrado em Química), Curso de Pós-Graduação em Química, Departamento de Química, UFPE, Recife.
- [Toko, 2000] TOKO, K. (2000). **Biomimetic Sensor Technology**. Cambridge University Press, p. 92-111.

- [Triola, 2005] TRIOLA, M. F. (2005). **Introdução à Estatística**. Rio de Janeiro: Livros Técnicos e Científicos.
- [Turin e Yashii, 2003] TURIN, L., YASHII, F. (2003). **Structure-Odor Relationships: A modern Perspective**. In: Handbook of Olfaction and Gustation, 2 ed. DOTY, R. L. (Ed), , p. 275-294
- [Turin, 1996] TURIN, L. (1996). **A Spectroscopic Mechanism for Primary Olfactory Reception**. In: Chemical Senses, vol. 21/6, p. 773-791.
- [UB, 2002] UNIVERSIA BRASIL. **Nariz Eletrônico - Software que Detecta Odores à Base de Poliméricos foi Desenvolvido pela UFPE e pela UNICAP**. Disponível em <http://www.universia.com.br/html/materia/materia_ibi.html>. Acesso em 20 dez. 2006.
- [Vilela, 2006] VILELA, A. L. K. (2006). **O Olfato**. Disponível em <<http://www.afh.bio.br/sentidos/Sentidos7.asp>>. Acesso em: 20 dez. 2006.
- [Warren, 1994] WARREN, S.S. (1994). **Neural Networks and Statistical Model**. In: Proceedings of the Nineteenth Annual SAS Users Group International Conference, April.
- [WHO, 1999] WORLD HEALTH ORGANIZATION. (1999). Department of Noncommunicable Disease Surveillance. **Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications**. Geneva. Disponível em <<http://www.who.int/en/>>. Acesso em 18 dez. 2006.
- [Wilkins e Hatman, 1964] WILKENS, W. F. e HATMAN, A. D. (1964). **An Electronic Analog for the Olfactory Processes**, In: Annals of the New York Academy of Sciences, vol. 116, p. 608-612.
- [Wright, 1977] WRIGHT, R. H. (1977). **Odor and molecular vibration: neural coding of olfactory information**. In: Journal of Theoretical Biology, vol. 64, p. 473-502.
- [Yao e Freeman, 1990] YAO, Y. e FREEMAN, W. (1990). **Model of Biological Pattern Recognition with Spatially Chaotic Dynamics**. In: Neural networks, vol. 3, p. 153-170

- [Zanchettin e Ludermir, 2005] ZANCHETTIN, C. e LUDERMIR, T. B. (2005). **Sistemas Neurais Híbridos para Reconhecimento de Padrões em Narizes Artificiais**, In: Sba Controle & Automação, vol.16, n. 2, p.159-172.
- [Zanchettin et al., 2003] ZANCHETTIN, C., LUDERMIR, T. B. e YAMAZAKI, A. (2003). **Classification of Gases from the Petroliferous Industry by an artificial Nose with Neural Networks**. In: Joint 13th International Conference on Artificial Neural Networks and 10th Internacional Conference on Neural Information Processing, Istanbul, Turkey, p.208-211.