



FC

FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO

DEPARTAMENTO DE  
MATEMÁTICA APLICADA

# As Redes Neurais na Análise de Tráfego com o GPS

Dissertação de Mestrado em Posicionamento e Navegação por Satélite

Maria Alexandrina da Silva Meneses



FC

FACULDADE DE CIÊNCIAS  
(UNIVERSIDADE DE COIMBRA)

DEPARTAMENTO DE  
MATEMÁTICA APLICADA

# As Redes Neurais na Análise de Tráfego com o GPS

Dissertação de Mestrado em Posicionamento e Navegação por Satélite

Maria Alexandrina da Silva Meneses

# As Redes Neurais na Análise de Tráfego com o GPS

Maria Alexandrina da Silva Meneses

Dissertação de Mestrado em Posicionamento e Navegação por Satélite sob a  
Orientação Científica da Professora Doutora Isabel Maria Teixeira Vergueiro

Pereira Osório

Faculdade de Ciências da Universidade do Porto

## **Júri das Provas de Mestrado**

Presidente: Doutor José Joaquim de Sousa Pereira Osório

Vogais: Doutor Afzal Suleman

Doutora Teresa Maria de Gouveia Torres Feio Mendonça

Doutor José Alberto Álvares Pereira Gonçalves

Doutora Isabel Maria Teixeira Vergueiro Pereira Osório

Março 2003

## **AGRADECIMENTOS**

Várias pessoas e entidades contribuíram, de forma diversa, para que este trabalho pudesse ser levado a bom termo. Assim, quero aqui expressar os meus agradecimentos:

- À Professora Doutora Isabel Pereira Osório pela orientação desta tese, empenho e apoio demonstrado ao longo de todo o trabalho.
  
- Ao Professor Doutor José Pereira Osório pelos comentários e sugestões dadas no tratamento de algumas questões matemáticas que aqui foram abordadas.
  
- À Câmara Municipal de Santa Maria da Feira, nomeadamente ao Vereador do Pelouro do Planeamento e Urbanismo, José Manuel Oliveira, pelo fornecimento da base cartográfica digital e disponibilização de meios técnicos para a realização deste trabalho.
  
- À Faculdade de Ciências da Universidade do Porto pela disponibilização de meios para a realização deste trabalho.
  
- Aos colegas da Câmara Municipal e amigos que generosamente aceitaram participar neste trabalho, bem como, aos que contribuíram de forma sistemática no desenvolvimento de alguns aspectos, pelos comentários e sugestões.
  
- Ao Rui pela forma como me acompanhou ao longo da realização deste trabalho, pelo incentivo e inestimável colaboração.

## RESUMO

O trabalho apresentado nesta dissertação tem como objectivo um estudo que visa prever os percursos efectuados pela maioria dos condutores, dentro de uma cidade, utilizando como ferramenta de análise as redes neuronais artificiais e o Sistema Global de Posicionamento por Satélite (GPS). Uma rede neuronal artificial não é mais do que um instrumento de análise e programação baseado no cérebro humano, que permite a aquisição de conhecimento após uma fase de aprendizagem. Escolhendo um par origem-destino, o estudo visa responder a questões tipo: quais as ruas mais percorridas às 16h, quais os arcos congestionados face às condições atmosféricas ou, ainda, quais as preferências dos condutores face à percepção que têm do trânsito no momento da condução.

No primeiro capítulo aborda-se a temática das redes neuronais sob o ponto de vista histórico, como tiveram início, como foram sendo melhoradas desde os primeiros estudos até à actualidade, bem como aplicações práticas em diversas áreas que vão desde a medicina aos transportes. No segundo capítulo são abordadas as principais características das redes neuronais (funções de activação, arquitectura, aprendizagem e recorrência), os perceptrões e alguns métodos de optimização numérica. No terceiro capítulo é abordado o algoritmo de Retropropagação, tipos de processamento, critérios de paragem, generalização e outras variantes do algoritmo (optimização). O quarto capítulo é dedicado à contribuição do GPS no trabalho que foi desenvolvido, nomeadamente os receptores utilizados, o tipo de medidas realizadas e a forma como foram processadas. Neste capítulo é, ainda, tratado o modo de utilização do GPS tanto para a actualização da cartografia digital disponível, incluindo as características que serão usadas nas redes neuronais, como para o registo dos percursos efectuados livremente por diferentes condutores em diferentes condições. A implementação da rede neuronal é abordada no quinto capítulo, respondendo a questões como: o porquê da escolha de um modelo em detrimento de outro, a escolha do número de neurónios, da arquitectura e do número de camadas intermédias.

O trabalho é concluído com um conjunto de possíveis melhoramentos ao sistema apresentado.

São, ainda, apresentados no final deste trabalho as referências bibliográficas, a bibliografia e os apêndices.

## **ABSTRAT**

The work presented in this thesis aims to study methods to foresee the road courses made by the majority of drivers inside a city, by using as tool of analysis the artificial neural networks and the Global Positioning System (GPS). An artificial neural network is no more than an instrument of analysis and programming based on the human brain, which allows the acquisition of knowledge after a learning period. Choosing a pair origin – destination, the study aims to answer questions like: which are the most used roads at 16h00 p.m., which are the congested arcs concerning the atmospheric conditions or which are the drivers preferences concerning their traffic preception at the driving moment.

In the first chapter the neural networks thematic is specified under the historical point of view, how did they begin, how they have been worked out since the first studies until today, as well as practical applications in several areas from medicine to transports.

In the second chapter it is specified the main characteristics of the neural networks (activation functions, architecture, learning and recurrence), the perceptrons and some numerical optimization methods.

The third chapter deals with the back-propagation algorithm, kinds of processing, stopping criteria, generalization and other algorithm variations (optimization).

The forth chapter is dedicated to the contribution of GPS to this work, namely GPS used receivers, types of measures accomplished and how they were processed. This chapter also treats GPS working methodologies, collected data precision, road courses made by the drivers, digital cartography and geographical information system used.

The neural network implementation is specified in the fifth chapter, answering questions like: why to choose a model in prejudice of another, the selection of the neuron number, the architecture and the number of intermediate layers.

The conclusion embraces a set of possible improvements on the presented system.

At the end of this work the bibliographical references, bibliography and appendix are given.

## ÍNDICE

|                                                            |           |
|------------------------------------------------------------|-----------|
| <b>Agradecimentos</b> .....                                | <b>3</b>  |
| <b>Resumo</b> .....                                        | <b>4</b>  |
| <b>Abstrat</b> .....                                       | <b>5</b>  |
| <b>Índice</b> .....                                        | <b>6</b>  |
| <b>Índice de figuras</b> .....                             | <b>8</b>  |
| <b>Índice de tabelas</b> .....                             | <b>10</b> |
| <b>Siglas</b> .....                                        | <b>11</b> |
| <b>Acrónimos</b> .....                                     | <b>11</b> |
| <b>1. INTRODUÇÃO ÀS REDES NEURONAIIS ARTIFICIAIS</b> ..... | <b>12</b> |
| <b>1.1 Introdução</b> .....                                | <b>13</b> |
| <b>1.2 Um pouco de História</b> .....                      | <b>15</b> |
| 1.2.1 Neurónio de <i>McCulloch-Pitts</i> .....             | 15        |
| 1.2.2 Regra de <i>Hebb</i> .....                           | 16        |
| 1.2.3 Perceptrões .....                                    | 17        |
| 1.2.4 Adaline .....                                        | 17        |
| 1.2.5 Memória Associativa .....                            | 18        |
| 1.2.6 <i>Hopfield</i> .....                                | 19        |
| 1.2.7 Retropropagação .....                                | 19        |
| <b>1.3 Aplicações das Redes Neurais Artificiais</b> .....  | <b>20</b> |
| 1.3.1 Medicina .....                                       | 20        |
| 1.3.2 Controlo .....                                       | 21        |
| 1.3.3 Processamento do Sinal .....                         | 21        |
| 1.3.4 Processamento de imagens .....                       | 22        |
| 1.3.5 Robótica .....                                       | 22        |
| 1.3.6. Transportes.....                                    | 22        |
| <b>2. PERCEPTRÃO</b> .....                                 | <b>24</b> |
| <b>2.1 Introdução</b> .....                                | <b>25</b> |
| <b>2.2 Caracterização das redes</b> .....                  | <b>25</b> |
| 2.2.1 Arquitectura .....                                   | 25        |
| 2.2.2 Funções de Activação.....                            | 26        |
| 2.2.3 Aprendizagem.....                                    | 27        |
| 2.2.3.1 Aprendizagem Supervisionada.....                   | 28        |

|            |                                                                  |           |
|------------|------------------------------------------------------------------|-----------|
| 2.2.3.2    | Aprendizagem Não Supervisionada .....                            | 29        |
| 2.2.4      | Estilos de treino .....                                          | 29        |
| <b>2.3</b> | <b>O Perceptrão .....</b>                                        | <b>30</b> |
| 2.3.1      | Teorema de convergência do perceptrão .....                      | 30        |
| 2.3.2      | Perceptrões de multicamadas .....                                | 32        |
| <b>2.4</b> | <b>Métodos de optimização.....</b>                               | <b>32</b> |
| 2.4.1      | Método do gradiente decrescente .....                            | 33        |
| 2.7.2      | Método de <i>Newton</i> .....                                    | 34        |
| 2.7.3      | Método de <i>Gauss-Newton</i> .....                              | 35        |
| <b>3.</b>  | <b>RETROPROPAGAÇÃO .....</b>                                     | <b>37</b> |
| <b>3.1</b> | <b>O Algoritmo de Retropropagação .....</b>                      | <b>38</b> |
| <b>3.2</b> | <b>O termo momento .....</b>                                     | <b>44</b> |
| <b>3.3</b> | <b>Aspectos práticos .....</b>                                   | <b>45</b> |
| 3.3.1      | Seleccção da arquitectura da rede e dos vectores de entrada..... | 46        |
| 3.3.2      | Tipos de processamento .....                                     | 46        |
| 3.3.3      | Pré-processamento.....                                           | 47        |
| 3.3.4      | Pós-processamento .....                                          | 47        |
| 3.3.5      | Número de neurónios na camada intermédia .....                   | 47        |
| 3.3.6      | Parâmetros de aprendizagem e termo momento .....                 | 48        |
| 3.3.7      | Mínimos locais .....                                             | 48        |
| 3.3.8      | Critérios de Paragem .....                                       | 48        |
| <b>3.4</b> | <b>Variantes do modelo de Retropropagação .....</b>              | <b>49</b> |
| 3.4.1      | Método do Gradiente Conjugado.....                               | 50        |
| 3.4.2      | Método Quasi-Newton .....                                        | 52        |
| <b>3.5</b> | <b>Generalização.....</b>                                        | <b>54</b> |
| 3.5.1      | Tamanho da amostra .....                                         | 55        |
| 3.5.2      | Método da Validação Cruzada .....                                | 56        |
| 3.5.3      | Método de Eliminação de Pesos .....                              | 57        |
| 3.5.4      | Método de Regularização .....                                    | 57        |
| <b>4.</b>  | <b>UTILIZAÇÃO DO GPS .....</b>                                   | <b>58</b> |
| <b>4.1</b> | <b>Introdução .....</b>                                          | <b>59</b> |
| <b>4.2</b> | <b>Actualização da cartografia digital.....</b>                  | <b>62</b> |
| <b>4.3</b> | <b>Registo de percursos entre a origem-destino .....</b>         | <b>68</b> |
| <b>4.4</b> | <b>Resultados obtidos.....</b>                                   | <b>74</b> |
| <b>5.</b>  | <b>APLICAÇÃO PRÁTICA DAS REDES NEURONAIS.....</b>                | <b>77</b> |

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| 5.1 Introdução .....                                               | 78  |
| 5.2 Criação da rede neuronal .....                                 | 79  |
| 5.3 Análise dos percursos efectuados.....                          | 89  |
| 6. CONCLUSÕES E LINHAS DE ORIENTAÇÃO FUTURA.....                   | 96  |
| 7. REFERÊNCIAS BIBLIOGRÁFICAS E BIBLIOGRAFIA.....                  | 101 |
| APÊNDICES .....                                                    | 108 |
| Apêndice 1 – Teorema de convergência do perceptrão.....            | 109 |
| Apêndice 2 – Método do gradiente decrescente .....                 | 110 |
| Apêndice 3 – Normalização dos vectores de entrada e de saída ..... | 111 |

## ÍNDICE DE FIGURAS

|                                                                                              |                                     |
|----------------------------------------------------------------------------------------------|-------------------------------------|
| Fig. 1.1 - Redes neuronais biológicas [16].....                                              | 14                                  |
| Fig. 1.2 - Neurónio de McCulloch-Pitts [6].....                                              | 16                                  |
| Fig. 1.3 - Modelo de rede neuronal do tipo perceptrão .....                                  | 17                                  |
| Fig. 1.4 - Rede neuronal do tipo perceptrão de multicamadas .....                            | 20                                  |
| Fig. 2.1 - Função degrau .....                                                               | 26                                  |
| Fig. 2.2 - Função rampa .....                                                                | 26                                  |
| Fig. 2.3 - Funções sigmóide .....                                                            | 27                                  |
| Fig. 2.4 - Funções sigmóide anti-simétricas .....                                            | 27                                  |
| Fig. 3.1 - Neurónio j de uma rede neuronal com uma camada intermédia .....                   | 38                                  |
| Fig. 3.2 - Neurónios j e k de uma rede neuronal com duas camadas intermédias .....           | 41                                  |
| Fig. 3.3 - Comportamento do Algoritmo de Retropropagação com e sem o termo momento [14]..... | 45                                  |
| Fig. 3.4 - Aceleração do Algoritmo de Retropropagação usando o termo momento [13].....       | 45                                  |
| Fig. 3.5 - Validação cruzada.....                                                            | 56                                  |
| Fig. 4.1 - Área de estudo .....                                                              | <b>Erro! Marcador não definido.</b> |
| Fig. 4.2 - Tipo de pavimento.....                                                            | 62                                  |
| Fig. 4.3 - Estado de conservação do pavimento .....                                          | 63                                  |
| Fig. 4.4 - Largura da via .....                                                              | 63                                  |
| Fig. 4.5 - Sinalização: lombas, paragem de autocarro, passadeiras e semáforos.....           | 64                                  |
| Fig. 4.6 - Sinalização: sinais de trânsito.....                                              | 64                                  |
| Fig. 4.7 - Constrangimentos na via : carros estacionados em segunda fila.....                | 65                                  |
| Fig. 4.8 - Constrangimentos na via – conflitos com peões .....                               | 65                                  |
| Fig. 4.9 - Estacionamento .....                                                              | 66                                  |
| Fig. 4.10 - Vias de sentido único .....                                                      | 66                                  |
| Fig. 4.11 - Presença de veículos pesados .....                                               | 67                                  |
| Fig. 4.12 - Ruas sem saída ou não conduzem ao destino pretendido .....                       | 67                                  |

|                                                                                                                                             |    |
|---------------------------------------------------------------------------------------------------------------------------------------------|----|
| Fig. 4.13 - Percentagem de condutores que trabalham em Sta. Maria da Feira .....                                                            | 70 |
| Fig. 4.14 - Condições atmosféricas .....                                                                                                    | 70 |
| Fig. 4.15 - Horário de realização dos percursos.....                                                                                        | 70 |
| Fig. 4.16 - Estado do pavimento face às condições climatéricas.....                                                                         | 71 |
| Fig. 4.17 - Percepção do trânsito.....                                                                                                      | 71 |
| Fig. 4.18 - Variação dos comprimentos dos percursos.....                                                                                    | 72 |
| Fig. 4.19 - Tempo de percurso .....                                                                                                         | 72 |
| Fig. 4.20 - Arcos mais percorridos .....                                                                                                    | 73 |
| Fig. 4.21 - Distribuição dos percursos e tempos médios por horário .....                                                                    | 74 |
| Fig. 4.22 - Precisão, PDOP e número de satélites visíveis.....                                                                              | 75 |
| Fig. 4.23 - Número de satélites visíveis .....                                                                                              | 76 |
| Fig. 5.1 - Percentagem de erro na classificação em função do número de neurónios utilizado na<br>camada intermédia .....                    | 79 |
| Fig. 5.2 - Percentagem de erro obtido na classificação em função da taxa de aprendizagem.....                                               | 80 |
| Fig. 5.3 - Percentagem de erro na classificação em função do termo momento .....                                                            | 80 |
| Fig. 5.4 - Desempenho da rede com a taxa de aprendizagem variável .....                                                                     | 81 |
| Fig. 5.5 - Análise da rede neuronal arco a arco .....                                                                                       | 82 |
| Fig. 5.6 - Percentagem de erro na classificação total dos arcos.....                                                                        | 82 |
| Fig. 5.7 - Percentagem de erro na classificação: percurso 4 e 10 .....                                                                      | 83 |
| Fig. 5.8 - Percurso 4 (a vermelho) e 10 (a verde) realizado pelos condutores .....                                                          | 84 |
| Fig. 5.9 - Resposta dada pela rede neuronal para o percurso 4 (a vermelho) e 10 (a verde).....                                              | 84 |
| Fig. 5.10 - Comparação entre o algoritmo Retropropagação e Retropropagação com momento ....                                                 | 86 |
| Fig. 5.11 - Comparação entre algoritmo Gradiente Conjugado e Quasi-Newton.....                                                              | 86 |
| Fig. 5.12 - Comparação entre a utilização de um algoritmo com optimização numérica (Quasi-<br><i>Newton</i> ) e outro sem optimização ..... | 86 |
| Fig. 5.13 - Erro obtido na previsão da percentagem de arcos percorridos.....                                                                | 88 |
| Fig. 5.14 - Previsão da rede neuronal para o grupo 2 .....                                                                                  | 88 |
| Fig. 5.15 - Percentagem de erro na previsão dos arcos mais percorridos (exemplo dos arcos 69 a<br>90).....                                  | 88 |
| Fig. 5.16 - Arcos congestionados: percepção do trânsito elevada.....                                                                        | 89 |
| Fig. 5.17 - Arcos congestionados entre as 14h e as 15.30h .....                                                                             | 90 |
| Fig. 5.18 - Arcos mais percorridos entre as 10h e as 11h (percepção do trânsito: pouco e médio) ..                                          | 90 |
| Fig. 5.19 - Arcos mais percorridos entre as 16h e as 18h30m.....                                                                            | 91 |
| Fig. 5.20 - Arcos mais percorridos face às condições atmosféricas (chuva ou sol) e hora do dia<br>(diurno ou nocturno) .....                | 91 |
| Fig. 5.21 - Arcos alternativos aos congestionados (percentagem de utilização inferior a 3).....                                             | 92 |
| Fig. 5.22 - Arcos percorridos em piso do tipo paralelo, em dias de chuva e entre as 16h e as<br>18h30m .....                                | 92 |
| Fig. 5.23 - Percentagem de pessoas que percorreram o arco em dias de piso seco e a hora de ponta<br>.....                                   | 93 |

|                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------|----|
| Fig. 5.24 - Percentagem de pessoas que percorreram os arcos em dias de piso seco e não a hora de ponta .....    | 93 |
| Fig. 5.25 - Percentagem de pessoas que percorreram os arcos em dias de piso molhado e não a hora de ponta ..... | 94 |

## ÍNDICE DE TABELAS

|                                                                                              |    |
|----------------------------------------------------------------------------------------------|----|
| Tabela 4.1 - Exemplo da base de dados levantada pelo operador durante os trajectos.....      | 68 |
| Tabela 4.2 - Quadro descritivo do número de vezes que cada condutor realizou o percurso..... | 69 |
| Tabela 4.3 - Distribuição dos percursos por horário.....                                     | 74 |

## **SIGLAS**

DGPS – Differential GPS

FKP – Elächenkorrekturparameter

GPS – Global Positioning System

GSM – Global System for Mobile Communications

IGP – Instituto Geográfico Português

RNA – Redes Neurais Artificiais

RNB – Redes Neurais Biológicas

RTCM – Radio Technical Commission for Marine Services

RTK – Real Time Kinematics

SMS – Short Message Service

VRS – Virtual Reference Station

## **ACRÓNIMOS**

NASA – National Aeronautics and Space Administration

DOP – Dilution Of Precision

## **1. INTRODUÇÃO ÀS REDES NEURONAIS ARTIFICIAIS**

## 1.1 INTRODUÇÃO

Redes Neuronais ou Redes Neuronais Artificiais (RNA) são ferramentas de análise, adaptadas para linguagem de programação matemática e baseadas na estrutura neuronal do cérebro humano, que permitem adquirir conhecimento após uma fase de aprendizagem. Trata-se de um sistema de processamento de informação que contém as mesmas características que as Redes Neuronais Biológicas (RNB), ou seja, a informação é processada através de simples elementos denominados neurónios.

O sistema nervoso é composto por células nervosas denominadas neurónios. Os neurónios, por sua vez, são compostos por dendritos e axónios. Os dendritos não são mais do que ligações que permitem aos neurónios receberem a informação que provém de outras células nervosas. Os axónios são extensões dos neurónios que servem de canal de transporte unidireccional entre os neurónios e o botão sináptico. Cada célula tem apenas um axónio, que por sua vez pode ter diversas ramificações [16]. Na Fig. 1.1 estão exemplificados os elementos que compõem uma célula nervosa, bem como outros elementos que permitem ao sistema nervoso transmitir as mensagens recebidas. Trata-se do botão sináptico, da sinapse e da zona de transferência. Denomina-se por botão sináptico a extremidade do axónio, local onde a carga eléctrica proveniente do neurónio é transformada em sinal químico, por sinapse a zona onde ocorre a comunicação química e por zona de transferência a zona sensível aos sinais transmitidos pelos axónios. Os neurónios recebem informações de outras células nervosas e avaliam se transmitem ou não a informação que receberam. Um neurónio pode estar ligado a 10 000 outros neurónios, ou, mesmo mais [16].

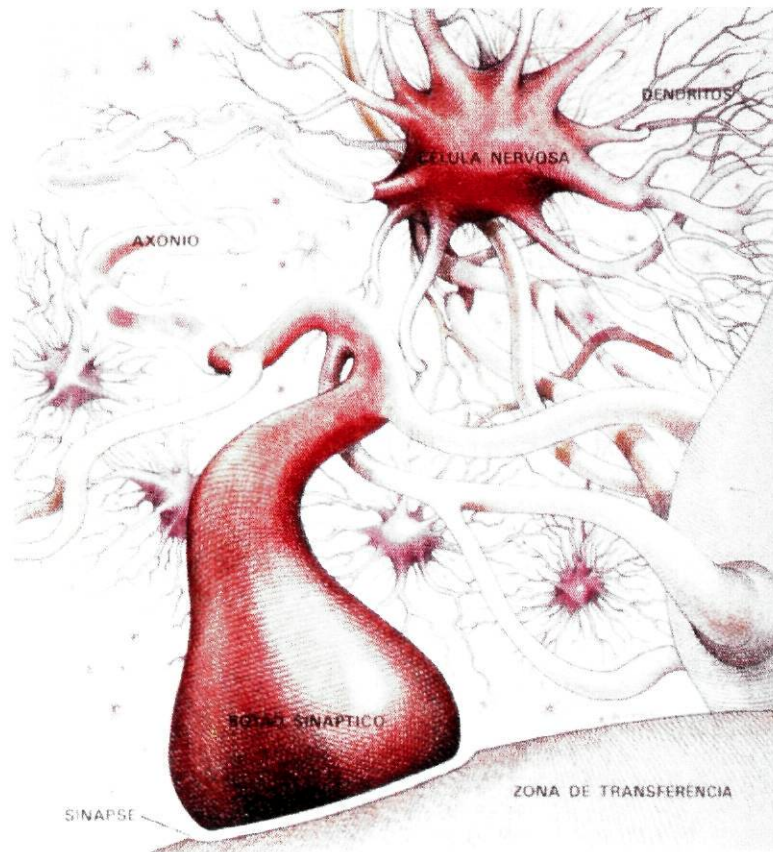


Fig. 1.1 - Redes neuronais biológicas [16]

Apesar das RNA serem construídas à semelhança das RNB, subsistem diferenças significativas entre estes dois tipos de redes. Numa RNA as ligações entre neurónios são compostas por pesos que correspondem à excitação ou inibição das ligações e, por isso, os pesos podem tomar valores positivos ou negativos. Os estados de um neurónio – excitação e inibição – são transmitidos entre os vários neurónios de uma rede, através de um valor que representa o seu actual nível de estado. Nas RNB, a valores elevados de activação correspondem elevados batimentos e a variação da frequência ou do batimento equivale à activação do nível de estado. Nas RNB existem diversos tipos de neurónios, enquanto que numa RNA, geralmente, só é utilizado um tipo de neurónios, podendo, no entanto, em certas redes, utilizarem-se dois ou três tipos diferentes de neurónios. A quantidade de neurónios de uma RNB é muito elevada quando comparada com a de uma RNA. Assim, enquanto que uma RNA pode conter entre uma dúzia e algumas centenas de neurónios, numa RNB estima-se que os neurónios rondem os 10-500 biliões [3].

## 1.2 UM POUCO DE HISTÓRIA

Antes de iniciar um estudo mais detalhado sobre as RNA, considerou-se conveniente abordar o tema sob o ponto de vista histórico: como surgiram, quem as inventou, quais os cientistas que mais contribuíram para o seu desenvolvimento, quais os algoritmos e aplicações mais utilizadas.

### 1.2.1 Neurónio de *McCulloch-Pitts*

Em meados de 1940, o neurofisiologista *McCulloch* e o matemático *Pitts* publicaram um dos mais famosos artigos científicos sobre RNA. Nesse artigo, os cientistas desenvolveram teoremas relacionados com modelos neuronais inspirados no cérebro humano. Teoremas esses que se baseiam nas seguintes condições: a activação do neurónio é binária; um certo número de sinapses (número esse independente de anteriores actividades e da posição do neurónio) deve ser excitado num determinado período, de forma a excitar um neurónio em qualquer momento; o único atraso significativo no interior de um sistema nervoso é o atraso sináptico; uma sinapse inibidora impede a excitação do neurónio nesse momento; a estrutura de uma rede neuronal não se altera com o tempo. Um neurónio que obedeça a estas cinco conclusões denomina-se 'neurónio de *McCulloch-Pitts*' [3].

Os sinais a introduzir numa rede neuronal são apresentados à entrada, cada sinal de entrada é multiplicado pelo seu peso, indicando assim a sua influência na saída. Após a soma ponderada dos sinais, o resultado obtido determina o nível de actividade. Se o nível de actividade for superior a um dado número, a unidade produz uma resposta positiva (excitação) na saída; se o nível de actividade for inferior a unidade produz uma resposta negativa (inibição). Trata-se de uma RNA estática, ou seja, os pesos não são actualizados.

Sejam  $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$  um vector de entrada da rede neuronal, de dimensão  $m$ ,  $\mathbf{W}$  a matriz de pesos, de dimensão  $s \times m$ , correspondente a uma camada de  $s$  neurónios,  $x_0$  uma activação virtual com valor unitário,  $w_{k0}$  o respectivo peso (desvio sistemático) correspondente ao  $k$ -ésimo neurónio, com  $k = 1, \dots, s$  e, finalmente,  $\mathbf{y} = [y_1, y_2, \dots, y_s]^T$  o vector da camada de saída [6].

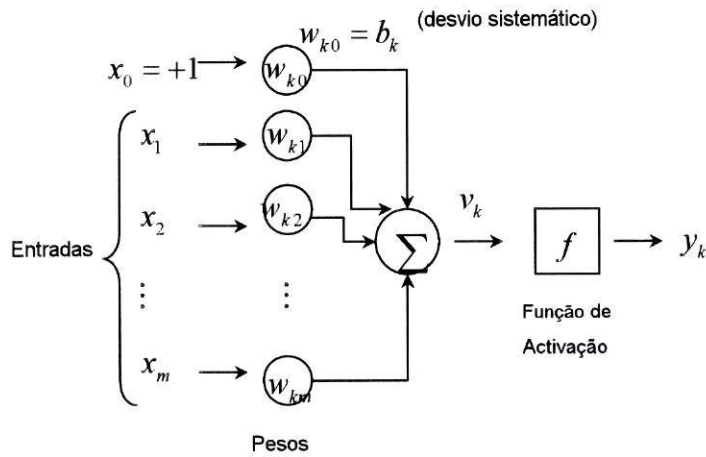


Fig. 1.2 - Neurónio de McCulloch-Pitts [6]

As componentes de ordem  $k$  dos vectores  $v$  e  $y$  são dados pelas seguintes expressões, respectivamente:

$$v_k = \sum_{j=1}^m w_{kj} x_j + b_k \quad [1.1]$$

$$y_k = f(v_k) \quad [1.2]$$

### 1.2.2 Regra de Hebb

Alguns anos mais tarde, o psicólogo *Donald Hebb* introduziu um novo conceito: a actualização dos pesos numa RNA dependente da sua função de activação. Nasceu, assim, a primeira lei de aprendizagem das RNA, que diz que, se dois neurónios forem activados em simultâneo, então, a robustez da conexão entre as suas ligações deve ser aumentada, isto é, se a saída de um neurónio A num sistema nervoso estimula repetidamente um segundo neurónio B a disparar, o caminho entre A e B torna-se mais eficiente no transporte do estímulo [4].

Considerando os neurónios  $i$  e  $j$  e os respectivos pesos  $w_{ij}$ , a Regra de *Hebb* é dada por:

$$\Delta w_{ij} = \eta S_i S_j, \quad \text{onde} \quad [1.3]$$

$S_i$  – Representa o sinal de saída transmitido pelo neurónio  $i$  para o neurónio  $j$

$S_j$  – Representa o sinal de saída do neurónio  $j$

$\eta$  – Representa uma constante de proporcionalidade indicando o valor da ligação

$\Delta w_{ij}$  – Representa a aprendizagem

Quando as quantidades  $s_i$  e  $s_j$  são grandes, a relação entre os neurónios  $i$  e  $j$  é fortalecida, gerando aprendizagem. Apesar do trabalho de *Hebb* ser inovador no ajuste dos pesos e na teoria de aprendizagem, ele não incluiu o conceito de observar os efeitos da combinação de múltiplas entradas num simples neurónio [1].

### 1.2.3 Perceptrões

EM 1958, *Frank Rosenblatt* definiu uma nova estrutura de rede neuronal: o Perceptrão. O Perceptrão foi descrito como uma 'máquina de aprendizagem' devido às suas capacidades de aprendizagem na classificação de certos padrões modificando apenas as suas ligações [3].

O Perceptrão mais conhecido é composto por uma camada de entrada, conectada através de caminhos com pesos ajustáveis, a uma outra camada de neurónios. A regra de aprendizagem do Perceptrão utiliza um processo iterativo para proceder ao ajuste dos pesos [4].

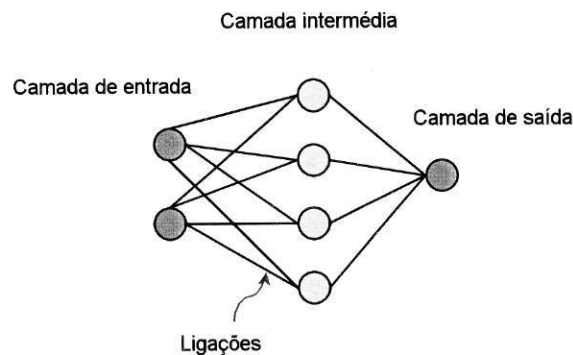


Fig. 1.3 - Modelo de rede neuronal do tipo perceptrão

### 1.2.4 Adaline

Em 1960, *Bernard Widrow* e *Marcian Hoff* publicaram um dos mais importantes artigos relacionados com a tecnologia das RNA, denominado '*Adaptive Switching Circuits*' [3]. *Widrow* e *Hoff* desenvolveram uma nova regra de aprendizagem que está directamente relacionada com a regra

referida no parágrafo anterior. Essa regra de aprendizagem é conhecida como Regra Delta. Enquanto na regra de aprendizagem do Perceptrão, os pesos são ajustáveis iterativamente, sempre que a resposta à unidade for incorrecta, na Regra Delta, os pesos são ajustados de forma a reduzir a diferença entre as saídas da rede e a saída desejada. Este facto resulta num erro médio quadrático menor e numa maior capacidade de generalização da rede. Este tipo de RNA é designada pelo acrónimo Adaline (*Adaptive Linear Neuron*) [4].

A Regra Delta, também conhecida como Regra dos Mínimos Quadrados, tem como base o método dos mínimos quadrados, ou seja, o algoritmo procede a um contínuo ajuste dos pesos até que o erro entre o alvo e a saída calculada seja reduzido. O principal objectivo deste método consiste em exprimir a diferença entre o actual valor da saída e o valor alvo em termos dos valores de entrada e dos pesos. O erro médio quadrático entre a saída alvo e a actual saída é dado por:

$$\begin{aligned} E &= \frac{1}{2}(t_i - a_i)^2 \\ &= \frac{1}{2}(t_i - f(w_i x_i))^2 \end{aligned}$$

sendo  $t_i$  os valores alvos e  $a_i$  os valores da saída calculada pelo algoritmo. O erro do vector gradiente é dado por:  $\Delta E = -(t_i - a_i)f'(w_i x_i)$  [8].

### 1.2.5 Memória Associativa

*Marvin Minsky* e *Seymour* lançaram um livro, em 1969, subordinado ao tema Perceptrões. Nesse livro, os cientistas classificam os Perceptrões como elementos sem valor científico e os teoremas de convergência desenvolvidos, até então, sem razão de ser. O livro incide, essencialmente, em elementos que os Perceptrões não conseguiam solucionar e não sobre o que eles conseguiam resolver. Apesar dos ânimos terem desfalecido, as investigações à volta das RNA continuaram durante a década de 70.

O cientista *Teuvo Kohonen* desenvolveu uma estrutura de RNA denominada memória associativa [4]. Trata-se de uma rede neuronal com capacidade de aprendizagem, isto é, sem realizar um treino supervisionado. Até essa época, todas as redes neuronais utilizavam o treino da rede, ou seja, a rede era supervisionada. O sistema que *Kohonen* desenvolveu permite realizar representações dos sinais externos fora e dentro do sistema neuronal, sem ser necessário a intervenção humana.

Surgiram, assim, os primeiros sucessos na utilização de RNA no reconhecimento de voz e padrões [1, 3].

*Stephen Grossberg* introduziu alguns conceitos revolucionários nas RNA. Ele desenvolveu a chamada teoria de ressonância adaptável. Essa teoria diz que se um neurónio, num conjunto de muitos neurónios, é excessivamente excitado, então, os neurónios vizinhos vão ser afectados de uma forma inibidora. *Grossberg* contribuiu também para a teoria da memória (como os padrões podem continuar activos mesmo após terem parado as entradas na rede) e realizou estudos na área das funções de transferência, nomeadamente na função logística [3].

### 1.2.6 Hopfield

Em 1982, *John Hopfield* desenhou uma nova estrutura de rede neuronal, utilizando conceitos definidos na década de 70. A rede neuronal de *Hopfield* era constituída por camadas de neurónios que estavam interligados entre si, entre as camadas de entrada e saída. O conceito principal residia no facto de todos os neurónios transmitirem sinais para a frente e para trás num ciclo fechado, até estabilizar a rede. *Hopfield* não utilizava o método de actualização dos pesos definidos no Perceptrão como forma de melhorar as saídas; ele apenas tirava partido da transmissão de sinais, entre neurónios, para trás e para a frente [1].

### 1.2.7 Retropropagação

Na década de 70 é apresentado, por *Werbos*, um método de propagação da informação sobre os erros desde as saídas até às entradas de uma RNA. Anos mais tarde, este mesmo método foi re-inventado, independentemente, por *David Parker*, em 1985, e por *LeCun*, em 1986. Mais tarde ainda, já na posse do trabalho desenvolvido pelos anteriores colegas, os cientistas *David Rumelhart* e *James McClelland* aperfeiçoaram-no e publicitaram-no. O método foi denominado Retropropagação [4].

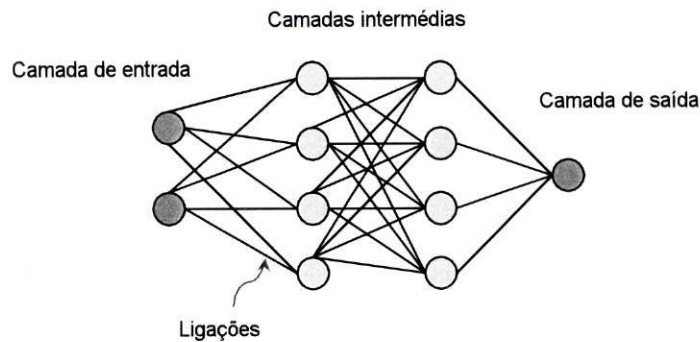


Fig. 1.4 - Rede neuronal do tipo perceptron de multicamadas

O elemento essencial neste método reside no facto de existirem camadas intermédias de neurónios. Assim, no método de Retropropagação existe uma camada de neurónios de entrada, uma ou mais camadas intermédias e uma camada de saída. Neste método cada camada tem as suas funções bem definidas. Após a descoberta do algoritmo de Retropropagação, estudaram-se métodos de acelerar a convergência do algoritmo. Alguns cientistas desenvolveram estudos utilizando um método de actualização dos pesos baseado no cálculo de funções de densidade de probabilidade.

### 1.3 APLICAÇÕES DAS REDES NEURONAIS ARTIFICIAIS

As RNA são ferramentas de análise interdisciplinar podendo ser utilizadas em diversas áreas bem distintas, tais como, Medicina, Controlo, Processamento de Sinal e de Imagem, Robótica e Transportes.

#### 1.3.1 Medicina

Um dos exemplos de aplicação das RNA desenvolvidos na área da Medicina consiste no treino de uma rede neuronal com memória associativa, de forma a que ela guarde registos de pacientes, diagnósticos, registo de medicamentos, registos de tratamentos utilizados, entre outros. O objectivo desta rede consiste em armazenar padrões de comportamento de forma a encontrar o melhor tratamento a aplicar a um determinado paciente [4]. A detecção de picos num electroencefalograma é outra das possíveis aplicações das RNA na Medicina.

### 1.3.2 Controlo

As aplicações das RNA na área do controlo incluem funções de controlo de processos industriais. Estão a ser desenvolvidos algoritmos de RNA, que, uma vez instalados nos automóveis, vão fazer com que os mesmos realizem percursos mais suaves, através da instalação de um controlador neuronal que funciona como um sensor e detecta vibrações nas rodas dos carros, nos seus eixos e na suspensão. As aplicações das RNA estendem-se também à aterragem e descolagem de aviões. Na área da aviação, grandes inovações foram conseguidas com a criação dos autopilotos neuronais. O sistema de simulação de voo da NASA utiliza uma versão modificada do algoritmo de Retropropagação, que gera sinais correspondentes às entradas de um sensor que mede cerca de 120 variáveis cinemáticas e dinâmicas, tais como, velocidades e acelerações lineares e angulares. Estes autopilotos neuronais conseguiram reduzir os erros de altitude obtidos por autopilotos normais, de sensivelmente 9 para 1 metro [1].

### 1.3.3 Processamento do Sinal

Uma das aplicações das RNA, na área do Processamento de Sinal, consiste na eliminação do ruído das linhas telefónicas, utilizando para tal a rede Adaline. Outra aplicação, realizada nesta área, consiste na classificação de sinais provenientes de um radar (separando o sinal do ruído).

As redes neuronais têm sido aplicadas no reconhecimento da fala. O reconhecimento da fala envolve três fases: a fase em que se procede a uma amostragem do discurso e se extrai as palavras; a fase em que o processador encontra a probabilidade das palavras ditas no discurso coincidirem com as que existem no vocabulário; por fim, a fase em que o processador procura o sentido da frase para as diversas palavras existentes no vocabulário. De uma forma geral, o tempo é um factor crucial no reconhecimento de voz. Mas, os sistemas com extensos vocabulários requerem elevadas capacidades de memória. De forma a resolver este problema de falta de memória no acesso ao tempo, foram criadas redes neuronais que utilizam métodos de reconhecimento de fonemas e sílabas. O problema do som de uma pessoa a falar se confundir com outros possíveis sons veio dificultar o reconhecimento da voz por parte da rede neuronal. Para ultrapassar este problema, foram criadas redes neuronais que têm como função separar o ruído da fala, através da visualização dos movimentos faciais [8].

### 1.3.4 Processamento de imagens

Na área de reconhecimento de padrões, foram desenvolvidos algoritmos que permitem o reconhecimento de caracteres (estas RNA são muito utilizadas na Ásia, onde as linguagens contêm inúmeros caracteres), a detecção de caracteres errados no sítio errado, a detecção de padrões ou texturas numa linha de montagem não detectáveis ao olho humano, o reconhecimento de voz, detecção de sismos, análise de ecos provenientes de um sonar ou de um radar, entre outros [1]. O reconhecimento visual de imagens é particularmente utilizado na manipulação automática de objectos e correspondência de imagens processadas – imagens ampliadas, comprimidas, transformadas. O reconhecimento visual é utilizado em áreas tão distintas como na identificação e segurança, no reconhecimento de mãos e impressões digitais [8].

### 1.3.5 Robótica

A robótica é outras das áreas onde as redes neuronais são utilizadas. Num simples robot estacionário composto por um braço e por sensores capazes de detectar objectos estão implícitas aplicações de redes neuronais, na realização de tarefas e na manipulação e reconhecimento de objectos. Os robots mais sofisticados são não estacionários e executam múltiplas tarefas, tais como, controlo de movimentos, orientação e equilíbrio, tarefas que envolvem o evitar colisões, identificação de objectos, reconhecimento e geração de voz, comunicações, entre outras [8].

### 1.3.6. Transportes

A década de 90 assistiu a um renovar do interesse, por parte da comunidade científica, na aplicação das redes neuronais a estudos de transportes. Alguns desses estudos utilizam como ferramentas de apoio os SIG's, o GPS, a Electrónica, o Processamento de Imagem, etc. Na área dos transportes, a maior parte dos estudos realizados compreendem o comportamento dos condutores ao volante de um automóvel, no entanto, outras áreas são também objecto de análise, tais como:

1. A estimação de parâmetros – são utilizadas as quantidades de fluxos de tráfego para estimar parâmetros complicados de medir, tais como, matriz origem-destino e sua inversa, tempo de percurso, entre outros.

2. Manutenção de pavimentos – as redes neuronais têm sido aplicadas em duas vertentes: na detecção de pavimento degradado, usando processamento de imagem, e no prognóstico de intervenção nas vias. Este último exige não só uma previsão da degradação do pavimento face ao fluxo de tráfego e às condições atmosféricas da região, como também o conhecimento do investimento que se deseja efectuar.
3. Classificação/ detecção de veículos – este tipo de estudos é muito aplicado nas auto-estradas, onde são colocadas câmaras de vídeo e sensores sendo, posteriormente, realizado o processamento dos respectivos dados com recurso às redes neuronais.
4. Transporte de mercadorias – um dos problemas mais complexos na área dos transportes diz respeito à optimização das rotas de veículos, principalmente no que concerne à sua parametrização. Em problemas deste género são utilizadas as redes neuronais do tipo 'máquina de *Boltzmann*'.
5. Previsão de tráfego – a previsão de fluxos de tráfego pode ser realizada de duas formas: previsão para fluxos para meses ou anos e previsão para alguns minutos. Neste último tipo de previsão o objectivo consiste em utilizar os resultados para alimentar um sistema controlador de tráfego.
6. Controlo de tráfego, transportes aéreos, marítimos, metro – é outra das áreas onde as redes neuronais têm sido aplicadas, mas não com tanta insistência como as anteriores.

O algoritmo de Retropropagação tem vindo a ser muito utilizado na área de transportes devido ao seu baixo custo em termos de *software* de programação e à programação em si, quando comparado com outro tipo de redes neuronais [2].

## 2. PERCEPTRÃO

## 2.1 INTRODUÇÃO

As RNA são consideradas por alguns autores como sistemas de processamento caracterizadas pelas suas capacidades de aprendizagem e de generalização a partir de padrões ou outro tipo de informação. São, também, caracterizadas pela sua semelhança com o mundo biológico e foram sendo desenvolvidas dando origem a diferentes modelos. Assim, neste capítulo, e sem se pretender ser exaustivo, será feita a caracterização das redes neuronais de acordo com a sua arquitectura, função de activação, técnicas de aprendizagem e estilos de treino. Seguidamente, será estudado o percepção, encarado como uma generalização do neurónio de 'McCulloch-Pitts' e o capítulo termina com os métodos de optimização da função de custo introduzida no percepção.

## 2.2 CARACTERIZAÇÃO DAS REDES

### 2.2.1 Arquitectura

De forma a atingir determinados objectivos um neurónio pode não ser o suficiente, sendo necessário acrescentar (em paralelo) mais neurónios, dando origem a uma camada. A forma como os neurónios estão organizados nas camadas, as suas ligações e os pesos associados, denomina-se arquitectura da rede. As redes neuronais podem ser classificadas como redes com uma camada ou com múltiplas camadas. Nesta denominação, os vectores de entrada da rede não são considerados como uma camada, uma vez que os mesmos não entram nos cálculos da rede [4]. Geralmente, os neurónios numa mesma camada têm um comportamento semelhante, sendo ele determinado pela sua função de activação e pelas suas ligações e respectivos pesos em cada camada. As redes neuronais com uma camada são utilizadas em problemas de classificação e associação de padrões. Numa rede neuronal composta por várias camadas, a função de transferência em cada camada pode ser diferente, mas, de uma maneira geral, dentro de cada camada os neurónios tem a mesma função de transferência [4]. As RNA com múltiplas camadas são mais poderosas que as redes com uma única camada, sendo as primeiras muito usadas na resolução de problemas mais complexos [5].

### 2.2.2 Funções de Activação

As funções são escolhidas de acordo com o objectivo a atingir com a rede e podem ser lineares ou não lineares. As operações mais elementares numa rede neuronal são as que envolvem a soma do produto dos pesos pelos valores de entrada, aplicando de seguida a respectiva função de activação (para os vectores de entrada a função de activação é a identidade).

A função identidade também poderá ser utilizada nas camadas intermédias, mas, de uma maneira geral, utilizam-se funções não lineares, uma vez que, em RNA com múltiplas camadas usando uma função de activação linear, o resultado na saída é idêntico ao resultado usando uma rede neuronal com uma camada. A função degrau (*Threshold function* ou *Heaviside function*) é muito usada na classificação de elementos por categorias [4, 6]. A função rampa comporta-se como a função linear entre os valores  $-0.5$  e  $0.5$ ; para valores superiores a  $0.5$  ela produz um  $1$  na saída e para valores inferiores a  $0.5$  produz um  $0$  na saída.

$$f(v) = \begin{cases} 1 & \text{se } v \geq 0 \\ 0 & \text{se } v < 0 \end{cases} \quad [2.1]$$

$$f(v) = \begin{cases} 1 & \text{se } v \geq 0.5 \\ v & \text{se } -0.5 < v < 0.5 \\ 0 & \text{se } v \leq -0.5 \end{cases} \quad [2.2]$$

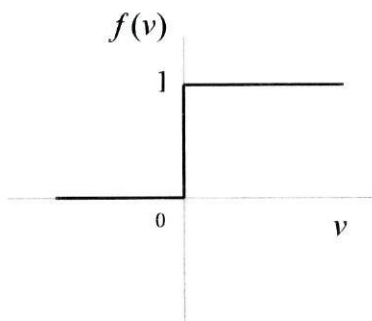


Fig. 2.1 - Função degrau

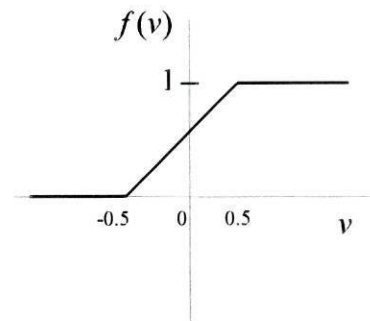


Fig. 2.2 - Função rampa

As funções do tipo sigmóide são muito utilizadas em aplicações de RNA, especialmente no algoritmo de Retropropagação, em parte, devido ao facto de serem diferenciáveis. Estas funções tomam o aspecto de um S. Estas funções podem ser definidas como funções crescentes cujo comportamento oscila entre uma função linear e uma função não linear. Uma das funções sigmóide mais utilizadas é a Função Logística que é definida pela seguinte expressão:

$$f(v) = \frac{1}{1 + e^{-av}} \quad [2.3]$$

sendo  $a$  o parâmetro de declive da curva. A derivada da função é dada por:

$$f'(v) = af(v)(1 - f(v)) \quad [2.4]$$

Esta função toma apenas valores positivos na saída, ou seja, transforma as entradas (que tomam valores de menos infinito a mais infinito) em saídas com valores entre zero e um.

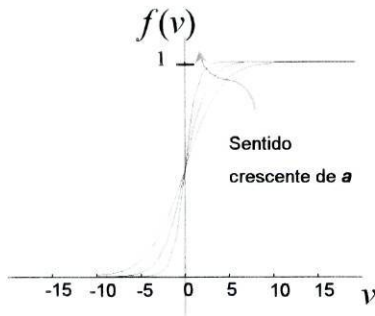


Fig. 2.3 - Funções sigmóide

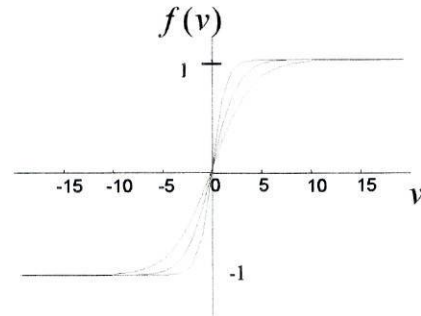


Fig. 2.4 - Funções sigmóide anti-simétricas

No limite, à medida que  $a$  tende para infinito, a função tem um comportamento idêntico ao da função degrau, isto é, enquanto que a função degrau assume valores 0 ou 1, a função sigmóide assume valores contínuos entre 0 e 1.

Por vezes, torna-se necessário, em diversos tipos de aplicação, ter uma função de activação cujos valores estejam compreendidos entre  $-1$  e  $1$ . Nestes casos, a função sigmóide toma uma forma anti-simétrica com respeito à origem. A função de activação é dada por:

$$f(v) = \frac{2}{1 + e^{-av}} - 1 \quad [2.5]$$

Atendendo à forma da função sigmóide, por vezes utiliza-se a função tangente hiperbólica [4, 6, 8].

### 2.2.3 Aprendizagem

A capacidade de aprendizagem das RNA é uma das suas características mais importantes, à qual os investigadores têm dedicado grande parte do seu tempo, ao longo destes anos de investigação. Sendo a aprendizagem uma capacidade que o Ser Humano desenvolve e coloca em prática diariamente, à medida que os mais variados desafios surgem, não é de estranhar que tais capacidades sejam desejadas nas RNA. Nas RNA o conceito de aprendizagem é baseado na teoria

comportamentalista, ou seja, o processo de aprendizagem obedece a um estímulo produzindo uma resposta [8].

### **2.2.3.1 Aprendizagem Supervisionada**

Na aprendizagem supervisionada considera-se que a resposta correcta na saída é conhecida, para cada valor de entrada.

Na técnica à qual se atribui a designação Aprendizagem Supervisionada, a rede neuronal é treinada através da inserção de um estímulo que produz uma determinada resposta. A resposta obtida é comparada com a saída desejada. Se a resposta obtida for diferente da resposta esperada o respectivo erro é determinado e usado para o cálculo dos ajustamentos a fazer à rede, de forma a minimizar os erros, de acordo com um critério previamente estabelecido [8]. As RNA mais simples utilizam este tipo de aprendizagem que é também muito usado na classificação de padrões.

### **Aprendizagem Reforçada**

Este tipo de aprendizagem utilizada na classificação de padrões, difere da Aprendizagem Supervisionada pelo facto da rede não calcular se a saída está ou não próxima do modelo desejado. Enquanto que, como já foi referido, na Aprendizagem Supervisionada os valores de saída são comparados com os alvos, gerando-se um erro que é utilizado no ajuste da rede, neste processo há, apenas, uma solução que, estando errada, leva a rede a reiniciar todo o processo até acertar na classificação. Quando esta técnica é utilizada no treino de uma rede neuronal devem ser definidos critérios de paragem de forma a que a rede termine o processamento ao atingir determinados valores; caso contrário, a rede continuará a treinar indefinidamente [8].

### **Aprendizagem Competitiva**

Tal como a Aprendizagem Reforçada, a Aprendizagem Competitiva é uma forma de aprendizagem supervisionada, distinguindo-se da anterior pelas suas características e arquitectura. A filosofia deste tipo de aprendizagem consiste, tal como o nome o indica, na competição entre os neurónios de saída de forma a produzirem um valor o mais próximo possível do valor do modelo desejado. O neurónio que produzir o valor mais próximo do modelo (para um determinado valor de entrada)

torna-se no neurónio dominante e os restantes neurónios deixam de produzir valores na saída. Cada neurónio é treinado de forma a responder a diferentes vectores de entrada, ou seja, esta técnica pode ser entendida como um processo de especialização por parte dos neurónios.

As variantes mais utilizadas deste tipo de técnica de aprendizagem são: a Regra Delta (ou dos Mínimos Quadrados), a Regra de Aprendizagem de *Hebb* e a Regra do Gradiente Decrescente [8].

### **2.2.3.2 Aprendizagem Não Supervisionada**

Neste tipo de aprendizagem não existe um alvo definido, isto é, os vectores de entrada não são comparados com um modelo treinado e definido previamente. A rede tem que aprender por si só os padrões, as irregularidades e as correlações entre os diversos elementos, de forma a poder classificá-los. Durante a fase de treino da rede, esta recebe vários vectores de entrada com características diferentes e vai organizando-os em classes. A entrada de um novo vector produz uma determinada saída, que é analisada e incluída numa determinada classe. Se essa classe não existir nas que já estão definidas, a rede neuronal cria uma nova classe. Enquanto a rede vai descobrindo as características dos vectores de entrada, ela vai modificando os seus parâmetros, ou seja, ela auto-organiza-se. As RNA com aprendizagem não supervisionada requerem uma maior orientação na definição das classes. A orientação por classes pode ser baseada em características do tipo: cor, forma, material, peso, entre outras. Se a rede neuronal não for bem orientada corre-se o risco de ela não conseguir atingir os objectivos pré-definidos [8].

### **2.2.4 Estilos de treino**

Uma rede neuronal pode ser classificada como recorrente ou não recorrente. As RNA recorrentes são caracterizadas pela existência de malhas de realimentação, ou seja, as camadas de saída estão interligadas com as camadas de entrada de dados. O algoritmo de Retropropagação é um exemplo da aplicação de redes neuronais do tipo recorrente com aprendizagem supervisionada.

Nas redes não recorrentes esta realimentação não existe. As interligações são realizadas num único sentido, da camada de entrada para a camada de saída. Este tipo de rede é muito utilizada nos perceptrões de multicamada [1]. Um modelo não recorrente é caracterizado por uma estrutura fixa de vectores de entrada, ou seja, a rede tem um comportamento estático. Por outro lado, as redes

com um modelo recorrente são caracterizadas por uma dinâmica própria durante a fase de processamento, uma vez que os vectores de saída entram de novo na rede até a mesma atingir um determinado erro. Devido a esta propriedade cíclica, a rede continua a processar os elementos até atingir a estabilidade [1].

## 2.3 O PERCEPTRÃO

O modelo de *McCulloch-Pitts* gerou interesse por parte de muitos investigadores, que, utilizando os seus princípios básicos, criaram um novo modelo onde funções de aprendizagem e de adaptação foram acrescentadas dando origem aos perceptrões. O perceptrão é construído em torno de um neurónio não linear (modelo de *McCulloch-Pitts*) e utiliza a Regra Delta para a actualização dos pesos da rede [6]. Os perceptrões são modelos que exigem supervisionamento na fase de aprendizagem, possuindo capacidade de realizar generalizações limitadas e conseguindo classificar padrões na presença de ruído nas entradas. A aprendizagem dos perceptrões depende do erro entre a saída calculada e a saída desejada para o modelo. Na fase anterior ao treino dos perceptrões é necessário ter em conta alguns factores, tais como, os vectores de entrada, os alvos a atingir, a taxa de aprendizagem  $\eta$  (que deve ser pequena e positiva) e o erro de paragem do processo de actualização dos pesos [8].

### 2.3.1 Teorema de convergência do perceptrão

Suponhamos que se pretende criar uma rede neuronal cujo objectivo seja classificar um conjunto de padrões em 2 classes. Considere-se o vector  $\mathbf{x}(n) = [x_0, x_1(n), x_2(n), \dots, x_m(n)]^T$  que representa o vector de entrada na iteração de ordem  $n$ ,  $x_0 = 1$  que representa a entrada do desvio sistemático cujo peso é dado por  $w_0(n) = b(n)$  e  $\mathbf{W}(n) = [w_0(n), w_1(n), w_2(n), \dots, w_m(n)]^T$  que representa a matriz coluna dos pesos. A combinação linear que corresponde à saída da rede é dada por:

$$\begin{aligned} v(n) &= \sum_{i=0}^m w_i(n)x_i(n) \\ &= \mathbf{W}^T(n)\mathbf{x}(n) \end{aligned} \quad [2.6]$$

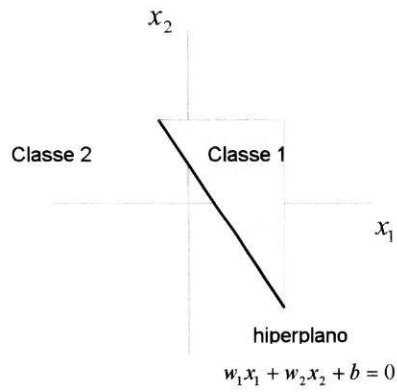


Fig. 2.5 - Hiperplano que separa duas classes linearmente separáveis

Para um determinado valor de  $n$ , a equação  $\mathbf{W}^T \mathbf{x} = 0$  representada num espaço de dimensão  $m$ , com coordenadas  $x_1, x_2, \dots, x_m$ , define um hiperplano que divide as duas classes, classe 1 e classe 2. Estas duas classes devem ser linearmente separáveis, ou seja, os dois padrões que vão ser classificados devem ser suficientemente separáveis um do outro, de forma a garantir que a superfície de decisão esteja contida no hiperplano (Fig. 2.5). (Consultar o apêndice para uma explicação mais detalhada sobre o algoritmo de classificação).

Na figura seguinte encontra-se representado um exemplo de duas classes que são linearmente separáveis e outro em que as classes não o são [6].

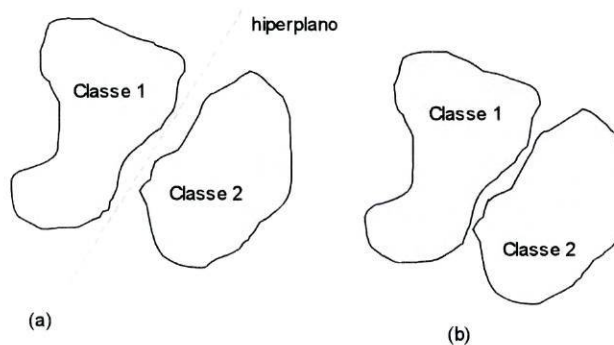


Fig. 2.6 - a) Duas classes linearmente separáveis; b) duas classes não linearmente separáveis

## Função de Custo

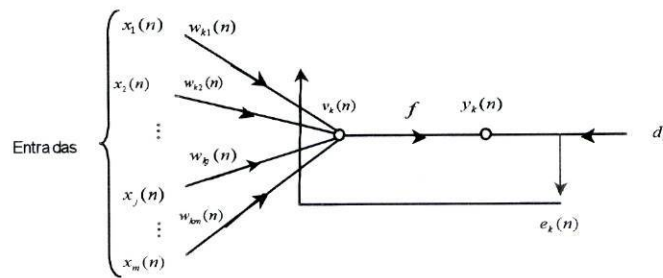


Fig. 2.7 - Erro da rede neuronal

Considere a Fig. 2.7. O sinal produzido na saída da rede pelo  $k$ -ésimo neurónio na  $n$ -ésima iteração,  $y_k(n)$ , é comparado com o valor alvo  $d_k$ , dando origem ao erro do sinal:  $e_k(n) = d_k - y_k(n)$ .

A função de custo,  $E(n)$ , é expressa em termos do erro do sinal  $e_k(n)$ :  $E(n) = \frac{1}{2} e_k^2(n)$  [6].

### 2.3.2 Perceptrões de multicamadas

Os perceptrões de multicamada representam outro modelo de perceptrões desenvolvidos numa estrutura hierárquica com topologia recorrente, com uma ou mais camadas entre a camada de entrada e a camada de saída. O número de camadas intermédias e o número de neurónios por camada são determinados mediante várias experiências, uma vez que dependem do tipo de aplicações a que a rede se destina [8]. As redes neuronais do tipo perceptrão de multicamadas serão abordadas mais em detalhe no capítulo seguinte através do algoritmo de Retropropagação.

## 2.4 MÉTODOS DE OPTIMIZAÇÃO

Algumas redes neuronais aplicam a técnica do gradiente decrescente no espaço dos pesos. A técnica tem como base o cálculo das derivadas parciais de uma função  $F(\mathbf{W}, \mathbf{x})$ , sendo  $\mathbf{W}$  a matriz linha dos pesos referentes a um neurónio e  $\mathbf{x}$  o vector de entrada.

Considere-se a função de custo  $E(\mathbf{W})$ , referente a um neurónio, contínua e diferenciável. Pretende-se encontrar a solução otimizada,  $\mathbf{W}^*$  que satisfaça a condição  $E(\mathbf{W}^*) \leq E(\mathbf{W})$ , ou seja, minimizar a função de custo  $E(\mathbf{W})$  com respeito aos pesos  $\mathbf{W}$ . Para tal, é necessário

encontrar  $\nabla E(\mathbf{W}^*) = 0$ , onde  $\nabla E(\mathbf{W})$  representa o gradiente da função de custo que é dado por:

$$\nabla E(\mathbf{W}) = \left[ \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \dots, \frac{\partial E}{\partial w_m} \right]^T \quad [2.7]$$

Com base neste conceito, foram desenvolvidas técnicas de aceleração da convergência do algoritmo, sendo de destacar o método do gradiente decrescente, o método de *Newton* e o método de *Gauss-Newton* [6].

#### 2.4.1 Método do gradiente decrescente

No método do gradiente decrescente os ajustes dos pesos são realizados no sentido decrescente do gradiente da função de custo, ou seja, na direcção oposta do vector gradiente.

Sendo  $\mathbf{g} = \nabla E(\mathbf{W})$  o vector gradiente e  $\eta$  a taxa de aprendizagem, os pesos são actualizados segundo a seguinte expressão:

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \eta \mathbf{g}(n) \quad [2.8]$$

A correcção a aplicar ao algoritmo na iteração seguinte é dada por:

$$\begin{aligned} \Delta \mathbf{W}(n) &= \mathbf{W}(n+1) - \mathbf{W}(n) \\ &= -\eta \mathbf{g}(n) \end{aligned} \quad [2.9]$$

O método do gradiente decrescente satisfaz a condição definida pela seguinte expressão:

$$E(\mathbf{W}(n+1)) < E(\mathbf{W}(n)) \quad [2.10]$$

A condição [2.10] pode ser reescrita sob a forma de desenvolvimento em série de Taylor de primeira ordem em torno de  $\mathbf{W}(n)$  de forma a aproximar  $E(\mathbf{W}(n+1))$ ,

$$E(\mathbf{W}(n+1)) \cong E(\mathbf{W}(n)) + \mathbf{g}^T(n) \Delta \mathbf{W}(n) \quad [2.11]$$

Substituindo a equação [2.9] na equação [2.11] obtêm-se:

$$\begin{aligned} E(\mathbf{W}(n+1)) &\cong E(\mathbf{W}(n)) - \eta \mathbf{g}^T(n) \mathbf{g}(n) \\ &= E(\mathbf{W}(n)) - \eta \|\mathbf{g}(n)\|^2 \end{aligned} \quad [2.12]$$

Analisando a equação [2.12] verifica-se que o método do gradiente decrescente converge para a solução otimizada muito lentamente e a taxa de aprendizagem exerce uma influência profunda na

convergência do algoritmo: se  $\eta$  for muito pequena, a trajectória do gradiente decrescente é suave demais; se  $\eta$  for muito grande, a trajectória definida pelo algoritmo é muito irregular; se  $\eta$  exceder um certo valor crítico, o algoritmo torna-se instável [6].

Existem dois métodos possíveis para determinar a taxa de aprendizagem,  $\eta$ . Um dos métodos consiste em minimizar a função  $E(\mathbf{W})$  em relação a  $\eta_k$  em cada iteração, ou seja, minimizar ao longo de uma linha,  $\mathbf{W}(n) - \eta \mathbf{g}(n)$ . O outro método possível para determinar  $\eta_k$  consiste em seleccionar valores fixos de  $\eta_k$  ou seleccionar através da utilização de uma variável (por exemplo,  $\eta_k = \frac{1}{k}$ ) [5]. (Consultar o apêndice para uma explicação mais detalhada sobre o método do gradiente decrescente).

### 2.7.2 Método de Newton

O método de *Newton* consiste em minimizar a aproximação quadrática da função de custo em torno de um determinado ponto  $\mathbf{W}(n)$ , sendo esta minimização efectuada a cada iteração. A expansão da série de *Taylor* de segunda ordem da função de custo em torno do ponto  $\mathbf{W}(n)$  é dada por:

$$\begin{aligned} \Delta E(\mathbf{W}(n)) &= E(\mathbf{W}(n+1)) - E(\mathbf{W}(n)) \\ &\cong \mathbf{g}^T(n) \Delta \mathbf{W}(n) + \frac{1}{2} \Delta \mathbf{W}^T(n) \mathbf{H}(n) \Delta \mathbf{W}(n) \end{aligned} \quad [2.13]$$

onde  $\mathbf{H}(n)$  representa a matriz *Hessiana* de  $E(\mathbf{W})$ ,

$$\begin{aligned} \mathbf{H} &= \nabla^2 E(\mathbf{W}) \\ &= \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 w_2} & \cdots & \frac{\partial^2 E}{\partial w_1 w_m} \\ \frac{\partial^2 E}{\partial w_2 w_1} & \frac{\partial^2 E}{\partial w_2^2} & \cdots & \frac{\partial^2 E}{\partial w_2 w_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial w_m w_1} & \frac{\partial^2 E}{\partial w_m w_2} & \cdots & \frac{\partial^2 E}{\partial w_m^2} \end{bmatrix} \end{aligned} \quad [2.14]$$

A equação [2.13] é minimizada quando  $\mathbf{g}(n) + \mathbf{H}(n) \Delta \mathbf{W}(n) = 0$ , o que equivale a:

$$\Delta \mathbf{W}(n) = -\mathbf{H}^{-1}(n) \mathbf{g}(n) \quad [2.15]$$

Os pesos são actualizados segundo a expressão seguinte:

$$\begin{aligned}\mathbf{W}(n+1) &= \mathbf{W}(n) + \Delta\mathbf{W}(n) \\ &= \mathbf{W}(n) - \mathbf{H}^{-1}(n)\mathbf{g}(n)\end{aligned}\quad [2.16]$$

O método de *Newton* converge assintoticamente e não tem um comportamento de ziguezague, como no caso do método do gradiente decrescente. No entanto, a matriz *Hessiana* deve ser positiva para todos os seus elementos, de forma a garantir que o método funcione. Mas tal garantia não é possível de dar, para cada iteração do algoritmo, portanto torna-se necessário introduzir algumas modificações [6].

A matriz *Hessiana* desempenha um papel muito importante no estudo e aplicação de redes neuronais a problemas do mundo real, uma vez que os vectores próprios da matriz têm uma influência significativa na dinâmica de aprendizagem do algoritmo. A inversa da matriz *Hessiana* permite distinguir os parâmetros que são verdadeiramente relevantes para rede e permite obter uma base para a optimização do algoritmo usando outras variantes do algoritmo de Retropropagação [6].

### 2.7.3 Método de *Gauss-Newton*

Este método é aplicado a uma função de custo que representa a soma dos quadrados dos erros entre a saída desejada e a saída alvo de uma rede neuronal.

Considerando  $E(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^n e^2(i)$  uma função de custo, onde o factor  $\frac{1}{2}$  é utilizado apenas para simplificar as equações e  $e(i)$  representa o erro na iteração de ordem  $i$ . A linearização de  $e(i)$  relativamente à matriz  $\mathbf{W}$  procurada é dada por:

$$e'(i, \mathbf{W}) = e(i) + \left[ \frac{\partial e(i)}{\partial \mathbf{W}} \right]_{\mathbf{W}=\mathbf{W}(n)} (\mathbf{W} - \mathbf{W}(n)), \quad i = 1, 2, \dots, n \quad [2.17]$$

Sendo  $\mathbf{e}(n) = [e(1), e(2), \dots, e(n)]$  o vector de dimensão  $n$  cujas componentes são os erros desde a primeira até à iteração de ordem  $n$  e, sendo  $\mathbf{J}(n)$  a matriz Jacobiana do vector  $\mathbf{e}(n)$ , dada por:

$$\mathbf{J}(n) = \begin{bmatrix} \frac{\partial e(1)}{\partial w_1} & \frac{\partial e(1)}{\partial w_2} & \cdots & \frac{\partial e(1)}{\partial w_m} \\ \frac{\partial e(2)}{\partial w_1} & \frac{\partial e(2)}{\partial w_2} & \cdots & \frac{\partial e(2)}{\partial w_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial e(n)}{\partial w_1} & \frac{\partial e(n)}{\partial w_2} & \cdots & \frac{\partial e(n)}{\partial w_m} \end{bmatrix}_{\mathbf{W}=\mathbf{W}(n)} \quad [2.18]$$

A equação [2.17] pode ser reescrita da seguinte forma:

$$\mathbf{e}'(n, \mathbf{W}) = \mathbf{e}(n) + \mathbf{J}(n)(\mathbf{W} - \mathbf{W}(n)) \quad [2.19]$$

Os pesos são actualizados segundo a expressão  $\mathbf{W}(n+1) = \arg \min_{\mathbf{w}} \left\{ \frac{1}{2} \|\mathbf{e}'(n, \mathbf{W})\|^2 \right\}$ , sendo a norma euclidiana dada por:

$$\frac{1}{2} \|\mathbf{e}'(n, \mathbf{W})\|^2 = \frac{1}{2} \|\mathbf{e}(n)\|^2 + \mathbf{e}^T(n) \mathbf{J}(n) (\mathbf{W} - \mathbf{W}(n)) + \frac{1}{2} (\mathbf{W} - \mathbf{W}(n))^T \mathbf{J}^T(n) \mathbf{J}(n) (\mathbf{W} - \mathbf{W}(n)) \quad [2.20]$$

Diferenciando a equação [2.20] e igualando a zero, obtem-se a expressão que permite calcular a actualização dos pesos:

$$\mathbf{J}^T(n) \mathbf{e}(n) + \mathbf{J}^T(n) \mathbf{J}(n) (\mathbf{W} - \mathbf{W}(n)) = 0 \quad [2.21]$$

A expressão que representa o método de *Gauss-Newton* é pela equação [2.22]:

$$\mathbf{W}(n+1) = \mathbf{W}(n) - (\mathbf{J}^T(n) \mathbf{J}(n))^{-1} \mathbf{J}^T(n) \mathbf{e}(n) \quad [2.22]$$

Enquanto que no método de *Newton* é necessário o cálculo da matriz *Hessiana*, no método de *Gauss-Newton* é requerido o cálculo da matriz *Jacobiana* do erro. Para o cálculo da actualização dos pesos é necessário que a matriz  $\mathbf{J}^T(n) \mathbf{J}(n)$  seja não singular e positiva. Como não existem garantias que tais condições sejam possíveis de realizar, é prática usual acrescentar a matriz diagonal à matriz anterior, ou seja,  $\mathbf{J}^T(n) \mathbf{J}(n) + \delta \mathbf{I}$ , onde  $\delta$  representa uma constante de valor pequeno e positivo. O efeito que esta modificação produz na prática consiste numa redução substantiva no número de iterações. Assim, a equação que define a actualização dos pesos, com este método, é dada pela equação [2.23]:

$$\mathbf{W}(n+1) = \mathbf{W}(n) - (\mathbf{J}^T(n) \mathbf{J}(n) + \delta \mathbf{I})^{-1} \mathbf{J}^T(n) \mathbf{e}(n) \quad [2.23]$$

### 3. RETROPROPAGAÇÃO

### 3.1 O ALGORITMO DE RETROPROPAGAÇÃO

O Algoritmo de Retropropagação foi desenvolvido por *Paul Webos* em 1974 e re-descoberto, independentemente, por *Rumelhart* e *Parker* na década de 80. Este algoritmo tem sido largamente usado como algoritmo de aprendizagem em redes neuronais do tipo multicamadas com topologia recorrente, compostas por uma ou mais camadas intermédias.

A grande diferença entre este algoritmo e os outros modelos considerados reside no processo de cálculo e actualização dos pesos durante a fase de treino. A grande dificuldade dos perceptrões de multicamada consiste no cálculo dos pesos nas camadas intermédias numa forma eficiente e que minimize o erro na saída. Quantas mais camadas intermédias tiver, mais difícil se torna o cálculo dos erros. O valor do erro na saída é fácil de calcular, pois é a diferença entre a saída obtida e a saída desejada, mas nas camadas intermédias a dificuldade é acrescida, pois não existe uma observação directa do erro entre as camadas. O algoritmo de Retropropagação veio preencher esta lacuna. Trata-se de um algoritmo cujas expressões matemáticas são baseadas num processo iterativo, sendo portanto acessível ao cálculo em computador [6].

Considere-se a Fig. 3.1. Seja  $j$  um neurónio na iteração  $n$ .

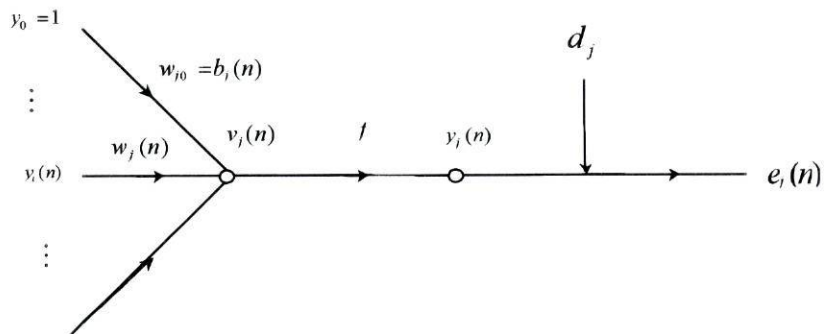


Fig. 3.1 - Neurónio  $j$  de uma rede neuronal com uma camada intermédia

O erro do sinal na saída do neurónio  $j$  na iteração  $n$  é definido por:

$$e_j(n) = d_j - y_j(n) \quad [3.1]$$

Considerando o erro causado pela função de custo para o neurónio  $j$  dado por  $\frac{1}{2} e_j^2(n)$  e o erro total da função dado pela soma de todos os erros,

$$E(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad [3.2]$$

onde  $C$  representa o conjunto de todos os neurónios na camada de saída da rede, o erro da função de custo para um conjunto de  $N$  padrões é dado pela média dos erros,

$$E_m = \frac{1}{N} \sum_{n=1}^N E(n) \quad [3.3]$$

O erro da função de custo em cada iteração,  $E(n)$  e a média dos erros,  $E_m(n)$  são funções dos parâmetros da rede (pesos e peso do desvio sistemático). A função  $E_m(n)$  representa a função de custo como uma medida do processo de aprendizagem. O objectivo da rede consiste em ajustar os parâmetros da rede de forma a minimizar  $E_m(n)$ . A função  $E_m(n)$  é também denominada de risco empírico ou custo médio [10].

Na Fig. 3.1 o neurónio  $j$  encontra-se ligado a outros neurónios através de uma camada intermédia. O produto  $v_j(n)$ , encontrado na entrada da função de activação associada ao neurónio  $j$ , é dado por:

$$v_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n) \quad [3.4]$$

sendo  $m$  o número total de entradas incluindo o peso do desvio sistemático ( $w_{j0} = b_j$  corresponde à entrada  $x_0 = 1$ ) aplicado ao neurónio  $j$ .

A função  $f$  aplicada na saída do  $j$ -ésimo neurónio na  $n$ -ésima iteração é dada por:

$$y_j = f_j(v_j(n)) \quad [3.5]$$

Tal como o método dos mínimos quadrados, o algoritmo de Retropropagação aplica uma correcção  $\Delta w_{ij}(n)$  aos pesos  $w_{ji}(n)$ , correcção essa que é proporcional às derivadas parciais  $\frac{\partial E(n)}{\partial w_{ji}(n)}$ .

Assim, pode-se exprimir o gradiente em ordem a  $w_{ji}(n)$  como um produto de derivadas parciais:

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad [3.6]$$

As derivadas parciais  $\frac{\partial E(n)}{\partial w_{ji}(n)}$  representam um factor de sensibilidade que determina a direcção de busca dos pesos  $w_{ji}(n)$ . A equação [3.6] adquire uma nova forma substituindo as derivadas parciais pelos seus respectivos valores, ou seja, diferenciando (de ambos os lados) a equação [3.2] em ordem a  $e_j(n)$ , a equação [3.1] em ordem a  $y_j(n)$ , a equação [3.5] em ordem a  $v_j(n)$  e a equação [3.4] em ordem a  $w_{ji}(n)$ , obtendo-se, respectivamente,  $\frac{\partial E(n)}{\partial e_j(n)} = e_j(n)$ ,  $\frac{\partial E(n)}{\partial y_j(n)} = -1$ ,

$\frac{\partial y_j(n)}{\partial v_j(n)} = f'(v_j(n))$  e  $\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n)$ . Substituindo estas expressões na equação [3.6], vem:

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = -e_j(n)f'(v_j(n))y_i(n) \quad [3.7]$$

A correcção  $\Delta w_{ji}(n)$  a aplicar ao peso  $w_{ji}$  é definida pela Regra de Delta:

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)} \quad [3.8]$$

onde  $\eta$  representa a taxa de aprendizagem do algoritmo de Retropropagação. Substituindo a equação [3.7] na [3.8] obtém-se,

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad [3.9]$$

onde o gradiente local  $\delta_j(n)$  é definido por

$$\begin{aligned} \delta_j(n) &= -\frac{\partial E(n)}{\partial v_j(n)} \\ &= e_j(n)f'(v_j(n)) \end{aligned} \quad [3.10]$$

Da equação [3.10] conclui-se que o gradiente local  $\delta_j(n)$  pode ser expresso em função do produto entre o erro no sinal  $e_j(n)$  e a derivada da função de activação  $f'(v_j(n))$ .

Analisando as equações [3.9] e [3.10] conclui-se que o gradiente local e a actualização dos pesos numa rede neuronal estão directamente ligados ao erro do sinal na saída  $e_j(n)$ . Assim, podem analisar-se dois casos distintos, dependendo da localização do neurónio  $j$  na rede, ou seja, se o neurónio  $j$  é um neurónio da camada de saída ou se pertence a uma camada intermédia [6].

Se o neurónio  $j$  pertence à camada de saída, então, calcula-se o erro na saída da rede  $e_j(n)$  e de seguida o gradiente local  $\delta_j(n)$ , aplicando para isso as equações [3.1] e [3.10], respectivamente.

Se o neurónio  $j$  pertence à camada intermédia de neurónios, então, a atribuição de penalidades ou recompensas a esse neurónio torna-se difícil, uma vez que o acesso a ele não é directo. Este problema é ultrapassado retropropagando os erros através da rede [6].

Considere-se a Fig. 3.2, sendo  $j$  um neurónio da camada intermédia.

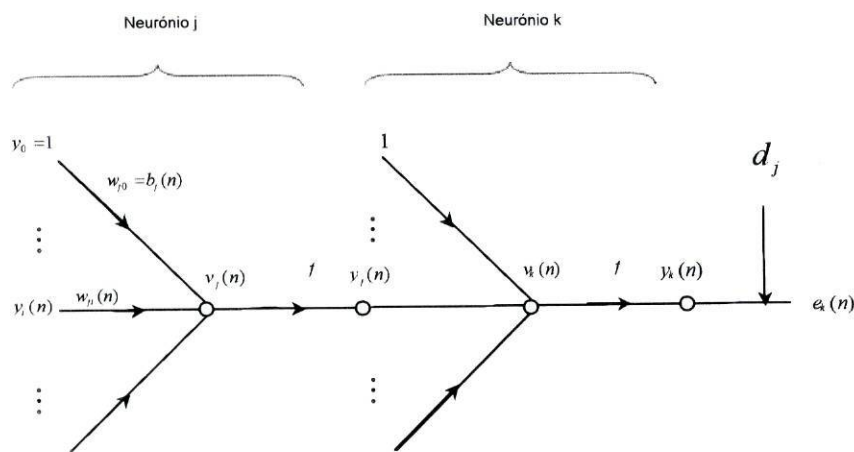


Fig. 3.2 - Neurónios  $j$  e  $k$  de uma rede neuronal com duas camadas intermédias

Da Fig. 3.2 obtém-se a função de custo,

$$E(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad [3.11]$$

onde  $k$  é o neurónio de saída e  $C$  representa o conjunto de todos os neurónios na camada de saída da rede, o gradiente local  $\delta_j(n)$  para um neurónio  $j$  numa camada intermédia é dado por:

$$\begin{aligned} \delta_j(n) &= - \frac{\partial E(n)}{\partial y_i(n)} \frac{\partial y_i(n)}{\partial v_j(n)} \\ &= - \frac{\partial E(n)}{\partial y_i(n)} f'(v_j(n)) \end{aligned} \quad [3.12]$$

derivando a equação [3.11] em ordem a  $y_j(n)$ , obtém-se

$$\begin{aligned}\frac{\partial E(n)}{\partial y_j(n)} &= \sum_k e_k(n) \frac{\partial e_k(n)}{\partial y_j(n)} \\ &= \sum_k e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)}\end{aligned}\quad [3.13]$$

da equação [3.1] tem-se

$$\begin{aligned}e_k(n) &= d_k - y_k(n) \\ &= d_k - f(v_k(n))\end{aligned}\quad [3.14]$$

e derivando a equação [3.14] em ordem a  $v_k(n)$  obtem-se,

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -f'(v_k(n))\quad [3.15]$$

onde  $v_k(n) = \sum_{j=0}^m w_{kj}(n)y_j(n)$ , sendo  $m$  o número de entradas aplicadas ao neurónio  $k$ . Assim,

derivando esta última equação em ordem a  $y_j(n)$  vem,

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n)\quad [3.16]$$

substituindo as equações [3.11] e [3.12] na equação [3.9] obtem-se a derivada parcial,

$$\begin{aligned}\frac{\partial E(n)}{\partial y_j(n)} &= -\sum_k e_k(n) f'(v_k(n)) w_{kj}(n) \\ &= -\sum_k \delta_k(n) w_{kj}(n)\end{aligned}\quad [3.17]$$

usando as equações [3.17] e [3.12] obtem-se a equação que representa a expressão matemática do algoritmo de Retropropagação :

$$\delta_j(n) = f'(v_j(n)) \sum_k \delta_k(n) w_{kj}(n)\quad [3.18]$$

onde  $j$  é um neurónio da camada intermédia e  $k$  um neurónio da camada de saída.

Analisando a equação [3.18] verifica-se que o primeiro factor do 2º membro,  $f'(v_j(n))$ , envolvido no conhecimento do gradiente local  $\delta_j(n)$  depende da função de activação associada ao neurónio  $j$ ; o segundo termo,  $\sum_k \delta_k(n) w_{kj}(n)$ , envolve o conhecimento do erro do sinal na saída  $e_k(n)$ , para todos os neurónios que estejam na camada imediatamente à direita do neurónio  $j$  e para os que se encontram directamente ligados ao neurónio  $j$ , bem como os respectivos pesos associados às

conexões dos neurónios envolvidos nas camadas [6]. A correcção  $\Delta w_{ji}(n)$  a aplicar à conexão do neurónio  $i$  para o  $j$  é definida pela Regra de Delta, que é dada pela seguinte equação:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \quad [3.19]$$

O algoritmo de Retropropagação pode ser resumido da seguinte forma:

- a) Antes de se iniciar o processo de treino da rede, deve-se seleccionar os padrões de entrada da rede, os alvos a atingir, a arquitectura da rede (número de neurónios, número de camadas intermédias, função de activação), o parâmetro de aprendizagem,  $\eta$ , e o critério de paragem.
- b) Iniciar os pesos da rede com valores aleatórios.
- c) Cálculo do produto dos pesos pelo vector de entrada da rede:

$$v_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n)$$

$$y_j = f_j(v_j(n)), \text{ saída calculada pela rede}$$

Se o neurónio  $j$  pertence à primeira camada intermédia, então,  $y_j(n) = x_j(n)$ , onde  $x_j(n)$  representa o elemento  $j$  do vector de entrada  $x(n)$ .

Se o neurónio  $j$  pertence à camada de saída, então,  $y_j(n) = o_j(n)$ .

O erro do sinal é dado por:  $e_j(n) = d_j - o_j(n)$ , sendo  $d_j$  o elemento  $j$  do vector alvo,  $d$ .

- d) Cálculo dos gradientes:

$$\delta_j(n) = \begin{cases} e_j(n) f'(v_j(n)), & \text{se } j \text{ for um neurónio na camada de saída} \\ f'(v_j(n)) \sum_k \delta_k(n) w_{kj}(n), & \text{se } j \text{ for um neurónio na camada intermédia} \end{cases}$$

- e) Ajuste dos pesos de acordo com a Regra de Delta:

$$w_{ji}(n+1) = w_{ji}(n) + \eta \delta_j(n) y_i(n) \quad [3.20]$$

### 3.2 O TERMO MOMENTO

A superfície de erro da função de custo depende de vários parâmetros a calcular pela rede neuronal, alguns deles muito complexos e com vários mínimos locais. De uma maneira geral, são necessários vários testes, com diferentes taxas de aprendizagem, para se conseguir otimizar a rede e encontrar o mínimo global. Para factores de aprendizagem muito grandes, podem-se obter comportamentos divergentes ou oscilações à volta do mínimo local [14]. Assim, quanto mais pequeno for o termo  $\eta$ , mais pequenas são as mudanças nos pesos de uma iteração para outra e mais suave será a trajectória no espaço dos pesos. Mas, por outro lado, mais pequena será a capacidade de aprendizagem da rede [6].

Se a taxa de aprendizagem  $\eta$  for muito alta, de forma a acelerar o processo de aprendizagem, então, serão realizadas grandes mudanças nos pesos de uma iteração para outra, tornando a rede instável. Um dos métodos utilizados para se evitar que a rede fique a oscilar num mínimo local, consiste em adicionar um novo termo à expressão que representa a actualização dos pesos da rede. O método consiste em aumentar progressivamente a taxa de aprendizagem (mas evitando que a rede fique instável) e modificar a Regra de Delta através da inserção do termo momento à equação [3.19]. Acrescentando o termo momento  $\alpha$ , à equação [3.19], obtém-se:

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j(n) y_i(n) \quad [3.21]$$

Do mesmo modo, a equação [3.20] pode ser re-escrita da seguinte forma:

$$w_{ji}(n+1) = w_{ji}(n) + \alpha [w_{ji}(n-1)] + \eta \delta_j(n) y_i(n) \quad [3.22]$$

A equação [3.22] denomina-se por Regra Delta Generalizada, uma vez que se trata da Regra Delta para o caso específico do termo momento ser nulo, ou seja,  $\alpha = 0$  [6].

Na Fig. 3.3 pode-se analisar o comportamento do algoritmo sem e com o termo momento, donde facilmente se percebe a razão pela qual o termo momento ajuda no processo de actualização dos pesos.

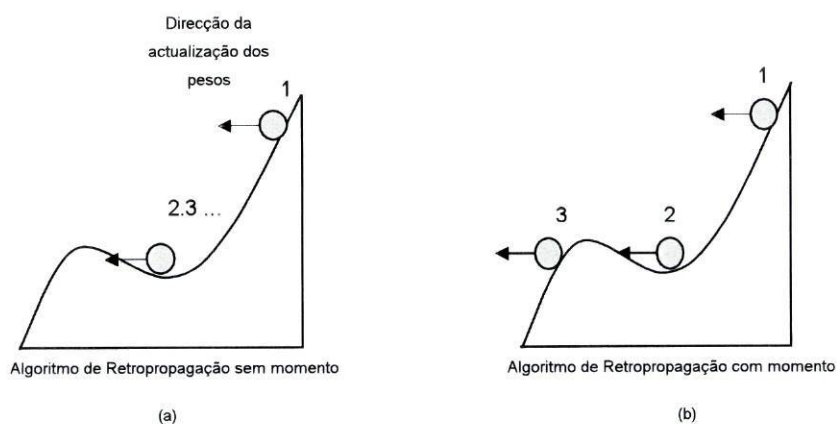


Fig. 3.3 - Comportamento do Algoritmo de Retropropagação com e sem o termo momento [14]

A inserção do termo momento no algoritmo de Retropropagação representa uma pequena modificação na actualização dos pesos. O termo momento tende a acelerar a convergência da rede evitando a oscilação da mesma e a sua 'queda' num mínimo local da superfície de erro. O algoritmo de Retropropagação assume que a taxa de aprendizagem é constante, mas, usando o termo momento, tal parâmetro é variável (depende da conexão da rede). Na Fig. 3.4 encontra-se representado o comportamento do ajuste dos pesos  $w_1$  e  $w_2$ , para o algoritmo de Retropropagação com e sem momento [13, 14].

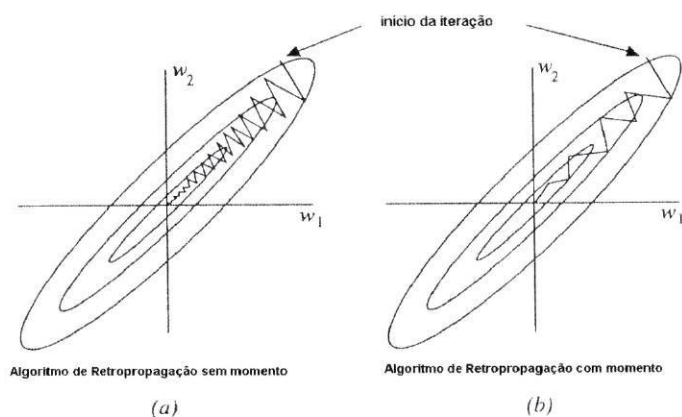


Fig. 3.4 - Aceleração do Algoritmo de Retropropagação usando o termo momento [13]

### 3.3 ASPECTOS PRÁTICOS

O algoritmo de Retropropagação contém duas características fundamentais que justificam por si só o facto dele ser largamente utilizado em vários tipos de aplicações: é fácil de programar no

computador e calcula o gradiente decrescente bem como actualiza os pesos da rede em modo sequencial, ou seja, actualiza os pesos padrão a padrão. Estas duas características são as responsáveis pelo facto deste algoritmo ser o mais conhecido e o mais utilizado dos algoritmos de perceptrões de múltiplas camadas com treino supervisionado [6].

### 3.3.1 Selecção da arquitectura da rede e dos vectores de entrada

Quando se pretende aplicar as redes neuronais a um determinado problema subsistem sempre dúvidas quanto à arquitectura que se deve utilizar para melhor resolver esse problema. De uma maneira geral, os programas desenvolvidos não indicam qual a melhor arquitectura para tratar um determinado problema, nem quais os padrões que estão a introduzir uma elevada quantidade de erro na rede. A busca da melhor solução, num determinado domínio, para a resolução de um determinado problema, tem sido um dos objectivos dos programadores. Esta característica foi implementada no programa *Statistica*. Assim, na maior parte dos programas, a arquitectura da rede que melhor se adapta a um determinado problema é encontrada após vários testes de possíveis variáveis [14].

### 3.3.2 Tipos de processamento

Denomina-se por época uma representação completa (iteração) do treino durante a fase de aprendizagem. O processo de aprendizagem mantém-se, época após época, até que os pesos estabilizem e o custo médio de todo o conjunto de treino convirja para um valor mínimo. O treino da rede neuronal com o algoritmo de Retropropagação pode ser executado através de um processamento em modo sequencial ou em *batch*.

O processamento em modo sequencial ou estocástico realiza-se quando os pesos da rede são actualizados à medida que um novo par {entrada, saída} é apresentado à rede. O treino sequencial é muito utilizado em aplicações em tempo real, devido ao facto de utilizar menos memória no seu processamento uma vez que os padrões são apresentados à rede par a par e os pesos são actualizados após o seu processamento. Este facto faz com que a rede tenha uma maior probabilidade de não cair num mínimo local, bem como, seja mais difícil estabelecer condições teóricas para a convergência do algoritmo. Uma das vantagens da utilização do método sequencial

consiste no facto de, ao trabalhar com dados extensos e redundantes, o algoritmo conseguir tirar partido, já que os dados são apresentados à rede par a par. Apesar das desvantagens do modo sequencial em detrimento do modo em *batch*, o mesmo é muito usado devido ao facto de ser um algoritmo simples de aplicar e proporcionar soluções em vários tipos de problemas com dificuldades diversas [6].

No processamento em *batch*, a actualização dos pesos é realizada após todos os pares {entrada, saída} serem apresentados à rede e processados em conjunto formando uma época. O treino em *batch* permite o cálculo da precisão do gradiente, sendo a convergência para um mínimo local garantida através do uso de condições simples. Este modo de processamento permite mais facilmente estabelecer comparações entre os diversos parâmetros escolhidos [6].

### 3.3.3 Pré-processamento

Devido ao efeito de saturação da maior parte das funções de activação (funções sigmóide) utilizadas no processamento das redes neuronais, é aconselhável que se escale os vectores de entrada para valores bem distribuídos pela função de activação e que os mesmos não tomem valores muito elevados para não abafarem a contribuição dos valores pequenos. De uma maneira geral, os vectores de entrada são escalados num intervalo de  $[0,1]$  ou  $[-1,1]$  [9, 14].

### 3.3.4 Pós-processamento

As operações de pós-processamento estão relacionadas com a aplicação de redes neuronais na classificação. Num processo de classificação de elementos convém denominar as saídas de uma forma simples; por exemplo, para um problema com duas classes, a saída da rede neuronal é dada por  $\{-1,1\}$ . Os vectores de saída de uma rede neuronal podem ser convertidos numa classe, bastando utilizar uma função degrau na saída da rede [14].

### 3.3.5 Número de neurónios na camada intermédia

Para a resolução da maior parte dos problemas com redes neuronais, a utilização de uma rede com duas camadas intermédias é mais do que suficiente para obter bons resultados. No que diz respeito ao número de neurónios a utilizar na primeira camada intermédia, ele deve rondar o número de

vectores próprios da matriz de covariância. No entanto, não existe uma regra bem definida para o número de neurónios a usar nas camadas. A última camada de neurónios, geralmente, é composta por menos elementos que a primeira [14].

### 3.3.6 Parâmetros de aprendizagem e termo momento

Como foi referido anteriormente, a taxa de aprendizagem,  $\eta$ , deve ser escolhida de tal forma que evite a rápida convergência do algoritmo (isto acontece quando a taxa de aprendizagem é muito grande) uma vez que esta faz com que a rede oscile à volta da solução ideal. Geralmente treina-se uma rede com valores de  $\eta$  abaixo de 0.5 [14].

O termo momento é usado para acelerar e estabilizar o algoritmo. Normalmente, o termo momento é escolhido entre 0.5 e 0.9 e é aconselhável que durante a fase de treino se teste com valores inferiores ao que inicialmente se escolher [12].

### 3.3.7 Mínimos locais

Como foi referido anteriormente, para se otimizar o resultado de uma rede neuronal, o treino deve ser efectuado e testado com vários parâmetros diferentes para a taxa de aprendizagem, momento e valores iniciais de pesos. O número de experiências necessárias,  $r$ , com diferentes valores iniciais de pesos para assegurar que a rede encontrou uma solução otimizada é dado por:

$$r = \frac{\ln(1 - \alpha)}{\ln(1 - p)} \quad [3.23]$$

onde  $p$  representa a percentagem e  $\alpha$  o nível de confiança. Supondo que se pretende assegurar um grau de confiança de  $\alpha = 99\%$  entre as melhores  $p = 20\%$  soluções, então serão necessárias

$$r = \frac{\ln(1 - 0.99)}{\ln(1 - 0.2)} \approx 20 \text{ experiências, ou seja, torna-se necessário repetir a experiência vinte vezes [14].}$$

### 3.3.8 Critérios de Paragem

Ao processar uma rede com o algoritmo de Retropropagação não existem critérios de paragem bem definidos. Existem apenas critérios adquiridos com a prática e a experiência do treino de redes neuronais que podem ser usados para determinar a paragem do modelo e o cálculo do ajuste dos

pesos. Quando o gradiente da função de custo  $W$  for zero para  $W = W^*$ , então,  $W^*$  representa um mínimo local ou global. Um dos critérios de convergência do algoritmo de Retropropagação, elaborado por *Kramer and Sangiovanni –Vincentelli* em 1989, diz que o algoritmo converge quando o valor absoluto da diferença do custo médio por época é suficientemente pequeno.

No entanto, existe outro critério de convergência utilizado no processamento de redes neuronais com o algoritmo de Retropropagação, a generalização. Após cada iteração, a rede é testada no que diz respeito à sua generalização. O processo de aprendizagem é interrompido quando a generalização é adequada ou quando efectuou um pico [6].

### 3.4 VARIANTES DO MODELO DE RETROPROPAGAÇÃO

O algoritmo de Retropropagação converge lentamente; no entanto, existem alguns métodos que permitem acelerar o processo de convergência e aprendizagem do algoritmo. Os principais problemas do algoritmo residem nos seguintes factos:

- a) Os parâmetros da função de custo devem ser ajustáveis e devem ter o seu valor individual – o algoritmo utiliza sempre o mesmo valor da taxa de aprendizagem e esse valor pode não ser o ideal para todos os valores da superfície de erro, ou seja, a taxa de aprendizagem para o peso 1 pode não ser a ideal para o peso 2.
- b) A taxa de aprendizagem deve variar de iteração para iteração – a superfície de erro tem um comportamento diferente de uma zona para outra. Para evitar tais diferenças de comportamento a taxa de aprendizagem deve ser diferente de uma iteração para outra.
- c) Quando as derivadas da função de custo em ordem aos pesos, após várias iterações consecutivas, tem o mesmo sinal, então, a taxa de aprendizagem para esse valor do peso deve ser aumentada – a função pode cair numa zona plana da superfície de erro devido ao facto da derivada da função de custo em ordem aos pesos conter o mesmo sinal durante iterações consecutivas e portanto apontar sempre na mesma direcção. Nesta situação, o número de iterações necessárias para movimentar a função através da superfície de erro poderá ser reduzido aumentando progressivamente a taxa de aprendizagem.

- d) Quando o sinal das derivadas da função de custo em ordem aos pesos altera durante várias iterações consecutivas, a taxa de aprendizagem para esse peso deve ser diminuída – quando a superfície de erro é muito irregular e um ponto da função se localiza num pico ou num vale, é possível que a derivada da função de custo em ordem aos pesos mude de sinal de uma iteração para outra. De forma a evitar que a função oscile, a taxa de aprendizagem do peso em questão deve diminuir.

Estas quatro características deram origem a outros algoritmos que nasceram do algoritmo de Retropropagação [6].

### 3.4.1 Método do Gradiente Conjugado

O método de *Newton* minimiza uma função quadrática num número finito de iterações, o que requer o cálculo e o armazenamento das derivadas de segunda ordem. Assim, quando o número de parâmetros é muito grande pode-se tornar pouco prático o cálculo das derivadas. Este problema tem um relevo especial quando as redes neuronais são aplicadas a situações concretas do mundo real. Nestes casos concretos é desejável resolver o problema em questão derivando apenas uma vez, mas com soluções que correspondem à segunda ordem das derivadas [5].

Considere-se a superfície de erro de uma rede neuronal do tipo perceptrão de multicamadas com aprendizagem supervisionada. Seja  $E(\mathbf{W})$  a função de custo do conjunto de treino da rede. A função  $E(\mathbf{W})$  pode ser escrita em termos da série de Taylor:

$$E(\mathbf{W}(n) + \Delta\mathbf{W}(n)) = E(\mathbf{W}(n)) + \mathbf{g}^T(n)\Delta\mathbf{W}(n) + \frac{1}{2}\Delta\mathbf{W}^T(n)\mathbf{H}(n)\Delta\mathbf{W}(n) + O^3(n) \quad [3.24]$$

onde  $\mathbf{g}(n)$  e  $\mathbf{H}(n)$  representam, respectivamente, o vector gradiente e a matriz *Hessiana*, e  $O^3(n)$  os termos de ordem superior a dois [6].

O método do gradiente conjugado pertence à classe de métodos de optimização de segunda ordem, também denominado método de direcção conjugada. Estes métodos são baseados na minimização de uma função quadrática do tipo:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \quad [3.25]$$

onde  $\mathbf{x} = \Delta\mathbf{W}(n)$ ,  $\mathbf{A} = \mathbf{H}(n)$ ,  $\mathbf{b} = \mathbf{g}(n)$  e  $c = E(\mathbf{W}(n))$ .

Minimizar a função quadrática  $f(\mathbf{x})$  e resolver o sistema linear de equações  $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$ , representam problemas equivalentes, ou seja, a função quadrática é minimizada assumindo para  $\mathbf{x}$  o valor único da expressão  $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$  [6]. Assim, dada uma matriz  $\mathbf{A}$  diz-se que um conjunto de vectores (diferentes de zero)  $\{\mathbf{p}_n\}$  é mutuamente conjugado de  $\mathbf{A}$  se a condição seguinte for satisfeita:

$$\mathbf{p}_n^T \mathbf{A} \mathbf{p}_j = 0, \text{ para todo o } n \text{ e } j \text{ tal que } n \neq j \quad [3.26]$$

Considere-se um conjunto de vectores que consiste nos vectores próprios da matriz  $\mathbf{A}$ . Sejam  $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$  e  $\{z_1, z_2, \dots, z_m\}$  os valores próprios e os vectores próprios, respectivamente, da matriz Hessiana de  $\mathbf{A}$ . Para se verificar que os vectores próprios são conjugados basta substituir  $\mathbf{p}_n$  por  $\mathbf{z}_n$  na equação [3.26]:

$$\mathbf{z}_n^T \mathbf{A} \mathbf{z}_j = \lambda \mathbf{z}_n^T \mathbf{z}_j = 0, \text{ para } n \neq j \quad [3.27]$$

A última igualdade é verdadeira uma vez que os vectores próprios de uma matriz simétrica são mutuamente ortogonais, portanto, os vectores próprios são ortogonais e conjugados.

Através de uma sequência linear ao longo de uma qualquer direcção conjugada  $\{p_1, p_2, \dots, p_m\}$  é possível determinar o mínimo exacto de uma função quadrática com  $m$  parâmetros, em cerca de  $m$  iterações. Para determinar as linhas de direcção conjugadas é necessário obter a condição de conjugação sem a matriz Hessiana. Para tal, calcula-se o gradiente da equação [3.25].

$$\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} \quad [3.28]$$

$$\nabla^2 f(\mathbf{x}) = \mathbf{A} \quad [3.29]$$

Combinando as equações [3.28] e [3.29], obtem-se a variação do gradiente na iteração  $k+1$ ,

$$\begin{aligned} \Delta \mathbf{g}_n &= \mathbf{g}_{n+1} - \mathbf{g}_n \\ &= (\mathbf{A}\mathbf{x}_{n+1} + \mathbf{b}) - (\mathbf{A}\mathbf{x}_n + \mathbf{b}) \\ &= \mathbf{A}\Delta \mathbf{x}_n \end{aligned} \quad [3.30]$$

Onde  $\Delta \mathbf{x}_n = (\mathbf{x}_{n+1} - \mathbf{x}_n) = \eta_n \mathbf{p}_n$  e  $\eta_k$  é escolhido de forma a minimizar a função  $f(\mathbf{x})$  na direcção de  $\mathbf{p}_n$ . Restabelecendo a condição de conjugação, obtem-se:

$$\eta_n \mathbf{p}_n^T \mathbf{A} \mathbf{p}_j = \Delta \mathbf{x}_n^T \mathbf{A} \mathbf{p}_j = \Delta \mathbf{g}_n^T \mathbf{p}_j = 0 \text{ para } k \neq j \quad [3.31]$$

A condição de conjugação foi restabelecida a partir das mudanças no gradiente ao longo das sucessivas iterações. A direcção de procura será conjugada se for ortogonal às mudanças do gradiente. Na primeira direcção de busca,  $\mathbf{p}_0$  é escolhido de forma arbitrária e  $\mathbf{p}_1$  pode ser qualquer vector ortogonal a  $\Delta \mathbf{g}_0$ , existindo portanto um variado conjunto de possíveis vectores. Normalmente, a busca é iniciada usando o método do gradiente decrescente,  $\mathbf{p}_0 = -\mathbf{g}_0$ . Assim, a cada iteração é necessário construir um vector  $\mathbf{p}_n$  que é ortogonal a  $\{\Delta \mathbf{g}_0, \Delta \mathbf{g}_1, \dots, \Delta \mathbf{g}_{n-1}\}$ , e é dado pela expressão seguinte:

$$\mathbf{p}_n = -\mathbf{g}_n + \beta_n \mathbf{p}_{n-1} \quad [3.32]$$

O escalar  $\beta$  pode ser escolhido entre uma série de métodos equivalentes. Dois dos métodos mais usados são os de *Fletcher* e *Reeves* e o *Polak* e *Ribière*, respectivamente:

$$\beta = \frac{\mathbf{g}_n^T \mathbf{g}_n}{\mathbf{g}_{n-1}^T \mathbf{g}_{n-1}} \quad [3.33]$$

$$\beta = \frac{\Delta \mathbf{g}_{n-1}^T \mathbf{g}_n}{\mathbf{g}_{n-1}^T \mathbf{g}_{n-1}} \quad [3.34]$$

### 3.4.2 Método Quasi-Newton

No algoritmo de Retropropagação o ajuste dos pesos é dado por:

$$\Delta \mathbf{W}(n) = -\eta \mathbf{g}(n) \quad [3.35]$$

onde  $\eta$  representa a taxa de aprendizagem e  $\mathbf{g}(n)$  o vector gradiente. Como o método do gradiente decrescente trabalha apenas com o gradiente  $\mathbf{g}(n)$  à volta do ponto  $\mathbf{W}(n)$  na superfície de erro, torna o método de implementação mais simples, mas, por outro lado, o algoritmo converge lentamente (o que pode ser fundamental em problemas mais complexos). A inclusão do termo momento no algoritmo representa uma tentativa de utilização de informação de segunda ordem acerca da superfície de erro. No entanto, este acréscimo nem sempre é benéfico para a aceleração do processo de aprendizagem.

A utilização de termos de segunda ordem vem melhorar o processo de aprendizagem do algoritmo, isto é, a utilização de uma aproximação quadrática da superfície de erro em torno do ponto  $\mathbf{W}(n)$ .

O ajuste dos pesos é dado pela equação [3.36]:

$$\Delta \mathbf{W}^*(n) = \mathbf{H}^{-1}(n) \mathbf{g}(n) \quad [3.36]$$

onde  $\mathbf{H}^{-1}(n)$  representa a inversa da matriz *Hessiana* e  $\Delta \mathbf{W}^*(n)$  o ajuste otimizado de  $\Delta \mathbf{W}(n)$ .

A equação [3.36] representa o que anteriormente foi denominado por método de *Newton*. O método de *Newton* apresenta alguns inconvenientes, tais como, o cálculo da inversa da matriz *Hessiana* (em certos casos) pode ser demorado, para se calcular a sua inversa é necessário que a matriz seja não singular (subsiste o problema da matriz *Hessiana* poder não ser linearmente independente) e se a função de custo não for quadrática não existem garantias de que o método de *Newton* converge.

Uma forma de ultrapassar estes problemas consiste na utilização de uma estimativa do gradiente. Esta modificação no método de *Newton* mantém a matriz *Hessiana* positiva e permite estimar a sua inversa. Usando esta estimativa é garantido que o método decresce pela superfície de erro abaixo. A este método dá-se o nome método *Quasi-Newton*.

O método *Quasi-Newton* utiliza informação de segunda ordem da superfície de erro sem requerer o cálculo da matriz *Hessiana*. Considere-se a actualização dos pesos dada pelo método de *Newton*:

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \eta(n) \mathbf{S}(n), \quad [3.37]$$

onde o vector que define a direcção do gradiente é dada por

$$\mathbf{s}(n) = -\mathbf{S}(n) \mathbf{g}(n)$$

sendo a matriz  $\mathbf{S}(n)$  positiva e ajustada de uma iteração para outra.

No método de *Quasi-Newton* o ajuste dos pesos é dado pela equação [3.38] sendo a expressão que representa o método dado pela equação [3.39].

$$\Delta \mathbf{W}(n) = \mathbf{W}(n+1) - \mathbf{W}(n) \quad [3.38]$$

$$\mathbf{q}(n) = \mathbf{g}(n+1) - \mathbf{g}(n) \quad [3.39]$$

A equação [3.39] pode ser escrita através do recurso a uma expressão aproximada:

$$\mathbf{q}(n) \cong \left( \frac{\partial}{\partial \mathbf{W}} \mathbf{g}(n) \right) \Delta \mathbf{W}(n) \quad [3.40]$$

A matriz *Hessiana* pode ser escrita sob uma forma aproximada usando  $\Delta\mathbf{W}(0), \Delta\mathbf{W}(1), \dots, \Delta\mathbf{W}(W-1)$  como vector de incremento dos pesos (sendo  $W$  linearmente independente) e  $\mathbf{q}(0), \mathbf{q}(1), \dots, \mathbf{q}(W-1)$  como o vector de incremento do gradiente,

$$\mathbf{H} \cong [\mathbf{q}(0), \mathbf{q}(1), \dots, \mathbf{q}(W-1)] [\Delta\mathbf{W}(0), \Delta\mathbf{W}(1), \dots, \Delta\mathbf{W}(W-1)]^{-1} \quad [3.41]$$

$$\mathbf{H}^{-1} \cong [\Delta\mathbf{W}(0), \Delta\mathbf{W}(1), \dots, \Delta\mathbf{W}(W-1)] [\mathbf{q}(0), \mathbf{q}(1), \dots, \mathbf{q}(W-1)]^{-1} \quad [3.42]$$

O método do Gradiente Conjugado e o método *Quasi-Newton* evitam o cálculo da matriz *Hessiana*, embora este último realize um cálculo aproximado da inversa da matriz *Hessiana*. O método de *Quasi-Newton*, numa região próximo do mínimo local, tende a comportar-se como o método de Newton, e, portanto, faz com que a rede convirja mais rapidamente do que com o método do Gradiente Conjugado [6].

O método de *Quasi-Newton* não é tão sensível em termos de precisão na fase de optimização do algoritmo, comparado com o método do Gradiente Conjugado. O método do Gradiente conjugado utiliza menos recursos computacionais, ao contrário do método de *Quasi-Newton* em que é necessário calcular aproximações da matriz *Hessiana* e da sua inversa, desta forma, o primeiro método é muito utilizado em redes neuronais cujos parâmetros a calcular são elevados. Assim, na maior parte das aplicações práticas, o método de *Quasi-Newton* é usado em redes de pequena dimensão [6].

### 3.5 GENERALIZAÇÃO

Um dos factores cruciais quando se treina uma rede neuronal consiste na análise da generalização da rede, ou seja, verificar se a rede aprendeu bem para um conjunto de padrões independentes do conjunto de treino. Estudos mais recentes sobre as redes de perceptrões de multicamada indicam que após um determinado ponto, as redes treinadas com o algoritmo de Retropropagação continuam a obter bons resultados com o conjunto de treino mas a solução do conjunto de teste começa a deteriorar-se. Quando isto acontece, diz-se que a rede entrou em 'excesso de treino'. Assim, existem dois factores importantes e distintos quando se analisa uma rede neuronal: o desempenho do conjunto de teste e a boa generalização do algoritmo. Ao desenhar uma rede

neuronal deve-se ter em consideração estes dois factores, uma vez que uma rede neuronal que não generalize bem não se adequa à aplicação prática de um determinado problema e uma rede que tenha um bom desempenho no conjunto de teste, não significa que esteja a generalizar bem [6, 9, 14].

Pode-se cair num de dois casos: se a rede utilizar poucos parâmetros de cálculo ela não consegue generalizar bem e, por outro lado, se ela tiver parâmetros em excesso, pode adaptar-se apenas àquele conjunto de teste e não a qualquer um outro. A generalização de uma rede neuronal depende do tamanho da amostra e do método de generalização usado.

### 3.5.1 Tamanho da amostra

Uma das questões mais importantes que envolvem o processamento de dados com as redes neuronais diz respeito ao tamanho da amostra e à quantidade de parâmetros a calcular. O tamanho do conjunto de treino influencia todo o desempenho da rede. Em qualquer problema a ser resolvido com o recurso às redes neuronais é necessário um elevado conjunto de dados.

O número de padrões,  $N$ , necessários para classificar um determinado conjunto com um erro  $\delta$  é, aproximadamente, dado por  $N > \frac{w}{\delta}$ , onde  $w$  representa o número de parâmetros da rede. Esta expressão indica que o número de padrões necessários num conjunto de treino aumenta linearmente com o número de parâmetros da rede [12].

Sendo  $\frac{N}{w}$  a razão entre o número de padrões e o número de parâmetros da rede, se  $w$  for muito grande a rede pode cair num caso de 'excesso de treino' (*over-fitting*); se o inverso se verificar, ela pode entrar num caso de 'falta de treino' (*under-fitting*). Subsiste uma regra empírica que diz que o número de padrões num conjunto de treino deve ser aproximadamente dez vezes superior ao número de parâmetros da rede,  $N \approx 10w$  [12].

O número de parâmetros a calcular numa rede neuronal está directamente ligado ao número de vectores de entrada, ao número de neurónios nas camadas intermédias e na saída.

Supondo que se tem uma rede neuronal com 6 vectores de entrada, uma camada intermédia com 4 neurónios e 2 neurónios na camada de saída, o total de parâmetros a calcular pela rede é de 38.

A expressão matemática que traduz este valor é dada por:

$$w = (d + 1) \times h + (h + 1) \times c \quad [3.43]$$

sendo  $d$  o número de vectores na camada de entrada na rede,  $h$  o número de neurónios na camada intermédia e  $c$  o número de neurónios na camada de saída [14].

### 3.5.2 Método da Validação Cruzada

Um dos métodos existentes para garantir boa generalização consiste na paragem antecipada do treino ou paragem com validação cruzada [12].

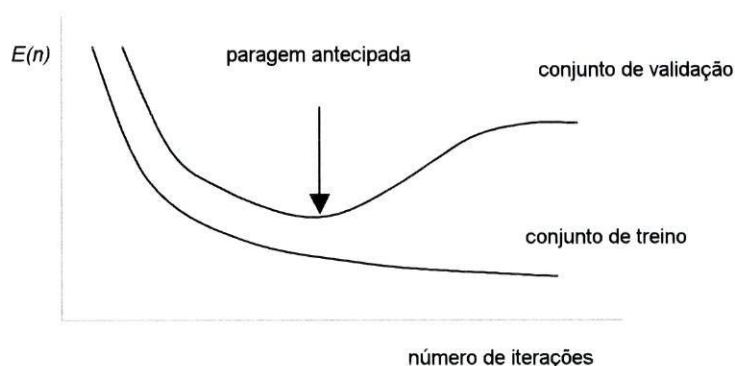


Fig. 3.5 - Validação cruzada

O método de paragem com validação consiste, basicamente, em dividir a amostra em três grupos: o grupo de treino, o de teste e o de validação. Os vectores são introduzidos na rede e o treino começa. Quando o erro do conjunto de validação começa a aumentar significativamente, a rede pára de treinar, evitando o 'excesso de treino'. O conjunto de validação deve conter cerca de dez por cento dos elementos do conjunto de treino. No entanto, este método contém um senão, o tamanho da amostra de treino. Geralmente, no treino de redes neuronais tem-se sempre um conjunto limitado de dados e, utilizando a paragem com validação cruzada, o conjunto de treino fica ainda mais pequeno. Apesar deste problema, este método continua a ser muito utilizado em aplicações reais [12].

### 3.5.3 Método de Eliminação de Pesos

Outra técnica que existe para garantir uma boa generalização do modelo consiste em eliminar os parâmetros que não estejam a contribuir para a solução ideal. Começa-se por ter uma rede neuronal com várias ligações e com vários neurónios e, sucessivamente, elimina-se as conexões menos importantes, ou seja, a rede fica com uma topologia mais apropriada para a resolução do problema em causa e que permita uma boa generalização. Em oposição à técnica de eliminação de parâmetros existe a técnica de construção da rede neuronal, ou seja, começa-se com uma rede de topologia pequena e sucessivamente aumenta-se a topologia de forma a que a rede generalize bem [6, 14].

### 3.5.4 Método de Regularização

Outra técnica que permite obter uma boa generalização consiste em acrescentar à função de custo um termo de regularização,

$$E_{total} = E + \lambda E_{reg} , \quad [3.44]$$

onde  $\lambda$  representa o parâmetro que controla a regularização,  $E_{reg}$  o termo de regularização e  $E_{total}$  o total da função de custo que será minimizada.

O termo de regularização,  $E_{reg}$ , é escolhido de forma a que a função que é gerada na saída da rede neuronal seja suavizada. O termo  $E_{reg}$  deve ter valores pequenos para vectores de pesos que gerem valores suavizados na saída e valores elevados para os que gerarem valores não suavizados na saída [6].

#### **4. UTILIZAÇÃO DO GPS**

## 4.1 INTRODUÇÃO

Como foi referido anteriormente, o trabalho apresentado tem como objectivo a elaboração de um estudo de fluxos de tráfego dentro de uma cidade, escolhendo um par origem-destino. O par escolhido para este trabalho foi Biblioteca Municipal-Centro Comercial. Escolheu-se este par origem-destino devido à curta distância entre eles (cerca de 3km) e devido ao facto de existirem várias alternativas para o destino. Ao realizar-se este trajecto, as ruas percorridas são, obrigatoriamente, as do centro da cidade, permitindo uma análise mais detalhada da dinâmica da mesma.

Para a realização deste trabalho, a existência de um Sistema de Informação Geográfica (SIG) de apoio a todo o processo torna-se fundamental. A realização deste estudo teve o apoio da Câmara Municipal de Santa Maria da Feira, nomeadamente, na disponibilização dum conjunto vasto de dados do seu SIG, referentes à cidade de Santa Maria da Feira. Assim, foram disponibilizados para a realização deste trabalho os seguintes dados:

1. Cartografia digital – voo fotogramétrico datado de 1993, Datum 73;
2. Bases de dados do SIG: rede viária, construções, serviços, escolas, desporto, saúde, religião, património, entre outros;
3. Programa informático de SIG - Geomedia Professional 4.0;

Da Faculdade de Ciências da Universidade do Porto foram disponibilizados os receptores GPS Pro XR e GeoExplorer III da Trimble, capazes de realizar medidas de códigos e de fases dinâmicas e o *software* para tratamento dos dados. Devido ao facto da cartografia digital disponível ser datada de 1993, encontrava-se desactualizada, tornando-se necessário proceder a uma actualização das bases de dados existentes. A base de dados foi actualizada com recurso ao GPS em modo diferencial. Assim, foram levantadas as seguintes características:

1. Eixos de via
  - Nome da rua
  - Tipo de pavimento
  - Estado de conservação do pavimento
  - Largura da estrada
2. Passadeiras
3. Sinalização Vertical

4. Paragens de autocarros
5. Lombas
6. Semáforos

Foram também actualizadas as seguintes características relativas à rede viária:

1. Zonas conflituosas: escolas, zonas comerciais e de serviços
2. Estacionamentos
3. Zonas onde existem viaturas em segunda fila
4. Ruas onde é normal passarem camiões
5. Zonas de conflito com peões

Para o apoio GPS na cidade, recorreu-se à estação de referência que se localiza no topo do Edifício I do Departamento de Administração Urbanística da Câmara Municipal. As coordenadas da estação de referência da Câmara Municipal são:

$$\varphi = 40^{\circ} \ 55' \ 32''.4654 \ N$$

$$\lambda = 8^{\circ} \ 32' \ 36''.5127 \ W$$

$$H = 201.654 \ m$$

Na figura seguinte encontra-se representada a área de actuação neste trabalho.

# MESTRADO EM POSICIONAMENTO E NAVEGAÇÃO POR SATÉLITE

Concelho de Santa Maria da Feira

## Área de Estudo

- Biblioteca e Centro Comercial
- Serviços gerais
- Mercado
- Igrejas e capelas
- Saúde
- Logradouros de saúde
- Elementos desportivos
- Logradouros desportivos
- Escolas
- Logradouros escolares
- Rede viária



Projeção do Gauss (Eipsovir Internacional) Datum 73 Escala 1:12.000 data Março 2003

## 4.2 ACTUALIZAÇÃO DA CARTOGRAFIA DIGITAL

Para o levantamento das características das ruas foram utilizados os receptores mencionados anteriormente (trata-se de receptores de uma frequência, sendo o Pro XR utilizado como estação de referência e o GeoExplorer como receptor móvel), um veículo ligeiro, um condutor e um operador. No levantamento dos eixos de via e suas características, a antena do receptor GeoExplorer foi colocada no tejadilho de um veículo e, à medida que o veículo ia percorrendo as ruas da cidade, o operador guardava no receptor as características da via a ser percorrida utilizando um dicionário previamente definido. Os elementos referentes à sinalização das ruas: passadeiras, lombas, paragens de autocarro, semáforos e sinais de trânsito foram levantados sob a forma de ponto, ficando o veículo estático cerca de 10 segundos em cada elemento. As características levantadas pelo receptor deram origem a vários mapas temáticos. As figuras seguintes exemplificam alguns:



Fig. 4.2 - Tipo de pavimento

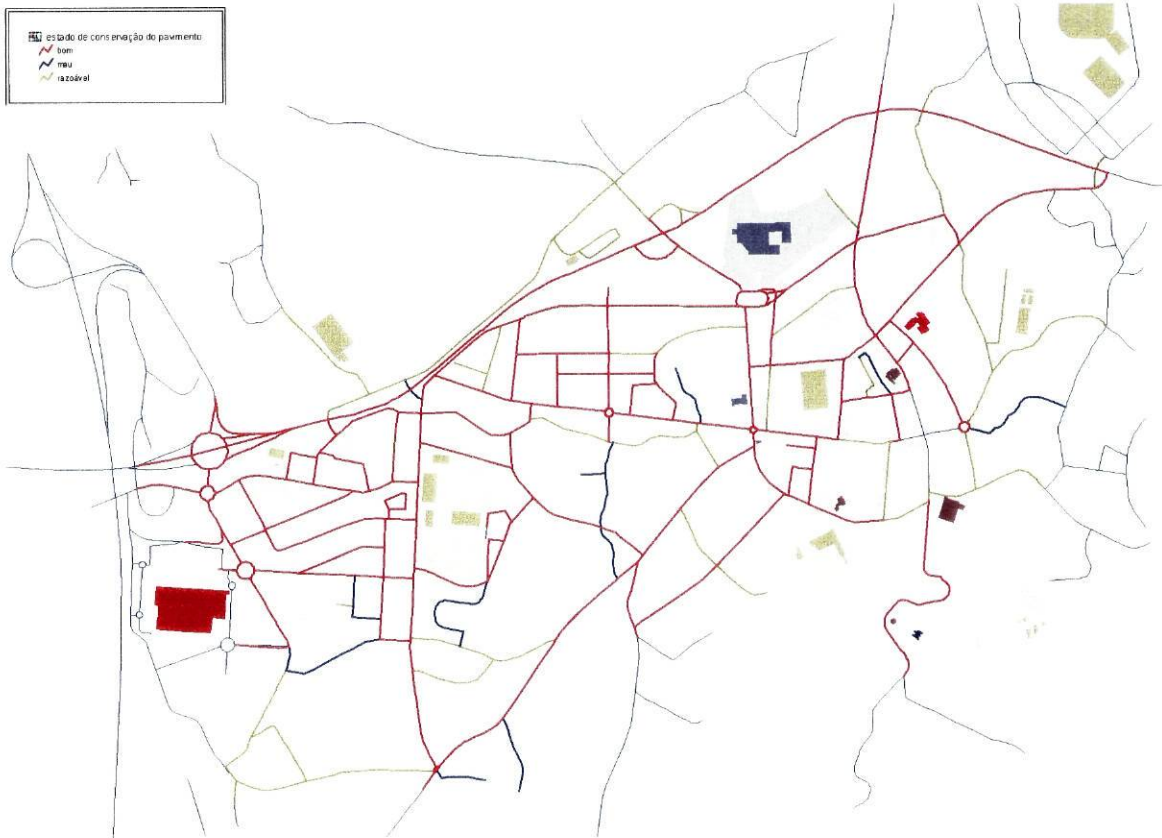


Fig. 4.3 - Estado de conservação do pavimento



Fig. 4.4 - Largura da via



Fig. 4.5 - Sinalização: lombas, paragem de autocarro, passadeiras e semáforos



Fig. 4.6 - Sinalização: sinais de trânsito



Fig. 4.7 - Constrangimentos na via: carros estacionados em segunda fila



Fig. 4.8 - Constrangimentos na via – conflitos com peões



Fig. 4.9 - Estacionamento



Fig. 4.10 - Vias de sentido único

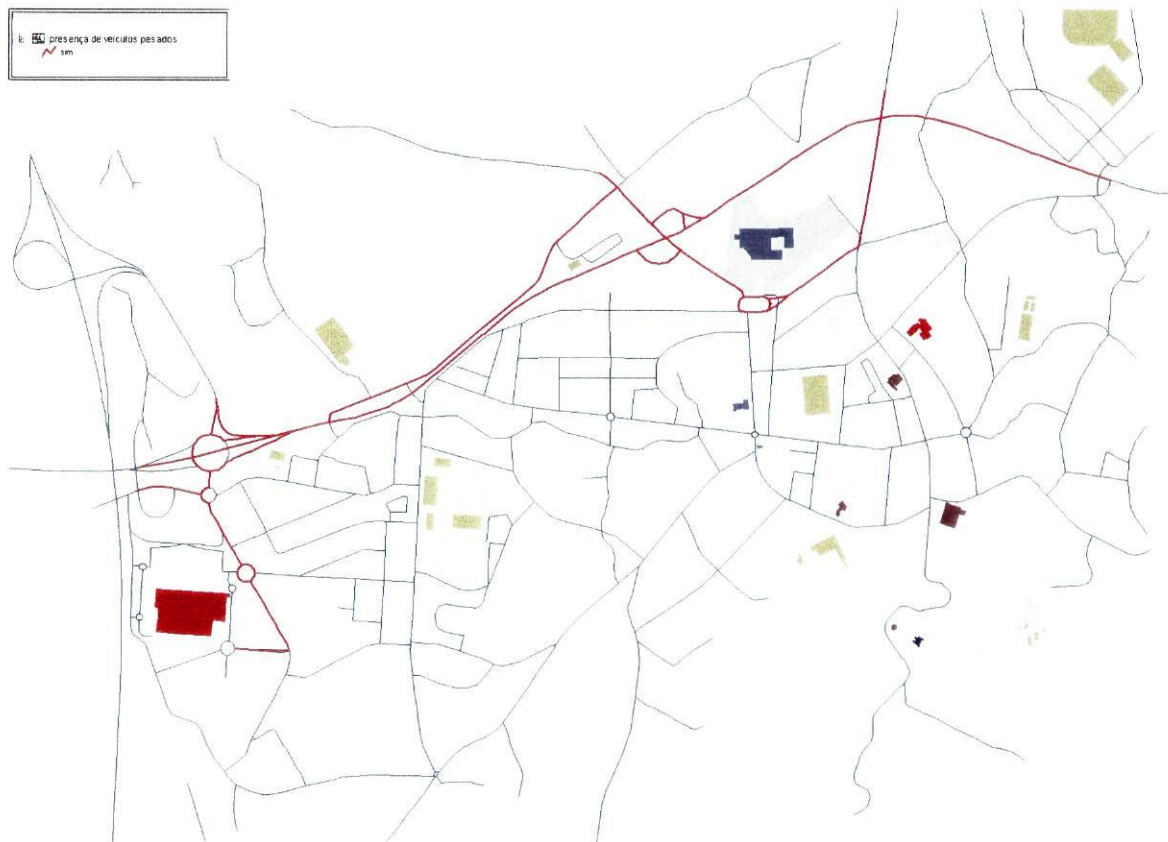


Fig. 4.11 - Presença de veículos pesados



Fig. 4.12 - Ruas sem saída ou não conduzem ao destino pretendido

### 4.3 REGISTO DE PERCURSOS ENTRE A ORIGEM-DESTINO

Para a realização deste estudo solicitou-se a colaboração de 29 pessoas, que gentilmente acederam a realizar o trajecto Biblioteca-Centro Comercial. Às pessoas foi apenas indicado que deveriam conduzir um veículo, com a antena do GPS colocada no tejadilho, entre a origem e o destino. O percurso por eles escolhido foi registado no receptor GPS, processado e lançado no SIG sendo o alvo deste estudo.

Os percursos foram realizados entre os dias 06 e 22 de Fevereiro de 2002, das 10h às 22.30h, durante a semana (apenas um dia foram realizados percursos ao fim de semana).

Durante o percurso o operador ia registando as seguintes características:

- Percepção do trânsito (indicada pelo condutor),
- Condições climatéricas do dia,
- Condições do pavimento face às condições climatéricas.

Ao todo foram realizados 104 percursos entre a origem e o destino. Na tabela seguinte encontram-se resumidas as características levantadas por 7 condutores. A cada condutor foi atribuído um número identificador (ID) do trajecto que efectuou, bem como as características que foram levantadas durante o percurso.

| ID | Condutor        | Data  | Hora Início | Hora Fim | Comprimento (m) | Condições Climatéricas | Estado do Pavimento | Percepção do trânsito |
|----|-----------------|-------|-------------|----------|-----------------|------------------------|---------------------|-----------------------|
| 1  | Isabel Milheiro | 16/02 | 15:04       | 15:10    | 2860,9          | sol                    | seco                | muito                 |
| 2  | João            | 16/02 | 15:16       | 15:22    | 2838,5          | sol                    | seco                | médio                 |
| 3  | Rui             | 16/02 | 15:30       | 15:34    | 3288,5          | sol                    | seco                | médio                 |
| 4  | Moita           | 16/02 | 15:40       | 15:46    | 2838,5          | sol                    | seco                | médio                 |
| 5  | Carlos Meneses  | 16/02 | 15:55       | 15:59    | 2540,6          | sol                    | seco                | médio                 |
| 6  | Zé Carlos       | 18/02 | 12:12       | 12:17    | 3106,0          | chuva                  | molhado             | pouco                 |
| 7  | Helder          | 18/02 | 12:23       | 12:27    | 3072,1          | sol                    | molhado             | médio                 |

Tabela 4.1 - Exemplo da base de dados levantada pelo operador durante os trajectos

Um factor essencial na escolha dos condutores a realizar o trajecto, diz respeito ao seu grau de conhecimento da cidade em estudo, à sua idade e sua cultura. Assim, na Tabela 4.2 encontram-se descritos o número de vezes que cada condutor efectuou o trajecto, bem como alguns dados relativos à sua profissão e local onde a exercem.

| Condutores      | Nº de vezes que o      |       | Profissão               | Local onde exercem a profissão |
|-----------------|------------------------|-------|-------------------------|--------------------------------|
|                 | percurso foi realizado | Idade |                         |                                |
| Alexandrina     | 5                      | 27    | Engenheira              | Sta. Maria da Feira            |
| Bastos          | 5                      | 49    | Desenhador              | Sta. Maria da Feira            |
| Cardoso         | 1                      | 23    | Corticeiro              | Fora da cidade                 |
| Carlos Almeida  | 8                      | 27    | Engenheiro              | Sta. Maria da Feira            |
| Carlos Meneses  | 5                      | 23    | Técnico informática     | Sta. Maria da Feira            |
| Catarina        | 1                      | 19    | Estudante 12º           | Fora da cidade                 |
| Cátia           | 4                      | 20    | Estudante universitária | Fora da cidade                 |
| Célia           | 4                      | 30    | Arquitecta              | Sta. Maria da Feira            |
| Daniela         | 1                      | 21    | Empregada de escritório | Fora da cidade                 |
| Filipe Amorim   | 2                      | 24    | Técnico informática     | Sta. Maria da Feira            |
| Filipe Milheiro | 4                      | 31    | Engenheiro              | Sta. Maria da Feira            |
| Helder          | 15                     | 21    | Administrativo          | Sta. Maria da Feira            |
| Isabel          | 1                      | 26    | Professora              | Fora da cidade                 |
| João            | 1                      | 23    | Estudante universitário | Fora da cidade                 |
| Marco           | 4                      | 25    | Engenheiro              | Sta. Maria da Feira            |
| Mário           | 1                      | 21    | Madeireiro              | Fora da cidade                 |
| Moita           | 1                      | 23    | Estudante universitário | Fora da cidade                 |
| Paulo           | 1                      | 22    | Corticeiro              | Fora da cidade                 |
| Pedro           | 9                      | 22    | Administrativo          | Sta. Maria da Feira            |
| Pedro Santos    | 1                      | 22    | Motorista               | Fora da cidade                 |
| Pinheiro        | 6                      | 26    | Arquitecto              | Sta. Maria da Feira            |
| Rita            | 8                      | 25    | Geógrafa                | Sta. Maria da Feira            |
| Rosário         | 2                      | 31    | Téc. Superior de RI     | Sta. Maria da Feira            |
| Rui             | 1                      | 26    | Operário fabril         | Fora da cidade                 |
| Sónia           | 3                      | 27    | Urbanista               | Sta. Maria da Feira            |
| Teixeira        | 1                      | 26    | Arquitecto              | Sta. Maria da Feira            |
| Tozé            | 1                      | 26    | Urbanista               | Sta. Maria da Feira            |
| Vânia           | 5                      | 25    | Arquitecta              | Sta. Maria da Feira            |
| Zé Carlos       | 3                      | 27    | Estudante universitário | Fora da cidade                 |

Tabela 4.2 - Quadro descritivo do número de vezes que cada condutor realizou o percurso

Da Fig. 4.13 pode-se constatar que 59% dos condutores, pelo facto de exercerem a sua actividade profissional na cidade, conhecem muito bem o trajecto a realizar bem como possíveis alternativas em caso de congestionamento. Os restantes condutores, apesar de não exercerem a sua profissão na cidade, são naturais do Concelho e habitualmente deslocam-se à Biblioteca ou ao Centro Comercial.

Durante a realização dos percursos as condições climatéricas, em Santa Maria da Feira, foram sempre muito semelhantes, predominando o bom tempo (sol), e portanto, o estado do pavimento face às condições climatéricas encontrou-se 94.2% das vezes, seco, como se pode verificar nas Fig. 4.14 e 4.16.

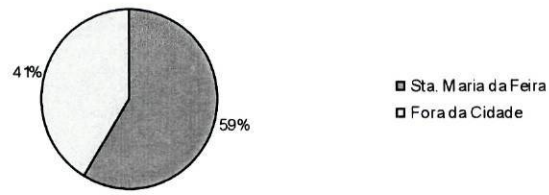


Fig. 4.13 - Percentagem de condutores que trabalham em Sta. Maria da Feira

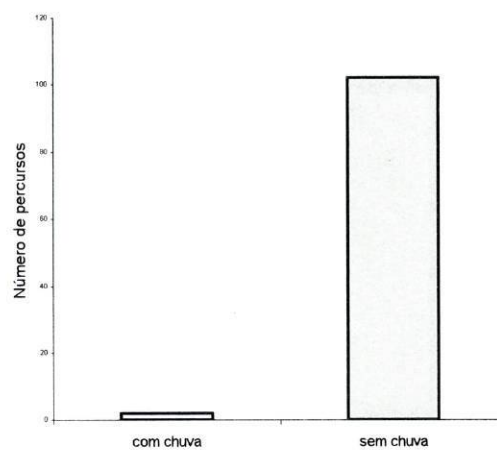


Fig. 4.14 - Condições atmosféricas

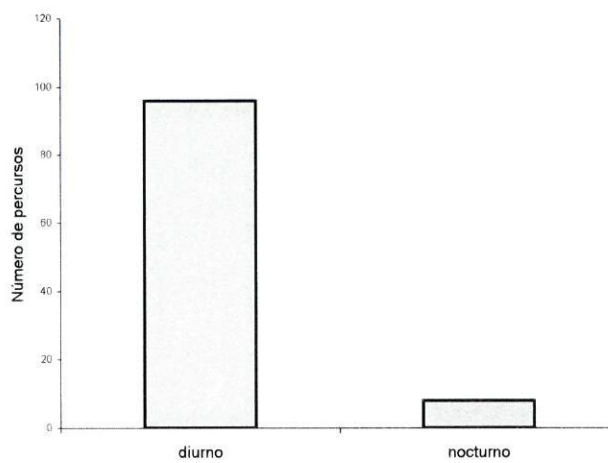


Fig. 4.15 - Horário de realização dos percursos

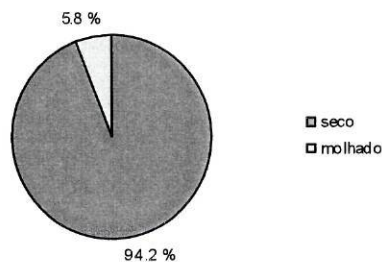


Fig. 4.16 - Estado do pavimento face às condições climáticas

Outra das características levantadas durante a realização dos percursos foi a percepção do trânsito que o condutor tinha, alguns segundos depois de ter iniciado o trajecto. A percepção que o condutor tem do trânsito durante a viagem é um factor essencial para se perceber se tal facto foi suficiente para que o condutor escolhesse percorrer outras ruas (por norma com menos trânsito).

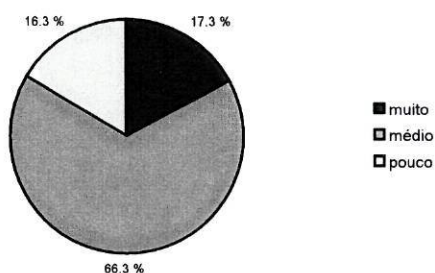


Fig. 4.17 - Percepção do trânsito

Existem várias possibilidades de aceder ao destino em estudo. Algumas obrigam a passar pelas escolas básica e secundária, outras obrigam a passar por zonas conflituosas em termos de fluxo de tráfego, por zonas comerciais e de serviços. Desta forma, é de esperar que os comprimentos dos percursos variem de condutor para condutor. Assim, a maior parte dos condutores (48.1%) realizaram percursos entre os 3040 e os 3540 metros, apenas 13.5% dos condutores realizaram percursos superiores a 3540 metros. Tal facto deve-se, por um lado, à diversidade de alternativas para evitar o congestionamento e, por outro lado, ao facto de muitos dos condutores quando se deparam com a tarefa, optam por utilizar as mesmas estradas que normalmente usam quando se deslocam de casa para o trabalho e vice-versa.

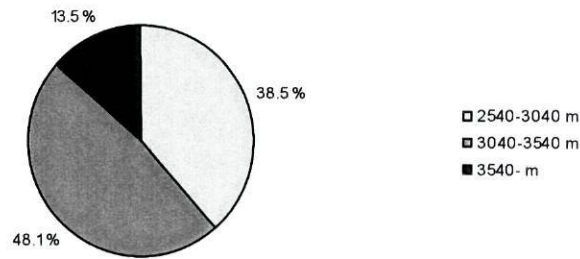


Fig. 4.18 - Variação dos comprimentos dos percursos

Associado ao comprimento dos percursos estão os tempos que cada condutor demorou a realizar o trajecto. Uma vez que o trajecto é pequeno, o tempo que cada condutor demorou a efectuar-lo está directamente ligado com a percepção do trânsito, com as condições climatéricas, com o estado de conservação do pavimento, com as zonas conflituosas, entre outras. Verificou-se que 30.8% e 40.4% demoraram, respectivamente, 5 e 6 minutos a entrar no estacionamento do Centro Comercial.

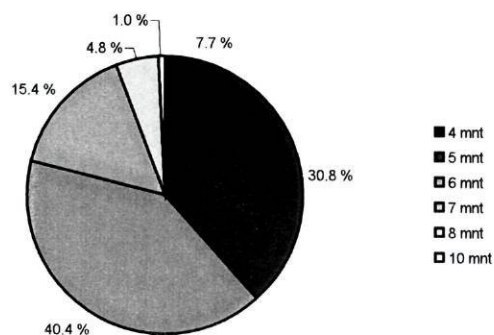


Fig. 4.19 - Tempo de percurso

Analisando os vários trajectos que os condutores realizaram verifica-se que o trajecto seguido mais vezes compreende a passagem pela Avenida Dr. Belchior Cardoso Costa (em sentido Sul-Norte), pela Avenida 25 de Abril, pela Rua Comendador Sá Couto e pela Avenida Dr. Francisco Sá Carneiro. A Fig. 4.20 demonstra que, partindo da Biblioteca, os condutores preferiram a direcção Sul-Norte (a cor castanha), passando nos semáforos. Outro trajecto que reúne muitos adeptos é o

que passa pelas ruas António de Castro Corte Real, pelas escolas, entrando no Centro Comercial pela Rua Domitília Carvalho.

O dia e a hora da realização dos percursos foram condicionados à disponibilidade dos condutores de efectuarem o trajecto. A maior parte dos condutores mostrou-se disponível para essa tarefa durante a manhã e a tarde. Assim, apenas 9.6% mostraram disponibilidade para efectuarem o percurso na hora de almoço, 8.7 % entre as 17.30 e as 20h e 5.8 % na hora de jantar. Na tabela seguinte encontra-se registada a distribuição dos percursos pela hora do dia, bem como a média de tempo de percurso por horário.



Fig. 4.20 - Arcos mais percorridos

| Designação    | Horário (h) | Nº de percursos | % Por horário | Tempo por horário (mnt) | Média (mnt) |
|---------------|-------------|-----------------|---------------|-------------------------|-------------|
| Manhã         | 9,30-12,0   | 38              | 36,5          | 226                     | 5,9         |
| Almoço        | 12,0-14,0   | 10              | 9,6           | 52                      | 5,2         |
| Tarde         | 14-17,30    | 41              | 39,4          | 249                     | 6,1         |
| Hora de ponta | 17,30-20    | 9               | 8,7           | 47                      | 5,2         |
| Jantar/noite  | 20-22,30    | 6               | 5,8           | 32                      | 5,3         |

Tabela 4.3 - Distribuição dos percursos por horário

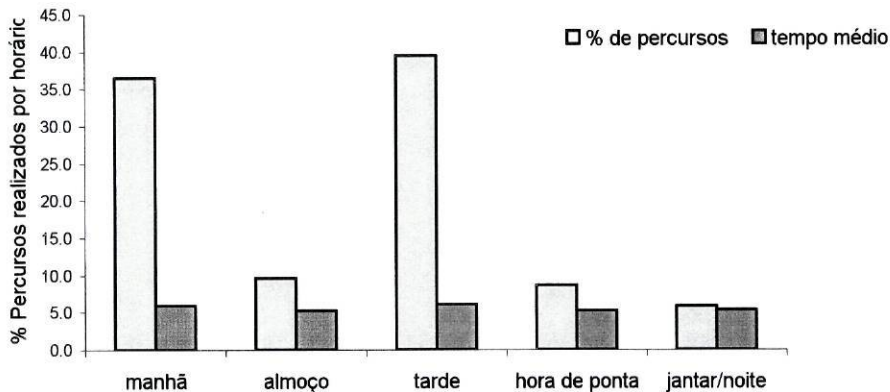


Fig. 4.21 - Distribuição dos percursos e tempos médios por horário

#### 4.4 RESULTADOS OBTIDOS

O GPS foi utilizado na determinação das posições dos elementos que compõem uma rede viária, como é o caso da sinalização das vias, da existência de lombas, semáforos, entre outros, e no registo dos percursos que os condutores efectuaram. Os dados foram processados em GPS modo diferencial, utilizando, para tal, a estação de referência da Câmara Municipal, onde foi colocado o receptor Trimble Pro XR.

A posição da antena móvel é obtida a partir das medidas de pseudo-distâncias a pelo menos 4 satélites comuns aos dois receptores (móvel e fixo), para se conseguir posicionamento tri-dimensional. O sistema, a ser utilizado em ambiente urbano, perde muito da sua eficácia devido aos frequentes bloqueios dos sinais GPS por parte das construções e outros obstáculos existentes nestes ambientes [7]. Desta forma, as medidas de fase ficam praticamente inutilizadas quando o veículo se encontra em movimento, pois a fixação da ambiguidade inicial (necessária sempre que haja interrupção de sinal) exige que o veículo fique imobilizado num período que oscila entre os 30 segundos e os 5 minutos. Esta paragem, numa cidade, é difícil de concretizar.

A precisão do posicionamento com o GPS depende não só da precisão de cada medida de pseudo-distância, à qual está associada o correspondente desvio padrão  $\sigma_r$ , mas também à distribuição geométrica dos satélites [15]. A relação entre estes dois factores é traduzida através do factor DOP que é dado pela seguinte expressão.

$$\sigma_x = xDOP \times \sigma_r \quad [4.3]$$

onde  $x$  pode tomar a forma de H em caso de posicionamento horizontal (2D), V em caso de posicionamento vertical, P em caso de posicionamento tridimensional (3D), T no caso da determinação do tempo e G no caso de se obter posicionamento a quatro dimensões ( $X, Y, Z, Tempo$ ). Normalmente, um GDOP abaixo de 6 é considerado bom [7].

Na Fig. 4.22 encontra-se representada a distribuição do PDOP durante a realização dos percursos do dia 06/02/2002. Os valores de PDOP são valores médios para cada percurso. Pode-se constatar que, num ambiente urbano, a perda de satélites visíveis é um problema, uma vez que diminui o PDOP e consequentemente diminui a precisão no posicionamento.

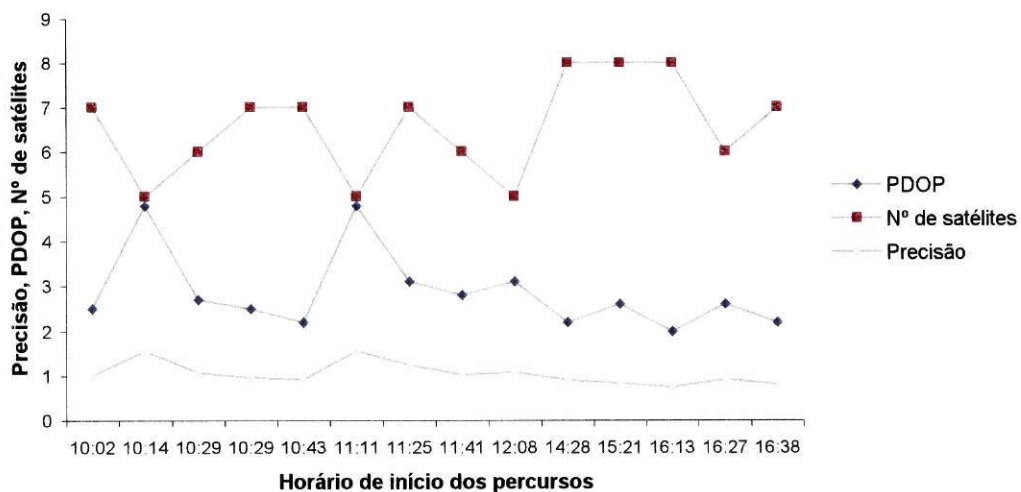


Fig. 4.22 - Precisão, PDOP e número de satélites visíveis

Para evitar que os sinais vindos de satélites próximos do plano do horizonte que contêm reflexões e ruído, sejam recebidos no receptor, configurou-se este para a recepção de sinais com uma máscara de elevação superior a 15°.

Numa cidade com edifícios muito altos ('canyons' urbanos), o problema da obstrução do sinal toma proporções elevadas, sendo necessário recorrer a outras formas de posicionamento. No caso de Santa Maria da Feira, como pode ser analisado pela Fig. 4.23, predominam as moradias e as novas avenidas largas, portanto, este problema não é considerado muito grave, já que em 42 % dos trajectos realizados pelos condutores foram visíveis, em média, 6 satélites.

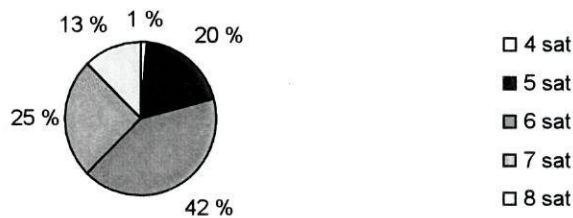


Fig. 4.23 - Número de satélites visíveis

Os levantamentos para este trabalho foram realizados em modo cinemático, registando códigos e fases. No levantamento da sinalização da rede viária, após a colocação da antena no veículo e a configuração do receptor, iniciou-se o trabalho parando cerca de dez segundos em cada sinal, registando posições a uma cadência de segundo. Verifica-se que, após a correcção diferencial, com a estação de referência a 3 km e utilizando o método de processamento de códigos e fases (modelo de *Klobuchar* para a ionosfera e modelo de *Hopfield* para a troposfera), as melhores soluções resultam das medidas de código, como já era de esperar, atendendo ao que foi anteriormente exposto. O desvio padrão dos pontos levantados por este método variam entre 0.8 e 1.5 metros (para um nível de confiança de  $1\sigma$ ).

Para o registo dos percursos foram utilizadas apenas medidas de código, consoante o que ficou exposto.

## **5. APLICAÇÃO PRÁTICA DAS REDES NEURONAIS**

## 5.1 INTRODUÇÃO

O objectivo deste trabalho consiste em treinar uma rede neuronal que permite prever a percentagem de pessoas que vão utilizar determinados arcos face a um conjunto de características previamente estabelecidas. Deste modo será possível não só prever os congestionamentos de tráfego em determinados arcos mas também fornecer aos condutores percursos alternativos.

Como foi referido no capítulo anterior, o GPS teve um papel fundamental quer no levantamento das características da rede viária quer no registo dos percursos que cada condutor livremente realizou.

As características levantadas foram analisadas e, como nem todas são independentes, face á pequena dimensão do número de percursos, foram utilizadas na rede neuronal apenas seis:

1. Percepção do trânsito – dada pelo condutor em cada percurso,
2. Estado de conservação do pavimento – levantado com o GPS e introduzido no SIG,
3. Velocidades praticadas em cada arco – levantado com o GPS e introduzido no SIG,
4. Hora do dia – registada no ficheiro de dados de cada percurso,
5. Estado do pavimento face às condições atmosféricas – registado no ficheiro de dados pelo operador e introduzido no SIG,
6. Número de entroncamentos – determinado no SIG.

Os vectores foram normalizados entre  $[-1, 1]$  e o método de normalização utilizado para os vectores de entrada e dos alvos encontram-se descritos em pormenor no apêndice.

O conjunto de dados foi dividido em três partes: treino, validação e teste final da rede. O conjunto de treino é composto por 62 percursos, sendo os restantes divididos em partes iguais para validação e teste. O trajecto escolhido para estudo engloba, no total, 281 arcos da rede viária da cidade. Para o tratamento dos dados foi criada uma rede neuronal com uma camada intermédia de neurónios e uma camada de saída contendo apenas um neurónio.

Foram feitos dois tipos de estudo. Numa primeira parte, a rede neuronal foi treinada arco a arco, sendo o vector de entrada constituído pelas características já referidas e a resposta da rede a passagem ou não passagem do condutor no arco. Se cada um dos condutores tivesse realizado o percurso muitas vezes, sob condições diferentes, o resultado da rede permitiria prever o comportamento de cada condutor face a um conjunto de características fixado.

Se a amostra dos condutores caracterizasse bem o universo dos utilizadores das ruas de Santa Maria da Feira, o problema estava resolvido. Como nenhuma destas condições é totalmente verificada, optou-se, numa segunda fase, por um tratamento mais impessoal dos dados. Com efeito, todos os arcos foram tratados simultaneamente, os vectores de estudo continuaram a ser seis características, o estado de conservação do pavimento foi substituído por conflitos com peões e a saída é a percentagem de condutores que percorreram cada arco. Neste caso, para o treino da rede foram utilizados dois estilos diferentes: uma rede estática, com todos os vectores de entrada concorrentes e, posteriormente, uma rede dinâmica, para a qual foram criadas quatro sequências obtidas com as combinações das características 4 e 5.

O estudo da sensibilidade da rede, face ao número de neurónios da camada intermédia, a taxa de aprendizagem e o valor do momento é apresentado nos parágrafos seguintes.

## 5.2 CRIAÇÃO DA REDE NEURONAL

A escolha do número de neurónios a introduzir na camada intermédia teve em conta a expressão [3.43], para encontrar o número ideal de neurónios, tendo-se verificado que a utilização de 6 neurónios produzia uma percentagem de erro inferior à que resultava de considerar 4, 5 ou 7 neurónios. No gráfico da Fig. 5.1 pode-se analisar a variação do erro proveniente do comportamento dos neurónios na camada intermédia, em função do número de neurónios (exemplo do arco 23).

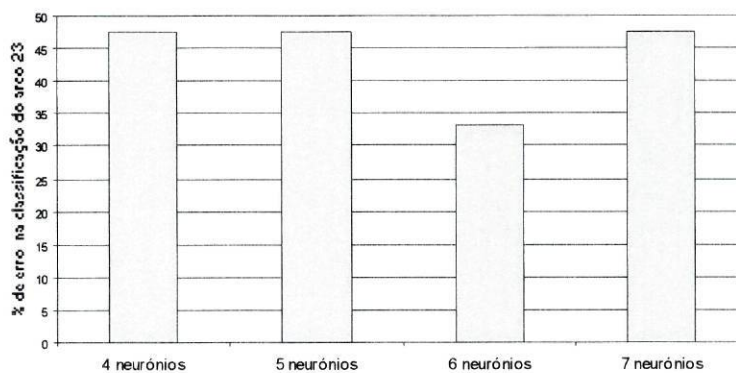


Fig. 5.1 - Percentagem de erro na classificação em função do número de neurónios utilizado na camada intermédia

A escolha da melhor taxa de aprendizagem a utilizar para um dado problema nem sempre é uma tarefa fácil. A taxa de aprendizagem não deve ser muito grande pois provocará instabilidade na

aprendizagem e se for muito pequena a aprendizagem será muito lenta. A Fig. 5.2 representa os vários valores da taxa de aprendizagem que foram utilizados no teste da rede neuronal.

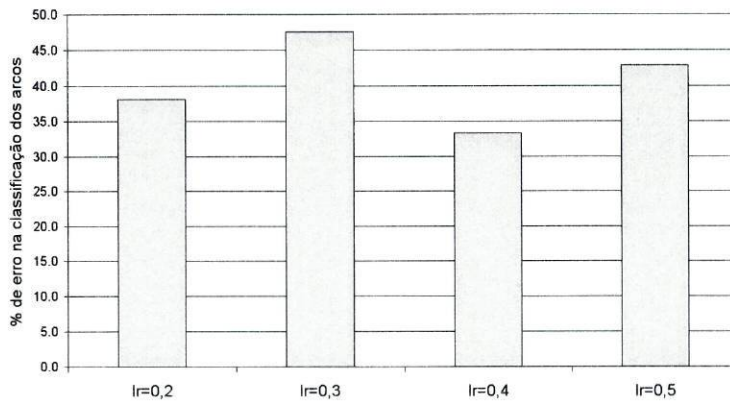


Fig. 5.2 - Percentagem de erro obtido na classificação em função da taxa de aprendizagem

Adicionando o termo momento ao algoritmo de Retropropagação a convergência do algoritmo torna-se mais rápida. O momento actua como um filtro 'passa-baixo' (*low-pass filter*) ignorando pequenos elementos na superfície de erro evitando assim que a função fique 'presa' num mínimo local. Na Fig. 5.3 encontram-se representadas as percentagens de erro obtidas com diferentes valores para o momento.

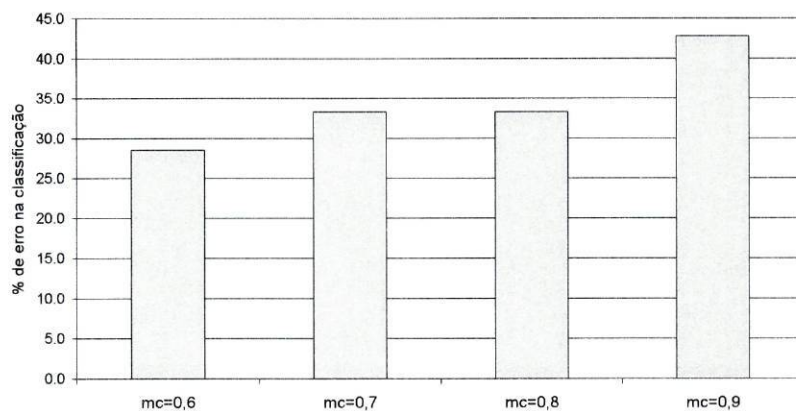


Fig. 5.3 - Percentagem de erro na classificação em função do termo momento

Como foi referido anteriormente, um dos métodos utilizados como critério de paragem foi o da validação cruzada. Durante o treino da rede com este método verificou-se que o mesmo não era o mais adequado para esta rede, uma vez que o desempenho da mesma convergia com um erro

médio quadrático elevado. Assim, após vários testes na tentativa de melhorar o desempenho da rede, optou-se por treinar a mesma com o método de regularização. Os resultados obtidos foram mais satisfatórios que os obtidos com o método anterior. A rede foi treinada com diversas funções de activação e com diferentes variantes do algoritmo de Retropropagação.

Um dos algoritmos utilizados permite a actualização da taxa de aprendizagem ao longo do treino da rede, através da definição de valores (mínimo e máximo) para o incremento da mesma. Após o cálculo do erro na saída, se o novo erro for superior ao erro calculado anteriormente por uma margem definida, os novos pesos e desvios sistemáticos não são considerados, fazendo com que a taxa de aprendizagem seja diminuída, do contrário, ou seja, se o novo erro for inferior ao anterior a taxa de aprendizagem é aumentada.

Os testes realizados com este último algoritmo resultaram num erro, relativamente ao número de pessoas que efectivamente passaram num dado arco, que varia entre os 15% e os 40%. Os erros obtidos devem-se essencialmente à pequena dimensão da amostra. Na Fig. 5.4 encontra-se demonstrado o desempenho da rede com a taxa de aprendizagem variável.

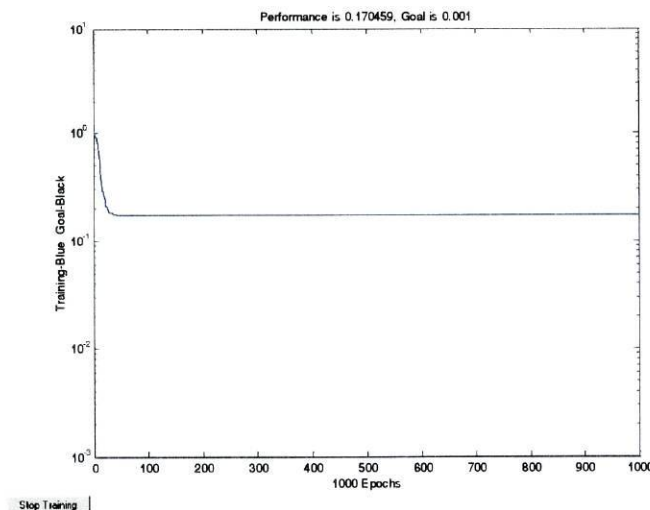


Fig. 5.4 - Desempenho da rede com a taxa de aprendizagem variável

Os exemplos apresentados a seguir dizem respeito a uma arquitectura composta por uma camada intermédia com 6 neurónios, uma taxa de aprendizagem de 0.4 e um termo momento de 0.7.

Como seria de esperar, a rede neuronal consegue aprender a 100% quando nenhum condutor passa pelo arco ou vice-versa. No entanto, ela não tem um comportamento tão favorável quando apenas dois ou três condutores passam pelo arco a ser treinado. Como se pode verificar pela Fig. 5.5, nos casos em que pelo menos dez condutores passam no arco a rede consegue classificar o arco (passou ou não passou no arco) com um sucesso superior a 55%.

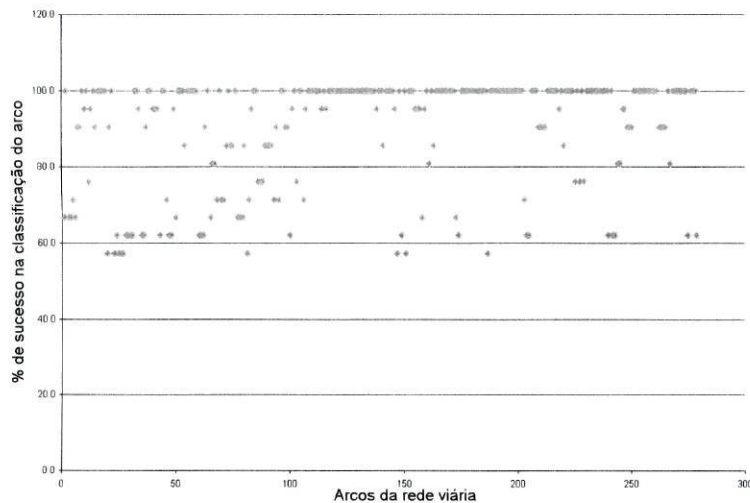


Fig. 5.5 - Análise da rede neuronal arco a arco

Analisando os resultados da rede neuronal, percurso a percurso, verifica-se que os erros obtidos são inferiores a 15%. Na figura seguinte encontram-se demonstrados dois exemplos de percursos (3 e 15), bem como o seu comportamento face ao algoritmo utilizado.

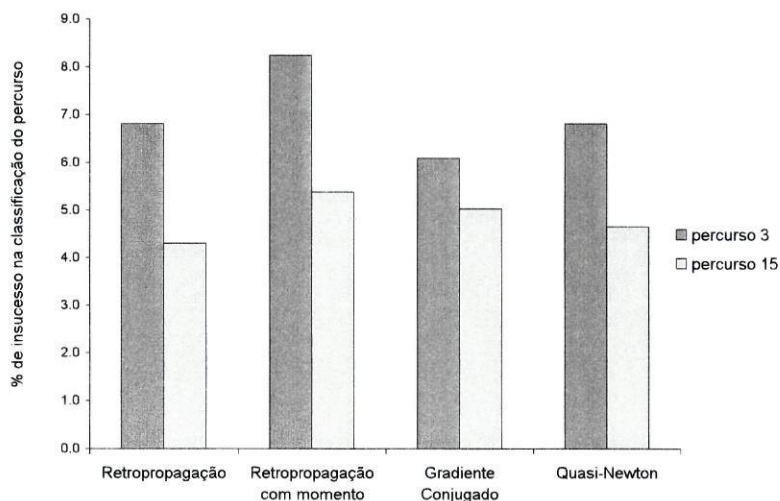


Fig. 5.6 - Percentagem de erro na classificação total dos arcos

Os erros analisados na Fig. 5.6 dizem respeito à classificação total dos arcos, isto é, se o arco foi ou não percorrido pelo condutor, no entanto, não indicam qual o erro na classificação dos percursos tendo em conta apenas os arcos que foram percorridos pelos condutores. Esta análise pode ser verificada na Fig. 5.7. Neste exemplo encontram-se demonstrados os percursos 4 e 10. O percurso 10 é classificado com um sucesso de 29% no algoritmo Quasi-Newton e 48% no algoritmo de Retropropagação. Já no percurso 4 a rede tem um comportamento mais positivo conseguindo um sucesso na classificação superior a 65% em ambos os algoritmos.

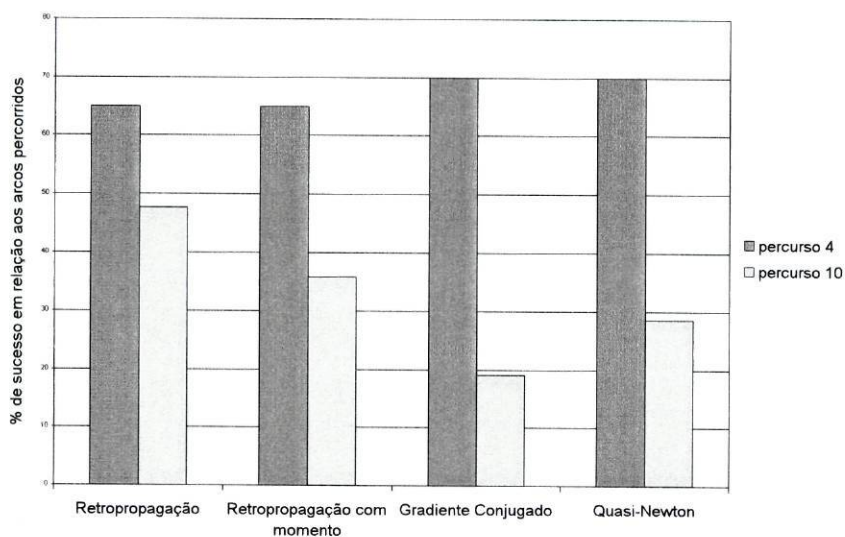


Fig. 5.7 - Percentagem de erro na classificação: percurso 4 e 10

Analisando a Fig. 5.8 compreende-se porque a classificação falhou no percurso 10 (a verde) e não no percurso 4 (a vermelho). Tal deve-se ao facto do percurso 10 começar por utilizar os arcos a sul e depois utilizar arcos mais a norte, não seguindo os padrões da maior parte dos condutores que quando decidem o trajecto mais a sul não passam pelos arcos a norte. O percurso 4 opta por percorrer arcos muito transitáveis.



Na Fig. 5.9 pode-se analisar os arcos que a rede neuronal atribui como percorridos pelos condutores. O percurso 4 é aquele a que a rede neuronal consegue atribuir um maior número de arcos percorridos, apesar de errar na classificação dos arcos a sul perto do Centro Comercial, uma vez que o condutor optou por percorrer os arcos mais a norte (Avenida Dr. Francisco Sá Carneiro). Através da observação da Fig. 5.9 não se consegue decifrar que arcos foram percorridos pelo condutor do percurso 10, apenas se pode dizer que o condutor entrou na área do Centro Comercial vindo de Norte (nó da A1).

O comportamento da rede neuronal ao longo do percurso 4 (realizado às 14.30h) foi muito estável, isto é, através da análise dos resultados da rede pode-se afirmar que o condutor partiu da origem em direcção ao destino passando pela Avenida Dr. Belchior Cardoso Costa, pela Rua António de Castro Corte Real, Rua Comendador Sá Couto e pela Rua Ferreira de Castro, ficando indefinido se o condutor optou pela Avenida Dr. Francisco Sá Carneiro ou Rua Domitília Carvalho.

Comparando os algoritmos de Retropropagação e Retropropagação com momento usando apenas arcos que foram percorridos pelo menos 10 vezes pelos condutores, verifica-se que o termo momento associado ao algoritmo de Retropropagação faz com que o mesmo tenha um comportamento superior em comparação à sua não utilização. Em alguns casos a diferença entre a utilização ou não, é superior a 25%. Verificou-se também que o algoritmo converge mais rapidamente utilizando o termo momento. Enquanto que a comparação entre os dois últimos métodos é quase imediata, o mesmo não se pode dizer em relação aos métodos Gradiente Conjugado e *Quasi-Newton*. Como se pode verificar pela Fig. 5.11 o comportamento entre estes dois métodos é muito semelhante, no entanto o método *Quasi-Newton*, na maior parte das vezes, é ligeiramente superior. Verifica-se que a utilização de um método de optimização numérica melhora em muito o desempenho da rede.

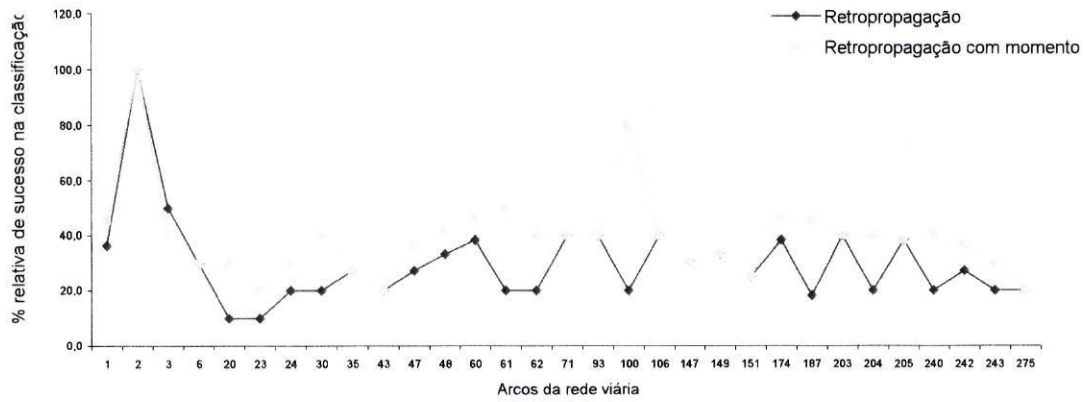


Fig. 5.10 - Comparação entre o algoritmo Retropropagação e Retropropagação com momento

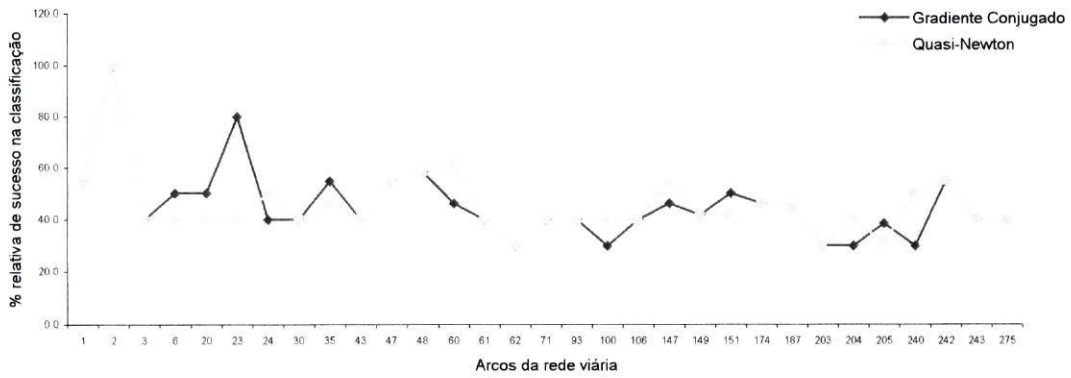


Fig. 5.11 - Comparação entre algoritmo Gradiente Conjugado e Quasi-Newton

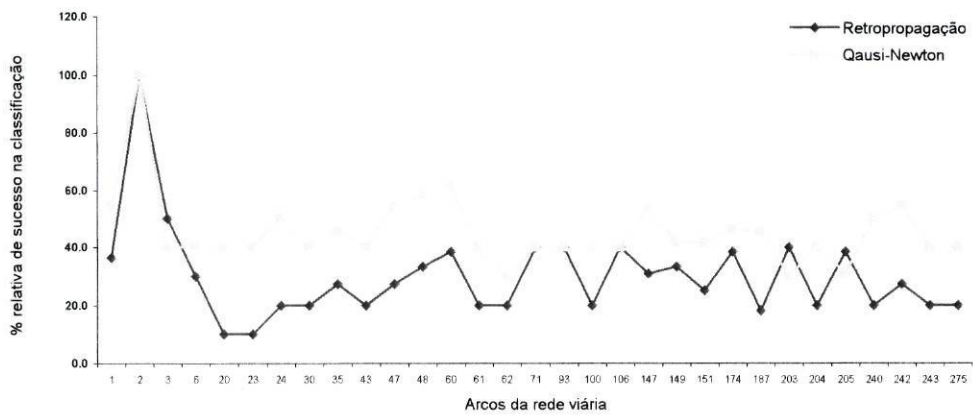


Fig. 5.12 - Comparação entre a utilização de um algoritmo com otimização numérica (Quasi-Newton) e outro sem otimização

Tendo em vista um aperfeiçoamento dos resultados, optou-se por realizar outra análise, criando-se uma nova rede neuronal. Esta segunda análise tem como objectivo o estudo dos arcos mais percorridos na cidade de Santa Maria da Feira, a diferentes horas do dia e com diferentes condições climatéricas. A rede neuronal criada para este estudo é composta por 6 vectores de entrada - hora do dia, estado do pavimento face às condições atmosféricas, tipo de pavimento, conflitos com peões, velocidades praticadas e número de entroncamentos de cada arco – e por 182 arcos da rede viária da cidade. O alvo da rede é a percentagem de pessoas que percorreram esse arco nas mesmas condições.

O treino da rede foi realizado com recurso a vários algoritmos, escolhendo-se para demonstração a variante do algoritmo de Retropropagação que elimina as pequenas variações dos pesos e desvios sistemáticos. Ao contrário da rede neuronal anterior, para este estudo utilizou-se dois tipos de processamento, o primeiro com uma rede estática composta por vectores de entrada concorrentes e o segundo com uma rede dinâmica processada em modo sequencial. Para o processamento em modo sequencial, dividiu-se a amostra em 4 grupos distintos: grupo1 – piso seco e hora de ponta; grupo 2 – piso seco e não hora de ponta; grupo 3 – piso molhado e hora de ponta; grupo 4 – piso molhado e não hora de ponta.

Analisando os dados recolhidos e os grupos criados, verifica-se que a amostra não contém uma distribuição homogénea, uma vez que o grupo 1 é composto por 21 percursos, o grupo 2 por 77, o grupo 3 por zero e o grupo 4 por 21. Desta forma, a análise da rede será realizada essencialmente com o grupo 2. Foi criada uma rede neuronal com uma arquitectura composta por 7 neurónios na camada intermédia e um na de saída. Os valores obtidos resultam numa previsão dos arcos mais percorridos com um erro inferior a 20 %, na maior parte dos arcos estudados.

A Fig. 5.14 representa os resultados obtidos no treino da rede e a Fig. 5.13 o erro obtido na classificação. Enquadrando estes erros na localização geográfica dos arcos, verifica-se uma tendência da rede para atribuir menor percentagem de pessoas nos arcos mais percorridos pelos condutores e, por outro lado, um excesso de pessoas nos arcos menos percorridos. Na Fig. 5.15 encontra-se demonstrado esta evidência.

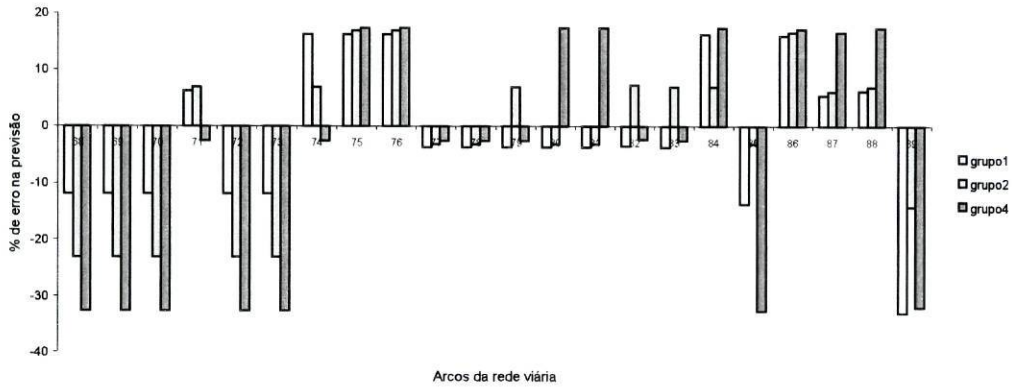


Fig. 5.13 - Erro obtido na previsão da percentagem de arcos percorridos

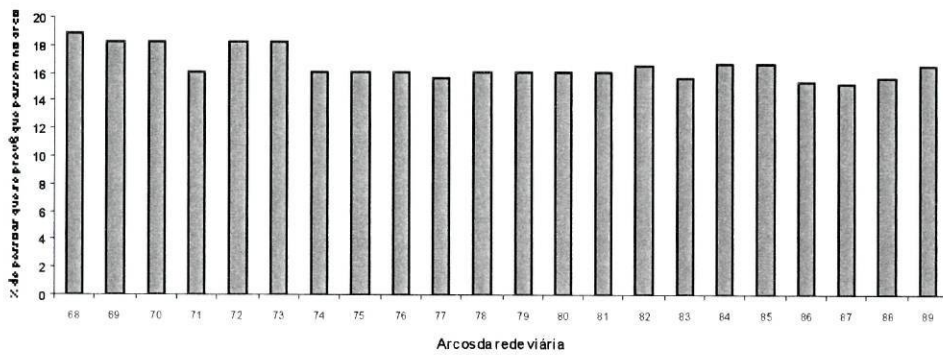


Fig. 5.14 - Previsão da rede neuronal para o grupo 2

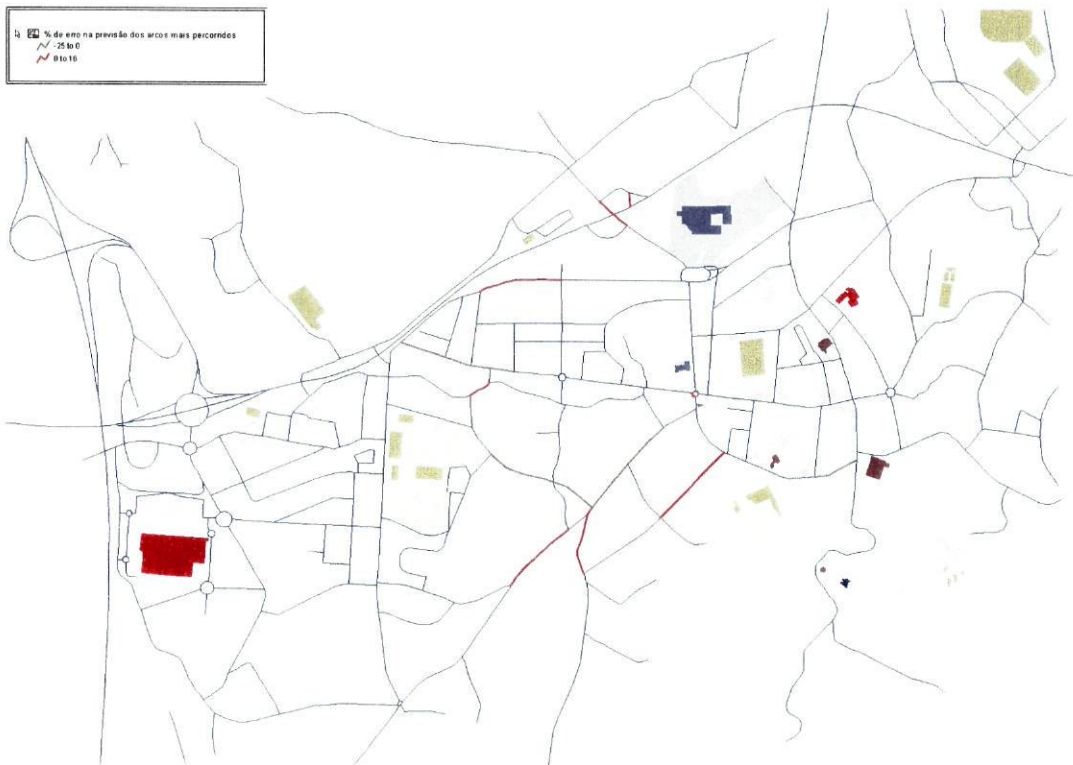


Fig. 5.15 - Percentagem de erro na previsão dos arcos mais percorridos (exemplo dos arcos 69 a 90)

### 5.3 ANÁLISE DOS PERCURSOS EFECTUADOS

A percepção do trânsito dada pelo condutor permite-nos saber quais os arcos que, naquela hora, foram considerados como muito congestionados. Na Fig. 5.16 encontram-se os arcos que foram percorridos pelos condutores números 8, 9 e 21, respectivamente, às 14.30h, 15.30h e 18.30h, os quais indicaram a percepção do trânsito como sendo elevada. Analisando a figura, verifica-se que os arcos por eles percorridos correspondem à Avenida 25 de Abril, Rua Ferreira de Castro e Rua Dr. Crispim Borges de Castro. A Fig. 5.17 pretende demonstrar quais os arcos mais percorridos entre as 14h e as 15.30h, verificando-se que coincidem com aqueles que têm um maior congestionamento de trânsito.

Na Fig. 5.18 apresentam-se os arcos mais percorridos entre as 10h e as 11h, verificando-se que os mesmos foram percorridos com pouco trânsito. Verifica-se, também, a tendência preferencial dos condutores para percorrerem a zona sul da cidade quando querem evitar congestionamentos ou quando não é hora de saída ou entrada de alunos nas Escolas EB 2/3 e Secundária. Assim, percorrendo a zona sul da cidade, os condutores chegam ao destino mais depressa.



Fig. 5.16 - Arcos congestionados: percepção do trânsito elevada



Fig. 5.17 - Arcos congestionados entre as 14h e as 15.30h



Fig. 5.18 - Arcos mais percorridos entre as 10h e as 11h (percepção do trânsito: pouco e médio)

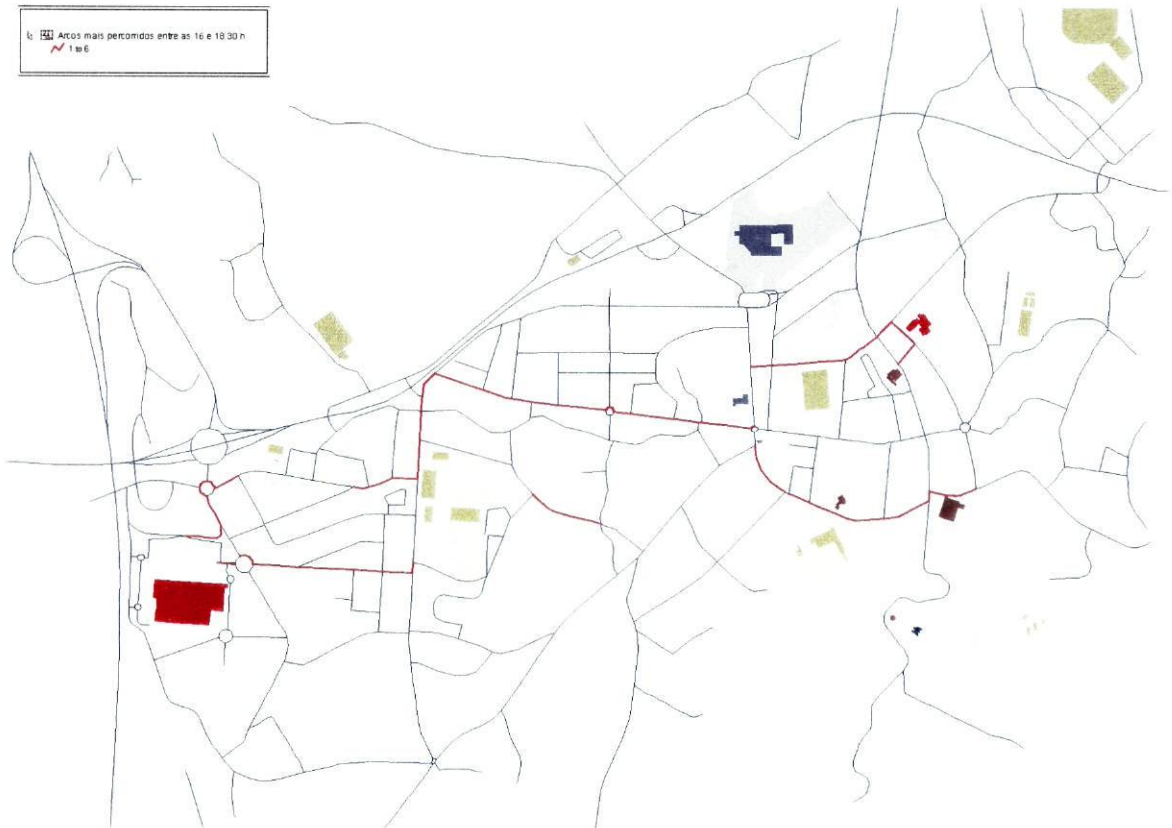


Fig. 5.19 - Arcos mais percorridos entre as 16h e as 18h30m

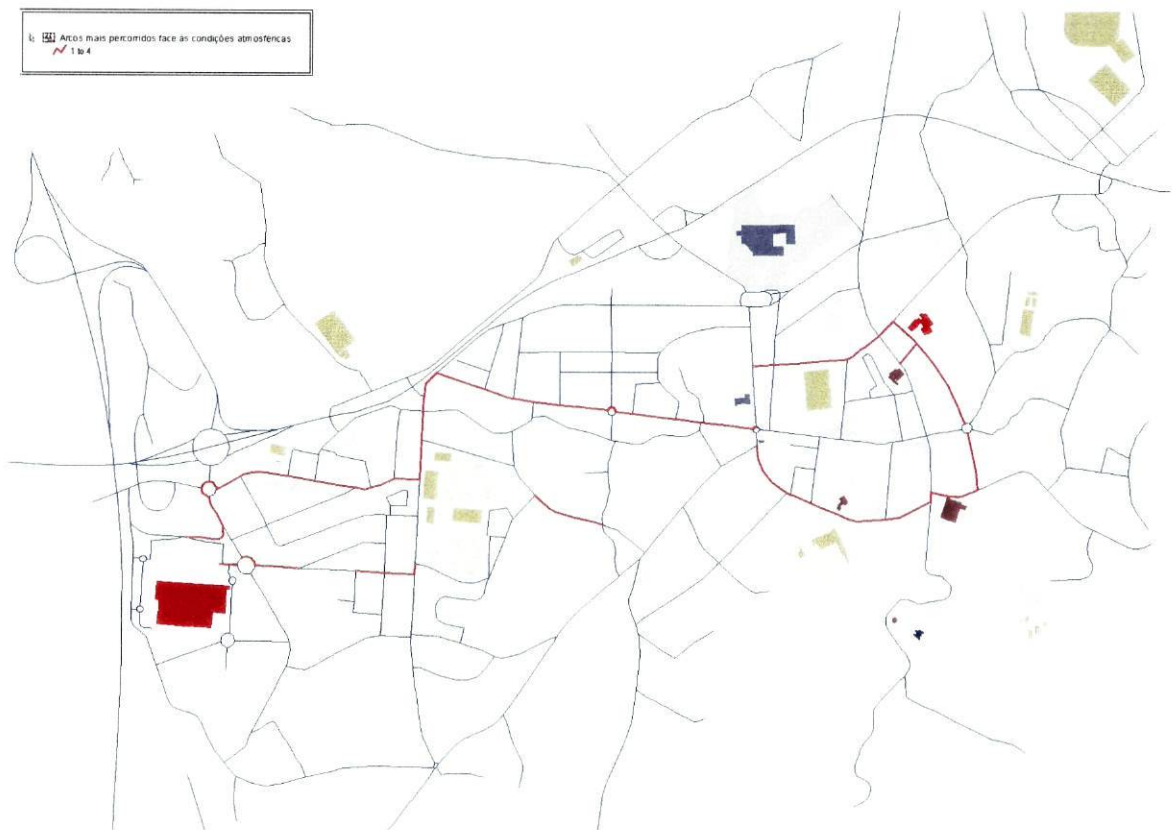


Fig. 5.20 - Arcos mais percorridos face às condições atmosféricas (chuva ou sol) e hora do dia (diurno ou noturno)



Fig. 5.21 - Arcos alternativos aos congestionados (percentagem de utilização inferior a 3)



Fig. 5.22 - Arcos percorridos em piso do tipo paralelo, em dias de chuva e entre as 16h e as 18h30m

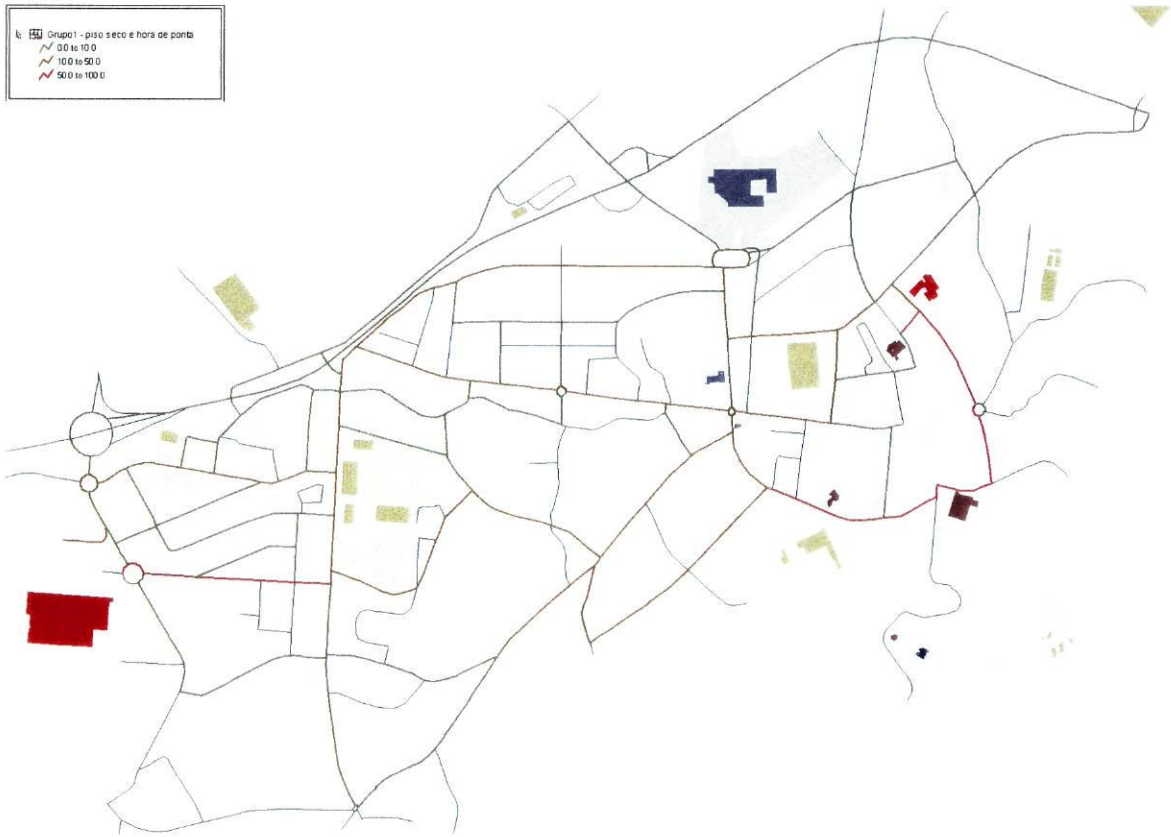


Fig. 5.23 - Percentagem de pessoas que percorreram o arco em dias de piso seco e a hora de ponta

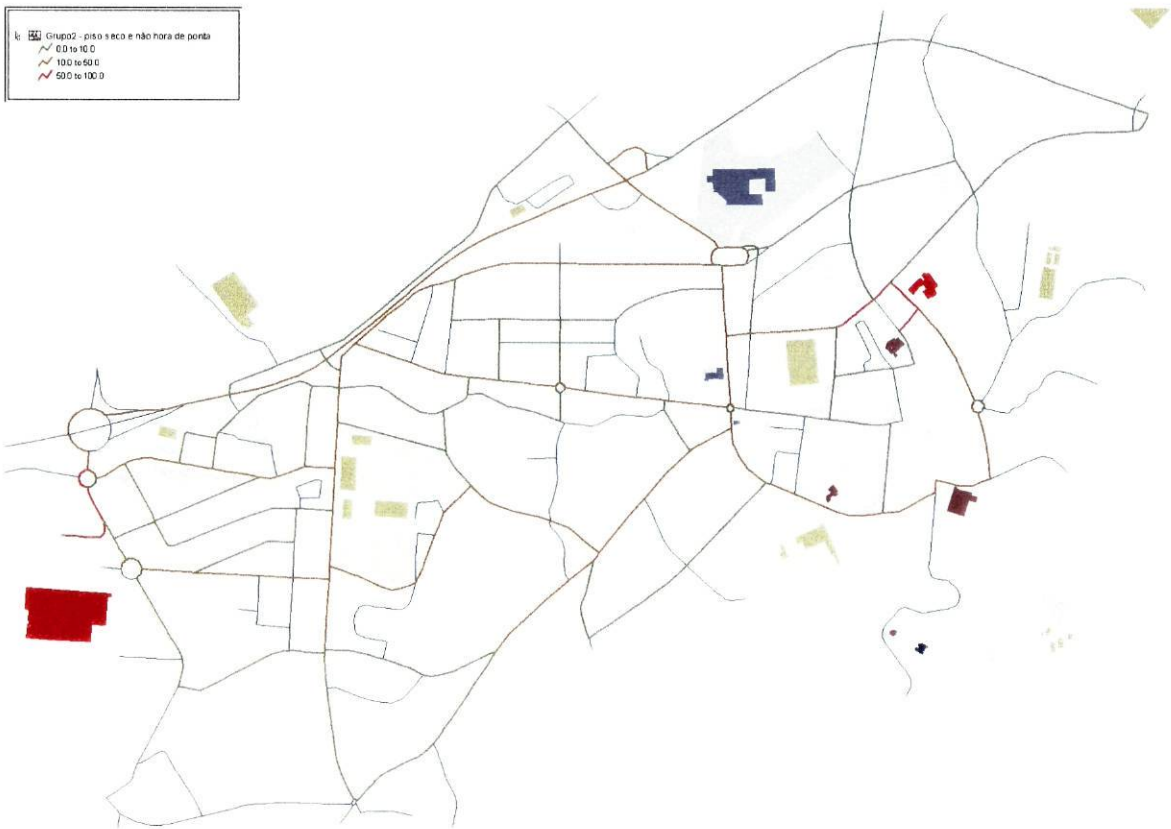


Fig. 5.24 - Percentagem de pessoas que percorreram os arcos em dias de piso seco e não a hora de ponta

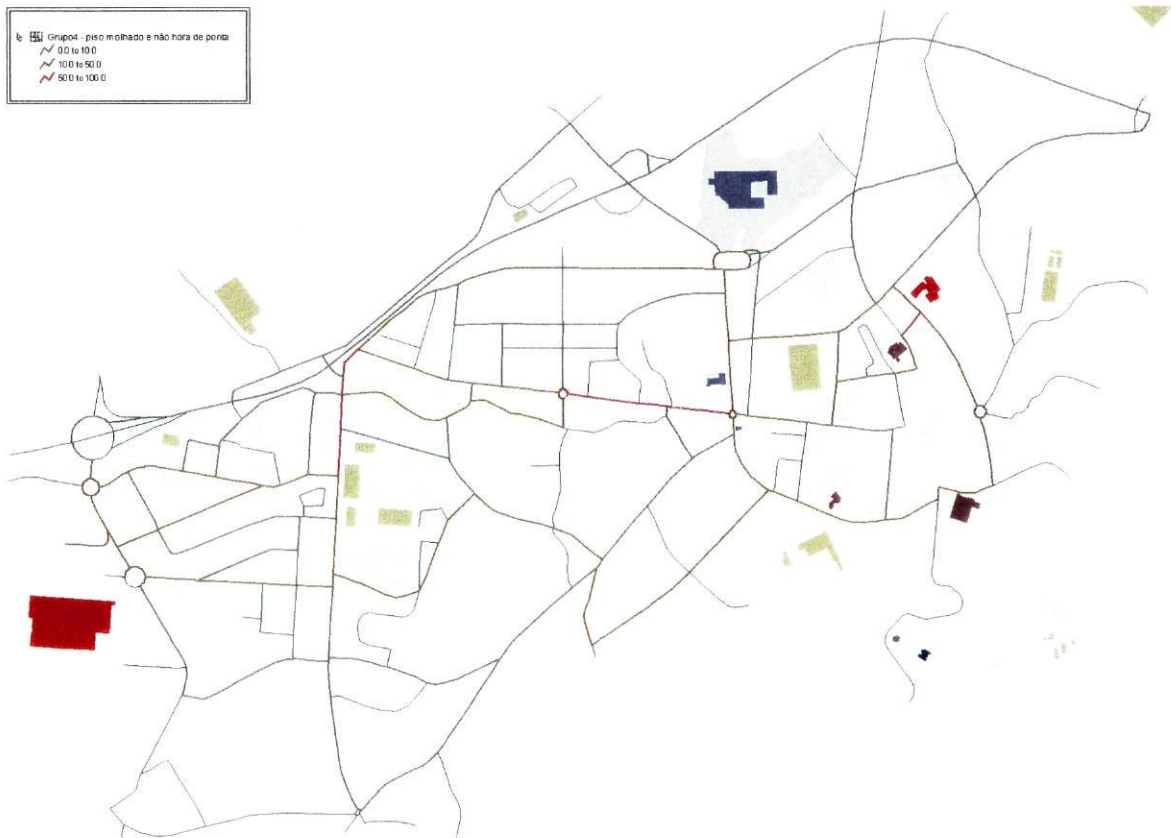


Fig. 5.25 - Percentagem de pessoas que percorreram os arcos em dias de piso molhado e não a hora de ponta

Verifica-se uma tendência dos condutores para:

1. Evitar a passagem em frente às Escolas durante as horas de entrada e saída de alunos;
2. Ao iniciar a marcha, seguir pelo trajecto que normalmente fazem ao ir para casa ou para o trabalho (por exemplo: percorrer a EN 223, Avenida Dr. Belchior Cardoso Costa);
3. Preferir a passagem na Avenida Comendador Sá Couto e na Avenida Dr. Francisco Sá Carneiro (local onde existem muitos bares e restaurantes) uma vez que a deslocação para tal rua é frequente;
4. Após vários percursos, alguns condutores escolhem trajectos alternativos aos anteriormente realizados;
5. Os condutores, ao depararem com obstáculos na via, recorrem a trajectos alternativos; da Fig. 5.21 verifica-se que os arcos assinalados a vermelho foram percorridos apenas em 3% dos percursos realizados pelos condutores;
6. Através do conhecimento das características da rede viária pode-se concluir que os condutores não foram muito influenciados por características, como, estado de

conservação do pavimento, tipo e largura das vias, uma vez que a maior parte das mesmas se encontrava em bom estado (as ruas foram pavimentadas dois meses antes do levantamento de campo para este trabalho).

Os arcos considerados como congestionados devem ser enquadrados na média dimensão da cidade de Sta. Maria da Feira, sendo esses congestionamentos pouco elevados quando comparados com os de grandes metrópoles como Porto ou Lisboa.

## **6. CONCLUSÕES E LINHAS DE ORIENTAÇÃO FUTURA**

Neste trabalho foram treinados vários tipos de redes neuronais com o objectivo de prever o percurso que a maioria dos condutores fará, entre dois pontos distintos de uma cidade, dando como parâmetros de entrada algumas características da rede viária e do dia, tais como, a percepção do trânsito (dada pelo condutor), o estado de conservação do pavimento, as velocidades praticadas em cada arco, a hora do dia e o estado de conservação do pavimento face às condições atmosféricas. Para isso, foi necessário começar por actualizar a cartografia digital existente, recorrendo ao GPS e à técnica DGPS com pós-processamento. Para isso, foram utilizados dois receptores de uma frequência, que permitem medir códigos e fases. O receptor Trimble Pro XR foi colocado na estação de referência e o GeoExplorer III foi utilizado como receptor móvel.

As características de linha (eixos de via) foram levantadas com medidas de código, enquanto que as características do tipo ponto (paragens de autocarro, sinalização, semáforos, entre outras) foram levantadas com medidas de fase. No processamento dos dados obteve-se um erro que oscila entre 0.8 e 1.5 metros. O GPS revelou-se uma ferramenta indispensável na realização deste trabalho, desde a recolha de dados de forma a completar a informação já existente, até à definição dos percursos realizados pelos condutores, passando pela aquisição de informação para a base de dados do SIG.

Os dados cartográficos levantados foram introduzidos no Sistema de Informação Geográfica, o qual teve um papel fundamental na união entre os dados levantados com o GPS e os dados obtidos com a rede neuronal. O SIG disponível para a realização deste trabalho encontrava-se à escala 1/1.000, o que implica uma tolerância que ronda os 20 cm em planimetria e 40 cm em altimetria, donde se conclui que os erros obtidos na actualização da cartografia não são os mais adequados para um SIG a esta escala. Tendo em atenção que o receptor GeoExplorer III foi utilizado como receptor móvel devido ao seu interface amigável com o utilizador e devido ao facto de, em Fevereiro de 2002, não existir ainda uma rede DGPS portuguesa, hoje em dia tal não se justifica. Como o receptor Pro XR tem capacidade para captar sinais RTCM (*Radio Technical Commission for Marine Services*) a utilização deste receptor, para a realização dos levantamentos, seria o ideal (no contexto de um receptor de uma só frequência). Neste caso, o erro esperado rondaria os 30cm, sendo, portanto, ajustada a sua utilização na realização deste tipo de levantamentos, uma vez que facilitaria em muito a tarefa de tratamento dos dados e conseqüente actualização do SIG.

A técnica DGPS em tempo-real (RTK - *Real Time Kinematics*) poderia ser utilizada para a realização dos levantamentos. Neste caso os resultados seriam bastante precisos com erros abaixo dos 10cm. Uma experiência interessante seria comparar as técnicas RTCM, RTK com as mais recentes VRS [*Virtual Reference Station*] e FKP [*Flächenkorrekturparameter*]. Estas duas últimas técnicas baseiam-se no conceito de rede RTK. Enquanto que a técnica VRS se baseia na densificação da rede de estações de referência, através da modelação e criação de dados de estações que simulem uma posição de referência próxima da posição móvel, a técnica FKP baseia-se no cálculo de parâmetros de correcção globais, para uma dada área abrangida por uma rede de estações de referência, cabendo a cada receptor móvel o cálculo dos seus parâmetros de correcção local.

Para criar os vectores de entrada foram realizados percursos por pessoas escolhidas dentro de determinados estratos sociais, com diferentes idades e com um bom conhecimento sobre a rede viária da cidade de Santa Maria da Feira. Procurou-se que as pessoas repetissem o percurso a diferentes horas do dia e com diferentes condições atmosféricas. Parte destes percursos foram utilizados para treinar a rede e outra parte para validação e teste da mesma. Os resultados obtidos, usando diferentes tipos de redes neuronais, foram analisados em pormenor, como ficou patente no capítulo anterior, dos quais se pode concluir que a rede criada consegue prever quais os arcos percorridos pela maioria dos condutores, face às condições iniciais, com uma percentagem de erro que ronda os 20%. Considera-se este valor um pouco elevado, mas, apesar do grande número de testes realizados, variando os métodos de treino da rede e os diferentes parâmetros que a caracterizam, verificou-se ser impossível baixá-lo. Esta incapacidade de melhorar o desempenho da rede deve-se a circunstâncias diversas, nomeadamente, a pequena dimensão da amostra de condutores, o número de repetições da experiência ser também pequeno, mesmo unitário em alguns casos e, finalmente, o curto período em que os receptores GPS estiveram disponíveis, obrigando a que todos os dados tivessem de ser recolhidos nos meses de Janeiro e Fevereiro. Infelizmente, neste período, como já foi indicado, as condições atmosféricas foram geralmente muito semelhantes e pouco propícias a situações críticas para o trânsito automóvel.

Para que uma rede neuronal efectue todas as tarefas para que foi desenvolvida é necessário ter um conjunto vasto de elementos de treino que devem ser actualizados periodicamente. Torna-se fundamental otimizar a recolha da informação a carregar na rede de forma que a mesma seja introduzida sem grandes tarefas intermédias. Para tal, devem ser realizados estudos que visem analisar o grau de desactualização dos dados, de maneira a estipular o intervalo de monitorização da rede. No caso concreto deste trabalho, considera-se que o intervalo de tempo para uma nova recolha de dados deverá ser anual, uma vez que os dados recolhidos no campo (após um ano desde a primeira aquisição de informação) não sofreram uma mudança substancial. Considera-se pertinente desenvolver métodos de aquisição de informação adicional à rede (por exemplo do tipo velocidades médias nos arcos, fluxos nos arcos, saturação da via, tempos de viagem entre dois pontos), através do recurso a câmaras vídeo, colocadas em cruzamentos estratégicos, do uso de sensores, de viaturas equipadas com GPS e outros sensores que permitam a recolha de dados referentes a acelerações e velocidades.

Considera-se também importante aproveitar o facto da Câmara Municipal ter instalado nas suas viaturas equipamentos GPS e utilizar os motoristas camarários como condutores que contribuam para a recolha de informação sobre a rede viária e sobre as preferências de percursos de cada um. O sistema de gestão de frotas de veículos da Câmara Municipal poderá ser melhorado acrescentando um módulo de previsão de fluxos de tráfego e outro referente à desactualização do estado de conservação do pavimento.

Este estudo poderá, ainda, ser melhorado com o recurso à optimização de percursos com o fim de aconselhar os condutores sobre o melhor trajecto a efectuar dependendo dos objectivos iniciais. As aplicações para este tipo de estudos são diversas e abrangem várias áreas de intervenção como por exemplo:

- Protecção civil – guiar os bombeiros desde o quartel até aos locais de incêndio, evitando ruas congestionadas e percursos longos e em más condições;
- Medicina – da mesma forma, guiar ambulâncias desde os locais de sinistro até ao hospital ou unidade de saúde mais próxima;

- Frotas de veículos – dada uma origem e um destino, descobrir qual o trajecto mais rápido para lá chegar, tendo em atenção as características da rede viária e da população.

Estas áreas de intervenção melhorarão a qualidade e os serviços, uma vez que conseguem atingir os seus objectivos com mais rapidez e eficácia. Mais ainda, numa Câmara Municipal, as redes neuronais poderão ser utilizadas noutro tipo de aplicações (para além da área dos transportes) tais como:

- Na previsão da expansão urbana (necessária para os Planos Directores Municipais);
- Na previsão de consumos de água ao longo de um ano (por exemplo: no Verão e no Inverno);
- Na previsão de consumos de energia nas vias públicas.

## **7. REFERÊNCIAS BIBLIOGRÁFICAS E BIBLIOGRAFIA**

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Chester, M. – *Neural Networks. A Tutorial*. Prentice-Hall. USA. 1993.
- [2] Dougherty, M. – *A review of neural networks applied to transport*. Transportation Research C Vol. 3, Nº 4. 1995, pg. 247-260.
- [3] Eberhart, R. C. and Dobbins, R. W. – *Neural Network. PC Tools. A Pratical Guide*. Academic Press, Inc. California, 1990.
- [4] Fausett, L. – *Fundamentals of Neural Networks. Architectures, Algorithms, and Applications*. Prentice-Hall. USA. 1994.
- [5] Hagan, M. T. et al. – *Neural Network Design*. PWS Publishing Co. 1996.
- [6] Haykin, S. – *Neural Networks. A Comprehensive Foundation*. 2ª Edition. Prentice Hall.1999.
- [7] Hofmann-Wellenhof, B. et al. – *Global Positioning System. Theory and Praticce* . 3ª Edition; Springer-Verlag Wien, 1994.
- [8] Kartalopoulos, S. V. – *Understanding Neural Networks and Fuzzy Logic: Basic Concepts and Applications*. IEEE Press. 1996.
- [9] Lin, C. and Lee, C. S. G. – *Neural Fuzzy Systems. A Neuro-Fuzzy Synergism to Intelligent Systems*. Prentice Hall. 1996.
- [10] Marques, J. S. – *Reconhecimento de padrões. Métodos estatísticos e neuronais*. IST Press. Coleção Ensino da Ciência e da Tecnologia. Setembro 1999.
- [11] Neves, J. J. S. E. – *Partilha dinâmica de recursos em redes de comunicações*. Dissertação de Doutoramento orientada pelos Professores Mário Jorge Moreira Leitão e Luís Borges de Almeida e apresentada na Faculdade de Engenharia da Universidade do Porto em Fevereiro de 1996.
- [12] Principe, J. C. et al. – *Neural and Adaptive Systems: Fundamentals through Simulations*. John Wiley & Sons, Inc. 2000.
- [13] Rojas, R. – *Neural Networks. A Systematic Introduction*. Springer. 1996.
- [14] Sá, J. P. M. – *Pattern Recognition. Concepts, methods and applications*. Springer. 2001.
- [15] Seeber, G. – *Satellite Geodesy: Foundations, Methods and Applications*. Walter de Gruyter Berlin, 1993.
- [16] Selecções do Reader's Digest – *ABC do Corpo Humano*. 1ª Edição. Outubro 1987.

**BIBLIOGRAFIA**

- [1] Adler, J. L. and Blue, V. J. – *A cooperative multi-agent transportation management and route guidance system*. Transportation Research C Vol. 10. 2002, pg. 433-454.
- [2] Adrados, C. et al. – *Global Positioning System (GPS) location accuracy improvement due to Selective Availability removal*. C. R. Biologies Vol. 325. 2002, pg. 165-170.
- [3] Al-Nuaimy, W. et al. – *Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition*. Journal of Applied Geophysics Vol. 43. 2000, pg. 157-165.
- [4] Arentze, T. A. and Timmermans, H. J. P. – *A spatial decision support system for retail plan generation and impact assessment*. Transportation Research C Vol. 8. 2000, pg. 361-380.
- [5] Bell, T. – *Automatic tractor guidance using carrier-phase differential GPS*. Computers and Electronics in Agriculture Vol. 25. 2000, pg. 53-66.
- [6] Berberidis, K. and Palicot, J. – *A frequency domain quasi-Newton algorithm*. Signal Processing Vol. 47. 1995, pg. 235-238.
- [7] Bhattacharjee, D. et al. – *Modeling the effects of traveler information on freeway origin-destination demand prediction*. Transportation Research C Vol. 9. 2001, pg. 381-398.
- [8] Bodri, B. – *A neural-network model for earthquake occurrence*. Journal of Geodynamics Vol. 32. 2001, pg. 289-310.
- [9] Cascetta, Ennio et al. – *A model of route perception in urban road networks*. Transportation Research B Vol. 36. 2002, pg. 577-592.
- [10] Chang, D. and Islam, S. – *Estimation of soil physical properties using remote sensing and artificial neural network*. Remote Sensing of Environment. 2000, pg. 534-544.
- [11] Chang, G. and Su, C. – *Predicting intersection queue with neural network models*. Transportation Research C Vol. 3, N° 3. 1995, pg. 175-191.
- [12] Chen, H. and Bell, M. – *Instrumented city database analysts using multi-agents*. Transportation Research C Vol. 10. 2002, pg. 419-432.
- [13] Chisholm, G. – *Virtual Reference Station (VRS) Systems for Improved Efficiencies in Port Operations*. Trimble. 2002.

- [14] Choi, E. and Cicci, D. A. – *Analysis of GPS positioning problems*. Applied Mathematics and Computation Vol. 140. 2003, pg. 37-51.
- [15] Cocard, M. et al. – *Airborne laser altimetry in the Ionian Sea, Greece*. Global and Planetary Change Vol. 34. 2002, pg. 87-96.
- [16] Dagli, C. H. – *Artificial Neural Networks for Intelligent Manufacturing*. Chapman & Hall. London. 1994.
- [17] Dia, H. – *An agent based approach to modelling driver route choice behaviour under the influence of real-time information*. Transportation Research C. 2002, pg. 331-349.
- [18] Emmerink, R. H. M. et al. – *The potencial of information provision in a simulated road transport network with non-recurrent congestion*. Transportation Research C Vol. 3, Nº 5. 1995, pg. 293-309.
- [19] Ericsson, E. – *Variability in urban driving patterns*. Transportation Research D Vol. 5. 2000, pg. 337-354.
- [20] Figueiredo, L. et al. – *Sistemas Inteligentes de Transportes*. Ingenium, 2ª Série, Nº 71. Nov/Dez. 2002, pg. 84-92.
- [21] Fitzgerald, E. and Bean, C. J. – *Sub-basalt imaging problems and the application of artificial neural networks*. Journal of Applied Geophysics Vol. 48. 2001, pg. 183-197.
- [22] Gardner, M. W. and Dorlinh, S. R. – *Artificial neural networks (the multilayer perceptron) – a review of applications In the atmospheric sciences*. Atmospheric Environment Vol. Vol. 32, Nº 14/15. 1998, pg. 2627-2636.
- [23] Hickman, M. D. et al. – *Passenger travel time and path choice implications of real-time transit information*. Transportation Research C Vol. 3, Nº 4. 1995, pg. 211-226.
- [24] Imrie, C. E. et al. – *River flow prediction using artificial neural networks: generalisation beyond the calibration range*. Journal of Hydrology Vol. 233. 2000, pg. 138-153.
- [25] Jauffred, F. and Bernstein, D. – *An alternative formulation of the simultaneous route and departure-time choice equilibrium problem*. Transportation Research Vol. 4, Nº 6. 1996, pg. 339-357.
- [26] Kim, J. and Oh, J. – *A land vehicle tracking algorithm using stand-alone GPS*. Control Engennering Prattice Vol. 8. 2000, pg. 1189-1196.

- [27] Kondo, Y. et al. - *Intelligent mobility enhancement system*. JSAE Review Vol. 21, 2000, pg. 231-234.
- [28] Koutsopoulos, H. et al. - *Travel simulators for data collection on driver behavior in the presence of information*. Transportation Research C Vol. 3, Nº 3. 1995, pg. 143-159.
- [29] Kropp, J. - *A neural network approach to the analysis of city systems*. Applied Geography Vol. 18, Nº 1. 1998, pg. 83-96.
- [30] Kwan, M. - *Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set*. Transportation Research C Vol. 8. 2000, pg. 185-203.
- [31] Lechner, W. and Baumann, S. - *Global Navigation Satellite Systems*. Computers and Electronics in Agriculture Vol. 25. 2000, pg. 67-85.
- [32] Lees, B. G. - *Neural networks applications in the geosciences: an introduction*. Computers & Geosciences Vol. 22, Nº 9. 1996, pg. 955-957.
- [33] Lundberg, J. B. - *Alternative algorithms for the GPS static positioning solution*. Applied Mathematics and Computation Vol. 119. 2001, pg. 21-34.
- [34] Mantri, S. and Bullock, D. - *Analysis of feedforward-backpropagation neural networks used in vehicle detection*. Transportation Research C Vol.3, Nº 3. 1995, pg. 161-174.
- [35] Martinelli, D. R. and Teng, H. - *Optimization of railway operators using neural networks*. Transportation Research C Vol. 4, Nº 1. 1996, pg. 33-49.
- [36] Martinez, J. L. et al. - *A new method for generating differential GPS corrections*. Control Engineering Practice Vol. 8. 2000, pg. 253-258.
- [37] McNoleg, O. - *The integration of GIS, Remote Sensing, Expert Systems and Adaptive Co-Kriging for the environmental habitat modelling of the highland haggis using object-oriented, fuzzy logic and neural network techniques*. Computers & Geosciences Vol. 22, Nº 5. 1996, pg. 585-588.
- [38] Meneses, M. A. S. - *Implementação do Sistema Municipal de Informação Geográfica em Santa Maria da Feira*. Ordem dos Engenheiros. Porto. Abril 2002.
- [39] Meneses, M. A. S. - *Sistema Municipal de Informação Geográfica*. Comissão de Coordenação da Região Norte. Porto. Maio 2002.

- [40] Miller, H. J. and Storm, J. D. – *Geographic information system design for network equilibrium-based travel demand models*. Transportation Research C Vol. 4, Nº 6. 1996, pg. 373-389.
- [41] Mohaghegh, S. et al. – *Design optimum frac jobs using virtual intelligence techniques*. Computers & Geosciences Vol. 26. 2000, pg. 927-939.
- [42] Monteiro, S. – *A rede DGPS Portuguesa*. Revista da Armada. Dezembro 2002.
- [43] Montenbruck, O. et al. – *A real-time Kinematic GPS sensor for spacecraft relative navigation*. Aerospace Science and Technology Vol. 6. 2002, pg. 435-449.
- [44] Morshed, J. and Kaluarachchi, J. J. – *Application of artificial neural network and genetic algorithm in flow and transport simulation*. Advances in Water Resources Vol. 22, Nº2. 1998, pg. 145-158.
- [45] Nijkamp, P. et al. – *Modelling inter-urban transport flows in Italy: a comparison between neural network analysis and logit analysis*. Transportation Research C Vol. 4, Nº 6. 1996, pg. 323-338.
- [46] Ochieng, W.Y. and Sauer, K. - *Urban road transport navigation: performance of global positioning system after selective availability* . Transportation Research Part C Vol. 10. 2002, pg. 171-187.
- [47] Omae, M. et al. – *Relative position measurement of neighboring vehicles using DGPS and inter-vehicle communication*. JSAE Review Vol. 22. 2001, pg. 75-80.
- [48] Peyret, F. et al. – *High-precision application of GPS in the field of real-time equipment positioning*. Automation in Construction Vol. 9. 2000, pg. 299-314.
- [49] Podnar, D. et al. – *Application of artificial neural networks to modelling the transport and dispersion of tracers in complex terrain*. Atmospheric Environment Vol. 36. 2002, pg. 561-570.
- [50] Sá, J. P. M. et al. – *Summer School in Neural Networks*. ISEP, Porto. Julho 2002.
- [51] Santos, L. – *Redes GPS de Estações de Referência – Soluções RTK*. Apresentação na Faculdade de Ciências da Universidade de Lisboa. Novembro 2002.
- [52] Shelley, B. and Stephenson, S. – *The use of artificial neural networks in complex stimulation and design*. Computers & Geosciences Vol. 26. 2000, pg. 941-951.
- [53] Sheu, J. and Ritchie, S. G. – *A new methodology for incident detection and characterization on surface streets*. Transportation Research C Vol. 6. 1998, pg. 315-335.

- [54] Shmuell, D. et al. – *Neural Network analysis of travel behaviour: evaluating tools for prediction*. Transportation Research C Vol. 4, Nº 3. 1996, pg. 151-166.
- [55] Tatem, A. J. et al. – *Super-resolution land cover pattern prediction using a Hopfield neural network*. Remote Sensing of Environment Vol. 79. 2002, pg. 1-14.
- [56] Taylor, M. A. P. et al. – *Integration of the global positioning system and geographical information systems for traffic congestion studies*. Transportation Research C Vol. 8. 2000, pg. 257-285.
- [57] Thill, J.-C. – *Geographic information systems for transportation in perspective*. Transportation Research C. 2000, pg. 3-12.
- [58] Tsao, H.-S. J. – *Traffic control for automated highway systems: a conceptual framework*. Transportation Research C Vol. 3, Nº 4. 1995, pg. 227-246.
- [59] Wahle, J. et al. – *The impact of real-time information in a two-route scenario using agent-based simulation*. Transportation Research C Vol. 10. 2002, pg. 399-417.
- [60] Wilson, J. N. – *Guidance of agricultural vehicles – a historical perspective*. Computers and Electronics in Agriculture Vol. 25. 2000, pg. 3-9.
- [61] Yang, Y. and Rosenbaum, M. S. – *Artificial neural networks linked to GIS for determining sedimentology in harbours*. Journal of Petroleum Science and Engineering Vol. 29. 2001, pg. 213-220.
- [62] Yin, H. et al. – *Urban traffic flow prediction using a fuzzy-neural approach*. Transportation Research C Vol. 10. 2002, pg. 85-98.
- [63] Yoon, S. and Lundberg, J. B. – *An integer ambiguity resolution algorithm for real-time GPS attitude determination*. Applied Mathematics and Computation Vol. 129. 2002, pg. 21-41.
- [64] Zarazaga-Soria, F. J. et al. – *Examples of vehicle location systems using CORBA – based distributed real-time GPS data and services*. Computers, Environment and Urban Systems Vol. 25. 2001, pg. 293-305.
- [65] Zito, R. et al. – *Global Positioning Systems in the time domain: how useful a tool for intelligent vehicle-highway systems?* Transportation Research C Vol. 3, Nº 4. 1995, pg. 193-209.

**APÉNDICES**

## APÊNDICE 1 – TEOREMA DE CONVERGÊNCIA DO PERCEPTRÃO

O algoritmo de classificação de duas classes pode ser resumido da seguinte forma: considere-se  $P_1$  e  $P_2$  dois subconjuntos do conjunto de treino  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ , pertencentes à classe 1 e à classe 2, respectivamente. A união destes dois subconjuntos representa o conjunto de treino. Considere-se os vectores  $\mathbf{p}_1$  e  $\mathbf{p}_2$  que se pretende classificar em duas classes linearmente separáveis, então, existe um vector de pesos em que:

$$\mathbf{W}^T \mathbf{x} > 0, \text{ para cada vector } \mathbf{x} \text{ que pertença à classe 1}$$

$$\mathbf{W}^T \mathbf{x} \leq 0, \text{ para cada vector } \mathbf{x} \text{ que pertença à classe 2}$$

Escolheu-se, arbitrariamente, que se o vector  $\mathbf{x}$  pertence à classe 2, então,  $\mathbf{W}^T \mathbf{x} = 0$ . O problema resume-se a encontrar um vector de pesos que garanta que a equação anterior é satisfeita. Assim, se um vector  $\mathbf{x}(n)$  é classificado correctamente pelo vector de pesos  $\mathbf{W}(n)$ , então, não é aplicada nenhuma correcção ao vector de pesos do perceptrão:

$$\mathbf{W}(n+1) = \mathbf{W}(n) \text{ se } \mathbf{W}^T \mathbf{x}(n) > 0 \text{ e } \mathbf{x}(n) \text{ pertence à classe 1}$$

$$\mathbf{W}(n+1) = \mathbf{W}(n) \text{ se } \mathbf{W}^T \mathbf{x}(n) \leq 0 \text{ e } \mathbf{x}(n) \text{ pertence à classe 2}$$

Doutro modo, o vector de pesos é actualizado da seguinte forma:

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \eta(n)\mathbf{x}(n) \text{ se } \mathbf{W}^T \mathbf{x}(n) > 0 \text{ e } \mathbf{x}(n) \text{ pertence à classe 2}$$

$$\mathbf{W}(n+1) = \mathbf{W}(n) + \eta(n)\mathbf{x}(n) \text{ se } \mathbf{W}^T \mathbf{x}(n) \leq 0 \text{ e } \mathbf{x}(n) \text{ pertence à classe 1}$$

em que  $\eta(n)$  é o parâmetro independente que controla o ajuste dos pesos a aplicar na iteração  $n$ .

Se  $\eta(n) = \eta > 0$  está-se perante um caso em que a adaptação a aplicar ao perceptrão é fixa [6].

## APÊNDICE 2 – MÉTODO DO GRADIENTE DECRESCENTE

Para uma taxa de aprendizagem constante  $\eta_n = \eta$  o valor máximo de  $\eta$  está directamente ligado à curvatura da função quadrática.

Supondo que se tem uma função quadrática  $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$  sendo o gradiente da função dado por  $\nabla f(\mathbf{x}) = \mathbf{A} \mathbf{x} + \mathbf{b}$ . Acrescentando à expressão o algoritmo do gradiente decrescente para uma taxa de aprendizagem constante obtém-se,

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{x}_n - \eta \mathbf{g}_n \\ &= \mathbf{x}_n - \eta (\mathbf{A} \mathbf{x}_n + \mathbf{b}) \\ &= [\mathbf{I} - \eta \mathbf{A}] \mathbf{x}_n - \eta \mathbf{b} \end{aligned}$$

Trata-se de um sistema linear dinâmico que será estável se os valores próprios da matriz  $[\mathbf{I} - \eta \mathbf{A}]$  forem menores que um em magnitude. Considere-se  $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$  e  $\{z_1, z_2, \dots, z_m\}$  os valores próprios e os vectores próprios, respectivamente, da matriz *Hessiana* de  $\mathbf{A}$ , então

$$\begin{aligned} [\mathbf{I} - \eta \mathbf{A}] \mathbf{z}_i &= \mathbf{z}_i - \eta \mathbf{A} \mathbf{z}_i \\ &= \mathbf{z}_i - \eta \lambda_i \mathbf{z}_i \\ &= (1 - \eta \lambda_i) \mathbf{z}_i \end{aligned}$$

Os vectores próprios de  $[\mathbf{I} - \eta \mathbf{A}]$  são os vectores próprios de  $\mathbf{A}$  e os valores próprios de  $[\mathbf{I} - \eta \mathbf{A}]$  são  $(1 - \eta \lambda_i)$ . A condição de estabilidade do método do gradiente decrescente baseia-se na seguinte expressão:  $|(1 - \eta \lambda_i)| < 1$ . Assumindo que a função quadrática tem um valor mínimo e que os seus valores próprios são positivos, tem-se  $\eta < \frac{2}{\lambda_{\max}}$ .

Conclui-se que o valor máximo da taxa de aprendizagem é inversamente proporcional ao valor máximo da curvatura da função quadrática. A curvatura dá-nos informação sobre a rapidez com que o gradiente se modifica [5].

### APÊNDICE 3 – NORMALIZAÇÃO DOS VECTORES DE ENTRADA E DE SAÍDA

Os vectores de entrada e de saída foram normalizados entre [-1, 1] segundo os seguintes critérios:

#### 1. Percepção do trânsito

$$\text{entrada: } \begin{cases} -1 & \text{pouco} \\ 0 & \text{médio} \\ 1 & \text{muito} \end{cases}$$

#### 2. Tipo de pavimento

$$\text{entrada: } \begin{cases} -1 & \text{paralelo} \\ 0 & \text{semi - penetração} \\ 1 & \text{betuminoso} \end{cases} \quad \text{betuminoso}$$

#### 3. Velocidades praticadas em cada arco

$$\text{entrada: } \begin{cases} -1 & < 50km/h \\ 0 & = 50km/h \\ 1 & > 50km/h \end{cases}$$

#### 4. Hora do dia

A hora do dia é a única entrada que se encontra normalizada de forma diferente. Tal, deve-se ao facto de que o excesso de normalização faz com que seja difícil ou mesmo impossível tirar conclusões sobre os deslocamentos em hora de ponta. Assim, escolheu-se agrupar os percursos por intervalos de 30 minutos e de seguida normalizar entre [0, 1]. Por exemplo: das 9h30m às 10h como sendo 9h30m e das 10h -10h30m, como sendo 10h.

$$\text{entrada: } \begin{cases} 9.50 & 9h30m - 10h \\ 10.00 & 10h0 - 10h30m \\ \vdots & \\ 22.50 & 22 - 22h30m \end{cases}$$

## 5. Condições do pavimento face às condições atmosféricas

$$\text{entrada: } \begin{cases} 1 & \text{seco} \\ -1 & \text{molhado} \end{cases}$$

## 6. Número de entroncamentos

$$\text{entrada: } \begin{cases} 1 & > 3 \text{ entroncamentos} \\ 0 & = 2 \text{ entroncamentos} \\ -1 & < 2 \text{ entroncamentos} \end{cases}$$

## 7. Conflitos com peões

$$\text{entrada: } \begin{cases} 1 & \text{existem} \\ -1 & \text{não existem} \end{cases}$$

Vector de saída:

$$\text{entrada: } \begin{cases} -1 & \text{não passou no arco} \\ 1 & \text{passou no arco} \end{cases}$$