



UAlg **FCT**

UNIVERSIDADE DO ALGARVE
FACULDADE DE CIÊNCIAS E TECNOLOGIA

**Desenho de modelos neuronais para a
detecção de sismos**

Ozias Moisés Vieira Barros

Dissertação para a obtenção de Grau de Mestre em
Engenharia Electrónica e Telecomunicações

Orientador: Prof. Doutor ANTÓNIO EDUARDO DE BARROS RUANO
Co-orientador: Mestre GUILHERME MADUREIRA



UAAlg FCT

UNIVERSIDADE DO ALGARVE
FACULDADE DE CIÊNCIAS E TECNOLOGIA

**Desenho de modelos neuronais para a
deteção de sismos**

Ozias Moisés Vieira Barros

Dissertação para a obtenção de Grau de Mestre em
Engenharia Electrónica e Telecomunicações

2013

Desenho de modelos neuronais para a deteção de sismos

Declaração de autoria de trabalho

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Ozias Moisés Vieira Barros

Copyright ©

A Universidade do Algarve tem o direito, perpétuo e sem limites geográficos, de arquivar e publicar este trabalho, através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

A minha mãe e em memória do meu pai.

Agradecimentos

Eu gostaria de expressar o mais sincero agradecimento ao meu orientador, Prof. Doutor ANTÓNIO EDUARDO DE BARROS RUANO, pela sua inestimável orientação, supervisão e apoio constante durante o desenvolvimento dessa dissertação. Essa orientação foi fundamental para poder atingir os objetivos definidos no início deste trabalho.

Também estou sinceramente grato ao Mestre Guilherme Madureira, por todos, os apoios prestados, seus conselhos e ajudas na resolução dos problemas encontrados, colocando a minha disposição todo o conhecimento adquirido durante esses anos a trabalhar nessa área.

Devo também agradecer a Fundação Calouste e Gulbenkian por me proporcionar bolsas de estudo, e a oportunidade de estudar durante esses fantásticos 5 anos na Universidade do Algarve.

Agradecimentos especiais são transmitidos a todos os meus professores e colegas, que durante esses anos tornaram mais agradável e memorável a minha experiência na UALG.

Agradeço ao Sérgio Silva, pelos apoios prestados durante o desenvolvimento do trabalho.

Meu reconhecimento final e mais sincero deve ir para para a minha mãe, irmãos, primos, tios, tias, sobrinhos, amigos, conhecidos, pelo incentivo e amor. Sem eles, esse trabalho nunca teria existido.

Resumo

A performance dos detetores sísmicos atualmente utilizados pode e deve ser melhorada.

Atualmente existem vários algoritmos para a deteção de sismos de forma automática, desde os sistemas simples baseados em STA/LTA, aos mais sofisticados baseados em reconhecimento de padrões. Este estudo pretende dar continuidade ao desenvolvimento de uma abordagem de deteção de eventos sísmicos ao nível da estação local, utilizando uma técnica bastante conhecida, chamada Máquina de Vetores de Suporte (SVM). SVM é amplamente utilizada em problemas de classificação, devido a sua boa capacidade de generalização.

Nesta experiência, a técnica baseada em SVM é aplicada em diferentes modos de operações. Os resultados mostraram que a técnica proposta dá excelentes resultados em termos de sensibilidade e especificidade, além de exigir um tempo de deteção suficientemente pequeno para ser utilizado num sistema de aviso precoce (early-warning system).

Começamos pela classificação de dados de forma Off-line, seguido da validação do classificador desenvolvido. Posteriormente, o processamento de dados é executado de forma contínua (On-line).

Os algoritmos foram avaliados em conjuntos de dados reais, provenientes de estações sísmicas da Rede de Vigilância Sísmica de Portugal, e em aplicações reais da área de Sismologia (simulação de funcionamento em ambiente real).

Apesar de apenas duas estações serem consideradas, verificou-se que utilizando a combinação de detetores, consegue-se uma percentagem de deteção idêntica para quando utilizado um único modelo (Abordagem OR) e o número de falsos alarmes para a combinação de modelos é quase inexistente (Abordagem AND).

Os resultados obtidos abrem várias possibilidades de pesquisas futuras.

Palavras-chave: Máquina de Vetores de Suporte, Redes Neurais, Deteção de Sismos, Sistemas de Alerta Antecipado, Inteligência Computacional, Aprendizagem Ativa.

Abstract

The performance of the seismic detectors currently used can and should be improved.

Nowadays there are several algorithms for detecting seismic events automatically, ranging from a simple systems based STA/LTA, to the more sophisticated ones using pattern recognition techniques.

This study intends to continue the development of an approach for the seismic events detection, using a well known technique, called Support Vector Machine (SVM). SVM is widely used in classification problems because of its good generalization ability.

In this experiment, the SVM based technique is applied to different modes of operations. The results showed that the proposed technique gives excellent results in terms of sensitivity and specificity, and require a time of detection sufficiently small to be used in an early-warning system.

We start with Off-line data classification, following by a validation of the developed classifier. Subsequently, the data is processed continuously, in On-line scheme.

The algorithms were tested on real data sets, from seismic stations of the Portugal Seismic Monitoring Network, and in real applications of Seismology (simulation of operation in real environment).

Despite only two stations were considered, it was shown that by using the combination of detectors the percentage of detections is similar to when using a single model (OR approach), and the number of false alarms for the combination of models is almost null (AND approach). The results obtained in this work have open possibilities for many future works .

Keywords: Support Vector Machines, Neural Networks, Seismic Detection, Early-Warning Systems, Computational Intelligence, Active Learning .

"Si ka badu ka ta biradu"

Eugénio Tavares

Conteúdo

Declaração de autoria de trabalho	v
Agradecimentos	ix
Resumo	xi
Abstract	xiii
Lista de Tabelas	xxi
Lista de Figuras	xxiv
Lista de Símbolos	xxiv
Lista de Abreviaturas	xxvi
1 Introdução	1
1.1 Objetivos	1
1.2 Motivação	2
1.3 Conceitos Básicos	3
1.3.1 Reconhecimento de Padrões	3
1.3.2 Métodos de Classificação	4
1.3.3 Abordagens de reconhecimento de padrões	6
1.4 Aprendizagem Ativa	7
1.4.1 Critério de paragem aprendizagem ativa	8
1.5 Tipos de implementação dos algoritmos	10
1.5.1 Implementação Off-line	10
1.5.2 Implementação On-line	10
1.6 Organização da dissertação	10
1.7 Contribuições	12
1.8 Resumo	13
2 Estado da Arte	15
2.1 Introdução	15

2.2	Métodos baseados na representação no tempo – Avaliação das amplitudes dos sinais	16
2.3	Métodos baseados em análises Multi-resolução e a sua combinação com outros métodos	20
2.4	Métodos sem aplicação de inteligência computacional	22
2.5	Métodos Recursivos	23
2.6	Métodos baseados em Redes Neurais Artificiais	25
2.7	Resumo	33
3	Máquina de Vetores de Suporte	35
3.1	Introdução	35
3.2	Características da SVM	36
3.3	História e Teoria de Base	37
3.4	A ideia fundamental	37
3.5	Noções básicas da teoria de aprendizagem estatística	38
3.5.1	A teoria da aprendizagem estatística	38
3.5.2	A Dimensão de Vapnik – Chervonenkis	40
3.6	Minimização do Risco Estrutural	40
3.7	Máquina de Vetores de Suporte	42
3.7.1	Classificador SVM Linear	42
3.7.2	Determinação do Hiperplano Ótimo	44
3.7.3	Hiperplano ótimo para padrões não-separáveis	49
3.7.4	Classificador SVM não linear	52
3.8	O algoritmo Kernel-Adatron	54
3.9	Considerações Finais	55
3.10	Resumo	56
4	Aplicações de técnicas de inteligência computacional para a detecção sísmica	57
4.1	Introdução	58
4.2	Detetor sísmico baseado em rede neuronal	58
4.2.1	Procedimento Experimental	60
4.2.2	Métodos de Treino e resultados	65
4.3	Comparação dos resultados obtidos com o Sistema de Detecção ao Nível da Estação Sísmica em Funcionamento no CND	67

4.3.1	Considerações sobre o sistema de avaliação atualmente em utilização nas estações sísmicas	67
4.4	Operação On-Line de um Detetor Sísmico Inteligente	68
4.4.1	Aplicação do sistema de janela deslizante	68
4.5	Resumo	69
5	Procedimento Experimental e Resultados	71
5.1	Introdução	72
5.2	Dados Utilizados	72
5.2.1	Dados Funcionamento Off-line	72
5.2.2	Normalização dos dados	73
5.3	Classificação em Off-line	74
5.3.1	Resultados com o detetor PESTR	74
5.3.2	Resultados com o detetor PVAQ	76
5.4	Classificação em On-line	77
5.4.1	Normalização dos dados	77
5.5	Resultados PESTR-2008	77
5.5.1	Resultados On-line com o detetor PESTR	78
5.6	Resultados PVAQ-2008	79
5.6.1	Resultados On-line com o detetor PVAQ	79
5.7	Tempo de Deteção sísmica	80
5.7.1	Tempo de Deteção com o detetor PESTR	80
5.7.2	Tempo de Deteção com o detetor PVAQ	82
5.8	Sistemas de deteção considerando os vários detetores implementados	84
5.8.1	Resultados considerando vários detetores	85
5.9	Discussão dos Resultados	87
5.10	Resumo	88
6	Conclusão	89
6.1	Introdução	89
6.2	Contribuições	89
6.3	Limitações	90
6.4	Trabalho Futuro	91
6.5	Resumo	91

A	Informações Adicionais	93
A.1	Rede de Vigilância Sísmica	93
	Bibliografia	103

Lista de Tabelas

4.1	Parâmetros da estação sísmica PVAQ na REDE SÍSMICA PORTUGUESA . . .	60
4.2	Resultados primeira experiência Off-line com MLP, dados PVAQ	66
4.3	Resultados primeira experiência Off-line com MLP, dados PVAQ	66
4.4	Resultados experiência On-line com dados PVAQ	69
4.5	Parâmetros da estação sísmica PESTR na REDE SÍSMICA PORTUGUESA . .	69
4.6	Resultados experiência On-line com dados PVAQ	69
5.1	Resultados obtidos aplicando aprendizagem ativa PESTR-2009. (SCALE) . . .	74
5.2	Resultados obtidos aplicando aprendizagem ativa PESTR-2009. (Escalamat) . .	75
5.3	Resultados obtidos aplicando aprendizagem ativa PVAQ-2007. (SCALE)	76
5.4	Resultados obtidos aplicando aprendizagem ativa PVAQ-2007. (Escalamat) . . .	76
5.5	Performance On-line do SVM para PESTR (Escalamat)	78
5.6	Performance On-line do SVM para PESTR (SCALE)	78
5.7	Performance On-line do SVM para PVAQ (Escalamat)	79
5.8	Performance On-line do SVM para PVAQ (SCALE)	79
5.9	Performance conjunta Catálogo PESTR	85
5.10	Performance conjunta Catálogo PVAQ	86
5.11	Performance conjunta Catálogo Geral (PVAQ + PESTR)	86
5.12	Performance conjunta Catálogo PESTR	86
5.13	Performance conjunta Catálogo PVAQ	87
5.14	Performance conjunta Catálogo Geral (PVAQ + PESTR)	87

Lista de Figuras

1.1	Modelo de Sistema de Reconhecimento de Padrões	4
1.2	Fluxograma para a classificação utilizando Aprendizagem Ativa	9
2.1	Taxa de deteção de XRTP e WDetect em Março 2001.	21
2.2	Comparação das taxas de deteção (sem classificação) entre DHMM e o modelo recursivo STA / LTA.	24
2.3	Performance da rede para o conjunto de dados do monte Vesúvio	28
2.4	Performance da rede para o conjunto de dados do monte Vesúvio	29
2.5	Esquema de uma árvore neuronal, correspondente 2 perceptron's com 2 saídas cada	30
2.6	Comparação dos resultados experimentais entre RNA e IANN	32
3.1	O hiperplano ótimo, separando exemplos positivos de negativos	38
3.2	Distribuição de pontos num espaço bidimensional	41
3.3	Subconjuntos de funções aninhadas, ordenados por dimensão VC	41
3.4	Classificação de dados linear	43
3.5	Classificação de dados linear, múltiplas soluções	44
3.6	Hiperplano de separação ótimo, para padrões linearmente separáveis	45
3.7	Hiperplano de separação num espaço linearmente não separável	49
3.8	Classificação não linear	52
3.9	Mapeamento do espaço de entrada num espaço de maiores dimensões	53
4.1	Extrato Catálogo sísmico PVAQ 2008.	62
4.2	Extração de segmentos de registo sísmico.	62
4.3	Segmento de registo sísmico e Espectrograma.	64
4.4	Extração de segmentos, cálculo de features e construção do universo de dados.	64
5.1	Tempo de Deteção para PESTR-EW (Escalonamento Escalamat).	80
5.2	Tempo de Deteção para PESTR-EW (Escalonamento SCALE).	81

5.3	Histogramas das magnitudes sísmicas PESTR (2008).	82
5.4	Tempo de Detecção para PESTR-EW (Escalonamento Escalamat).	82
5.5	Tempo de Detecção para PESTR-EW (Escalonamento SCALE).	83
5.6	Histogramas das magnitudes sísmicas PVAQ (2008).	83
5.7	Modos de operação utilizando os vários detetores desenhados.	84
A.1	Rede sismográfica de Portugal Continental	93

Lista de Abreviaturas

BP	Back Propagation
CND	Centro Nacional de Dados
DEP	Densidade Espectral de Potência
DOP	Degree of Dolarization
ERM	Empirical Risk Minimization.
EWS	Early Warning System
IC	Inteligência Computacional
LTA	Long Time Average
MLP	Multilayer Perceptron
RBF	Radial Base Function
RNA	Rede Neuronal Artificial
ROC	Receiver Operating Characteristic
RP	Reconhecimento de Padrões
SOM	Self Organizing Map
SRM	Structural Risk Minimization
STA	Short Time Average
SVM	Support Vector Machine (Máquina de Vetores de Suporte)
SV	Support Vector
VC	Dimensão de Vapnik-Chervonenkis

Capítulo 1

Introdução

1.1 Objetivos

Este estudo tem por objetivo continuar a investigação da utilização do método de classificação utilizando Máquinas de Vetores de Suporte (SVM) desenvolvido em [Madureira, 2009, Madureira and Ruano, 2009, 2012, Ruano and Madureira, 2011], para a classificação de dados sísmicos provenientes de estações sísmicas da Rede De Vigilância Sísmica¹ de Portugal (Apêndice A). Foi possível aplicar o procedimento utilizando dados de duas estações sísmicas,² a estação sísmica de Vaqueiros (PVAQ) e a estação sísmica de Estremoz (PESTR).

Como objetivo específico tem-se o desenho de modelos neuronais para cada uma das estações, testados em dois modos de operação, Off-line e On-line. Ainda é analisado o desempenho na tarefa de deteção considerando os vários detetores implementados conjuntamente. Para a validação, os resultados obtidos pelos classificadores implementados são confrontados com os obtidos pelo sistema atualmente em uso nas estações sísmicas, performances obtidos nos trabalhos de base, e outros métodos tradicionalmente usados para este fim.

¹ Rede De Vigilância Sísmica é responsável por processar toda a informação de eventos sísmicos perceptíveis pelo equipamento numa determinada região do globo [Madureira, 2009]

² A estação sísmica é a componente básica de uma rede de vigilância sismográfica. As estações sísmicas com tecnologia mais recente são normalmente constituídas por um sismómetro de banda larga e por um acelerómetro. A estes equipamentos de base juntam-se um digitalizador e os equipamentos necessários à transmissão de dados e ao fornecimento de energia. O equipamento deve estar colocado em instalações dedicadas para o efeito. (http://www.geofisicabrasil.com/geofisicabasica/99-sismologia/14_7_-sismos-terremotos.html)

1.2 Motivação

De acordo com [Mourad, José, and Bento Caldeira, 2008] em 17 de Abril de 1889, em Potsdam (Alemanha) foi feito o primeiro registo gráfico dos movimentos do solo produzidos por um sismo³ remoto (Japão), um dos mais significativos êxitos da história da sismologia. Esse fato assinala o começo da sismologia experimental, a ferramenta que forneceu os meios para a compreensão do fenómeno sísmico e da estrutura do interior da Terra.

Desde aí, grandes avanços têm sido produzidos com a finalidade de construir sistemas de deteção automática na análise de dados sísmicos. Contudo essa tarefa revelou ser de grande complexidade. Para a implementação de sistemas de alerta antecipado é preciso conseguir melhorias no detetor a funcionar ao nível da estação sísmica individual⁴, aumentando a fiabilidade dos resultados obtidos e diminuindo ao máximo possível o tempo de deteção de eventos. Dando seguimento aos trabalhos de base [Madureira, 2009, Madureira and Ruano, 2009, 2012, Ruano and Madureira, 2011], pretende-se criar modelos neuronais que facilitam o processo automático de deteção e análise sísmica.

Na literatura existem muitos trabalhos (referidos no Capítulo 2) com essa finalidade. Em alguns casos quando é utilizado aprendizagem automática utilizando dados sísmicos, muitos autores tentam prever quando e onde acontecerão os próximos eventos sísmicos e qual será a magnitude do mesmo, mas como os sismos apresentam padrões que ocorrem ao longo de centenas ou mesmo milhares de anos, levando em conta a data do primeiro registo sismograma podemos perceber que não há dados suficientes para atingir esse objetivo [Grubert and Vuong, 2012]. Este presente trabalho, pretende detetar sismos em sinais sísmicos, simulando todo um processo de deteção em modo contínuo, com uma grande redução no tempo de deteção de eventos (ver Capítulo 5-Resultados), que pode ser bastante útil em sistemas de alerta antecipado.

³Basicamente, sismo é a ocorrência de uma fratura subterrânea. As ondas elásticas geradas propagam-se por toda a Terra. As palavras sismo e terramoto são sinónimas, sendo que normalmente reserva-se o uso da palavra terramoto para a classificação de grandes sismos, e para os pequenos costuma-se usar abalo sísmico ou tremor de terra (<http://www.geofisicabrasil.com/geofisicabasica/99-sismologia/147-sismos-terremotos.html>)

⁴No funcionamento atual, deteção de um evento não significa simplesmente que um sinal sísmico é visto numa estação. Geralmente é preciso que a onda P identificável seja observada acima do nível de ruído em mais estações sísmicas.

1.3 Conceitos Básicos

1.3.1 Reconhecimento de Padrões

O reconhecimento de padrões (RP) corresponde ao processo de caracterização de dados em classes ou categorias. A identificação das classes baseia-se num conjunto de características ou atributos considerados fundamentais que são extraídos das informações disponíveis do problema em questão.

O RP tem como objetivo classificar informações, baseando em conhecimentos anteriores do problema ou extraídos das propriedades estatísticas de um conjunto de objetos; desta forma podemos dizer que o RP trata da classificação e descrição de objetos⁵.

[Duda, Hart, and Stork, 2001], descrevem o RP como sendo o “ campo interessado no reconhecimento de regularidades significativas em ambientes ruidosos ou complexos, feito por máquinas, ou a “ procura por uma estrutura em dados.”⁶

Um sistema de RP simples envolve normalmente:

- Extração de características dos objetos a classificar;
- Seleção das características mais discriminantes;
- Construção de um classificador;

Um modelo simples do sistema de RP consiste, basicamente, no sensor e no classificador. A saída do sensor é um conjunto de n medidas, cada uma correspondendo a um canal do sensor.

Às medidas do sensor são processados de forma a criar espaços de características, selecionando para isso as características mais discriminantes do objeto em estudo.

O classificador depois identifica com base nas características e das pré-especificadas classes, levando sempre em consideração uma regra de classificação apropriada, a classe que o padrão pertence (Fig. 1.1).

⁵Os objetos no problema de RP são genericamente denominados de padrões e podem ser caracteres, gráficos, sinais eletrônicos, células em biologia, sinais biológicos, sinais biomédicos, imagens, pixels, entre outros. (<http://www.cic.unb.br/lamar/te073/Aulas/ReconhecimentoPadroes.pdf>) .

⁶Frase original, [Duda, Hart, and Stork, 2001] *define it as a field concerned with machine recognition of meaningful regularities in noisy or complex environment.*

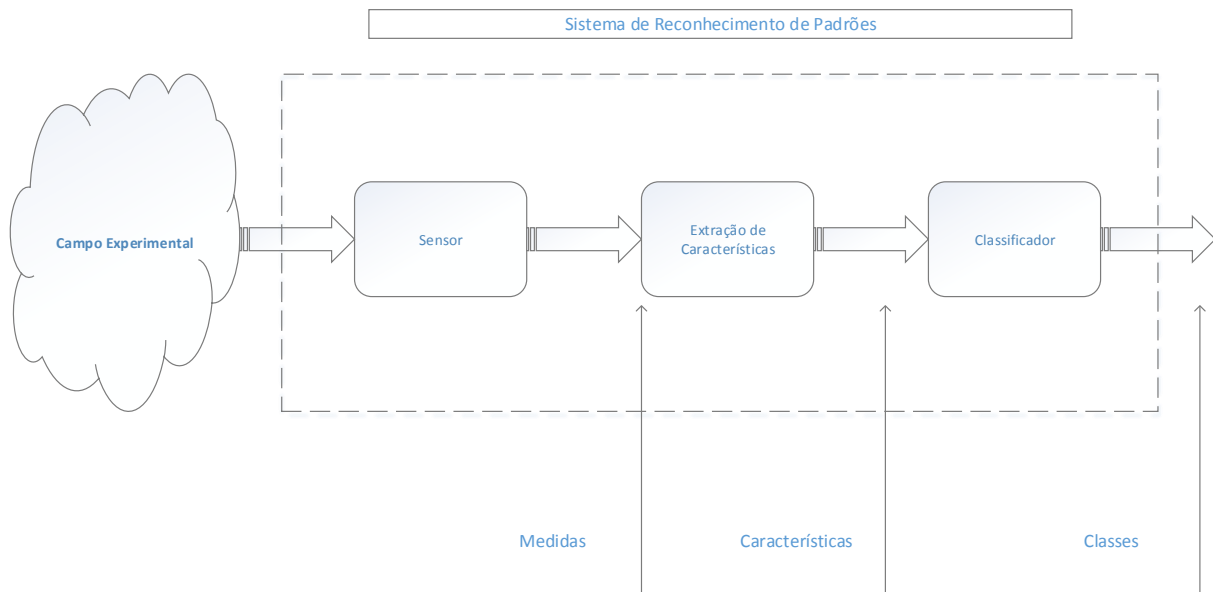


Figura 1.1: Modelo de Sistema de Reconhecimento de Padrões. Fonte: Adaptado de (<http://paginas.fe.up.pt/jmsa/recpad/>)).

1.3.2 Métodos de Classificação

No problema de RP há duas maneiras de classificar ou determinar a classe pertencente a um padrão:

1. Classificação supervisionada
2. Classificação não supervisionada

1.3.2.1 Métodos de classificação supervisionados

Nos métodos supervisionados a aprendizagem é de forma supervisionada, onde as amostras pertencentes ao conjunto de dados têm as classes bem definidas.

$$\mathfrak{S} = \{(x_1, k_1), (x_2, k_2), \dots, (x_n, k_n)\} \quad (1.1)$$

Sendo:

x_i corresponde ao padrão de ordem i

k_i representa a classe do padrão de ordem i

n , corresponde a quantidade de padrões existentes

Dessa forma podemos dizer que a aprendizagem supervisionada envolve um treino de um conjunto de dados já classificados, de onde é criada uma interface capaz de classificar novos dados, que definiriam a performance do algoritmo após o treino.

1.3.2.2 Métodos de classificação não supervisionados

No contexto de métodos não supervisionados, não existe um "orientador" responsável. Existe um problema a modelar idêntico ao apresentado anteriormente Secção 1.3.2.1, porém os k_i são desconhecidos.

Este tipo de problema é analisado por exemplo na área conhecida por Análise de Agrupamentos⁷.

Neste caso a margem entre classes, é determinada por uma fronteira que é desconhecida. As classes serão posteriormente determinadas com base nas similaridades existente entre os padrões dos conjuntos de dados. Sendo assim, o problema a ser tratado consiste não somente na classificação propriamente dita, mas também na própria definição das classes. Nesse problema o algoritmo tenta encontrar uma estrutura escondida em dados não classificados. A classificação é feita de forma automática, sem intervenção do utilizador, não sendo utilizadas informações prévias (targets); por isso, não necessitam de conjuntos para teste para determinar a performance da superfície de decisão construída para novos dados.

Ainda no contexto de métodos de classificação podemos definir mais duas categorias bastante importantes.

1.3.2.3 Métodos paramétricos

Nestes métodos o conjunto de treino é constituído por variáveis dependentes e independentes conhecidas, podendo existir parâmetros cujos valores são desconhecidos, que podem ser determinados utilizando o conjunto de treino. Geralmente os parâmetros do conjunto de treino têm interpretação física.

1.3.2.4 Métodos não paramétricos

Neste caso existe uma ausência completa ou quase completa, de conhecimentos a priori do problema em questão. As redes neuronais, sistemas Fuzzy, e SVM são exemplos de classificadores não-paramétricos⁸. Para o treino desses classificadores usam-se pares entrada saída, determinando a função de decisão que classificará os dados de entrada em uma das dadas classes.

⁷Análise de Agrupamentos do Inglês *Clustering Analysis*

⁸Estas técnicas não consideram informação da distribuição da população como no caso paramétrico; elas deixam os próprios dados definirem sua própria estrutura e serem capazes de encontrar esta estrutura explícita ou implicitamente.

1.3.3 Abordagens de reconhecimento de padrões

Os principais métodos ou abordagens são:

- **Abordagem estatística**

São utilizados conceitos da teoria de decisão estatística para decidir a classe que os padrões pertencem, assumindo que os padrões e classes são caracterizados por modelos estatísticos.

- **Abordagem sintática**

A estrutura dos padrões é descrita usando inter-relações de características descritoras básicas denominadas primitivas⁹. Expressam os padrões em termos de primitivas e operadores unidos por regras que descrevem o relacionamento espacial entre eles.

- **Abordagem neuronal**

É uma abordagem que procura determinar um mapeamento ótimo entre entradas e saídas inspirando em modelos de neurónios do cérebro¹⁰. As RNA assim como as técnicas estatísticas necessitam de um grande número de amostras para realizarem o treino de suas aproximações, já as técnicas sintáticas necessitam apenas de um conjunto de atributos que são constantes para cada padrão.

- **Abordagem difusa**

Abordagem que tem em conta o grau de incerteza por vezes inerente a características e a classificações, usando a teoria dos conjuntos difusos¹¹ para modelizar esse grau de incerteza¹². De acordo com o problema em questão, existirão técnicas que são mais apropriadas a utilização, ou até mesmos casos onde esses métodos são associados para produzir melhores resultados. Utilizando as abordagens anteriores, e outras associadas aos apresentados anteriormente, as técnicas de RP têm um grande leque de aplicação, podendo ser aplicados em áreas científicas e tecnológicas, industriais, médicas, agrícolas entre outros, tentando projetar *sistemas inteligentes que são um grande desafio para a tecnologia moderna*.

⁹<http://paginas.fe.up.pt/jmsa/recpad/>

¹⁰<http://paginas.fe.up.pt/jmsa/recpad/>

¹¹Conjuntos difusos do Inglês *fuzzy set*

¹²<http://paginas.fe.up.pt/jmsa/recpad/>

1.4 Aprendizagem Ativa

Aprendizagem ativa é um termo bastante popular em aprendizagem automática (aprendizagem-máquina)¹³ onde o algoritmo seleciona instâncias de dados a serem classificadas, pertencentes ao conjunto de teste, e os inclui no conjunto de treino.

Utilizando aprendizagem ativa geralmente espera-se que a seleção de dados, e a sua repartição entre conjunto de treino e de teste, conduza a uma melhor performance no conjunto de dados global. A quantidade de dados no conjunto de treino necessária para treinar um determinado método poder ser reduzida de forma significativa [Ruano and Madureira, 2011].

Um cenário típico de aprendizagem ativa, conforme descrito em [Tong and Koller, 2001], é constituído pelos seguintes componentes. Os dados são divididos, um conjunto de dados com os *targets* definidos X e um outro conjunto de dados U .

Geralmente são dados que ainda não possuem classes definidos, não foi aplicado ainda o processo de rotulação¹⁴.

Para o problema aqui tratado esses dados correspondem a dados do conjunto de teste, contendo informação que o algoritmo ainda não utiliza U . Após isso, existe um algoritmo de aprendizagem l que é treinado com dados de X , que já têm os *targets* bem definidos e uma função para a consulta q . A função de consulta q decide os casos de U (conjunto de teste) a introduzir no conjunto de treino X (em outros casos na literatura esses dados são rotulados nessa fase), que por sua vez será utilizada para o treino do algoritmo de aprendizagem. A função de consulta q nesse caso procede da seguinte forma:

Os exemplos mal classificados do conjunto U são incorporados no conjunto de treino X , e aleatoriamente são retirados o mesmo número de exemplos para o conjunto de validação, excluindo os padrões que definem o fecho convexo¹⁵ determinado com a totalidade dos dados [Madureira and Ruano, 2009, Ruano and Madureira, 2011]. O procedimento é esquematizado na Figura 1.2. Na maioria dos casos esse procedimento é repetido até atingir um determinado critério de paragem, sendo o procedimento iterativo.

Para o treino da SVM utilizado nesse trabalho o critério de paragem foi definido da seguinte forma:

¹³Aprendizagem de máquina, do Inglês *Machine Learning*

¹⁴ETIQUETAR definir o *target*, a classe a que pertence

¹⁵Fecho convexo do Inglês *Convex Hull*

1.4.1 Critério de paragem aprendizagem ativa

- O Classificador obtido é ótimo, ou seja o procedimento é repetido até a obtenção de um classificador perfeito, não havendo erros nem termos de especificidade ou sensibilidade (**R = 100% e S = 100%**), tudo isso para o universo de dados em análise.
- Se a performance do classificador deteriorar-se, a iteração é terminada.

Quando é aplicado aprendizagem ativa, a eficiência é medida geralmente de duas formas.

A mais popular corresponde a redução dos dados de treino necessários para obter uma determinada performance.

A segunda e a utilizada nesse trabalho é o aumento do desempenho para uma determinada quantidade de dados de treino.

A abordagem comum para o processo de aprendizagem ativa é a seleção aleatória de dados a incorporar nos dados de treino [Vlachos, 2004].

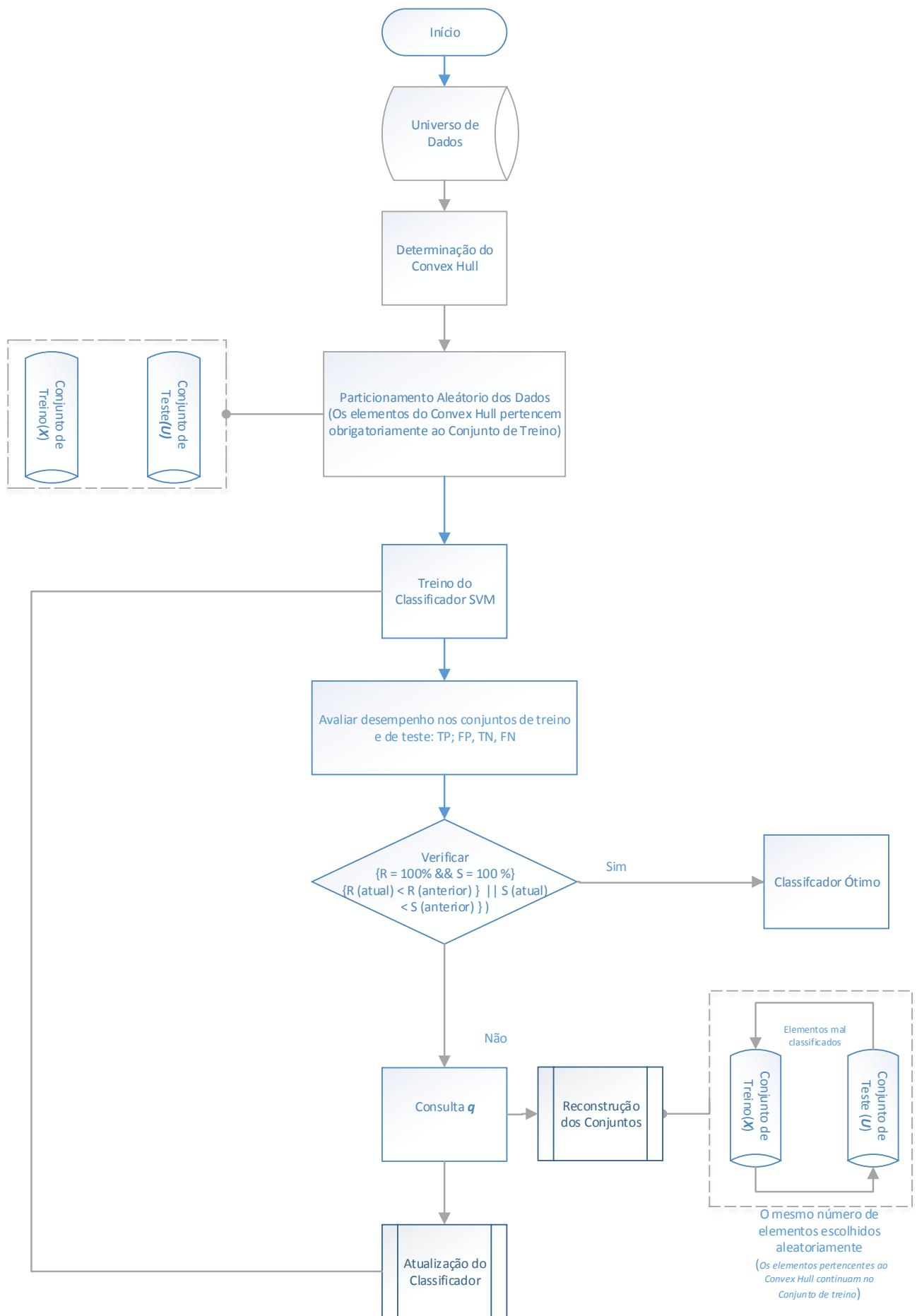


Figura 1.2: Fluxograma para a classificação utilizando Aprendizagem Ativa

1.5 Tipos de implementação dos algoritmos

Durante a classificação das informações sísmicas, procedeu-se a classificação em dois modos:

- Modo Off-line
- Modo On-line

È importante referir que quando utilizados os termos Off-line e On-line a referência é feita a forma como o modelo é utilizado, depois de desenhado.

1.5.1 Implementação Off-line

A implementação Off-line implica que algoritmo trabalha com um conjunto finito de dados, correspondentes a segmentos que foram anteriormente extraídos de um sinal sísmico. Esse universo de dados é utilizado depois no desenho do classificador desenvolvido (Nas experiências, esse universo de dados é dividido em 40% para teste e 60% para treino).

1.5.2 Implementação On-line

Quando referida essa implementação o funcionamento pode ser entendido como em linha (tempo real). A execução simula o funcionamento real, aplicando o classificador desenhado previamente a dados recolhidos continuamente pelo sistema de aquisição na estação sísmica.

1.6 Organização da dissertação

A dissertação está organizada em seis capítulos cujo conteúdo reflete o trabalho desenvolvido.

Os capítulos são os seguintes: Introdução, Estado da Arte, Máquina de Vetores de Suporte, Trabalhos anteriores, Experiências e Resultados, Conclusão. Seguidamente é apresentado uma breve descrição de cada um dos capítulos.

- **Capítulo 1** – No primeiro capítulo (presente capítulo) são apresentados os objetivos da dissertação e a motivação que levou a esse trabalho. São também expostos de forma resumida alguns conceitos básicos na classificação de padrões destacando os métodos de aprendizagem, o conceito de aprendizagem ativa e ainda a implementação de algoritmos On-line e Off-line. Por fim é apresentado a estrutura da dissertação.

- **Capítulo 2** – No segundo capítulo é apresentado uma revisão ao estado da arte sobre o problema de classificação de dados sísmicos, destacando os métodos mais populares e que apresentam melhores resultados.

Atualmente existe um amplo número de técnicas para o efeito, variando desde técnicas simples baseadas na razão de energia, a métodos mais sofisticados baseados no reconhecimento de padrões.

São também focados os trabalhos que utilizam técnicas baseadas em inteligência computacional.

É ainda discutido a técnica atualmente em uso nas estações sísmicas em Portugal baseado na razão de energia STA/LTA.

- **Capítulo 3** – O terceiro capítulo aborda o Classificador SVM. É analisada a sua formulação matemática e os aspetos a ter em conta na sua implementação.

Sendo a SVM o algoritmo de base desse trabalho é apresentado a proposta do algoritmo anteriormente implementado e empregue nesta dissertação, sendo em alguns caso adaptado e feito algumas alterações de forma a obter os melhores resultados possíveis.

- **Capítulo 4** – No quarto capítulo os trabalhos anteriores a qual essa dissertação dá seguimento são apresentados. São apresentados os testes e as experiências realizadas de acordo com algoritmos desenvolvidos anteriormente, apresentando sempre os resultados e análise do ponto de vista dos autores.

- **Capítulo 5** – No quinto capítulo, todos os testes, experiências efetuadas são apresentados, começando por apresentar a abordagem Off-line, e de seguida utilizando o classificador desenvolvido, implementação On-line.

Todas as experiências são realizadas de acordo com o algoritmo descrito no terceiro capítulo implementado para funcionar Off-line e On-line.

- **Capítulo 6** – No sexto e último capítulo são apresentados as conclusões, análise final e sugestões.

- **Apêndice A** – No apêndice são apresentados informações adicionais.

1.7 Contribuições

Parte do estudo desenvolvido nessa dissertação foi utilizado para a publicação de um artigo [Ruano, Madureira, Ozias, Khosravani, Ruano, and Ferreira, 2013] intitulado,

“A Support Vector Machine Seismic Detector for Early-Warning Applications”.

(keywords = early-warning systems, seismic detection, support vector machines)

Apresentado em, ***The 3rd IFAC International Conference on Intelligent Control and Automation Science (ICONS 2013), Chengdu, China, 2-4 September 2013.***

1.8 Resumo

Neste capítulo é feita uma introdução aos objetivos da dissertação, bem como as contribuições, e a organização da dissertação.

Foram discutidos alguns dos aspectos relativos a construção de sistemas inteligente, primeiramente foi analisado o problema de classificação de padrões, seguido de uma abordagem dos métodos de classificação, focando na classificação supervisionada e classificação não supervisionada. Foi também mostrado diferentes abordagens para o problema de reconhecimento de padrões.

È discutido a problemática de aprendizagem ativa, discutindo a sua importância e alguns aspectos a considerar na sua aplicação. Ainda foi apresentado os tipos de implementação dos algoritmos, focando nas abordagens utilizadas na realização implementações On-line e Off-line.

Capítulo 2

Estado da Arte

Em sismologia, técnicas de inteligência computacional (IC) têm sido utilizados para prever eventos sísmicos, controlo de estruturas civis, discriminação entre vários tipos de eventos sísmicos, determinação das fases sísmicas e para a deteção de sismos propriamente dito.

Na pesquisa da literatura encontramos vários trabalhos com resultados bastante interessantes tendo como principais objetivos, a deteção de eventos sísmicos, determinação de instantes de chegada das fases sísmicas, classificação de eventos sísmicos de diferentes tipos ou identificação de eventos sísmicos artificiais de origens conhecidos.

Esses estudos têm em comum o fato de todos residirem no processamento de dados sísmicos sendo que, alguns tratam dados que já foram previamente processados no sentido de identificarem sinal sísmico e outros identificam o sinal sísmico sem um pré-processamento.

2.1 Introdução

Neste capítulo são apresentados trabalhos com a finalidade específica de analisar dados sísmicos, principalmente baseando na inteligência computacional (RNA, SVM), de forma a detetar/classificar eventos sísmicos.

São apresentados não só trabalhos onde a RNA e SVM têm papéis de classificadores, más também, os que analisam informações sísmicas não recorrendo a inteligência computacional de forma a, obter melhor compressão de dados, caracterizar ou extrair informações a partir de eventos sísmicos, etc.

Também são analisados trabalhos realizados onde a deteção e classificação são feitas de forma automática, ou seja, todo o processamento é feito numa só fase. A maioria dos trabalhos concentram em três domínios, a deteção de pequenos sinais sísmicos, identificação das fases sísmicas e a discriminação entre sismos naturais e sismogramas de explosões artificiais.

2.2 Métodos baseados na representação no tempo – Avaliação das amplitudes dos sinais

Numa rede de estações sísmicas ou numa estação sísmica independente são produzidas grandes quantidades de informações que muitas vezes são difíceis de analisar em tempo real. Com a introdução da aquisição digitalmente de dados sísmicos, longos períodos de gravações e grandes quantidades de dados sísmicos a armazenar, tornaram-se problemas de carácter técnico. Para tornar mais eficaz todo o processo de recolha, processamento e armazenamento de dados, grande parte das estações sísmicas utilizam algoritmos baseados num modo de ativação. Um algoritmo de ativação permite detetar sinais sísmicos sempre que um suposto estado é alcançado, deixando esse estado quando o algoritmo declara o fim do evento.

Foram desenvolvidos vários algoritmos com essa finalidade, um dos mais famosos que emergiu dessa linha de pensamento foi [Stewart, 1977], este sistema ao nível da estação sísmica tanto pode funcionar através de software instalado no sistema de aquisição, ou ser implementado no próprio sistema. De referir que este detetor é usado na maioria das redes sísmicas internacionais, sendo standard para a rede sísmica nacional (Portugal). O sistema automático de deteção proposto em [Stewart, 1977] aplica um algoritmo STA/LTA que regista todas as mudanças de amplitudes dos ruídos sísmicos localmente na estação, e automaticamente ajusta a sensibilidade da estação para o nível atual de ruído.

Os cálculos geralmente são executados repetidamente em tempo real e de forma independente, ou seja, em todos os canais do registo sísmico os dados são analisados. O algoritmo STA/LTA processa sinais sísmicos em duas janelas de tempo, uma de curta duração (STA) e outra de longa duração (LTA).

O STA mede a amplitude instantânea do sinal, valor atual das últimas amostras processadas e procura identificar sismos. O LTA cuida da média atual da amplitude do ruído sísmico. Primeiramente a amplitude absoluta de cada uma das amostras de dados do sinal de entrada é determinado. Seguidamente a média das amplitudes absolutas em ambas as janelas são calculadas. Num passo adicional a proporção de ambos os valores, STA/LTA (ratio) é calculado. Essa razão é continuamente comparada com o valor limite definido pelo utilizador, STA/LTA nível de decisão.

Geralmente a transição dá-se depois do início do evento, por vezes a transição acontece perante a ocorrência da fase S, normalmente onde encontra-se mais energia. Por outro lado o abandono desse estado é antes do final do evento. De forma a garantir o registo da totalidade do evento, é preciso ajustar de forma ótima alguns parâmetros do trigger, tais como:

- Janela de duração do STA
- Janela de duração do LTA
- STA/LTA (trigger -Nível de decisão)
- STA/LTA (de - trigger -Nível de decisão)
- Frequência de corte do filtro passa-banda

Mesmo com esses parâmetros definidos de forma ótima, ainda é preciso levar em conta o fato da rede sísmica não depender exclusivamente desses parâmetros, existem parâmetros adicionais associados que só bem sintonizados permitem um ótimo registro de dados. De entre esses parâmetros associados, destacam-se a memória pré-vento (PEM) e a memória pós-evento (PET).

A definição dos parâmetros do trigger e os parâmetros associados dependem, do objetivo da aplicação, das condições do ruído no local, das propriedades do sismo no local a considerar e do tipo de sensor usado.

Por isso não existe uma regra genérica na definição dos mesmos. Esses parâmetros quase nunca são ótimos, sendo assim devem ser ajustados de forma gradual até encontrar uma configuração que melhor ajusta ao caso em estudo. Em geral, procura-se minimizar o número de falsos alarmes, e minimizar o número de não detecções.

Ao nível da estação sísmica geralmente esses detetores apresentam uma performance razoável, analisados de forma independente mas no contexto de rede sísmica o desempenho global do sistema é bastante maior. No entanto, o detetor STA/LTA clássico apresenta algumas limitações, como o facto de não identificar eventos sísmicos irreconhecíveis no espaço de sinal, a ausência de métodos para determinar o comprimento do sinal, etc.

Num esquema semelhante ao anterior, [Stewart, 1977] utilizou um envelope de dados (MDX), modificados a partir da derivada dos dados em que as alterações na inclinação do sinal são enfatizadas. Os valores MDX são encontrados primeiramente por estimativa da derivada (DX), nos i pontos consecutivos.

$$DX_i = X_i - X_{i-1} \quad (2.1)$$

Se o sinal da DX for constante durante menos de 8 valores consecutivos de i , então,

$$MDX_i = MDX_i + DX_{i-1} \quad (2.2)$$

Caso contrário,

$$MDX_i = DX_i \quad (2.3)$$

A série de tempo resultante é uma realização passa alto dos dados, onde o primeiro movimento oscilatório e a natureza dos dados são preservados e os sinais emergentes melhorados. Após isso, um conjunto de critérios de detecção e limites devem ser atendidos para declarar um evento.

[Allen, 1978] desenvolveu um detetor baseado num envelope que corresponde, ao quadrado de dados mais o quadrado da primeira derivada. Com isso conseguiu criar uma série de tempo, que inclui componentes de dados filtrados e não filtrados. O fluxo de dados processados é então submetido a uma série de testes lógicos e matemáticos para a identificação das fases e temporização.

Baseando no critério de energia transitório STA/LTA e envolventes dos sinais, [Patanè and Ferrari, 1999] implementaram um algoritmo que permite a detecção e localização de sismos locais. O algoritmo de detecção utiliza uma função característica baseada na envolvente do sinal, que depois é submetida a um detetor clássico STA/LTA. Fazendo uso de sistemas que operam em mainframes com uma boa capacidade de aquisição e processamento de dados conjuntamente com pacotes de software complexos para a análise automática e iterativa de dados sísmicos, são aplicadas análises espectrais e de polarização para identificar instantes de chegada das fases S e P.

[Sleeman and van Eck, 1999] propuseram um sistema para a detecção automática dos instantes de chegada da fase P do sinal sísmico. Determinaram o início do sinal através, da auto-regressão do sinal e do ruído, aplicando um critério denominado de *Aikake Information Criterion* utilizando o tempo de início como parâmetro.

A detecção consiste em duas etapas:

- Estimativa do início do evento sísmico através do detetor STA/LTA.
- Estimativa grosseira do instante de chegada da fase pelo detetor de fase.

O objetivo principal dos autores era fornecer rapidamente e de forma correta a localização dos sismos com magnitude superior a 5.0, permitindo determinar de forma automática e confiável o hipocentro. Este sistema chamado de AR-AIC foi testado com sucesso e permitiu fornecer informações de detecções precisas de forma automática num grande conjunto de dados experimentais. Conseguiram taxas de detecções de 71%, e uma diferença média com detecções manuais de 0,1 segundos.

[Withers, Aster, Young, Beiriger, Moore, and Trujillo, 1998] apresentaram uma comparação

de algoritmos para a detecção e identificação de fases sísmicas, de forma automática baseando no trigger. Os algoritmos analisados foram, STA/LTA, z-estatístico, frequência transitória, polarização, etc. Concluíram que nenhum foi claramente ótimo em todas as condições de, origem, recepção, ruído entre outros. O algoritmo STA/LTA foi o melhor computacionalmente, e forneceu melhores resultados em comparação com outros algoritmos.

[Sharma, Kumar, Mittal, and Shamshi, 2006] desenvolveram e implementaram um algoritmo STA/LTA, para gravação de 16-bit de dados sísmicos, usando um microprocessador 80C85 e os seus periféricos. Testaram o instrumento no laboratório e também no campo para períodos de longa duração. Para o valor de LTA muito grande, houve um problema de overflow. Implementaram o instrumento para garantir que não existam perdas de informações após a declaração de um evento positivo. Gravaram dados de pré-eventos, evento e pós-evento de forma sucessiva durante semanas, permitindo obter toda a informação relativamente a réplicas após o evento principal. O di-triggering foi incluído no STA/LTA desenvolvido por [Kumar, Sharma, Sharma, and Shamshi, 2009], durante o desenvolvimento do Sistema de aquisição de dados sísmicos, baseado na arquitetura de um PC de 24-bit. O algoritmo calcula a razão entre o STA e LTA para discriminar entre eventos positivos e negativos. Incorporaram sistema de tempo baseado em GPS para manter o tempo mais preciso. Na metodologia para a detecção de eventos sísmicos, o cálculo de STA e LTA, são realizados em tempo real. Primeiro amplitudes absolutas de cada amostra de dados do sinal de entrada são calculadas. Depois média da amplitude absoluta é determinada e, em seguida, adicionado ao cálculo do valor LTA. Após o cálculo do valor de LTA, um novo valor STA é determinado. Depois a média STA/LTA corrente é comparado com valor do trigger programado para a detecção de evento sísmico. Se o valor calculado não cumprir os critérios de disparo então LTA é atualizado como:

$$LTA = LTA - \text{primeirovalor}STA + \text{valorcorrente}STA \quad (2.4)$$

É calculado o novo valor STA e comparado com o trigger. Se o evento é negativo, o valor LTA é atualizado. No caso de evento positivo, o valor LTA é guardado para cálculos. A razão entre o LTA guardado e valores correntes de STA é comparada com o valor de di-trigger (programado pelo utilizador) para declarar final do evento sísmico.

2.3 Métodos baseados em análises Multi-resolução e a sua combinação com outros métodos

[Botella, Rosa-Herranz, Giner, Molina, and Galiana-Merino, 2003] desenvolveram um detetor sísmico (WDetect), em tempo real, baseado na razão STA/LTA com pré-filtragem por *wavelets* (DWT), aumentando a taxa de deteção e reduzindo os falsos alarmes em contraste com outros detetores sísmicos. Este detetor foi capaz de registar sinais recebidos em modo contínuo de todas as estações da rede sísmica (Alicante no Sudoeste de Espanha). Após aplicarem a DWT, reconstruíram o sinal com base na transformada inversa de *wavelet* (IDWT), mas apenas utilizando os coeficientes de alguns níveis seleccionados. O algoritmo funciona do seguinte modo:

1. O sinal é filtrado por um filtro, diferencial de, 1º ordem

$$dx(n) = [x(n) - x(n - 1)], \text{ onde } x(n) \text{ representa o sinal de entrada} \quad (2.5)$$

2. A razão STA/LTA é calculada de forma recursiva como:

$$STA(n) = STA(n - 1) + \frac{(dx(n) - STA(n - 1))}{T_{sta}} \quad (2.6)$$

Onde T_{sta} são as respetivas janelas usadas para calcular a média STA e LTA. O limiar mínimo para LTA é sempre definido como sendo o LTA no limite mínimo.

$$LTA(n) = LTA(n - 1) + \frac{STA(n) - STA(n - 1)}{T_{lta}} \quad (2.7)$$

3. O Valor do trigger ($\alpha(n)$), para eventos positivos é dado:

$$\alpha(n) = \text{frac}dx(n)LTA(n) \quad (2.8)$$

Quando $\alpha(n)$ é excedido um trigger é declarado e a predição entra no estado (2). De seguida o valor LTA da amostra é guardado como *preLTA*, para cálculos do trigger na etapa 2. O *preLTA* corresponde ao último valor LTA calculado.

No estado 2, se o valor limite for maior ou igual a $STA(n)/preLTA$ então, um evento é declarado positivo, dando inicio a gravação dos dados sísmicos, caso contrário é retornado a etapa 1. Obtiveram uma melhoria na deteção sísmica, como podemos confirmar com Figura 2.1, onde XDetect (Tottingham e Lee, 1989) e XRTP (Tottingham,1994).

	Seismic events (%)	Catalogued earthquakes (%)	False alarms (%)
XRTP-4-3	74.8	61.1	78.2
WDETECT-5-2	96.5	88.9	93.4
WDETECT-15-5-WVL	85.2	72.2	40.6
WDETECT-8-3-WVL	97.4	100.0	72.8
Improvement ratio WDETECT/XRTP	30.21	63.66	-6.91

Figura 2.1: Taxa de detecção de XRTP e WDetect em Março 2001. Fonte: [Botella, Rosa-Herranz, Giner, Molina, and Galiana-Merino, 2003]

Na validação do sistema a taxa de detecções corretas foi de 97.4% e de falsos alarmes de 72.8% melhorando significativamente o sistema em funcionamento. Consideraram o resultado desse algoritmo melhor do que o do simples STA/LTA algoritmo [Stewart, 1977]. Entretanto esse algoritmo não conseguiu identificar alguns eventos, onde a amplitude do sinal foi reduzido após aplicar a pré-filtragem por wavelets.

Num esquema parecido ao anterior, [Baranov, 2007] analisou várias formulações de detecções de eventos sísmicos, apresentando razões da ineficiência na formulação clássica e por fim propõe uma formulação adequada para o problema no espaço-tempo frequência. [Baranov, 2007] aplicou a transformada de *wavelet* ao sinal sísmico antes da submissão ao detetor STA/LTA. Com a aplicação da DWT determinou o comprimento do sinal sísmico e utilizando a expansão do sismograma, num pacote de ondas, eliminou as deficiências do sistema padrão permitindo a detecção de eventos sísmicos com baixa relação sinal ruído(SNR).

Do ponto de vista do autor, os métodos clássicos requerem pequenos cálculos, o que considerou uma vantagem para a sua utilização em sistemas de tempo real, mas por outro lado, a sua inaplicabilidade na determinação da duração de um evento e sua indiferença a composição do sinal são desvantagens, pelo que com a transformada *wavelet*, as desvantagens apontadas acima podem ser completamente removidas.

Para a eliminação dos problemas, o conceito de um envelope cumulativo do rácio STA/LTA foi introduzido, e um critério de determinação do comprimento do sinal sísmico proposto. Para aumentar a sensibilidade do detetor STA/LTA, procurou na representação do sinal no espaço de frequência/tempo, fazer a decomposição do registo em pacotes de wavelets, apresentando várias vantagens em comparação com os demais tipos de filtragem. O método consistiu no cálculo da relação STA/LTA, utilizando a decomposição dos conjuntos de wavelet das componentes verticais do sismograma digital. Com isso conseguiu a detecção de eventos sísmicos

que eram irreconhecíveis para o detetor STA/LTA clássico. Os principais parâmetros do algoritmo foram uma *wavelet* com um suporte compacto, e um detetor STA/LTA.

2.4 Métodos sem aplicação de inteligência computacional

[Joswig, 1990] implementou um detetor que utiliza ultrassonografias (detetor de ultrassom), para a construção de modelos gráficos na base de eventos sísmicos normalizados para a amplitude do ruído. Estes padrões são então comparados com os padrões de referência predefinidos (correspondendo a ambos os eventos sísmicos e ruído) que são deslocados ao longo da ecografia de forma semelhante ao processamento da janela de ultrassom, em cada etapa a medida de similaridade entre o segmento de ultrassom e o padrão de referência é calculada. Este processo é repetido para todos os padrões de referência disponíveis. O padrão mais semelhante é atribuído ao sinal analisado, e se a medida de similaridade exceder um certo limite, o segmento correspondente é declarado para ser detetado. O tipo de sinal detetado coincide com o tipo padrão de referência atribuído. Além disso, este método tornou possível determinar a duração de eventos sísmicos (comprimento do segmento do registo digital), que correspondia ao comprimento do padrão de referência.

Numa nova abordagem, [Massa, Ferretti, Spallarossa, and Eva, 2006] melhoraram o desempenho de um procedimento automático empregado para um sistema "quase em tempo real", para a localização de eventos sísmicos no Sudoeste de Itália. Para isso, desenvolveram um processo baseado na análise da similaridade de forma de onda usando apenas uma estação sísmica. Para detetar "*earthquake families*", utilizaram uma técnica de correlação cruzada num conjunto de ondas sísmicas registadas no período de 1985-2002, numa região localizada no sudoeste dos Alpes (Itália). Utilizaram cerca de 2700 eventos sísmicos, selecionados com base na relação sinal ruído(SNR).

Com a análise da similaridade da forma de onda, conseguiram classificar cerca de 65% dos eventos selecionados em 80 grupos *earthquake families*, localizados dentro da área considerada. Para cada grupo fizeram a seleção de um membro principal, representante que foi realocado manualmente.

Depois com a correlação cruzada com os elementos dos grupos, determinaram o grupo a que pertence cada novo evento a ser processado, e por fim atribuem a localização do representante do grupo ao novo elemento. Em relação a performance, de 104 eventos 50% das associações foram feitas corretamente.

Novamente fazendo uso da correlação de forma de onda, [Gibbons and Ringdal, 2006] explicaram a deteção de eventos sísmicos de baixas magnitudes. Disseram que sinais de

pequenos eventos são mais efetivamente detetados pela correlação cruzada do fluxo de dados de entrada com modelos de forma de onda. Disseram que sinais de pequenos eventos são mais efetivamente detetados pela correlação cruzada do fluxo de dados de entrada com modelos de forma de onda.

[Kueperkoch, Meier, Lee, Friederich, and Grp, 2010] apresentaram um algoritmo automático para determinar, o tempo de chegada da fase S de eventos, locais, regionais e telessismos, baseando-se na previsão autorregressiva (AR) da onda. As formas de ondas das componentes horizontais foram previstas usando um modelo escalar AR para gravações de várias componentes. Estimaram os coeficientes de AR de uma pequena janela em movimento utilizando, uma abordagem de mínimos quadrados, minimizando o erro de predição para a frente. Fizeram testes sintéticos com um único componente de dado, mostrando que a abordagem de mínimos quadrados produz resultados semelhantes, ou mesmo melhores do que algoritmos *YuleWalker* e algoritmos *Burgs*.

Discutiram a escolha do modelo de AR e mostraram que o erro de predição, correspondente ao modelo de AR aplicado a ambas as componentes horizontais é suficiente para detetar mudanças instantâneas em amplitude, fase, frequência e polarização. O erro de predição(rms) de ambas as componentes horizontais definiu a função característica aplicada no algoritmo para estimar o tempo de chegada da fase. O algoritmo proposto é também responsável pela avaliação automática da qualidade dos tempos de chegada estimados. Quatro critérios de qualidade foram usados para definir automaticamente o peso do tempo de chegada estimado, basearam-se em duas estimativas diferentes da inclinação da função característica, e em duas relações sinal ruído (SNR).

2.5 Métodos Recursivos

[Beyreuther and Wassermann, 2008a] apresentaram uma nova técnica para a da detecção automática e classificação de eventos sísmicos numa única etapa. Utilizaram os modelos ocultos de Markov (HMM) uma técnica originalmente desenvolvida para o problema de reconhecimento de voz mas muito prometedora em relação a detecção de eventos sísmicos induzidos pelo vulcão de acordo com [Beyreuther and Wassermann, 2008b].

Analisaram a sua utilização, num conjunto contínuo e limitado de dados, da rede sísmica de Bavaria. No final, apresentaram uma comparação dos resultados obtidos com os obtidos pelas técnicas de detecção simples, como por exemplo o STA/LTA. De forma geral, consideraram o desempenho dos modelos ocultos de Markov discretos (DHMM), como sendo bom para tarefas de classificação pré-trigger e apresentaram resultados razoáveis nas classificações no modo

contínuo. Em relação a performance do DHMM, comparando o com um detetor STA/LTA, apresentaram a seguinte tabela, como podemos ver na figura seguinte:

	DHMM		rec. STA/LTA	
Missed events	13	(19 per cent)	7	(10 per cent)
Detected events	56	(81 per cent)	62	(90 per cent)
Total earthquakes	69	69		

Figura 2.2: Comparação das taxas de deteção (sem classificação) entre DHMM e o modelo recursivo STA / LTA no período de um mês de forma contínua. Fonte: [Beyreuther and Wassermann, 2008a]

[Beyreuther, Carniel, and Wassermann, 2008] para problema de deteção e classificação de eventos, sísmicos vulcânicos/tectónicos na ilha de Tenerife, adotaram novamente metodologias de reconhecimento de fala. Como uma das motivações tinham a similaridade existente entre sinais sismogramas e sinais de fala. Nesse artigo, usaram HMM para detetar/classificar eventos sísmicos a partir de dados sísmicos registados de forma contínua. Com os HMM conseguiram boas taxas de deteções e relações sinal ruído (iteração entre o ruído sísmico vulcânico e tectónico), por isso consideram o sistema como sendo uma boa opção para sistemas de alerta antecipada(EWS). Analisaram resultados de um período de classificação contínuo de 1 mês começando em, 1 de Setembro de 2004 e terminando em 30 de Setembro de 2004. Conseguiram 74% de classificações corretas e 24% de classificações erradas.

Num novo estudo, [Beyreuther, Hammer, Wassermann, Ohrnberger, and Megies, 2012], apresentaram o projeto de um detetor sísmico probabilístico que opera numa única estação sísmica. Apesar de limitações, o desempenho do método desenvolvido foi pouco menor que o baseado na razão STA/LTA, a partir de informações de três ou mais estações. O detetor desenvolvido ofereceu uma alternativa valiosa e promissora, apresentando uma arquitetura probabilística baseada no HMM. Na análise apresentada em [Beyreuther, Hammer, Wassermann, Ohrnberger, and Megies, 2012] sabendo que a deteção de eventos sísmicos de fluxos contínuos de dados sísmicos, ser uma das questões chaves para um sistema sísmico de monitoramento automático ou semiautomático, geralmente as redes sísmicas são constituídas por um conjunto de pequenas estações, a maioria com algoritmos baseados nos níveis de decisão implementados em grandes números de estações ativas. Sendo assim, para o caso de algumas estações não estarem disponíveis, ou ocorrerem eventos pequenos, por exemplo em monitorização de vulcões ou estações hidrotérmicas, esses algoritmos baseados nos níveis de decisão frequentemente apresentam elevados números de falsos alarmes, por isso algoritmos de deteção, que mostram um desempenho razoável ao operarem mesmo numa única estação

sísmica são de bastante interesse.

2.6 Métodos baseados em Redes Neurais Artificiais

A identificação correta de eventos sísmicos continua a ser um aspecto crítico, para a confiabilidade na análise de dados sísmicos de forma automática permitindo projetar sistemas em tempo real. A classificação de sinais sísmicos tem sido tratada em quase todos os casos como sendo um processo de duas etapas. A tarefa é dividida, no problema de detecção simples e seguidamente, no problema de caracterização dos segmentos de tempo detetados em classes de eventos, como podemos confirmar nos trabalhos descritos de seguida.

Vários trabalhos têm como finalidade detetar diferentes tipos de eventos sísmicos com base nas características apresentadas por cada um, principalmente identificar sismos de origem natural e explosões artificiais.

[Dowla, Taylor, and Anderson, 1990] aplicaram uma RNA para discriminar entre eventos sísmicos naturais e explosões artificiais, usando distâncias corrigidas de informações espectrais das fases (Pg, Pn, Lg) de eventos sísmicos regionais. Implementaram a RNA na estação de trabalho usando uma MLP com o algoritmo *backpropagation* (BP), com uma estrutura simples (82 unidades de entrada, 1 unidade escondida, 2 unidades de saída). Atribuíram uma rede neuronal para cada estação sísmica de forma a regionalizar os dados. Para cada estação regional obtiveram um reconhecimento de 93% de sismos e explosões. Já para uma rede de 4 estações obtiveram um desempenho de mais de 97%.

[Wang and Teng, 1995] estudaram duas RNAs supervisionadas que diferem entre si nos tipos de entradas, no primeiro caso utilizaram a razão STA/LTA como entrada da rede, no outro caso utilizaram espectrogramas.

Em relação a resultados mostraram que a precisão das RNAs é melhor que os algoritmos convencionais (STA/LTA), especialmente para sinais com baixa SNR ou ruídos em forma de picos. Com as experiências comprovaram que estes sistemas superam em desempenho os algoritmos baseados no nível de decisão STA/LTA.

Novamente fazendo uso das RNAs, [Wang and Teng, 1997] estudaram a aplicação de uma MLP para a determinação dos instantes de chegada da fase S em sismos locais e regionais. Usaram dados produzidos por uma estação de três componentes e deles extraíram diferentes características, concretamente, a razão STA/LTA, a razão entre a potência horizontal e a potência total, coeficientes de auto-regressão e o menor eixo da elipsoide de polarização. Consideraram segmentos de registos sísmicos de 2.56 segundos de duração, que são previamente filtrados na banda de 2 a 8 Hz. Em relação aos resultados, os melhores foram 86% de

identificações corretas e 74% de determinações de tempos de chegada da fase S, com erro menor que 0,1 segundos.

Em 1996, [Dai and MacBeth, 1997] aplicaram uma RNA para identificar de forma automática, instantes de chegada das fases S e P a partir de um único componente de gravação. Apresentaram resultados de uma análise preliminar onde testaram a aplicação da RNA para eventos sísmicos após serem classificados. Como solução para o problema, utilizaram o grau de polarização (*“degree of polarization”*) para 3 componentes na série de tempo de um segmento, como entrada da RNA. O estudo teve como dados 877 registros pré-selecionados por um sistema de classificação, mas para o treino da rede apenas usaram nove instantes de chegada das fases P e ruído. Classificaram os eventos em 3 grupos, instantes de chegada das fases S, instantes de chegadas das fases P e ruído. Compararam a performance do sistema com a detecção manual. O sistema classificou corretamente 84% das fases P e 63% das fases S.

Já em 1997, [Dai and MacBeth, 1997] propuseram uma aproximação para determinar de forma automática, o instante de chegada das fases P e S, usando MLP treinada com o algoritmo BP, numa topológica com quarenta entradas e duas saídas. Recebiam na entrada valores absolutos da amplitude, extraídos de segmentos normalizados de 40 amostras de segmentos sísmicos. Definiram as saídas (0,1), determinadas aplicando uma janela deslizante nas duas saídas formando uma série temporal, depois identificaram as fases utilizando um nível de decisão e uma regra de máximo local. Em relação aos resultados, para as fases P as taxas de determinações bem-sucedidas variaram entre os 83% e 95%. Para o caso das fases S, as taxas variaram entre 75% e 61% para as determinações bem-sucedidas. Mas quando consideravam um erro de 10ms, os valores passaram a ser 63 % e 66% para os instantes de chegada das fases P, e para as fases S de 53% e 61%. Ainda neste ano os autores propuseram uma MLP, treinada com o algoritmo BP, que identificava as fases de sismos locais em sismogramas de três componentes.

O sistema desenvolvido, após selecionar características dos segmentos de registro contendo fases, P e S, e ruído, com base no DOP criaram um espaço de características com essas informações. Assim, tinham os valores entre 0 e 1, sendo os valores próximos do zero representativos do ruído, os valores perto do 1 representativos das fases S e valores intermédios representativos das fases P.

O sistema depois de treinado classificava corretamente entre 76,6% e 82,3% das fases P, e entre 60,5% e 62,6% das fases S.

[Tiira, 1999] treinou RNAs para reconhecerem sinais sísmicos utilizando um conjunto de

dados, previamente analisados e amostras de ruídos gravados em 3 estações sísmicas de curto período na Finlândia. No estudo, fez comparações entre diferentes configurações da mesma rede neuronal, e também diferentes tipos de redes neurais. Utilizou no estudo MLPs com topológicas de, uma camada escondida, duas camadas escondida, e também com uma camada linear de saída, sendo que a última topológica demonstrou claramente uma fraca performance em relação aos demais.

Também utilizou redes parcialmente recorrentes, que demonstraram uma performance inferior, aos MLPs, mas o autor propôs a sua utilização como um pós detetor, onde as entradas seriam as saídas de outros tipos de detetores combinando as melhores características de cada um destes.

A configuração que acabou por prevalecer foi a de uma MLP com uma camada escondida, demonstrando uma maior capacidade de detecção no nível mais baixo de decisão, que usava o algoritmo de aprendizagem LMS (Least Mean Square) de modo supervisionado. A entrada da rede consistiu em quatro janelas de valores diferentes de STA/LTA, calculados em sete bandas de frequência diferentes.

Com o objetivo de discriminar entre sismos locais e explosões na região do golfo de Suez (Egito), [Abu-Elsoud, Abou-Chadi, Amin, and Mahana, 2004] treinaram uma MLP com o algoritmo BP. Extraíram as características aplicando técnicas de predição linear. Trabalharam com um conjunto de dados constituído por 320 eventos utilizados para treino e validação, sendo 142 explosões e 172 sismos locais. Nos resultados obtiveram 93.7% de classificações corretas.

[Porto, Fogel, and Fogel, 2004] desenvolveram um algoritmo usando computação evolutiva (CE), para a detecção de sinais sísmicos em residências. Foi gerado um sinal de resposta tempo/frequência (TFR) por um conjunto de sensores sísmicos que também foram usados para monitorizar sinais gerados por uma pessoa, grupo de pessoas, e veículos de diferentes tipos.

Foram escolhidas redes neurais para a classificação, e computação evolutiva para projetar e treinar essas redes. A topologia da rede neuronal, função de ativação e pesos foram todos otimizados simultaneamente pelo algoritmo CE durante o processo de treino. Deram uma particular atenção aos algoritmos evolutivos aplicados a estrutura de dados alternativos (como conjuntos fuzzy, crisp) para mapear dados de entrada em saídas desejadas. O desempenho do sistema foi determinado como de 76% de classificações corretas.

[Benbrahim, Daoudi, Benjelloun, and Ibenbrahim, 2005] desenvolveram um discriminador entre sismos, regionais, naturais e explosões artificiais, recorrendo a uma MLP. Para extrair as informações utilizaram representações tempo-frequência (espectrogramas) e representações

de tempo. Experimentaram quatro algoritmos distintos. Num universo de dados constituído por 45 sismos regionais e 45 explosões obtiveram entre 82.5% e 88.75% de classificações corretas na validação das quatro variantes.

[Giudicepietro, Esposito, Martini, and Scarpetta] num relatório apresentado no, (*"Istituto Nazionale di Geofisica e Vulcanologia Sezione di Napoli (Osservatorio Vesuviano)"*), abordaram o problema de classificação de eventos sísmicos aplicando redes neuronais para eventos sísmicos registados na área vulcânica Napolitana (Ilha vulcânica Stromboli, Monte Vesúvio, Campo Flegrei). Tentaram implementar sistemas capazes de identificar, deslizamentos de terra, explosões, sinais microssismos entre outros. Consideraram a predição linear LPC (*Linear Predictive Coding*)(Características espectrais), e a parametrização da forma de onda (*Waveform Parameterization*)(Características temporais), para caracterizar no tempo e na frequência eventos sísmicos. Para a tarefa de classificação aplicaram uma MLP de forma supervisionada, e também apresentaram uma análise não supervisionada utilizando o Kohonen Self-Organizing Map (SOM). **Áreas de aplicação e resultados:**

(Phlegraean Fields) [Del Pezzo, Esposito, Giudicepietro, Marinaro, Martini, and Scarpetta, 2003] tem como o objetivo distinguir entre eventos artificiais (explosões submarinas geradas por pescadores na baía de Pozzouli) e os sismos Vulcânicos/tectónicos. Para a extração das características, aplicaram o método LPC fazendo a codificação usando apenas características espectrais. Na classificação utilizaram um método baseado numa MLP. Obtiveram uma taxa de classificação correta de 92% , demonstrando uma boa capacidade de generalização.

Vesuvius) Monte Vesúvio" [Scarpetta, 2005] distinguiu entre um conjunto de sinais registados numa estação sísmica em várias categorias.

Classes: sismos, explosões submarinas (artificiais), explosões em cavernas piro clásticos, explosões em cavernas calcárias ("cavernas = pedreiras"), trovões. Na fase de processamento usou as características espectrais e temporais. Comparativamente com o trabalho desenvolvido

Classification Task	Performance
earthquakes/quarry blasts - limestone (NL9)	100%
earthquakes/underwater explosions (CPV)	99%
earthquakes/thunders (BKE)	98%
earthquakes/quarry blasts - pyroclastic (TRZ)	95%

Figura 2.3: Performance da rede para o conjunto de dados do monte Vesúvio. Fonte :[Scarpetta, 2005]

em, [Del Pezzo, Esposito, Giudicepietro, Marinaro, Martini, and Scarpetta, 2003] , concluiu que a performance após juntar as características espectrais e temporais melhora.

(Stromboli) [Esposito, Giudicepietro, Scarpetta, D'Auria, Marinaro, and Martini, 2006] desenvolveram um sistema neuronal, para diferenciar sinais sísmicos em 3 tipos. *Explosões, deslizamento de terra e sinais micros sísmicos.*

Avaliaram a performance colocando os eventos em quatro classes: *Deslizamentos / explosões, microssismos / explosões, deslizamentos / microssismos, e juntando todas as classes deslizamentos /explosões/ microssismos.* A performance é dada de acordo com a Figura 2.4. No mesmo trabalho os autores ainda testaram um método não supervisionado, SOM.

Classification Task	Performance %
Explosion-quakes/Landslides	99.7%
Landslides/Microtremor signals	96.5%
Explosion-quakes/Microtremor signals	99.6%
Landslides/Explosion-quakes/Microtremor signals	97.2%

Figura 2.4: Performance da rede para o conjunto de dados do monte Vesúvio. Fonte :[Esposito, Giudicepietro, Scarpetta, D'Auria, Marinaro, and Martini, 2006]

Com este método conseguiram definir três conjuntos de dados consistentes (as três classes pretendidas).

[Gentili and Michelini, 2006] apresentaram um sistema que permitiu determinar de forma rápida e eficiente, instantes de chegadas das fases P e S, sendo o método aplicado baseado em redes neurais. Consideraram que os instantes de chegada das fases P e S eram linearmente separáveis. As fases foram determinadas utilizando apenas um *perceptron*. A rede neuronal adotada foi uma árvore neuronal, chamada IUANT2 caracterizada por possuir uma grande capacidade de generalização, permitindo determinar a melhor topologia e configurar os parâmetros iniciais da rede. Aplicaram este método a 342 sismos gravados por 23 estações. Em comparação com os métodos manuais, a técnica apresentada é geralmente robusta e capaz de identificar corretamente diferentes tipos de fases e as suas estimativas. Sendo as vantagens do método que propuseram, a velocidade de computação, grande percentagem de sismos detetados e precisão na deteção do início dos picos das fases. Em relação aos resultados os dados foram apresentados de acordo com várias situações. Com o sistema conseguiram resultados médios de $95.0 \pm 2.7\%$ de classificações corretas nos conjuntos de testes.

Ainda, [Gentili and Michelini, 2006] descreveram um sistema de rede neuronal, (utilizando

dois MLPs), chamado TAPNEI (*Time Arrival Picker for Northeastern*) (Figura 2.5) que permitiu detectar instantes de chegada das fases P e S no Nordeste de Itália. Usaram ao nível de cada estação sísmica o algoritmo STA/LTA para reconhecer os candidatos, os dados foram adquiridos nas estações e depois processados. Antes do processamento fizeram um pré-processamento baseado na função estatística curtose (*Kurt*).

O conjunto de dados correspondia a 7108 sismogramas de 1147 sismos ocorridos no nordeste de Itália (2000 - 2003). A maior parte das análises baseavam em 3 medidas estatísticas: Variância; Assimetria; Curtose. O sistema que implementaram reconheceu 89% e 67% dos tempos de chegada das fases, P e S. Com estes resultados concluíram ter um melhor sistema do que o sistema em utilização, mas ainda sem atingir a precisão da análise manual. [Masotti,

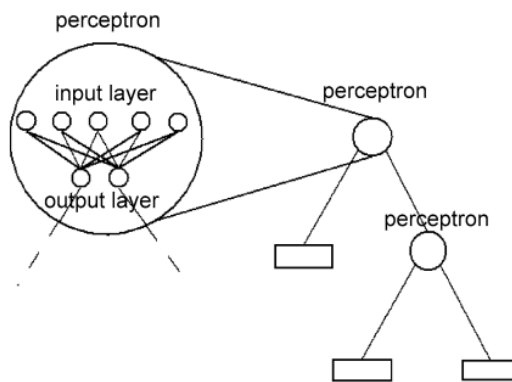


Figura 2.5: Esquema de uma árvore neural, correspondente 2 perceptron's com 2 saídas cada. Fonte :[Gentili and Michelini, 2006]

Falsaperla, Langer, Spampinato, and Campanini, 2006] aplicaram a SVM de modo a classificar dados de sismos vulcânicos, registados durante vários estágios de atividade do vulcão Etna na Itália. Treinado com um algoritmo supervisionado, o classificador aprendeu a reconhecer padrões pertencentes a quatro classes: *pré-eruptivos*; *fontes de lava*; *eruptivos*; *pós-eruptivos*; Durante a fase de treino e validação, utilizaram 425 padrões baseados na média das espectrogramas. Com o SVM classificaram corretamente $94,7 \pm 2.4\%$ dos dados. Desta forma, conseguiram menos do que 6% de classificações erradas, pelo que a abordagem para a discriminação de sismos vulcânicos usando SVM é bastante interessante.

[Curilem, Vergara, Fuentealba, Acuña, and Chacón, 2009] apresentaram um trabalho com o objetivo de construir um sistema capaz de classificar sinais sísmicos para o vulcão Villarrica (Chile). Três tipos de sinais foram considerados como os mais representativos da atividade deste vulcão: sinais de sismos energéticos; sismos; sinais de longos períodos. Para a classificação a amplitude do sinal e o conteúdo de frequência, foram as características mais importantes levadas em conta. Implementaram o classificador utilizando MLPs, cujas arquite-

turas e processos de formações foram otimizadas por meio de algoritmos genéticos. Procuraram a configuração mais adequada para a MLP com esta técnica, melhorando o desempenho da classificação, otimizando o número de camadas escondidas, as funções de transferências dos neurónios e o algoritmo de treino. Em relação aos resultados, mostraram que o classificador otimizado, atinge mais de 93% de exatidão, identificando sinais de cada um dos grupos.

[Langer, Falsaperla, Masotti, Campanini, Spampinato, and Messina, 2009] investigaram o desenvolvimento das características dos sismos e as suas relações com os regimes de atividade vulcânica, aplicando técnicas de reconhecimento de padrões. Apresentaram resultados para a classificação de forma, supervisionada e não supervisionada, num conjunto de dados constituído por 425 padrões de eventos sísmicos vulcânicos, registados entre julho e Agosto de 2001.

SVM e MLP, foram utilizados como classificadores de forma supervisionada. Para o treino da SVM e MLP consideraram quatro classes : *pré- eruptivos, fontes de lava, eruptivos, pós-eruptivos*; Com a SVM alcançaram uma performance de 94,8%, e com o MLP 81,9% . Realizaram a classificação não supervisionada através da análise de agrupamentos (CA), e SOM , conseguindo resultados bastante interessantes. Com isto concluíram que os métodos de classificação supervisionados e não supervisionados, podem ser ambos convenientemente implementados para a compreensão em profundidade das relações entre os dados sísmicos e fenómenos vulcânicos.

[Giacco, Esposito, Scarpetta, Giudicepietro, and Marinaro, 2009] aplicaram e compararam duas técnicas de forma supervisionada, nomeadamente o MLP e a SVM, para classificar sinais sísmicos registados na estação sísmica do monte de Stromboli. Primeiramente fizeram o pré-processamento dos dados para obter informações espectrais e temporais construindo vetores com um conjunto de componentes.

Implementaram as duas técnicas para diferenciar 3 classes de eventos sísmicos; *sinais microssismos vulcânicos, explosões/ sismos, deslizamentos de terra*. No primeiro método aplicaram uma rede de duas camadas MLPs, com uma função de erro de entropia cruzada (cross entropy) e função lógica de ativação. No segundo método utilizaram SVM, onde a constituição das várias classes foi por uma arquitetura para classificação multi-classes, com decomposição um contra todos com um núcleo gaussiano.

Com as experiências mostraram que apesar do MLP produzir resultados muito bons, a SVM apresenta muito melhor performance tanto em termos da melhor performance 99,5%, quer em termos da performance média 98,8%, obtidos para amostras, conjuntos de treinos e conjuntos de validação diferentes.

[Orlic and Loncaric, 2010] propuseram uma nova abordagem para a classificação de sismogramas. O objetivo era classificar sismogramas em dois grupos, um constituído por eventos sísmicos naturais e outro por eventos sísmicos artificiais (explosões). Utilizaram um algoritmo genético especialmente formulado para identificar características nos sismogramas de forma quase ótimos, em vez de usar os métodos correntes de extrair características (componentes espectrais e de frequência).

Num conjunto de sismogramas aplicaram o algoritmo genético para descobrir, um conjunto de características quase ideais. Este conjunto identificado pelo algoritmo foi então utilizado para o processo de classificação. Em relação aos resultados fizeram a verificação e validação do método, conseguindo taxa de classificação correta de 85%.

[Diersen, Lee, Spears, Chen, and Wang, 2011] utilizaram RNA e IANN para, a classificação dos segmentos de onda. Com isso eliminaram o custo computacional do processo de classificação eliminando também a necessidade de um perito para supervisionar as classificações. Conseguiram isto combinando a transformada de *wavelet* continua (CWT) e redes neuronais. Com a CWT analisaram formas de onda no domínio do tempo e frequência, depois aplicaram uma RNA para selecionar bons segmentos de janelas sísmicas dentro do algoritmo de tomografia de ondas completas. Fizeram experiências com dados sísmicos do sul de Califórnia mostrando que, esta técnica é promissora para a exatidão de classificação, e para a redução do tempo necessário para classificação dos dados. Os resultados tiveram como conjunto de dados 504 exemplos de eventos.

	ANN	IANN
Correct	500	502
False Negatives	0	0
False Positives	4	2
Total	504	504
Percentage Correct	99.21	99.60
Percentage Error	0.79	0.40

Figura 2.6: comparação dos resultados experimentais entre RNA e IANN Fonte :[Diersen, Lee, Spears, Chen, and Wang, 2011]

[Kislov and Gravirov, 2011] analisaram o problema de identificação de eventos sísmicos num nível elevado de ruído, a partir de dados de um único sensor ou um grupo local de sensores. Identificaram como sendo a diferença fundamental do sistema em relação aos existentes a análise dos dados num grande nível de ruído tecnológico, quando o nível de ruído excede o

nível do sinal sísmico.

Realizaram quatro blocos de experiências;

Primeiro, experiências computacionais para a identificação de fases baseando em dados reais de pequenos sismos com ruído de fundo. Segundo, experiências para determinar as alterações nas características dos sinais dependendo da localização do sensor. Terceiro, acumulação de dados em sismos desastrosos. Quarto, verificação da estabilidade das características dos eventos sísmicos de acordo com o sinal detectado.

Investigaram a seleção de informações relativamente as características dos sinais de forma a reduzir a quantidade de recursos necessários para o melhor funcionamento do sistema. Deram maior atenção a técnicas baseadas em redes neurais, para analisar a separabilidade das classes nos seus espaços de atributos. Tinham como interesse classificar sinais em apenas uma classe (classe de sismos), separando-os de todos os outros sinais que são considerados ruído.

Apresentaram várias configurações para os dados de entrada. Primeiramente utilizaram a amplitude do sinal (1024 valores) diretamente nos neurónios da camada de entrada. Na segunda representação, em adição a primeira inseriram o espectro do sinal a neurónios adicionais. Na terceira variante do sistema usaram a representação da WT como sinal de entrada para a rede neuronal, assim reduzindo a dimensionalidade da camada de entrada da RNA.

Consideraram o sistema que apresentou o menor número de erros o mais eficiente. Obtiveram melhor desempenho no *Perceptron* com uma camada escondida, e combinando a terceira variante, e a segunda variante na configuração.

2.7 Resumo

Neste capítulo é apresentado a pesquisa de estudos sem aplicação de Inteligência Computacional (métodos Algorítmicos) e estudos com aplicação de Inteligência Computacional, para o processo de deteção sísmica.

Primeiramente, começamos com trabalhos onde a partir de uma representação de dados no tempo, faz-se uma avaliação das amplitudes dos sinais como forma de classificar os eventos sísmicos.

Depois são descritos métodos que aplicam a análise multi-resolução (MRA) ou aproximação multi-escala (MSA) no processamento de informações sísmicas e também na sua combinação com outros métodos.

Após isso, apresentamos outros métodos que não aplicam técnicas fundamentadas no uso de inteligência computacional. Baseiam-se por outro lado no uso da correlação, auto-

regressão entre outras técnicas.

Também são apresentados métodos recursivos discretos, concretamente nos Modelos ocultos de Markov (HMM).

No final são descritos os métodos baseados na aplicação de inteligência computacional, com particular destaque para a utilização de RNA, SVM.

Capítulo 3

Máquina de Vetores de Suporte

No presente capítulo são discutidos os conceitos de base do funcionamento de um classificador designado por, Máquina de Vetores de Suporte (SVM)¹, técnica de Aprendizagem de Máquina que vem recebendo grande atenção nos últimos anos sendo aplicada em diferentes áreas.

Durante a análise e apresentação do algoritmo é importante referir, brevemente, a teoria que está na base da SVM, os principais aspetos a ter em conta de forma a conseguir os melhores desempenhos nas aplicações da mesma.

O algoritmo SVM constitui uma técnica com as bases na Teoria de Aprendizagem Estatística, descrita na Secção 3.5, e pretende construir um hiperplano como superfície de decisão, de forma a um conjunto de padrões sejam separados em classes diferentes. A SVM é utilizada em problemas que vão desde o reconhecimento de caracteres escritos, bioinformática, visão computacional a análise de séries temporais, apresentando quase sempre resultados superiores aos alcançados por técnicas similares nas várias aplicações.

Nessa dissertação é utilizado a SVM para a classificação de forma supervisionada, mas na literatura também encontram-se trabalhos onde às SVMs são utilizadas em tarefas de aprendizagem não supervisionada e regressão (ver Capítulo 1).

3.1 Introdução

Como apresentado no Capítulo 4 onde são descritos os trabalhos anteriores, foram testados várias soluções para a classificação de eventos sísmicos [Madureira, 2009]. Das soluções apresentadas houve a necessidade de seleccionar para utilização o classificador que apresentou melhores características em termos de, performance, simplicidade, capacidade generalização

¹Máquina de Vetores de Suporte do inglês *Support Vector Macinhes*.

e adaptação para desafios futuros como a classificação contínua dos dados [Ruano, Madureira, Ozias, Khosravani, Ruano, and Ferreira, 2013, Madureira, Ruano, and Ruano, 2013b, Madureira and Ruano, 2012, Ruano and Madureira, 2011].

O classificador SVM ² foi o escolhido para esse trabalho, pelo seu melhor desempenho na detecção de eventos sísmicos.

O classificador SVM consegue ultrapassar o problema existente em muitos classificadores, de retenção em mínimos locais durante a aprendizagem, contribuindo para o alcance da melhor solução dentro do domínio que se lhe sobpõe [Rodrigues, 2008]. O algoritmo consegue desempenhos robustos na tarefa de reconhecimento de padrões usando conceitos bem estabelecidos na teoria de otimização, sendo alguma das suas características apresentadas a seguir.

3.2 Características da SVM

Citando [Smola, Bartlett, Schölkopf, and Schuurmans, 2000], [Lorena, 2003] apresenta as principais características que tornam o uso das SVMs atrativo:

- **Boa capacidade de generalização**

Em geral os classificadores gerados por uma SVM apresentam boa capacidade de generalização.

A capacidade de generalização é a eficiência na classificação de dados que não pertencem ao conjunto utilizado no seu treino. Como no problema de detecção sísmica o classificador, quando treinado, é aplicado em contínuo a dados não usados no seu treino, esta capacidade é de grande interesse para a aplicação.

Dessa forma evita-se o overfitting³ (sobre-ajuste), situação onde o classificador desenvolvido torna-se muito especializado no conjunto de treino, diminuindo a performance quando confrontado com novos dados.

- **Robustez em grandes dimensões**

A SVM é robusta perante objetos de grande dimensão, como é o caso de imagens.

- **Convexidade da função objetivo**

Esta é uma vantagem sobre, por exemplo, as RNAs, onde há presença de mínimos locais na função objetivo a ser minimizada. Para a SVM é feita a otimização de uma

²Normalmente os classificadores SVM são também referidos como redes SVM, ou algoritmo SVM, sendo que nesse documento são utilizados ambos os termos sem nenhuma distinção.

³Overfitting é o termo, em aprendizagem de máquina, estatística e afins, para quando o modelo estatístico se ajusta em demasiado ao conjunto de dados/amostra <http://en.wikipedia.org/wiki/Overfitting>)

função quadrática que possui apenas um mínimo global.

- **Teoria bem definida**

Podemos verificar isso com análise feita a história e teoria de base da SVM apresentado de seguida.

3.3 História e Teoria de Base

Esta secção é baseada em [Haykin, 1999]. Vladimir Vapnik e Alexey Chervonenkis desenvolveram a teoria Vapnik – Chervonenkis, também conhecida como teoria VC⁴, durante 1960 – 1990 [Vapnik and Chervonenkis, 1971, Cortes and Vapnik, 1995].

A teoria VC caracteriza propriedades de máquinas de aprendizagem⁵, que lhes permitem generalizar bem para dados desconhecidos.

Este foi a base e o ponto de partida para a SVM, uma técnica de aprendizagem supervisionada para classificação e regressão (na literatura também é utilizado de forma não supervisionada).

Os autores [Cortes and Vapnik, 1995, Smola and Olkoph, 2004] apresentaram uma análise mais profunda da SVM, analisando um conjunto de aspetos como, teoria VC, aprendizagem estatística, margem máxima em hiperplanos ótimos, "truques" do Kernel⁶.

Esta teoria não recebeu grande atenção por parte da comunidade científica, de início. Apresentando as tendências e controvérsias a volta da utilização das SVMs [Hearst, Schölkopf, Dumais, Osuna, and Platt, 1998] mostram que apesar de sua teoria ser relativamente antiga, as primeiras aplicações são recentes.

3.4 A ideia fundamental

Para explicar como a SVM funciona seguem-se as abordagens apresentadas em [Cortes and Vapnik, 1995, Haykin, 1999], e começa-se por um caso de padrões separáveis que poderá surgir num contexto de classificação de padrões.

Nesse contexto, a ideia principal da SVM é construir hiperplanos como superfícies de decisão, de tal maneira que a margem de separação entre exemplos positivos e negativos seja maximizada. Para ter uma ideia clara, podemos analisar a Figura 3.1, onde é representado um hiperplano ótimo, separando exemplos positivos e negativos com a margem máxima. A

⁴teoria VC, corresponde a Vapnik – Chervonenkis teoria do Inglês *VC theory*

⁵Máquinas de aprendizagem, do inglês *Learning machines*

⁶"Truques" do Kernel, do Inglês *kernel tricks*

posição ótima do hiperplano é determinada somente pelos exemplos próximos do hiperplano (vetores de suporte (SV)⁷).

A SVM consegue essa propriedade desejada, seguindo princípios enraizados na teoria da aprendizagem estatística que é discutida na Secção 3.5. A SVM é uma implementação aproximada do conceito de minimização do risco estrutural (SRM)⁸ discutido na Secção 3.5.1.

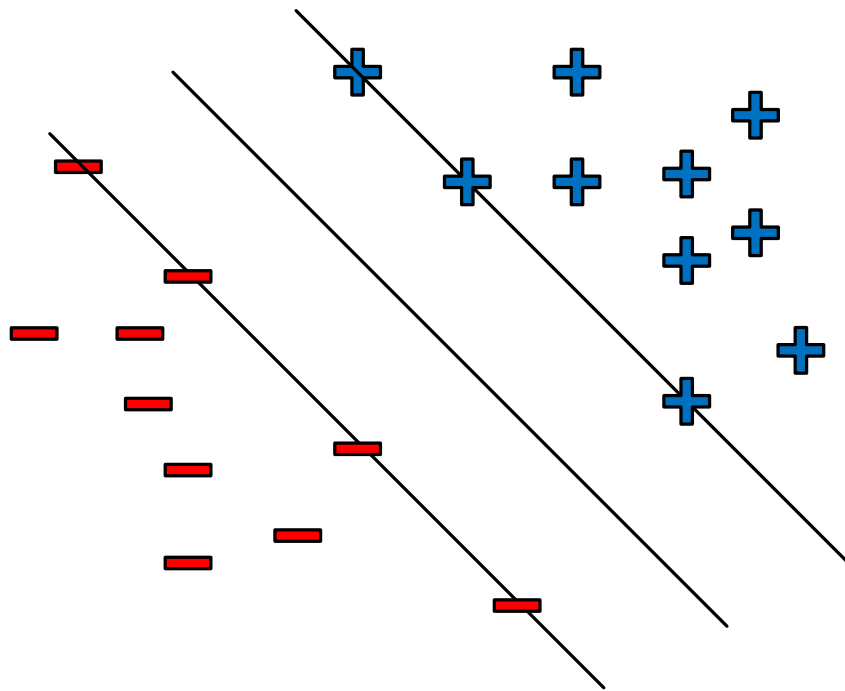


Figura 3.1: O hiperplano ótimo, separando exemplos positivos de negativos.

3.5 Noções básicas da teoria de aprendizagem estatística

Como referido anteriormente, às SVMs baseiam-se na teoria de aprendizagem estatística. Por isso, antes de entrar nos detalhes da SVM, é importante ter uma pequena perspetiva da teoria de aprendizagem estatística⁹.

3.5.1 A teoria da aprendizagem estatística

De acordo com [Burges, 1998b, Vapnik, 2006] dadas n observações cada uma composta por um conjunto de dados de entrada-saída pares $\{x_i, y_i, i = 1, 2, \dots, N\}$, onde x_i são os padrões

⁷Vetores de suporte do Inglês *Support vectors*

⁸Minimização do risco estrutural do Inglês *Structural Risk Minimization – SRM*)

⁹STL, *Statistical Learning Theory*

de entrada e y_i representa a classe do padrão de entrada x_i , num processo de aprendizagem, um classificador f é um mapeamento funcional $x_i \rightarrow f(x_i, \alpha)$ de um conjunto de padrões $x_i \in \mathfrak{R}^m$, regulados pelos parâmetros de aprendizagem ajustáveis α . Para uma rede neuronal por exemplo, α corresponde aos pesos e bias.

A teoria de aprendizagem estatística, tenta definir propriedades que permitem obter um classificador f ótimo. Obter o classificador ótimo significa obter uma função que caracteriza f de forma que quando o classificador treinado é testado consegue classificar corretamente os padrões.

Em relação aos pares de dados de entrada-saída, [Cortes and Vapnik, 1995] assumem que existe uma distribuição de probabilidade $P(x, y)$ desconhecida, com função densidade de probabilidade *pdf*, $p(x, y)$, que são variáveis aleatórias, gerados de forma independente e identicamente distribuídos (i.i.d)¹⁰. Com base nessa premissa, os autores definem a aproximação do erro de risco (R) que pode ser determinado em termos de parâmetros de aprendizagem α como

$$R(\alpha) = \int_{x,y} \frac{1}{2} |y_i - f(x_i, \alpha)| p(x, y) dx dy \quad (3.1)$$

O objetivo da aprendizagem supervisionada é minimizar o risco funcional $R(\alpha)$. De forma geral a distribuição de probabilidade é desconhecida [Haykin, 1999].

Mas [Vapnik, 1995] mostra como obter uma aproximação do risco seguindo o conceito de risco empírico (R_{emp}). Este princípio baseia-se apenas nos dados de treino, o que o torna perfeitamente adequado a filosofia de implementação de redes neuronais [Haykin, 1999, Burges, 1998b]. O risco empírico não é nada mais do que a taxa de erro médio dos pares de dados de treino [Vapnik, 1995],

$$R_{emp}(\alpha) = \frac{1}{2n} \sum_{i=1}^n |y_i - f(x_i, \alpha)| \quad (3.2)$$

O somatório apresentado na Equação 3.2 é referido como sendo a perda, que é dependente da gama de saída. Sendo assim, utilizam uma inequação que permite estabelecer uma fronteira para o risco. Escolhendo o parâmetro η de forma a $0 \leq \eta \leq 1$. Assim [Vapnik, 1995] apresenta a desigualdade (3.3), que define a fronteira superior do risco com uma probabili-

¹⁰Variáveis aleatórias independentes e identicamente distribuídas, do Inglês *Independent and identically distributed random variables (i.i.d)*

dade de $1 - \eta$, onde h é um inteiro não negativo chamado VC.

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h(\log(2n/h) + 1) - \log(\eta/4)}{n}} \quad (3.3)$$

A dimensão VC (descrita em 3.5.2) é uma medida quantitativa da aprendizagem. Note-se ainda que a Equação 3.3 não depende da função distribuição de probabilidade $P(x, y)$. Ainda analisando a Equação 3.3, podemos perceber que o objetivo é ter um processo de aprendizagem com o mínimo de risco.

3.5.2 A Dimensão de Vapnik – Chervonenkis

A dimensão VC é um valor escalar inteiro que mede a capacidade de um conjunto de funções de classificação (Secção 3.5.1), revelando-se muito importante na teoria da aprendizagem estatística pois determina o limite superior de risco de um modelo de classificação.

Dado um conjunto de n pontos que podem ser caracterizados em 2^n formas possíveis, é possível encontrar um membro de um conjunto de funções que consegue perfeitamente separar o conjunto de pontos. Logo existem no máximo 2^n combinações para n padrões. Para ter uma melhor noção do conceito apresentado anteriormente podemos analisar o exemplo da Figura 3.2, onde existe um conjunto de 3 pontos num espaço bidimensional (\mathfrak{R}^2). A função de separação é uma linha reta, e para todas as possíveis combinações dos pontos, podemos perceber que qualquer membro do conjunto de funções pode separar dois pontos (digamos classe 1) a partir de um ponto restante [Vapnik, 1995].

A dimensão VC(h) de um conjunto de funções $f\{(x, \alpha)\}$ é definida como sendo o número máximo de pontos de treino que podem ser distribuídos pelos membros do conjunto de funções [Burges, 1998b, Cortes and Vapnik, 1995].

Dessa forma para o exemplo apresentado na Figura 3.2 é $3 = (2 + 1)$. No caso do exemplo, é possível encontrar três pontos que podem ser distribuídos por um conjunto de funções, mas isso já não é possível para quatro pontos. Considerando hiperplanos em \mathfrak{R}^n diz-se que a dimensão VC de um conjunto de hiperplanos orientados em \mathfrak{R}^n é $n + 1$ [Burges, 1998a].

3.6 Minimização do Risco Estrutural

O princípio de minimização de risco estrutural [Burges, 1998b, Haykin, 1999, Vapnik, 1995], tem como principal objetivo escolher a complexidade do modelo ótimo para um conjunto de

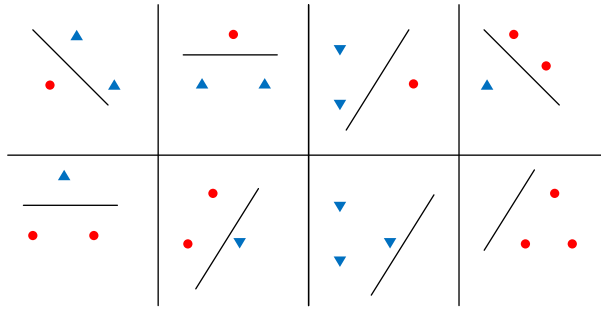


Figura 3.2: Distribuição de pontos num espaço bidimensional. Adaptado de: [Cortes and Vapnik, 1995]. Verifica-se que para todas as combinações possíveis, é possível traçar retas que separem perfeitamente as duas classes, nesse caso a classe círculo (azul), da classe triângulo.

exemplos de treino. Consideram-se conjuntos de estruturas (Figura 3.3), onde S_h é um espaço de soluções de dimensão VC h , de complexidade crescente definida:

$$S_1 \subset S_2 \subset S_3 \subset S_h \cdots \subset S_\infty$$

onde

$$h_1 < h_2 < h_3 \cdots h_h < h_\infty$$

As estruturas aumentam a medida que h aumenta, logo a capacidade também aumenta. O SRM consiste na resolução da seguinte equação:

$$\min R_{emp}(\alpha) + \sqrt{\frac{h(\log(2n/h) + 1) - \log(\eta/4)}{n}} \quad (3.4)$$

De onde se percebe que o SRM consiste na procura de um espaço de hipóteses que minimiza a combinação do risco empírico com o valor resultante da capacidade h .

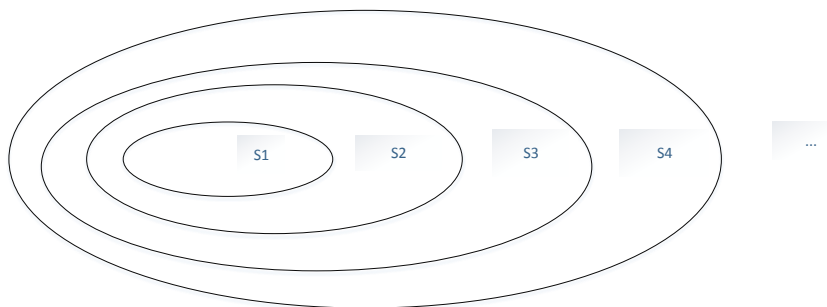


Figura 3.3: Subconjuntos de funções aninhadas, ordenados por dimensão VC. Adaptado de: [Cortes and Vapnik, 1995].

3.7 Máquina de Vetores de Suporte

Como referido anteriormente Secção 3.4, as SVMs podem ser utilizadas para a classificação de dados que são linearmente separáveis.

Para perceber de forma clara a estrutura e o funcionamento da SVM é preciso ainda definir quatro conceitos fundamentais: hiperplano de separação, máximo hiperplano de separação, funções do *kernel*, margem suave.

3.7.1 Classificador SVM Linear

Considerando um problema de classificação onde temos amostras de treino $\{(x_i, y_i)\}$, com $i = 1, \dots, N$, onde x_i representa o padrão de entrada da ordem i e o y_i é a respetiva saída desejada, (*target*), definindo o domínio para o problema, tem-se:

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\}$$
$$x \in \mathfrak{R}^n, y \in \{-1, 1\} \quad (3.5)$$

Um espaço será linearmente separável se é possível determinar pelo menos um par (w, b) tal que a Equação 3.6 consiga classificar corretamente todos os exemplos contidos no conjunto de dados.

Como apresentado na Figura 3.4, num espaço bidimensional a superfície de decisão que perfaz o hiperplano de separação será

$$w^T x + b = 0 \quad (3.6)$$

Onde x é a matriz de entrada, w é o peso ajustável e b corresponde ao vetor de *bias* (“termo compensador”).

Classificadores que separam os dados por meio de um hiperplano (Equação 3.6), são denominados lineares. A partir dessa Equação 3.6 o espaço de entrada é dividido em duas classes como apresentado nas Equações 3.7 e 3.8, sendo a classe $1(y_i = 1)$ dado pela Equação 3.7 e a classe $-1(y_i = -1)$ pela Equação 3.8 respetivamente.

$$w^T x + b > 0 \quad (3.7)$$

$$w^T x + b < 0 \quad (3.8)$$

Como referido anteriormente, desde que exista um par (w, b) que consiga separar corretamente os dados, temos um problema linearmente separável. Mas em geral há muitas combinações de pares de dados (w, b) capazes de separar corretamente os dados. Para cada par teremos uma equação, num caso como o anterior podemos ter um número infinito de soluções possíveis, ou seja, várias possibilidades para ser o hiperplano de separação [Vapnik, 1995], como podemos ver na Figura 3.5.

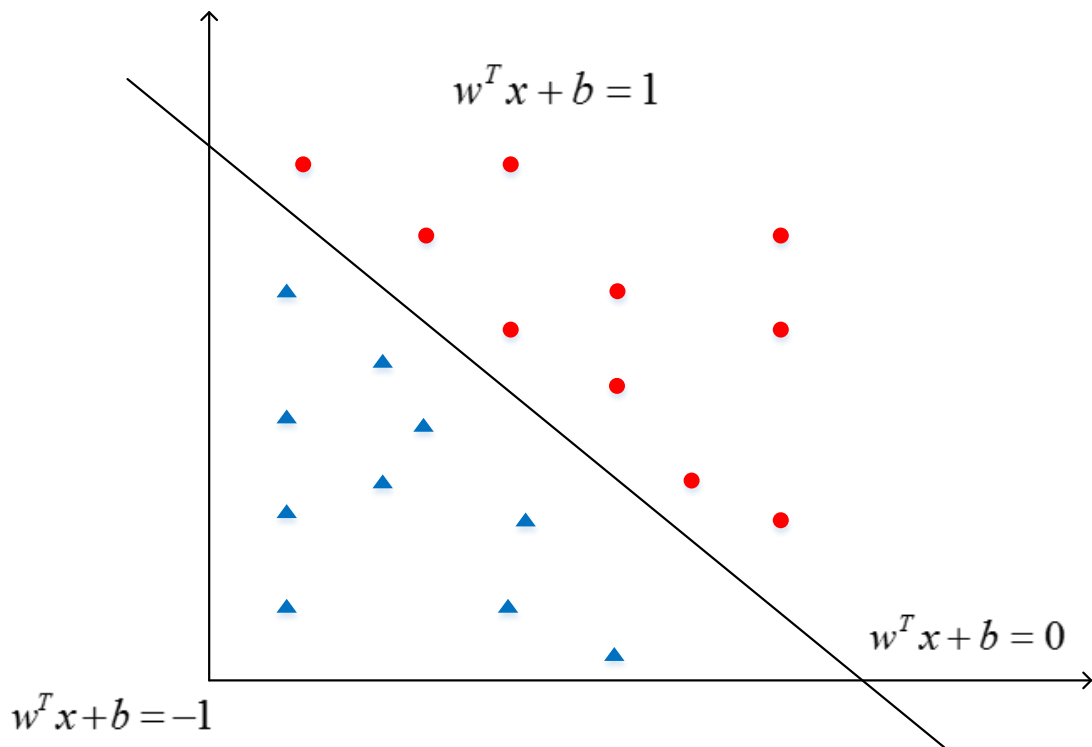


Figura 3.4: Exemplo de classificação de dados linear.

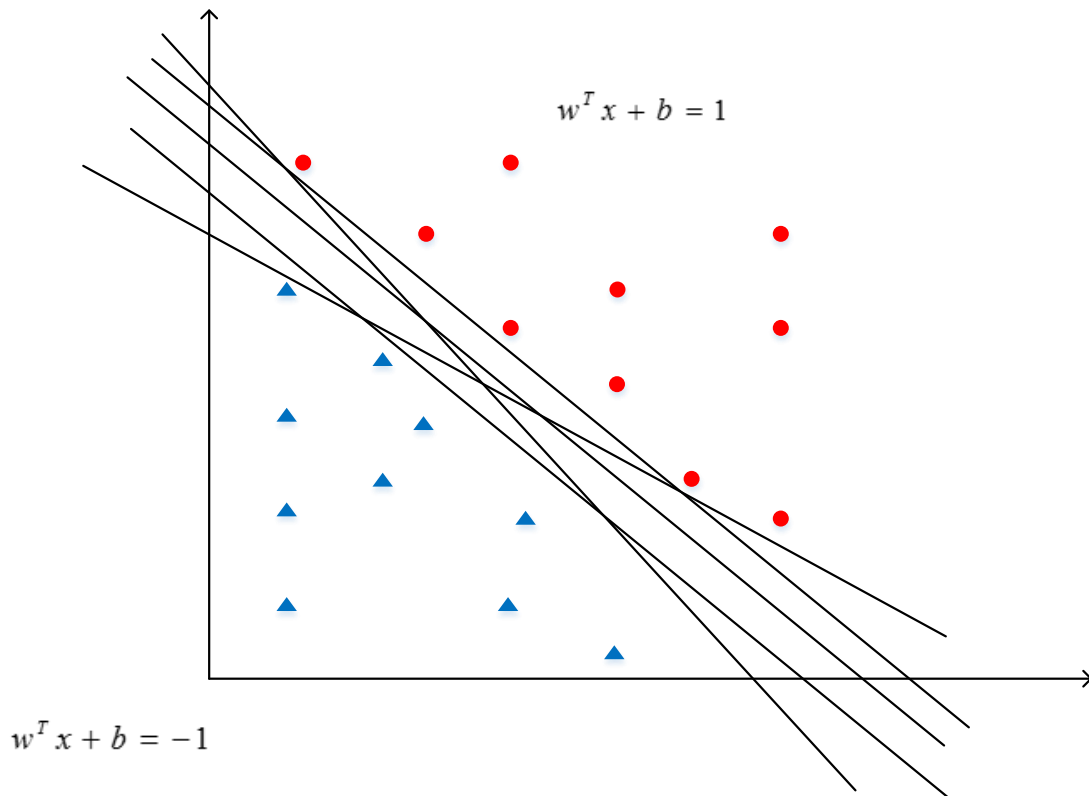


Figura 3.5: Exemplo de classificação de dados linear, múltiplas soluções.

3.7.2 Determinação do Hiperplano Ótimo

No exemplo apresentado na Secção 3.7.1, tivemos a oportunidade de ver que o hiperplano separador não é único. De um conjunto de vários possíveis hiperplanos, agora a questão que se coloca é, qual possui a capacidade de generalização?

Ou seja, conseguir depois do treino classificar não apenas os dados do treino, mas também padrões que não são conhecidos da fase de treino dentro do mesmo contexto de distribuição dos dados (Secção 3.2).

O objetivo da SVM é encontrar o hiperplano para a qual, a margem de separação é máxima. Nesta condição a superfície de decisão é referida como sendo o hiperplano ótimo. Podemos representar o hiperplano separador por meio de um produto interno, de acordo com a Equação 3.6.

Dum conjunto de classificadores que minimizam o risco empírico (Secção 3.6), deve-se escolher o classificador que possui a maior margem. O hiperplano que é caracterizado por essa margem é o ótimo.

[Ruano, 1996] mostra como determinar os parâmetros de forma a definir o hiperplano que gera a maior margem. Partindo da representação canónica do hiperplano apresentando nas

Equações 3.7 e 3.8, podemos expressar as equação da seguinte forma:

$$\begin{cases} w^T x_i + b \geq 1 & \text{se } y_i = 1. \\ w^T x_i + b \leq -1 & \text{se } y_i = -1. \\ i = 1, 2, 3, \dots, n. \end{cases} \quad (3.9)$$

Primeiramente é deduzido a distância de um ponto para o hiperplano de separação ótimo. Um hiperplano deve sempre, respeitar a restrição imposta pela Equação 3.9. A distância $d(w, b; x_i)$, distancia do ponto x_i ao hiperplano (w, b) é dada por:

$$d(w, b; x_i) = \frac{|w^T x + b|}{\|w\|} \quad (3.10)$$

A Figura 3.6 mostra um ponto qualquer x_i a uma distância d (Equação 3.10) de uma superfície, $w^T x + b = 0$. O hiperplano ótimo é alcançado maximizando a margem, ρ , sujeita as restrições da Equação 3.9. Na Figura 3.6 são definidos os limites que constituem a margem ρ .

O processo de determinação desses limites é apresentado posteriormente, mas de forma simples podemos dizer que são, definidos por pontos que são chamados de vetores de suporte (SV – dados que situam margem ótima determinada) na aprendizagem. Determinando a metade da distância entre os limites que definem a margem ρ temos o hiperplano separador.

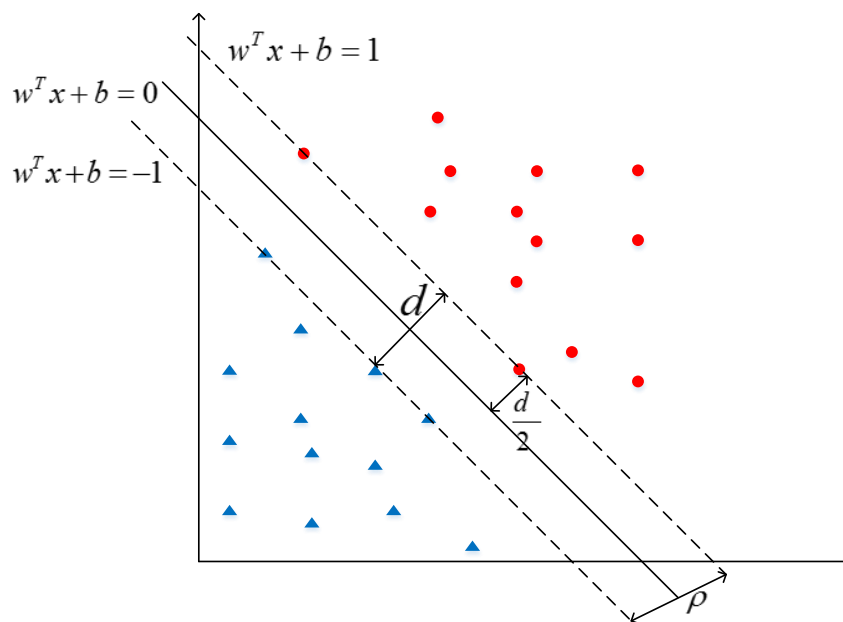


Figura 3.6: Representação da ideia de um hiperplano de separação ótimo, para padrões linearmente separáveis.

A margem ρ é então dada:

$$\begin{aligned}
 \rho(w, b) &= \min_{\{x_i: y_i = -1\}} d(w, b; x_i) + \min_{\{x_i: y_i = 1\}} d(w, b; x_i) \\
 &= \min_{\{x_i: y_i = -1\}} \frac{|w^T x_i + b|}{\|w\|} + \min_{\{x_i: y_i = 1\}} \frac{|w^T x_i + b|}{\|w\|} \\
 &= \frac{1}{\|w\|} \left(\min_{\{x_i: y_i = -1\}} |w^T x_i + b| + \min_{\{x_i: y_i = 1\}} |w^T x_i + b| \right) \\
 &= \frac{2}{\|w\|}
 \end{aligned} \tag{3.11}$$

De (3.11), vemos que maximizar a margem ρ corresponde a minimizar $\|w\|^2$. Podemos assim dizer que o hiperplano ótimo definido é único, e a condição ótima é alcançada minimizando a norma euclidiana do vetor peso w . Agora o problema de otimização pode ser formulado como:

$$\Phi(w) = \frac{1}{2} \|w\|^2 \tag{3.12}$$

Este problema restrito é chamado de problema primal¹¹. Essa função custo $\Phi(w)$ é uma função convexa de w e, portanto, pode-se encontrar sempre uma solução para o seu mínimo [Haykin, 1999].

A restrição da Equação 3.9 é de grande importância na simplificação da Equação 3.12. Para obter um hiperplano separador que é ótimo é preciso que os dados mais próximos de uma classe e os dados mais próximos de uma outra classe situem, a mesma distância do hiperplano.

Utilizando a Equação 3.12 na Equação 3.9, temos um problema de otimização com restrições que pode ser reescrito como:

$$\min \left(\|w\|^2 \right) \tag{3.13}$$

$$\text{Sob as restrições: } \begin{cases} y_i(w^T x_i + b) \geq 1 \\ i = 1, 2, 3, \dots, N. \end{cases}$$

O problema de otimização da Equação 3.13 pode ser resolvido usando multiplicadores de Lagrange¹².

¹¹Problema primal, do Inglês *primal problem*.

¹²Multiplicadores de Lagrange, do Inglês *Lagrange multipliers*.

Primeiro, introduz-se uma função Lagrangiana¹³17, definida em termos de w e b :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i(w^T x_i + b) - 1)$$

$$\min \left(\frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i (y_i(w^T x_i + b) - 1) \right) \quad (3.14)$$

Onde os α_i correspondem aos multiplicadores de Lagrange.

A solução do problema anterior é determinado diferenciando a função 3.14, respectivamente a w , b , e a maximização dos α_i . Com isso temos, as seguintes condições de otimização:

$$\text{Condição 1: } \frac{\partial}{\partial w} L(w, b, \alpha) = 0 \quad (3.15)$$

$$\text{Condição 2: } \frac{\partial}{\partial b} L(w, b, \alpha) = 0 \quad (3.16)$$

A aplicação da condição de otimização da Equação 3.15 dá o seguinte:

$$\sum_{i=1}^N \alpha_i y_i x_i \quad (3.17)$$

E da segunda condição Equação 3.16:

$$w = \sum_{i=1}^N \alpha_i y_i = 0 \quad (3.18)$$

É importante notar que no ponto de sela¹⁴, para cada multiplicador de Lagrange α_i , temos:

$$\alpha_i (y_i(w^T x_i + b) - 1) \quad i = 1, 2, 3, \dots, N \quad (3.19)$$

Dessa forma apenas aqueles multiplicadores que satisfazem exatamente a Equação 3.19 podem assumir valores não nulos.

Esta propriedade resulta das condições de Kuhn-Tucker da teoria de otimização [Haykin, 1999]. De acordo com a condição de Kuhn-Tucker, os valores não nulos são chamados de vetores de suporte.

O problema primal lida com uma função de custo convexa e com restrições lineares. De acordo com a teoria de um problema de otimização restrito como este, é possível construir um

¹³Função Lagrangiana, do Inglês *Lagrangian Function*.

¹⁴Ponto de sela, do Inglês *saddle point* (ponto de inflexão)

outro problema chamado problema dual, com o mesmo valor ótimo do problema primal, mas com os multiplicadores de Lagrange fornecendo a solução ótima. Utilizando os resultados das Equações 3.17 e 3.18, substituindo na Equação 3.14, é obtido o seguinte problema de otimização (denominado problema dual):

$$\mathbf{Max} \quad Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (3.20)$$

$$\text{sujeito às restrições: } \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0 \\ \alpha_i \geq 0 \quad i = 1, 2, 3, \dots, N. \end{cases}$$

È de notar que o problema dual é formulado inteiramente em termos de dados de treino. Além disso, a função $Q(\alpha)$ a ser maximizada depende apenas dos padrões de entrada na forma de um conjunto de produtos escalares, $\{x_i^T x_j\}_{(i,j)=1}^N$. Após obter a solução ótima de α (α^*), w^* ¹⁵ é determinada utilizando a Equação 3.18.

O hiperplano separador ótimo w^* , b^* será dado por:

$$w^* = \sum_{i=1}^N \alpha_i^* y_i x_i \quad (3.21)$$

$$b^* = -\frac{1}{2} \left[\max_{\{i|y_i=-1\}} (w^{*T} x_i) + \min_{\{i|y_i=1\}} (w^{*T} x_i) \right] \quad (3.22)$$

Os exemplos para os quais $\alpha_i^* > 0$ são os vetores de suporte, e assim podemos escrever a Equação 3.21 da seguinte forma:

$$w^* = \sum_{x_i \in SV} \alpha_i y_i x_i \quad (3.23)$$

Logo o hiperplano ótimo é apenas determinado pelos SVs. Com os parâmetros (w^*, b^*) o classificador é definido:

$$f(x) = \text{sgn} \left(\sum_{x_i \in SV} \alpha_i^* y_i (x_i^T x) + b^* \right) \quad (3.24)$$

Da Equação 3.24 verifica-se que a classificação de um novo padrão x requer apenas o calculo do produto interno entre x e cada SV. A Secção 3.7.3 estenderá os conceitos aqui

^{15*} Representa as soluções ótimas dos parâmetros

apresentados para conjuntos de treinos mas reais.

3.7.3 Hiperplano ótimo para padrões não-separáveis

O desenvolvimento apresentado na Secção 3.7.2, só funciona para casos onde os dados são linearmente separáveis. Porém, na realidade, normalmente não é isso que acontece.

Dado um conjunto de dados de treino, que não são linearmente separáveis, não é possível construir um hiperplano de separação sem existirem erros de classificação (Figura 3.7). Agora temos que lidar com conjuntos de treino mais gerais e a margem definida anteriormente determinando uma solução ótima (w^* b^*), classifica erradamente alguns padrões. Por isso recorre-se a uma suavização das margens, como podemos ver de seguida na Equação 3.25.

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, 3, \dots, N \quad (3.25)$$

Para resolver esse problema, são introduzidas novas variáveis escalares não negativas, $\{\xi_i\}_{i=1}^N$, conhecidas por variáveis de folga 3.7, na definição do hiperplano de separação. Como é apresentado de seguida:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, 2, 3, \dots, N \quad (3.26)$$

Objetivo agora é encontrar um hiperplano de separação onde o erro de classificação do

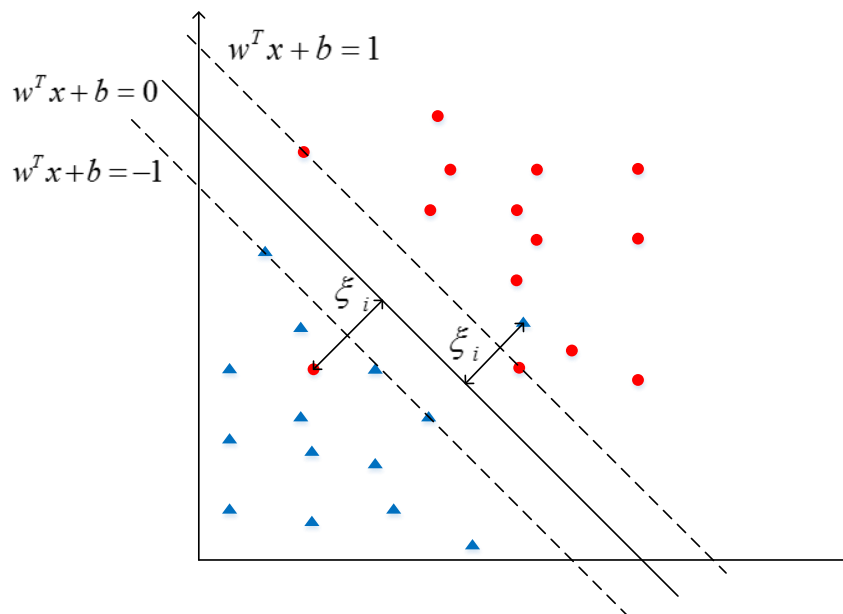


Figura 3.7: Representação da ideia de um hiperplano de separação num espaço linearmente não separável, onde as variáveis ξ_i são os erros relativos aos dados de treino de índice i .

próprio conjunto de treino é o menor possível. De acordo com [Cortes and Vapnik, 1995]

uma das formas de fazer isso é minimizar:

$$\Phi(\xi) = \sum_{i=1}^N I(\xi_i - 1) \quad (3.27)$$

A função $I(\xi)$ é uma função indicadora, definida por

$$I(\xi) = \begin{cases} 0, & \xi \leq 0 \\ 1, & \xi > 0 \end{cases} \quad (3.28)$$

A generalização do hiperplano separador ótimo é obtida através do vetor w , que minimiza o funcional de custo,

$$\Phi(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \quad (3.29)$$

Na Equação anterior 3.29, o primeiro termo minimiza a norma euclidiana do vetor de pesos que está relacionado com a maximização da região de separação da máquina de vetores de suporte, enquanto o segundo termo é um limite para os erros de classificação [Ruano, 1996]. O parâmetro C pode ser entendido como um parâmetro de regularização, e tem que ser escolhido pelo utilizador.

3.7.3.1 Determinação do hiperplano ótimo

A funcional de custo, Equação 3.29, é minimizada sujeito a restrições (Equação 3.30). Dessa forma, o problema primal para o caso de padrões não linearmente separável pode ser formulado como:

$$\min \left(\frac{w^T w}{2} + C \sum_{i=1}^N \xi_i \right) \quad (3.30)$$

$$\text{sujeito às restrições: } \begin{cases} y_i (w^T x_i + b) \geq 1 - \xi_i, & i = 1, 2, 3, \dots, N \\ \xi \geq 0 \end{cases}$$

Usando o método de multiplicadores de Lagrange, e procedendo de maneira similar á apresentada na Secção 3.7.1, podemos formular o problema dual para o caso de padrões não separáveis. Determinam-se os multiplicadores de Lagrange que maximizam a função objetivo:

$$\mathbf{Max} \quad Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (3.31)$$

$$\text{sujeito às restrições: } \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C \quad i = 1, 2, 3, \dots, N. \end{cases}$$

A partir do problema de otimização com restrições determinamos $\alpha^* = (\alpha_1^*, \alpha_2^*, \alpha_3^*, \dots, \alpha_N^*)$, e o par (w^*, b^*) que constituirá o hiperplano ótimo será então:

$$w^* = \sum_{i=1}^N \alpha_i^* y_i x_i \quad (3.32)$$

$$b^* = -\frac{1}{2} \left[\max_{\{i|y_i=-1\}} (w^{*T} x_i) + \min_{\{i|y_i=1\}} (w^{*T} x_i) \right] \quad (3.33)$$

Podemos perceber que a única diferença em relação ao caso de padrões linearmente separáveis é nas restrições onde temos um limite superior (C) a ser colocado aos multiplicadores de Lagrange.

3.7.4 Classificador SVM não linear

Na Secção 3.7.1 é considerado um problema de classificação linear; de seguida é feita uma maior aproximação a realidade considerando sobreposição de classes, na Secção 3.7.3. Apesar de melhoria do algoritmo, para corresponder a questões práticas ainda existem limitações. Há casos, como o apresentado na secção anterior, onde uma fronteira de decisão que não seja linear é mais adequada.

A Figura 3.8 ajuda a fazer uma análise geral as várias interpretações, começando pelos casos anteriores bastante mais simples. Gráfico (a), temos padrões separáveis usando uma linha de separação. No gráfico (b), podemos ainda utilizar uma linha para a classificação mas isto irá resultar em duas classificações erradas. No gráfico (c) da Figura, é apresentado uma nova fronteira de decisão que é capaz de separar as duas classes sem qualquer erro.

Neste tipo de problema não existe uma linha de separação linear mas é sim considerado um classificador linear no chamado espaço de características¹⁶ de maior dimensão.

É possível mapear os padrões de entrada num espaço de características de dimensão mais elevada. Após isso o SVM consegue construir um hiperplano separador ótimo no espaço de características e o problema pode ser abordado como um problema linearmente separável, conforme expresso na Figura 3.9.

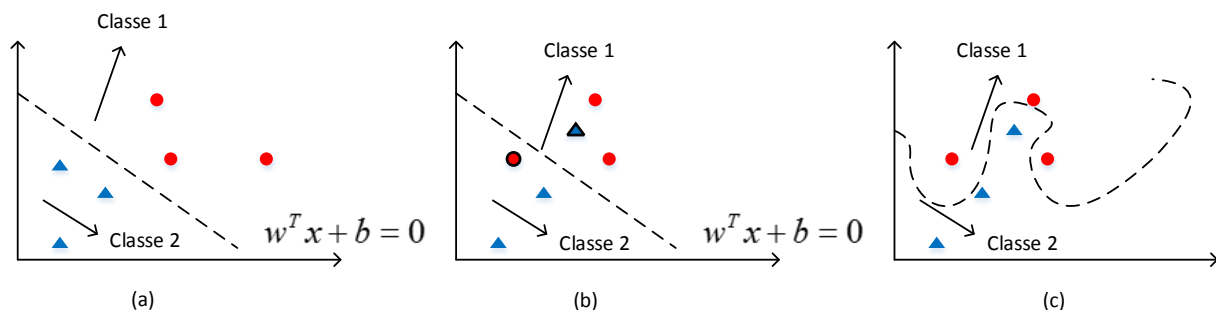


Figura 3.8: Exemplo de classificação não linear. Adaptado de [Vapnik, 1995]

¹⁶Espaço de características do Inglês *feature space*

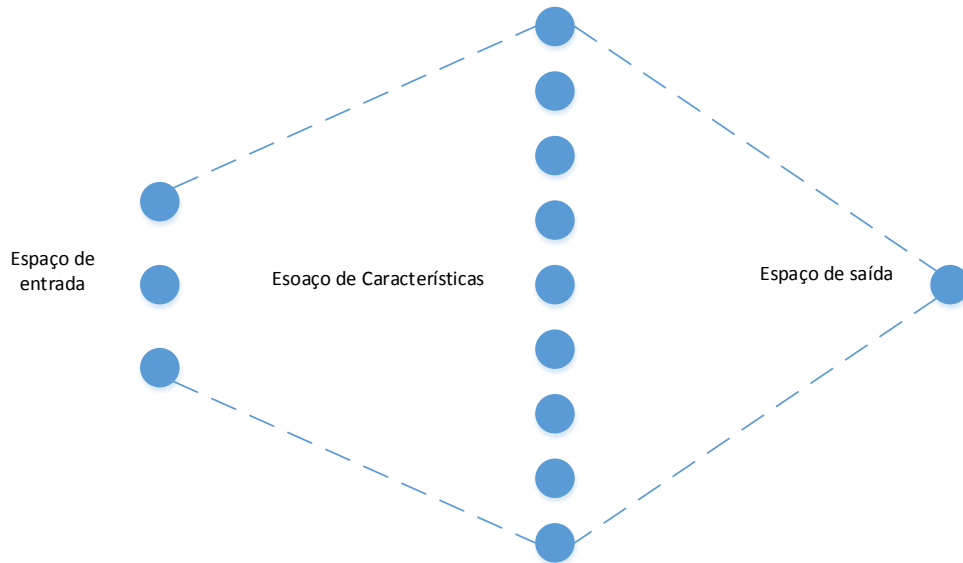


Figura 3.9: Combinação das etapas, mapeamento do espaço de entrada num espaço de maiores dimensões

Como os dados no espaço de características, podem ser separados utilizando fronteiras de decisão que são lineares, a solução encontrada para o SVM linear, descrita na Secção 3.7.1, será utilizado aqui, com as devidas alterações. Podemos utilizar a solução linear, substituindo na equação os dados de entrada pela sua representação no espaço de características. Desta forma a Equação 3.20 fica:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.34)$$

onde $K(x_i, x_j)$ representa uma função de kernel. As restrições continuam a ser as mesmas que na Equação 3.20.

3.7.4.1 Funções de Kernel Típica

Como exemplo de funções de kernel temos, o kernel linear (Equação 3.35), o kernel polinomial (Equação 3.36), o kernel de função de base radial (RBF) (Equação 3.37) e o kernel baseado na sigmoide (Equação 3.38).

$$K(x_i, x_j) = x_i \cdot x_j \quad (3.35)$$

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad (3.36)$$

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (3.37)$$

$$K(x_i, x_j) = \tanh(\rho_1 x_i \cdot x_j + \rho_2) \quad (3.38)$$

3.7.4.2 Escolha de um Kernel

Geralmente não é feita a escolha do kernel de forma determinística, apesar de existirem trabalhos onde são apresentadas técnicas que ajudam a tomar uma decisão dessa forma. A abordagem mais frequente na literatura é incluir múltiplas opções para o kernel, e depois comparar a performance obtida. O kernel que apresentar melhores resultados é escolhido por defeito. Neste trabalho a SVM tem como kernel funções de base radial¹⁷.

3.8 O algoritmo Kernel-Adatron

Após a apresentação dos fundamentos teóricos é descrito o algoritmo Kernel-Adatron, que implementa uma adaptação do algoritmo Adatron [Anlauf and BIEHL, 1989].

Nesse ponto o que damos maior importância em relação a SVM é a necessidade de resolver um problema de programação quadrática. Computacionalmente essa resolução não é uma tarefa simples e os custos podem ser elevados. Na literatura é possível encontrar vários trabalhos que visam simplificar este processo, propondo algoritmos para o treino de SVMs.

[Frie, Cristianini, and Campbell, 1998] apresentaram um método para o treino de SVMs que explora a técnica Kernel Adatron. Esse método apresenta um conjunto de vantagens, como a simplicidade de implementação e ajuste dos parâmetros dos Kernels. Várias outras

¹⁷Funções de base radial, do Inglês *Radial basic functions*

abordagens são apresentadas na literatura para simplificar o processo de treino das SVMs.

Algoritmo Kernel-Adatron

1. Inicializar $\alpha_i = 1$;

2. Calcular

$$Z_i = \sum_{j=1}^p \alpha_j y_j K(x_i, x_j)$$

3. Calcular $\gamma_i = y_i(Z_i - \theta)$

4. Sendo $\delta\alpha_i = \eta(1 - \Upsilon_i)$, modifica-se os multiplicadores α_i

- Se $(\alpha_i + \delta\alpha_i) \leq 0$, $\alpha_i \rightarrow 0$
- Se $(\alpha_i + \delta\alpha_i) > 0$, $\alpha_i \rightarrow 0$

5. Calcular o termo de desvio

$$b \frac{1}{2} (\min(z_i^+) + \min(z_i^-))$$

em que z_i^+ são os padrões da classe indicada por +1 e z_i^- são os padrões da classe indicada por -1

6. Se o número máximo de iterações do conjunto for ultrapassado, então parar; se não voltar ao ponto 2.

No final o classificador será obtido pela seguinte função:

$$f(x) = \text{sgn} \left(\sum_{x_i \in SV} \alpha_i^* y_i K(x_i, x) + b^* \right)$$

O kernel (x_i, x) pode ser qualquer que satisfaça as condições de Mercer, por exemplo funções de base radial (RBF) como o usado para esses dados.

3.9 Considerações Finais

A obtenção de um classificador utilizando SVM, envolve a escolha de um kernel, além dos parâmetros dessa função de kernel, e do algoritmo para encontrar a solução do problema de otimização quadrática com restrições.

A escolha do kernel, e dos parâmetros a ele associados têm efeito no desempenho do classificador obtido. Existem algumas técnicas para a determinação do modelo a utilizar, incluindo a função de kernel os parâmetros associados e o algoritmo para a determinação do hiperplano ótimo. [Vapnik and Chapelle, 1999, Chapelle, Vapnik, Bousquet, and Mukherjee, 2000, Chapelle and Cristianini, 2002] mostram como determinar automaticamente vários parâmetros para reconhecimento de padrões utilizando SVM, através da minimização de algumas estimativas do erro de generalização da SVM. Utilizaram para o efeito, um algoritmo gradiente descendente¹⁸, definindo o método conhecido por *span estimat*. [Jaakkola and Haussler, 1999] mostram como comparar vários kernels, e fazer testes com base na validação cruzada¹⁹, determinando a capacidade de generalização do kernel.

3.10 Resumo

Pela descrição apresentada anteriormente podemos verificar que a metodologia SVM é bem fundamentada, tendo como base o princípio da minimização estrutural de risco baseado na teoria da dimensão VC. A máquina de vetor de suporte tem a capacidade de alcançar performances na tarefa de classificação de padrões muito próximas do ótimo. Durante este capítulo foram apresentados conceitos de base para a implementação da SVM, bem como o desenvolvimento da SVM desde problemas de classificação lineares, e a sua extensão para casos não lineares onde é preciso uma projeção para espaços de maior dimensão para obter um hiperplano de separação linear.

¹⁸Gradiente descendente do Inglês *gradient descent*

¹⁹Validação cruzada do Inglês *Cross Validation*

Capítulo 4

Aplicações de técnicas de inteligência computacional para a detecção sísmica

Neste capítulo são apresentados os trabalhos realizados anteriormente, na tarefa de detecção de eventos sísmicos com base na inteligência computacional.

Começa-se por analisar as primeiras experiências efetuadas utilizando duas redes neurais (MLPs e SVMs), treinadas de forma supervisionada. A análise aqui apresentada dá grande importância aos detalhes das implementações dos algoritmos (ver Capítulo 1) utilizados, de notar que nas experiências em modo Off-line foram testados as duas redes neurais, mas já no caso On-line utilizou-se a rede neuronal que apresentou maior capacidade de generalização (SVM), como referido anteriormente.

È mostrado a evolução da técnica, os resultados e diferentes aproximações utilizados para fazer a classificação dos dados. Nas experiências em modo Off-line, foram caracterizados registos sísmicos de várias estações sísmicas processando os dados após, a recolha e o pré-processamento necessário dos dados. Já no funcionamento em modo contínuo é simulado o funcionamento em ambiente real correspondendo ao caso contínuo de detecção de sismos. O foco da análise é um conjunto de artigos científicos publicados na literatura e uma dissertação de Mestrado.

4.1 Introdução

Os estudos realizados anteriormente serviram como base e ponto de partida, para o desenvolvimento desta dissertação. Todos os estudos aqui referenciados, e as soluções estudadas, permitiram ter as primeiras impressões, acerca, da natureza, das dificuldades, e do potencial desse domínio. Nesse trabalho seguindo a lógica dos trabalhos anteriores, foram usados, classificadores neurais de aprendizagem supervisionada para a classificação de informação sísmica. Uma das principais conclusões retirada do estudo desses trabalhos, é o fato dos classificadores supervisionados reduzirem consideravelmente o erro apresentado pelo método em uso nas estações sísmicas. Para além da conclusão anterior também verificou-se a redução do tempo de identificação dos eventos sísmicos como poderemos ver mais a frente.

Tanto a redução do tempo de detecção, como a maior fiabilidade apresentada pelos classificadores desenvolvidos possibilitam, o uso futuramente desses classificadores nas estações sísmicas¹, principalmente aqueles que tem como finalidade reduzir com o máximo de fiabilidade o tempo necessário após a chegada da fase P para fazer a detecção de eventos sísmicos.

De assinalar ainda um aspeto bastante importante nesses trabalhos, procurou-se sempre, estabelecer comparações, análises dos métodos desenvolvidos e outros que são considerados potenciais.

Esses estudos devido a influência direta sobre o trabalho aqui desenvolvido no âmbito dessa dissertação foram analisados separadamente dos apresentados no estado da arte (Capítulo 2).

4.2 Detetor sísmico baseado em rede neuronal

No estudo do estado da arte (Capítulo 2) tivemos a oportunidade de perceber que a aplicação de inteligência computacional no domínio da detecção sísmica, é um problema atual e cheio de desafios a serem ultrapassados. Por isso, o trabalho aqui apresentado tem sido desenvolvido e experimentado há vários anos.

Do conjunto de motivos apontados por [Madureira, 2009] para o desenvolvimento desses trabalhos, se destacam alguns pontos importantes. Primeiramente a necessidade de agilizar todo o processo de detecção, tornando mais fiável a detecção e análise sísmica (diminuindo o número de falsos alarmes, e atingir um erro nulo, ou quase nulo nas não detecções). Por fim o

¹As estações sísmicas com tecnologia mais recente são normalmente constituídas por um sismómetro de banda larga e por um acelerómetro. A estes equipamentos de base juntam-se um digitalizador e os equipamentos necessários à transmissão de dados e ao fornecimento de energia. O equipamento deve estar colocado em instalações dedicadas para o efeito. <http://www.ipma.pt/pt/enciclopedia/geofisica/estacoes.sismo/index.html> (Última visita em 25/08/13)

crescimento do número de estações que integram as redes sismográficas, obrigando a uma otimização da deteção e análise das informações sísmicas, de forma a tornar possível uma resposta em tempo útil.

Para produzir um sistema capaz de funcionar ao nível da estação sísmica desempenhando um papel semelhante aos algoritmos baseados na razão de energia STA/LTA em funcionamento nas estações sísmicas de Portugal, [Madureira and Ruano, 2009] utilizaram dois tipos de redes neuronais, treinados de modo supervisionado.

Nesse estudo utilizaram um universo de dados constituído por 2903 padrões, extraídos de registos da estação sísmica PVAQ, da rede de vigilância sísmica do Instituto Português do Mar e da Atmosfera (IPMA)².

Adotando uma metodologia experimental durante o estudo, os autores apresentaram várias soluções para o problema. De acordo com a abordagem os resultados variaram. Utilizando diferentes parâmetros e configurações para as etapas de treino e teste, conseguiram implementações com performances quase perfeitas, quer em termos de deteção ou diminuição de falsos alarmes.

Como principal objetivo tinham, construir um sistema capaz de distinguir segmentos. A discriminação a ser feita é entre segmentos com informação sísmica, de todos os outros tipos de sinal. Após a segmentação, escolheram para construir as features, frequências selecionadas extraídas de um espectrograma calculadas para cada segmento de tempo.

Para as experiências com as MLPs, foram distribuídos os dados de forma a formar vários subconjuntos, 60% destinados ao treino da rede neuronal, 20% para o teste da rede, e outros 20% para a validação. Todo o processo de distribuição de dados foi feito de forma aleatório, levaram em consideração apenas o número de padrões de cada classe no conjunto de treino (propuseram distribuir a mesma percentagem de dados de cada classe no conjunto de treino).

Para o caso da SVM criaram dois conjuntos de dados, um para o treino e outro para o teste. Fizeram a distribuição com aproximadamente o mesmo número de elementos, mantendo a proporção de casos positivos e negativos equilibrada. Após isso adotaram uma estratégia de aprendizagem ativa (ver Capítulo 1), onde os dados mal classificados na fase de teste foram incluídos no conjunto de treino.

Os melhores resultados em termos de sensibilidade e especificidade para esse conjunto de dados, variaram entre os 98% e 100%.

De seguida, de forma resumida são apresentados os pontos mais importantes do procedimento experimental para a construção do detetor sísmico com base na rede neuronal [Madu-

²Nos trabalhos anteriormente é referido Instituto de Meteorologia de Portugal (IP).

reira and Ruano, 2009, Ruano and Madureira, 2011, Madureira and Ruano, 2012, Madureira, Ruano, and Ruano, 2013b].

4.2.1 Procedimento Experimental

Nesta Secção descreve-se a implementação prática do sistema de deteção sísmica. Recordando, o sistema é capaz de distinguir segmentos de registos sísmicos contendo sinal originado de eventos sísmicos locais e regionais, de todas as outras situações (considerados de uma certa forma como sendo ruídos, eventos distantes entre outros).

Como pretenderam distinguir duas classes, o objetivo principal portanto foi construir um classificador binário, capaz de após treinado, atribuir uma de duas classes a um novo segmento sísmico registado.

As duas classes são: *Classe 1*: Sismos³ naturais locais e regionais. *Classe 2*: Todos os outros casos⁴ (sismos induzidos/ruído⁵)

4.2.1.1 Dados ao nível da estação sísmica

Estação Sísmica

Nas primeiras experiências utilizaram dados registados na estação sísmica PVAQ.

Nome	Distrito ou Ilha	Localidade	Trans.	Lat.	Long.	Alt.	Sensores	Canais	Propriedade
PVAQ	Faro	Vaqueiros	S	37°24.22'N	07°43.04'W	200	3T-120s	HH*,BH*,LH*,HN*	IPMA

Tabela 4.1: Parâmetros da estação sísmica PVAQ na REDE SÍSMICA PORTUGUESA

* - Canais verticais, norte-sul e este-oeste, **S** - satélite, **T** - telefone, **R** - rádio, **L** - local

Informação Rede sísmica Portuguesa (Atualizado: 20-10-2011)⁶. Para mais informação relativamente a estação sísmica PVAQ consultar [Madureira, 2009].

³Um sismo é um fenómeno de vibração brusca e passageira da superfície da Terra, resultante de movimentos subterrâneos de placas rochosas, de atividade vulcânica, ou por deslocamentos (migração) de gases no interior da Terra, principalmente metano.

⁴Dentro dos outros casos temos por exemplo, quando existem explosões artificiais ou sinais provocados por qualquer outro fenómeno, esses casos foram todos agrupados num grupo independente.

⁵Estes são sismos associados à ação humana quer direta ou indiretamente. Podem-se dever à extração de minerais, água dos aquíferos ou de combustíveis fósseis, devido à pressão da água das albufeiras das barragens, grandes explosões ou a queda de grandes edifícios. Apesar de causarem vibrações na Terra, estes não podem ser considerados sismos no sentido lato, uma vez que geralmente dão origem a registos ou sismogramas diferentes dos terramotos de origem natural.<http://en.wikipedia.org/wiki/Mining>

⁶<http://www.ipma.pt/pt/enciclopedia/geofisica/estacoes.sismo/index.html>(Última visita em 25/08/13)

4.2.1.2 Dados sísmicos registados utilizados nas experiências

Relativamente ao ano de 2007 registaram 2903 exemplos, onde 502 representaram a classe dos sismos (classificados como sismos pelos sismólogos do CND⁷, e as fases sísmicas registadas em PVAQ) e outros 2401 representaram a classe dos classificados como não sísmicos.

Os dados foram registados em ficheiros diários num formato comprimido, que constituíam um total de 325 ficheiros, donde 13 dias Julianos encontravam em falta devido a problemas ao nível da estação (em alguns casos existem problemas nas estações, não funcionam da forma correta impedindo registo durante o período de inatividade). Para além desses ficheiros também foram registados os instantes que ocorrem os disparos do sistema de deteção automática noutros ficheiros, que também utilizaram para fins de segmentação.

4.2.1.3 Segmentação

A precisão e robustez do processo de segmentação do sinal são fundamentais para a construção de features, e estimação adequada dos instantes de início e fim dos eventos sísmicos. Fatores que influenciam diretamente a performance do sistema de deteção posteriormente construído.

Todos os eventos sísmicos processados no CND foram registados num catálogo sísmico, que contém várias informações, como podemos ver na Figura 4.1. No catálogo encontram-se, a classificação tendo em conta a distância do hipocentro, a data, as fases identificadas e as respetivas horas e as estações onde foram identificadas as fases.

Com base nestas informações foram determinaram os instantes de início dos segmentos que constituíram o conjunto de exemplos positivos.

Em relação a duração dos segmentos, levaram em conta o fato de como se pretende detetar, sismos locais e regionais incluindo as fases P e S, a diferença entre essas fases não excede os 105 segundos em eventos até 1000 km de distância, foi fixado a duração dos segmentos de 120 segundos. O processo de extração dos segmentos de registo sísmico esquematiza-se na Figura 4.2. Para mais informações sobre a utilização do catálogo e dos instantes e tempo para a segmentação dos registos, consultar [Madureira, 2009].

⁷Centro Nacional de Dados

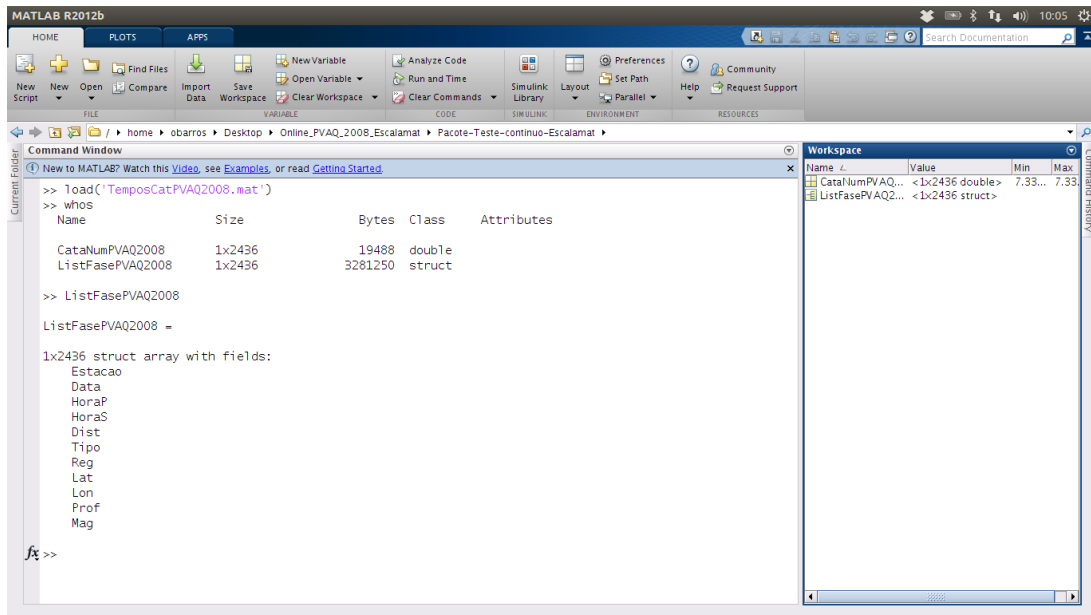


Figura 4.1: Esquema Extrato Catálogo sísmico PVAQ 2008.

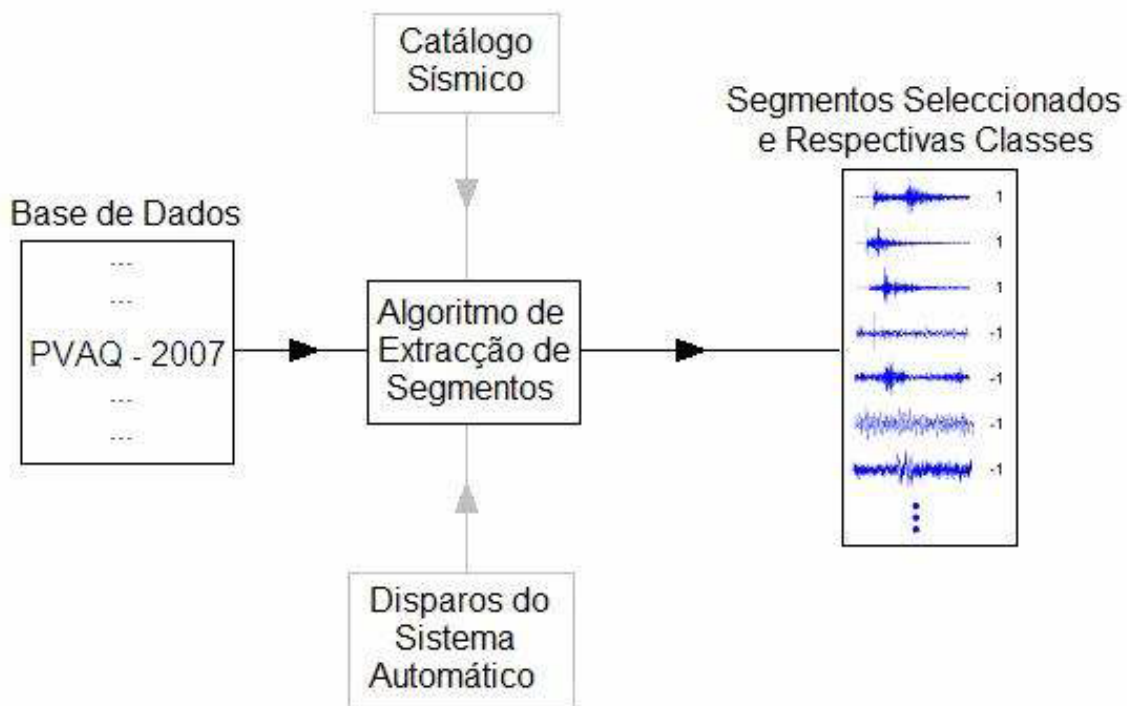


Figura 4.2: Esquema de Extração de segmentos de registo sísmico. Fonte: [Madureira, 2009]

Determinação das Features (Dados de entrada)

Para cada segmento calcularam a Densidade Espectral de Potência (DEP) da seguinte forma: Começa-se por estimar a DEP utilizando a média do periodograma [Welch, 1967], considerando apenas as frequências positivas, suavizando depois em cada décimo de década.

Seja $P(f)$ o valor da DEP num conjunto de frequências F . Partindo do mínimo de $F(f_{\min})$ criou-se uma sucessão de frequências separadas por um décimo da década,

$$f_k = f_{\min} 10^{\frac{k-1}{10}}, \quad k = 1, 2, 3, \dots \quad (4.1)$$

Depois, F é dividido em vários subconjuntos D_k ,

$$D_k = \{f\} : f_k \leq f_{k+1}, \quad f \in F, \quad k = 1, 2, 3, \dots \quad (4.2)$$

A DEP suavizada é então dada por:

$$P_s(f) = \frac{1}{\#D_k} \sum_{f \in D_k} P(f) \quad (4.3)$$

Após a segmentação, todos os segmentos de 120 segundos foram divididos, em 5 intervalos não sobrepostos. Para cada intervalo extraíram os valores da potência em seis frequências selecionadas; 1, 2, 4, 8, 10 e 15 Hz, construindo padrões com 30 elementos cada. Em algumas experiências antes de determinar a DEP filtraram o sinal, utilizando um filtro digital passa-alto do tipo Butterworth, de ordem 5 e frequência de corte 0.5 Hz.

A Figura 4.3 apresenta todo o processo da extração das features, e a Figura 4.4 apresenta todo o processo da construção do universo de dados utilizado na fase de treino, e no teste (também para o caso da validação para as MLPs) do classificador desenvolvido. No esquema geral o filtro apresentado nem sempre foi utilizado, como já referido. Todas as experiências apresentadas informam sobre a utilização ou não do filtro nos resultados apresentados.

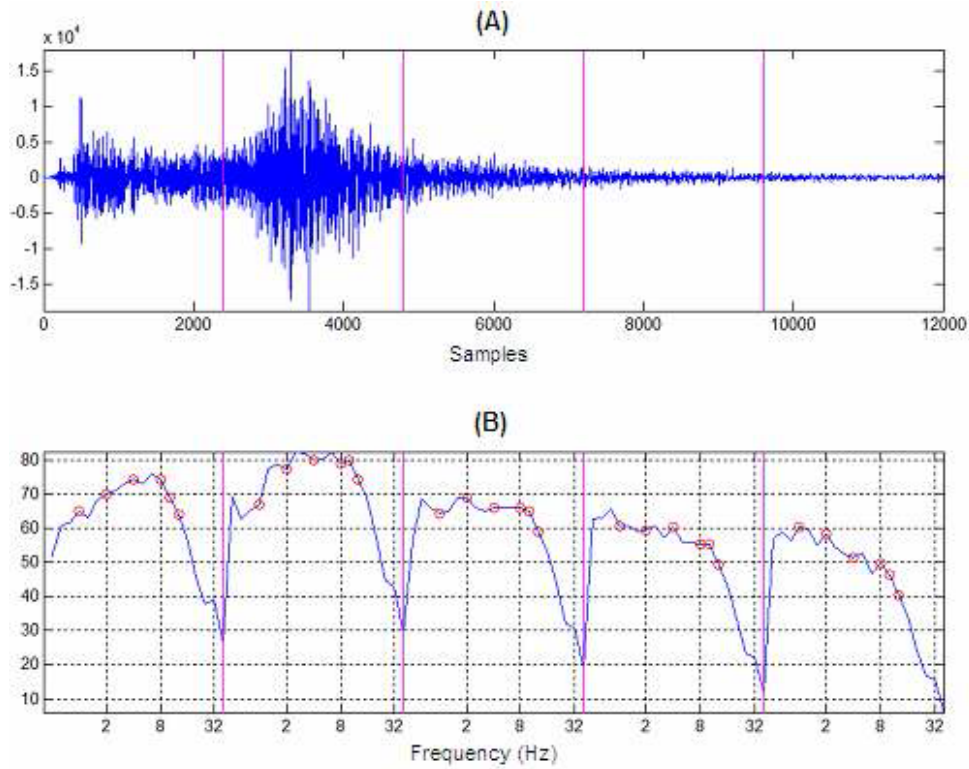


Figura 4.3: (A) 120 segundos de registo sísmico (B) Espectrograma: Fonte: [Madureira and Ruano, 2009]

Processo completo de determinação das features e construção do universo de dados

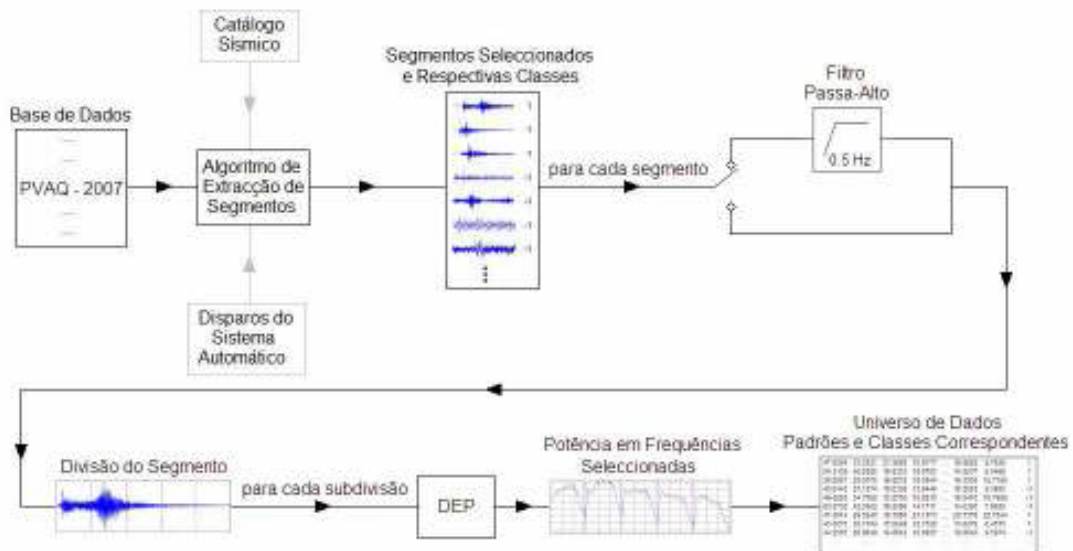


Figura 4.4: Extração de segmentos, cálculo de features e construção do universo de dados. Fonte: [Madureira, 2009]

4.2.1.4 Resultados

Nos estudos, os resultados foram apresentados em termos de Sensibilidade (**Sensitivity ou Recall**) (**R**), e Especificidade (**Specificity**) (**S**).

Sensibilidade definido como:

$$R = \frac{VP}{VP + FN} \quad (4.4)$$

Especificidade definido como:

$$S = \frac{VN}{VN + FP} \quad (4.5)$$

VP → Número de Verdadeiros Positivos

FP → Número de Falsos Positivos

FN → Número de Falsos Negativos

VN → Número de Verdadeiros Negativos

S → Taxa de exemplos negativos corretamente classificados

R → Taxa de exemplos positivos corretamente classificados

Os resultados calculados representavam soluções de um problema multiobjectivo que denomina-se Soluções Não-Dominadas ⁸.

Ou seja pretende-se:

Seja x uma solução admissível de um problema multiobjectivo ($x \in X$). A solução x diz-se não dominada se e só se não existir outra solução admissível y ($y \in X$) que domine x .

Pretende-se maximizar a sensibilidade e especificidade, de forma a que, nenhuma outra solução, exceda os dois parâmetros simultaneamente.

4.2.2 Métodos de Treino e resultados

Numa primeira fase foram treinadas MLPs com: diferentes topologias; diversas condições e parametrização; Seguidamente treinaram SVM-RBF com diferentes valores de variância, aplicando uma estratégia de aprendizagem ativa. Após isso utilizaram os conjuntos de treino e teste obtidos nas experiências com a SVM-RBF aplicando aprendizagem ativa, e procederam a novos treinos com a MLP. Treinaram a SVM-RBF, começando por alterar a variância, depois aplicaram uma estratégia de aprendizagem ativa, onde o conjunto de treino e teste são construídos iterativamente. Para mais informações relativamente a aprendizagem ativa, consultar a apresentação de forma resumida no Capítulo 1 e consultar [Cohn, Atlas, and Ladner, 1994, Ruano and Madureira, 2011].

⁸Soluções Não-Dominadas do Inglês *Nom-Dominated Solutions* (ND)

4.2.2.1 MLP resultados primeira experiência

Os melhores resultados em termos de especificidade e sensibilidade variaram entre os 91% e 100%. Selecionando um dos modelos que apresentou o menor número de classificações erradas temos uma solução ND:

Topologia	Filtro	R	S	R (ALL)	S (ALL)
[6 3]	Sim	95.83	99.38	94.82	98.96

Tabela 4.2: Resultados primeira experiência Off-line com MLP, dados PVAQ

Na tabela 4.2 os resultados da primeira coluna representam o número de neurónios da primeira camada, e o número de neurónios da segunda camada escondida respetivamente. Depois é referido se foi utilizado ou não filtro durante o processamento dos dados. Por fim, são apresentado por S e R as percentagens de Sensibilidade e Especificidade, sendo que S(ALL) e R(ALL), representam percentagens calculadas relativamente a união de todos os conjuntos; (Treino + Teste + Validação para ás MLPs) e (Treino + Teste para ás SVMs). Para esta fase de testes obtiveram valores de Sensibilidade e Especificidade superiores a 95%. Por não terem verificado diferenças significativas entre os resultados para quando aplicaram filtros, e os casos sem aplicação, nas experiências seguintes procederam sem utilizar filtro.

4.2.2.2 SVM resultados primeira experiência

Começaram por determinar o fecho convexo. De lembrar que para a SVM, não são consideradas topologias. Com a aprendizagem ativa atingiram desempenhos perfeitos para o conjunto de dados em estudo; (S = 100%, R = 100%).

4.2.2.3 MLP resultados segunda experiência

A experiência nessa fase foi igual a aplicada para na primeira experiência. Neste caso os dados utilizados para treinar a MLP resultaram da classificação utilizando a SVM após aprendizagem ativa. O conjunto de teste utilizado para testar a SVM foi particionado em conjuntos de treino e validação, seguindo os procedimentos normais dos testes com MLPs. Os resultados que obtiveram são representados na tabela seguinte;

Topologia	R	S	R (ALL)	S (ALL)
[6 4]	94.51	99.79	96.61	99.50

Tabela 4.3: Resultados primeira experiência Off-line com MLP, dados PVAQ

4.3 Comparação dos resultados obtidos com o Sistema de Detecção ao Nível da Estação Sísmica em Funcionamento no CND

De forma a fazer uma comparação entre o sistema desenvolvido e o sistema em utilização, ao nível da estação sísmica, garantiram que os sistemas apresentam os mesmos aspetos em termos de funcionamento e avaliação do desempenho. Para isso [Madureira, 2009] apresenta um conjunto de considerações, e imposições a efetuar de forma a criar bases que tornem possível a comparação entre os dois sistemas. Primeiramente destacam o fato do estudo só ser válido para o trabalho apresentado na Secção 4.2.

4.3.1 Considerações sobre o sistema de avaliação atualmente em utilização nas estações sísmicas

- Todos os eventos identificados na estação PVAQ 2007, estão representados no universo de dados, ou seja, consideraram todos os exemplos positivos disponíveis.
- A partir dos dados registados e informações disponíveis no catálogo sísmico, concluíram que o sistema apresentou uma Sensibilidade de 50%.
- Para a Sensibilidade foi considerado que quando um evento é detetado no CND e a fase é identificada pelos sismólogos, e não é detetado pelo sistema de deteção na estação sísmica, ocorre um falso negativo.
- Em relação a Especificidade escolheram de forma arbitrária 50% de falsos positivos, por isso a análise dessa relação não foi valida em termos absolutos.

4.3.1.1 Análise comparativa da Sensibilidade

Para os testes com as várias topologias das MLPs os resultados foram sempre superiores a 94%, sendo que, para o melhor caso foi superior a 95%, Secção 4.2.2.1 (Tabela 4.2). Esta performance é bastante superior aos 50% do sistema em utilização. Acontece o mesmo com as experiências utilizando MLP com a partição de dados originados da aprendizagem ativa do SVM Secção 4.2.2.3(Tabela 4.3), para esse caso os melhores resultados são superiores a 96%. Para o caso das SVMs as performances foram superiores a todas as abordagens, tanto o método baseado na razão de energia em utilização nas estações sísmicas, como os resultados alcançados quando utilizado MLP. Para as SVMs, após a aprendizagem ativa a classificação para o universo de dados em estudo é perfeita Secção 4.2.2.2.

4.3.1.2 Análise comparativa da Especificidade

Relembrando das considerações Secção 4.3.1 atribuíram um valor de Especificidade 50%, para o sistema em funcionamento de forma a, fazer uma comparação dos desempenhos dos sistemas para o universo de dados considerado. Para as experiências com MLPs os resultados foram sempre aproximadamente 99%, sendo o melhor resultado obtido na Secção 4.2.2.3(Tabela 4.3) Tabela onde o valor é 99,5%. Para o caso da SVM nas primeiras iterações os resultados foram aproximadamente 99%, atingindo classificações perto dos 100% nas iterações seguintes.

4.4 Operação On-Line de um Detetor Sísmico Inteligente

Após o processo de classificação da informação sísmica no modo Off-line onde foram testados um conjunto limitado de dados, o objetivo que agora colocaram é perceber de forma clara a capacidade do classificador continuar a apresentar um excelente desempenho.

Para isso aplicando um sistema de janela deslizante, onde é simulado o funcionamento do classificador num ambiente real, os dados da estação sísmica PVAQ foram analisados. Esses dados em nenhuma fase foram utilizados no treino do classificador a utilizar, sendo assim completamente desconhecidos.

Como apresentado na Secção 4.3 o classificador SVM, com o processo de aprendizagem ativa para o conjunto de dados em estudo apresenta uma classificação perfeita. Devido a melhor performance relativamente as várias topologias do MLP (Secção 4.3) escolheram para ser utilizado no processo contínuo. Ainda relembrando da capacidade de generalização das SVMs, encontrando sempre um mínimo global para o problema de otimização quadrática, e melhor performance na deteção de sismos, a escolha tinha mas do que fundamentos para ser considerada boa.

4.4.1 Aplicação do sistema de janela deslizante

A janela de 120 segundos percorre em intervalos de 50 amostras todas as amostras sísmicas consideradas. Correspondendo a avanços de 0.5 segundos respetivamente. Cada intervalo de 120 segundos é submetido para a classificação, sendo a classificação resultante analisada perante a informação do catálogo sísmico de forma a verificar se a classificação foi correta ou não (basicamente é definido a classe de resultados que o segmento classificado pertence, de acordo com os parâmetros VP,VN,FP,FN, (ver Secção 4.2.1.4) que depois são utilizados para determinar os valores em termos de sensibilidade e especificidade).

4.4.1.1 Resultados contínuos para PVAQ – 2008

[Madureira, Ruano, and Ruano, 2013a] avaliaram os primeiros 256 dias do ano de 2008, os dias Julianos, 3, 200 e 201 não foram considerados devido a problemas ao nível da estação sísmica. Em relação a performance os valores de R e S para o período considerado foram, 97,7% e 98,7% respetivamente.

R	S
97.7	98.7

Tabela 4.4: Resultados experiência On-line com dados PVAQ

4.4.1.2 Resultados contínuos para PESTR – 2008

[Madureira, Ruano, and Ruano, 2013a] aplicaram o classificador obtido com os dados sísmicos de PVAQ para o ano de 2007 continuamente, aos dados sísmicos dos primeiros 185 dias do ano de 2008 para a estação sísmica de PESTR.

Nome	Distrito ou Ilha	Localidade	Trans.	Lat.	Long.	Alt.	Sensores	Canais	Propriedade
PESTR	Évora	Estremoz	S	38°52.03.41'W	07°43.04'W	410	3T-120s	HH*,BH*,LH*,HN*	IPMA

Tabela 4.5: Parâmetros da estação sísmica PESTR na REDE SÍSMICA PORTUGUESA

Para esse caso a performance foi:

R	S
88.4	99.4

Tabela 4.6: Resultados experiência On-line com dados PVAQ

Verifica-se uma ligeira diminuição na performance do (**R**) pois os dados utilizados no treino Off-line tiveram origem de uma outra estação, e várias propriedades como a constituição do solo, localização da estação, influenciam no tipo de informação sísmica que é registado.

4.5 Resumo

Neste capítulo dando seguimento a análise extensiva apresentada no estado da arte são apresentados ,estudos efetuados anteriormente no âmbito do projeto de implementação de detetores sísmicos com base na utilização de inteligência computacional. Começa-se por analisar a implementação do detetor sísmico baseado em rede neuronal analisando os métodos utilizados, dando ênfase aos dados e resultados obtidos. Foram também analisados alguns aspetos da implementação On-line, onde é simulado o funcionamento em ambiente real. Também

foram discutidos os resultados obtidos, nomeadamente as performances On-line, Off- line, analisando de forma breve as considerações efetuadas durante os estudos.

Capítulo 5

Procedimento Experimental e Resultados

Neste capítulo são apresentados os resultados obtidos, tendo como base os estudos apresentado para a implementação de um classificador do tipo SVM (Capítulo 4, e Capítulo 3) para estações sísmica da rede sísmica nacional (Portugal). Após desenhar modelos neuronais para as estações sísmicas, é analisado o desempenho do detetor em funcionamento Off-line e em On-line.

Por fim, com os modelos desenvolvidos é avaliado o desempenho dos vários modelos funcionando como um todo, constituindo uma única rede. Para as estações sísmicas PESTR e PVAQ são criados vários classificadores SVMs, de onde depois são escolhidos as SVMs que apresentam melhor performance em modo contínuo. Durante as experiências, foram analisados os resultados de acordo com vários parâmetros; a complexidade do Fecho Convexo; o escalonamento utilizado durante o pré-processamento dos dados; o modo de teste em questão; os dados sísmicos; o catálogo sísmico utilizado;

5.1 Introdução

Os resultados de seguida apresentados foram obtidos aplicando o método descrito no Capítulo 4. Foram testados dados provenientes de duas estações, PVAQ e PESTR. São apresentados os resultados On-line para as duas estações e Off-line para as duas estações, em termos de Especificidade e Sensibilidade (Secção 4.2.1.4).

Os testes são feitos utilizando duas abordagens, primeiramente são testados, dados dos conjuntos de treino que de forma geral não apresentam erros, depois são testados a totalidade dos dados.

Para a totalidade dos dados (**ALL**) já são introduzidos informações desconhecidas ao sistema, pelo que em alguns casos acontecem erros de classificação, que depois utilizando uma estratégia de aprendizagem ativa são reduzidos de forma a conseguir classificadores perfeitos para o universo de dados em estudo. A abordagem é utilizada de forma iterativa até conseguir um classificador perfeito para este conjunto de dados

(**R = 100% & S = 100%**).

5.2 Dados Utilizados

Antes do processamento dos dados, existe uma fase de extração de segmentos, onde os ficheiros diários registados num formato específico são extraídos. Para fazer a extração dos dados defini-se o início e o fim dos eventos. Para mais informação relativamente ao processo de extração, pode-se consultar, [Madureira, 2009].

5.2.1 Dados Funcionamento Off-line

Estação Sísmica PVAQ

Utilizando, features extraídas dos dados sísmicos registados na estação sísmica PVAQ no ano de 2007, criou-se um universo de dados para esta estação, constituído por 2903 segmentos. Para cada segmento é atribuído a respetiva classe pertencente. Desse Universo de dados, 502 constituíram a classe positiva (sismos locais ou regionais), e os outros 2401 classificados como não sendo sismos constituíram a classe negativa.

Estação Sísmica PESTR

Utilizando features extraídas dos registos sísmicos da estação sísmica PESTR, no ano de 2009, o Universo de dados é constituído por 3606 elementos, com 2594 representando a classe negativa e 1012 representando a classe positiva.

5.2.2 Normalização dos dados

De forma breve é apresentado as duas abordagens utilizadas para fazer o escalonamento. Durante a fase de pré-processamento dos dados, há aplicação de operações de escala definidos de seguida.

5.2.2.1 Função de escala Escalamat

Todos os resultados apresentados no Capítulo 4 seguem esta abordagem. Efetua o escalonamento de cada linha da matriz de entrada, de modo a que a diferença entre o máximo e o mínimo seja 2. Seguidamente centra-se a linha base em 0, subtraindo-lhe a própria média.

A operação Escalamat é definido pelas Equações 5.1 e 5.2.

Primeiro é escalonado cada linha de forma a diferença entre o máximo e mínimo seja 2;

$$X_{sc} = \frac{2\left(X - \frac{max+min}{2}\right)}{max - min} \quad (5.1)$$

Após a operação de escala a linha é centrada em zero, subtraindo-lhe a média.

$$X_{sc} = X_{sc} - \mu(X_{sc}) \quad (5.2)$$

Onde X representa o vetor a escalar, max representa o valor máximo do vetor, min representa o valor mínimo do vetor e o X_{sc} representa o vetor resultado após o processo de escala. De forma a aplicar a operação a toda matriz que representa o universo de dados, é preciso aplicar a operação para cada linha.

5.2.2.2 Função de escala ESCALE

No novo procedimento, o escalonamento é feito para cada feature separadamente.

Para cada coluna representando as features, os valores são escalonados entre 1 e -1. Desta forma a gama de valores para cada X_i é de -1 a 1, sendo x_i a coluna de índice no universo de dados. Após isso é retornado os vetores com os respetivos ganhos K e b ; valores que posteriormente são utilizados para fazer o escalonamento no modo On-line.

Primeiro determinam-se os ganhos da seguinte forma:

$$\begin{aligned} K &= X_{max} - X_{min}/(max - min) \\ b &= X_{max} - K * min \end{aligned} \quad (5.3)$$

Utilizando os parâmetros calculados na Equação 5.3, define-se a coluna escalonada como sendo:

$$X_i = X \cdot k + b \quad (5.4)$$

Onde X representa o vetor coluna a escalonar, max representa o valor máximo do vetor, min representa o valor mínimo do vetor e o X_i representa o vetor resultado após o processo de escala. Ainda o X_{max} e X_{min} , representam vetores colunas inicializados com os valores 1 e -1, respetivamente. De forma a aplicar a operação a toda matriz que representa o universo de dados, é preciso aplicar a operação para todas as colunas.

5.3 Classificação em Off-line

Para cada tabela com os resultados é identificado o escalonamento utilizado pelo nome da função que faz a operação de escala.

5.3.1 Resultados com o detetor PESTR

Os resultados obtidos com o método descrito anteriormente, aplicado a estação sísmica de PESTR são representados nas Tabelas 5.1 e 5.2, o classificador é testado para os dados de treino e depois para a totalidade dos dados.

Os resultados apresentados correspondem aos SVMs que depois apresentaram as melhores performance para a classificação de dados simulando o ambiente real, onde aplicando a técnica de janela deslizante, os dados são classificados.

Iteração	R	S	R (ALL)	S (ALL)
1	100.00	100.00	96.95	99.22
2	100.00	100.00	99.73	99.53
3	100.00	100.00	100.00	99.96
4	100.00	100.00	100.00	100.00

Tabela 5.1: Resultados obtidos aplicando aprendizagem ativa PESTR-2009. (SCALE)

Iteração	R	S	R (ALL)	S (ALL)
1	100.00	100.00	99.01	99.73
2	100.00	100.00	99.80	99.96
3	100.00	100.00	100.00	99.98
4	100.00	100.00	100.00	100.00

Tabela 5.2: Resultados obtidos aplicando aprendizagem ativa PESTR-2009. (Escalamat)

De acordo com os a Tabela 5.1 e Tabela 5.2, podemos verificar que tanto em termos de, valores de especificidade como em termos de valores de sensibilidade, as classificações são perfeitas quando o conjunto de teste em estudo é o mesmo que o conjunto de dados utilizado no treino (R e S, teste com o conjunto de treino).

Após este teste, tendo em conta que o modelo foi desenvolvido porque é de interesse classificar dados novos, ou seja, informações adquiridas de forma contínua e desconhecidas pelo sistema, é de grande interesse saber a performance em dados desconhecidos, por isso, são efetuados testes com a totalidade de dados. É apenas apresentado a performance do classificador em relação a reunião de todos os dados disponíveis. Levando em consideração os valores de R(ALL) e S(ALL), para os dois escalonamentos utilizados, podemos verificar que com a aplicação da aprendizagem ativa num conjunto de iterações consegue-se um classificador perfeito, sendo que geralmente isso acontece com 4 iterações. Esse resultado é também obtido levando em conta outros fecho convexos, de forma mais concreta apresentando diferentes números de vértices.

Os resultados apresentados correspondem a SVMs que apresentaram melhores performances para as classificações de dados simulando o ambiente real (resultados mais a frente). A função de escala para esse caso foi o *SCALE* os dados são normalizados por coluna, cada feature separadamente. Para esse caso os resultados tanto em termos de especificidade e sensibilidade apresentam melhorias, podemos verificar que no pior caso que corresponde a primeira iteração no processo de aprendizagem ativa, antes de atingir os 100% de classificações corretas pretendido, os valores da performance em termos de R e S são superiores a 99%.

5.3.2 Resultados com o detetor PVAQ

Iteração	R	S	R (ALL)	S (ALL)
1	100.00	100.00	94.17	97.25
2	100.00	100.00	99.88	99.79
3	100.00	100.00	99.92	100.00
4	100.00	100.00	100.00	100.00

Tabela 5.3: Resultados obtidos aplicando aprendizagem ativa PVAQ-2007. (SCALE)

Iteração	R	S	R (ALL)	S (ALL)
1	100.00	100.00	98.13	99.40
2	100.00	100.00	99.64	99.90
3	100.00	100.00	99.70	99.95
4	100.00	100.00	100.00	100.00

Tabela 5.4: Resultados obtidos aplicando aprendizagem ativa PVAQ-2007. (Escalamat)

Da análise das tabelas anteriores (Tabela 5.3 e 5.4), podemos perceber que, da mesma forma que para as experiências realizadas com os dados da estação sísmica PESTR, o classificador treinado com os dados sísmicos da estação PVAQ, para a fase de teste levando em conta os dados utilizados no treino não apresentam erros de classificação. Com isso, é atingido sempre uma percentagem de 100%, tanto para a sensibilidade como para especificidade.

Os resultados não apresentam grandes diferenças em relação aos apresentados anteriormente, para o caso de PESTR. Os valores de especificidade e sensibilidade variam majoritariamente numa gama dos 99% aos 100%.

Classificadores perfeitos para o conjunto de dados em estudo são atingidos com 3/4 iterações geralmente. Esses resultados são também obtidos levando em conta outros fechos convexos, apresentando diferentes números de vértices. Não há novidades nos resultados obtidos, pelo que podemos perceber que, o sistema apresenta um comportamento bastante linear.

A função de escala para esse caso foi o *Escalamat*, ou seja dados normalizados por linha, cada padrão separadamente (30 features). Para esse caso os resultados apresentaram melhorias. No pior caso, que corresponde a primeira iteração no processo de aprendizagem ativa antes de atingir os 100% de classificações corretas pretendido, os valores da performance em termos de R e S são superiores a 99%.

5.4 Classificação em On-line

5.4.1 Normalização dos dados

5.4.1.1 Escalonamento utilizando a função Escalamat

O procedimento para a normalização dos dados, usando a função Escalamat, depende apenas do padrão em questão, os conjuntos de 30 features determinados para cada segmento (120 segundos) de dados quando simulado o funcionamento em ambiente real, aplicando a técnica de janela deslizante no registo contínuo de dados.

A janela de 120 segundos que desliza em intervalos de 50 amostras, correspondendo a deslocação de 0.5 segundos, antes de ser submetido ao classificador é normalizado utilizando o mesmo processo que no caso contínuo.

Como o escalonamento em si não depende de informações á priori ou posteriori, para cada segmento de nova informação é executado a normalização, de modo que a diferença entre o máximo e o mínimo seja 2, após isso o vetor (features do padrão) é centrado fazendo uma diferença em relação a própria média.

5.4.1.2 Escalonamento utilizando a função SCALE

Para o caso onde a normalização é feita com base na função SCALE, o procedimento é diferente. Existe uma ligação entre o processamento dos dados para o caso Off-line e o On-line. Informações obtidas para escalonar os dados durante a fase de treino/teste são utilizados para quando simulado o desempenho em ambiente real.

Após determinar o vetor de ganho k , e o vetor de ajuste b , como apresentado na Secção 5.2.2 esses parâmetros são utilizados para escalonar os segmentos de dados processados continuamente. Para cada feature, o processo de escalonamento é idêntico ao aplicado para o caso Off-line, apenas os parâmetros para a normalização é que são os obtidos anteriormente na fase de treino/teste, de forma a todo o processo de classificação ser coerente.

5.5 Resultados PESTR-2008

Para cada ano são apresentados resultados e o escalonamento efetuado.

Também são apresentados histogramas, onde é possível, ter a noção das amplitudes sísmicas no espaço de tempo considerado e o tempo de deteção.

Existem dias onde não ocorrem eventos, esses dias são classificados com um resultado de -1, em termos de especificidade ou sensibilidade. Nos dias que não ocorrem eventos, as

classificações com -1 não apresentam qualquer influência na performance geral do sistema durante o período de tempo considerado, visto serem ignorados para efeitos de cálculos. Para os cálculos, são sempre considerados de forma geral os *Verdadeiros positivos/negativos vs Falsos positivos/negativos*, esses valores para quando não existem eventos são nulos.

5.5.1 Resultados On-line com o detetor PESTR

5.5.1.1 Escalamat – PESTR 2008 – On-line

R	S
91.04	97.04

Tabela 5.5: Performance On-line do SVM para PESTR (Escalamat)

5.5.1.2 SCALE – PESTR 2008 – On-line

R	S
93.85	96.95

Tabela 5.6: Performance On-line do SVM para PESTR (SCALE)

Como podemos perceber levando em conta os resultados apresentados nas Tabelas 5.5 e 5.6, quando o método de classificação da informação sísmica utiliza a função SCALE para fazer a normalização dos dados é obtido uma melhoria considerável nos resultados, em termos de especificidade, o que significa que, a quantidade de sismos identificados corretamente aumentou, diminuindo o número de falsos alarmes, sendo este parâmetro de elevada importância para o problema em análise.

A análise é apresentada como sendo anual, mas nem todos os dias julianos, encontram-se representados.

Dos 366 dias do ano de 2008, informações relativamente a 301 dias é que foram analisados, por isso, quando é referido ao termo anual tratam-se desses dias. Faltam ainda 65 dias, começando no mês de outubro, sendo que as informações sísmicas relativamente ao mês de novembro são inexistentes, apresentando falhas também no mês de dezembro. Em termos de especificidade os resultados são parecidos, não existe muita diferença, mas em relação ao Recall podemos verificar que existem melhorias significativas.

5.6 Resultados PVAQ-2008

Para o PVAQ os resultados da classificação On-line seguem a mesma abordagem que para PESTR. Após classificar 366 (301) dias do ano de 2008, foram obtidos de forma geral os seguintes resultados:

5.6.1 Resultados On-line com o detetor PVAQ

5.6.1.1 Escalamat – PVAQ 2008 – On-line

R	S
97.54	99.63

Tabela 5.7: Performance On-line do SVM para PVAQ (Escalamat)

5.6.1.2 SCALE – PVAQ 2008 – On-line

R	S
97.46	99.56

Tabela 5.8: Performance On-line do SVM para PVAQ (SCALE)

Como podemos perceber com os resultados apresentados nas Tabelas 5.7 e 5.8, ao contrário do que acontece no caso de PESTR, para os dados da estação sísmica PVAQ, a escolha do escalonamento não apresenta melhorias nos resultados alcançados. Como sabemos os dados variam de estação para estação, dependendo do tipo de ruído presente, da localização da estação, das características do solo onde se encontra a estação, os dados de PVAQ diferente do que acontece com PESTR devem ter propriedades que tornam o escalonamento utilizado menos importante, apresentando para ambos os casos uma excelente performance.

Como já referido, dos 366 dias do ano de 2008, informação relativamente a 301 dias é que foram analisados, por isso, quando é referido ao termo anual tratam-se desses dias. Faltam ainda 65 dias, começando no mês de outubro, sendo que as informações sísmicas relativamente ao mês de novembro são inexistentes, apresentando falhas também no mês de dezembro, da mesma forma que para o caso de PESTR.

5.7 Tempo de Detecção sísmica

5.7.1 Tempo de Detecção com o detetor PESTR

Para uma abordagem EW convém levar em conta o instante de tempo necessário, para identificar os eventos. Com a análise do histograma podemos perceber até que ponto reduzindo a informação sísmica passada a rede neuronal é obtido uma excelente performance, reduzindo o tempo de deteção.

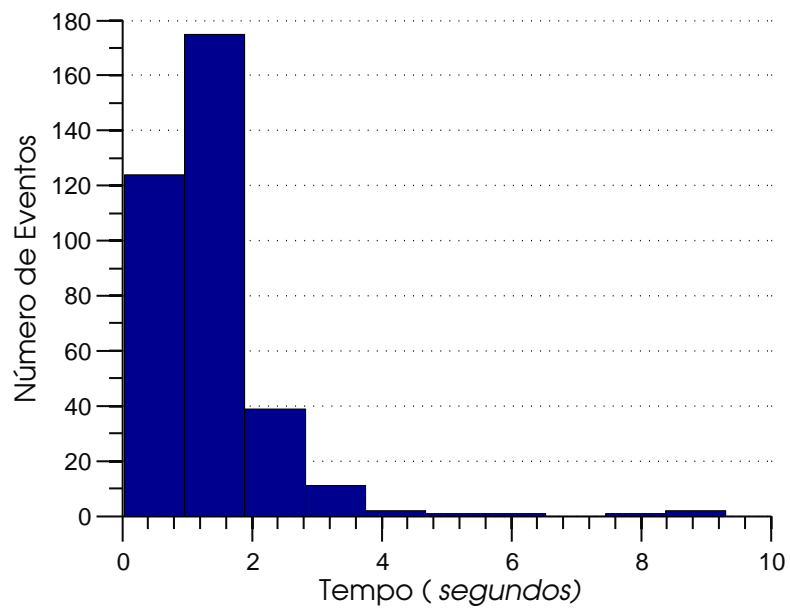


Figura 5.1: Tempo de Detecção para PESTR-EW (Escalonamento Escalamat)

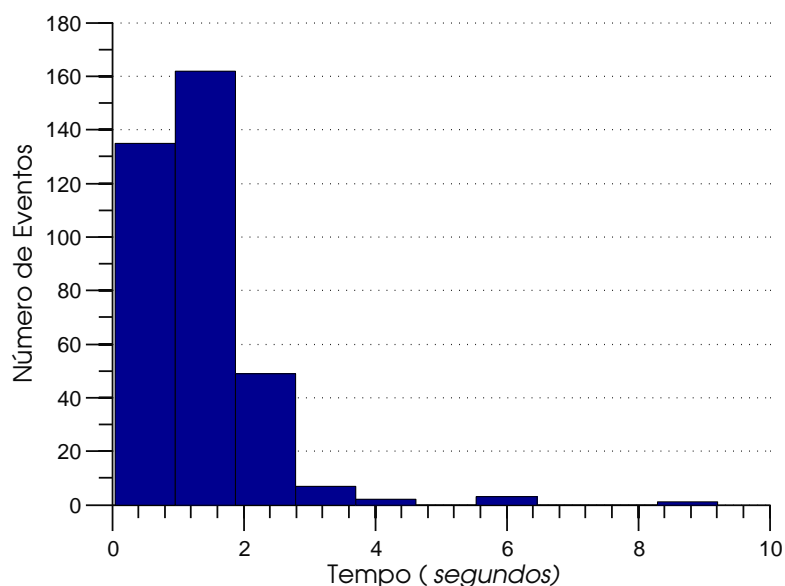


Figura 5.2: Tempo de Detecção para PESTR-EW (Escalonamento SCALE)

O classificador apresentado anteriormente foi treinado com segmentos de 120 segundos de informação sísmica. De acordo com [Ruano, Madureira, Ozias, Khosravani, Ruano, and Ferreira, 2013], o tempo de deteção médio para PVAQ e PESTR foi 88 segundos e 110 segundos respetivamente. Estes tempos são demasiados longo para serem aplicados em sistemas de alerta antecipado (EWS). De forma a reduzir os tempos de deteção, é utilizado menos informações sobre o evento sísmico, e mais informações entre a transição entre o ruído e o evento deve ser utilizado. Assim apresentaram um novo modelo de extração de features, utilizando um conceito de sobreposição de janelas, de pequenas e médias durações.

Agora podemos ver que o tempo médio de deteção foi apenas 1 segundo. Esta redução verifica-se para ambos as operações de escala (Figura 5.2 e Figura 5.1), e para dados sísmicos de qualquer uma das estações (Figura 5.3 e Figura 5.4).

A magnitude sísmica, que avalia a quantidade de energia libertada pelo hipocentro durante a atividade sísmica, sendo uma grandeza de grande importância para a análise de dados sísmicos.

Podemos verificar que para o caso considerado a magnitude varia dentro de uma gama bem definida, a intensidade mínima é registada em 0.5 e o máximo em 4.7.

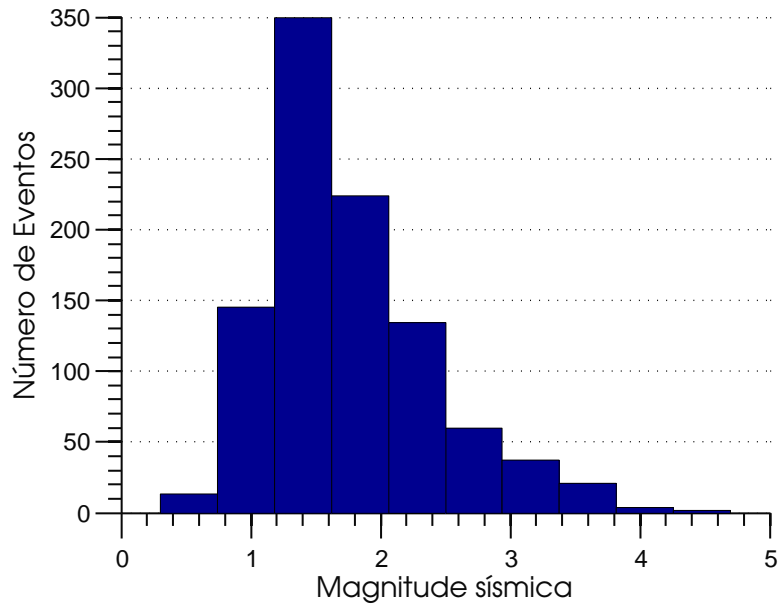


Figura 5.3: Histogramas das magnitudes sísmicas PESTR (2008).

5.7.2 Tempo de Detecção com o detetor PVAQ

Para a estação sísmica de PVAQ, aplicando o novo processo de extração das features o tempo médio de deteção foi reduzido, para apenas 1 segundo, como podemos ver nas Figuras 5.4 e 5.5.

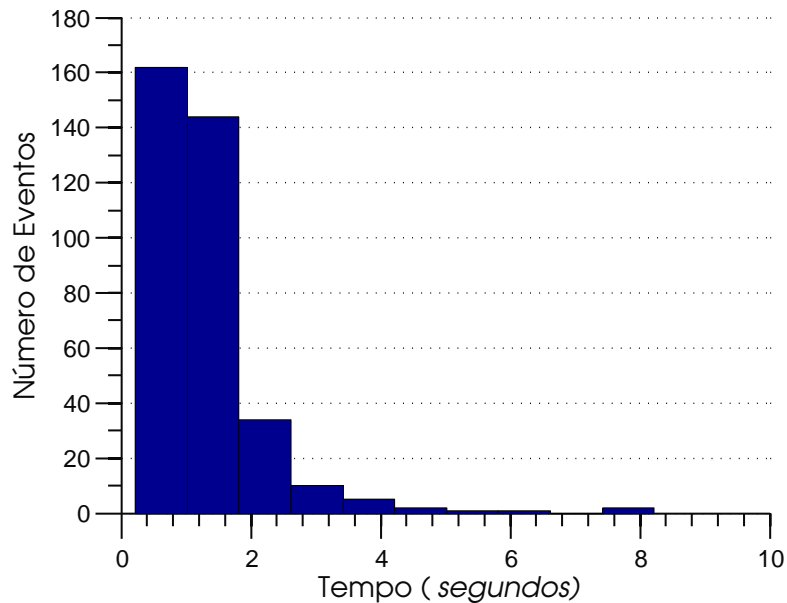


Figura 5.4: Tempo de Detecção para PESTR-EW (Escalonamento Escalamat).

Para o ano de 2008 em relação a a magnitude sísmica, a intensidade mínima é registada em

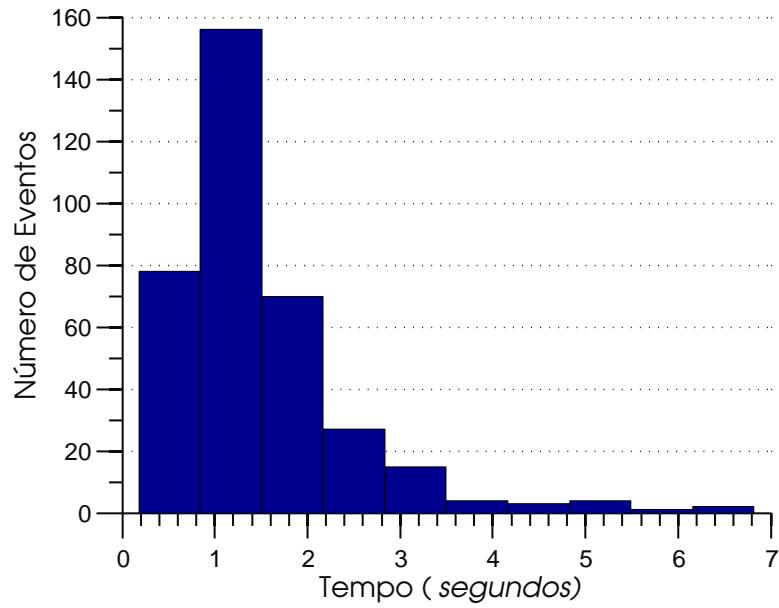


Figura 5.5: Tempo de Detecção para PESTR-EW (Escalonamento SCALE).

0.5 e o máximo em 4.7 (Figura 5.6).

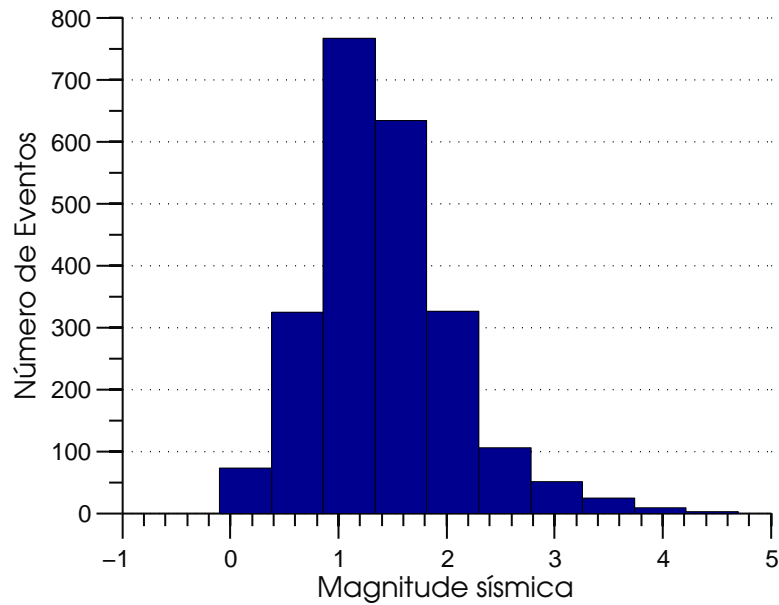


Figura 5.6: Histogramas das magnitudes sísmicas PVAQ (2008).

5.8 Sistemas de deteção considerando os vários detetores implementados

Após a implementação dos classificadores, estuda-se a sua utilização conjunta na deteção de eventos. Testaram-se duas situações: um evento é declarado quando for detetado por ambos os detetores (designado por abordagem AND), ou quando pelo menos um dos detetores deteta o evento (abordagem OR).

Quando não for detetado por nenhum dos detetores, é um caso de ruído, ou seja, não ocorreu nenhum evento. As duas situações são representadas na Figura 5.7.

A primeira abordagem como apresentado na Figura 5.7 (a) , utiliza a função de operação OR, ou seja, quando qualquer um dos detetores ao avaliar um segmento sísmico identifica a ocorrência de um evento, é declarado um caso de deteção (V), caso nenhum dos detetores identifica sinal sísmico no segmento em avaliação é registado como um caso de ruído, não houve ocorrência de sismo.

Na segunda abordagem (Figura 5.7 (b)) a função de operação é AND, neste caso um evento só é declarado quando for identificado simultaneamente nas duas estações. Quando o segmento em questão é declarado como sendo sismo, tanto pelo classificador PESTR como pelo classificador PVAQ estamos perante um caso de evento sísmico é declarado a ocorrência de um (V).

Utilizou-se uma abordagem experimental na escolha dos catálogos sísmicos utilizados para os testes do funcionamento dos vários detetores simultaneamente. Numa primeira fase utilizaram-se os catálogos de PESTR e PVAQ sem nenhuma alteração.

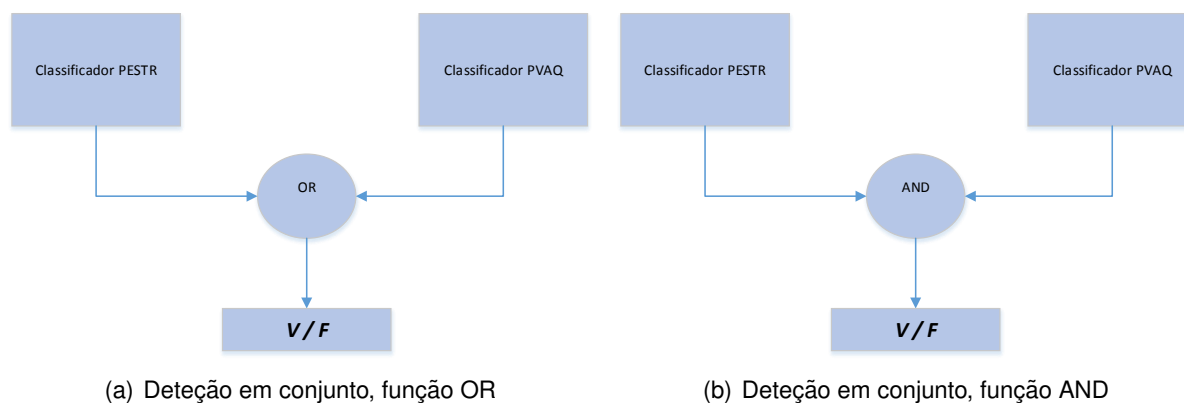


Figura 5.7: Modos de operação utilizando os vários detetores desenhados.

O catálogo de PVAQ, contendo tempos relativos aos eventos sísmicos do catálogo dos Vaqueiros contém mais do que o dobro de eventos registados, comparado com o catálogo de PESTR.

No caso de PESTR foram registados os instantes de chegada de 998 eventos, onde a magnitude sísmica varia de acordo com a histograma apresentado na Figura 5.3.

Para PVAQ foram registados 2436 eventos com uma amplitude a variar de acordo com o histograma apresentado na Figura 5.6.

As escolhas do catálogo sísmico bem como o modo de operação, influenciam a performance obtida. Para além da função utilizada (AND ou OR) para definir a classe que pertence cada evento é importante a escolha do catálogo.

Por isso após o teste com os dois catálogos base, sem nenhuma alteração, houve a necessidade de garantir que todos os registos comuns a ambas as estações são considerados. Sendo assim foi avaliado o desempenho levando em conta todos os eventos registados nos catálogos das respetivas estações sísmicas.

Resumidamente identificam-se os três catálogos utilizados.

- PESTR – Contém toda a informação sísmica registada na estação sísmica dos Estremoz no ano de 2008 (998 registos).
- PVAQ – Contém toda a informação sísmica registada na estação sísmica de Vaqueiros nos ano de 2008 (2436 registos).
- Geral (PVAQ + PESTR) – Eventos dos dois catálogos sísmicos, todas as informações contidas em ambas as estações encontram-se registadas (3212 registos).

5.8.1 Resultados considerando vários detetores

5.8.1.1 Resultados operação OR

Catálogo 1 - PESTR

R	S
90.13	97.20

Tabela 5.9: Performance conjunta Catálogo PESTR

Catálogo 2 - PVAQ

R	S
95.47	97.31

Tabela 5.10: Performance conjunta Catálogo PVAQ

Catálogo 3 - Geral (PVAQ + PESTR)

R	S
91.95	97.33

Tabela 5.11: Performance conjunta Catálogo Geral (PVAQ + PESTR)

Podemos verificar que quando avaliado o sistema com o catálogo sísmico de PVAQ, utilizando a função de operação OR, a quantidade de eventos sísmicos detetados corretamente (Especificidade Tabela 5.10) equivale aos obtidos utilizando um único modelo.

Utilizando a abordagem AND os falsos alarmes são quase nulos (Sensibilidade Tabela 5.14). Ainda nesta abordagem, analisando a sensibilidade podemos perceber que a ocorrência de falsos alarmes apresenta uma taxa menor do que os 0.5% . Quanto utilizado a operação AND reduz-se para quase nulo os falsos alarmes, mas por outro lado o número de deteções diminui consideravelmente.

5.8.1.2 Resultados operação AND

Catálogo 1 - PESTR

R	S
82.84	99.51

Tabela 5.12: Performance conjunta Catálogo PESTR

Catálogo 2 - PVAQ

R	S
78.01	99.45

Tabela 5.13: Performance conjunta Catálogo PVAQ

Catálogo 3 - Geral (PVAQ + PESTR)

R	S
76.86	99.52

Tabela 5.14: Performance conjunta Catálogo Geral (PVAQ + PESTR)

Os melhores resultados considerando os parâmetros de avaliação R e S, são obtidos quando utilizado a operação OR. De forma mais concreta, quando utilizado essa operação face aos tempos relativos aos eventos sísmicos do catálogo PVAQ.

Consegue-se uma percentagem de não deteção idêntica para quando utilizado um único modelo, na abordagem OR. O número de falsos alarmes para a combinação de modelos é quase inexistente, na abordagem AND.

5.9 Discussão dos Resultados

Ao contrário do que acontece no caso de PESTR, para os dados da estação sísmica PVAQ, a escolha do escalonamento não apresenta melhorias nos resultados alcançados. Isto pode não ter acontecido devido as, características dos dados sísmicos registados ou da própria estação em si, tendo em conta a sua localização, a quantidade de ruído no sinal registado, ou da influência do próprio catálogo, visto para cada estação as informações contidas diferem.

Durante o procedimento experimental, na fase de teste em modo Off-line, quando aplicado a estratégia de aprendizagem ativa, foram determinados vários fechos convexos, onde variavam o número de vértices. Como a performance do classificador escolhido variava dependendo do fecho convexo escolhido, primeiramente houve uma fase de experiência com todos os classificadores onde foram escolhidos os que apresentaram melhor performance On-line.

Quando os dois classificadores são testados como um todo, ao avaliar a performance face aos catálogos sísmicos variados, verificou-se que utilizando a combinação de detetores, consegue-se uma percentagem de deteção na abordagem OR, idêntica aos obtidos utilizando

um único modelo. Ainda o número de falsos alarmes para a combinação de modelos é quase inexistente para a abordagem AND.

Comparando o funcionamento dos modelos neuronais desenhados, com os apresentados anteriormente podemos ver que o comportamento é linear, desta forma a comparação da performance aqui obtida com a apresentada pelo sistema de detecção atualmente em funcionamento ao nível da estação sísmica continua a ser válido (Comparação dos resultados obtidos com o Sistema de Detecção ao Nível da Estação Sísmica em Funcionamento no CND Secção 4.3). A performance é muito superior, aos 50% apresentados em termos de Sensibilidade e Especificidade pelo sistema atual. Os resultados alcançado com a SVM no modo Off-line foram perfeitos após 4 iterações, e para o teste em modo contínuo, haviam poucas não detecções (Valores de Especificidade) e quase nenhum falso alarme (Valores de Sensibilidade).

Por último, de referir que a simulação do processo de detecção de eventos, para quando simulado o funcionamento em ambiente real, exige, em média 1 á 2 segundos para determinar com precisão a classe do evento. Este tempo aumenta com a quantidade de informação utilizada no processamento (variação do tamanho de janelas e cura e média duração). A utilização do método num suporte de detecção físico, diminuirá a ordem temporal normal de detecção 120 segundos de forma significativa.

5.10 Resumo

Neste capítulo são apresentados os resultados obtidos durante as experiências. São apresentados os resultados nos dois modos de operação aqui utilizados, analisando a performance para cada estação e os parâmetros levados em consideração.

Capítulo 6

Conclusão

6.1 Introdução

Esta dissertação apresentou um estudo da aplicação de um tipo de classificador SVM para a análise de problemas de classificação de eventos sísmicos. Nesta secção são apresentados, os pontos importantes do desenvolvimento deste trabalho, contribuições do trabalho, direções de trabalho futuro de forma a melhorar o sistema aqui apresentado e alguma limitações encontradas durante o desenvolvimento.

6.2 Contribuições

Tendo como base um classificador SVM desenhado anteriormente, foi desenhado classificadores para as estações de PVAQ e PESTR da rede sísmica nacional (Portugal). Foi verificado o desempenho dos classificadores em funcionamento Off-line e On-line, e ainda verificado o desempenho do sistema considerando os vários detetores implementados. Confirmando os resultados obtidos anteriormente, verificou-se que com essa abordagem o desempenho do sistema é muito superior ao sistema em uso nas estações da rede sísmica. Para além disso a abordagem EW permitiu reduzir drasticamente o tempo necessário para detetar os eventos, com isso tornando possível a sua utilização em sistemas de alerta antecipado de eventos sísmicos.

As experiências realizadas também podem ser ajustadas por meio de parâmetros das SVMs. O estudo realizado abre assim diversas possibilidades de testes, que não foram realizados devido a limitações tempo e serão propostos como trabalhos futuros.

As pesquisas bibliográficas conduzidas e apresentadas no Estado da Arte (Capítulo 2) podem ser consideradas uma outra contribuição desta dissertação. Embora existam diver-

os trabalhos relacionados com a utilização de redes neuronais em processamento de sinais sísmicos, procurou-se reunir os principais, tendo em conta os métodos e resultados obtidos.

Como dito anteriormente na Secção 1.7, parte do estudo desenvolvido nessa dissertação foi utilizado para a publicação de um artigo [Ruano, Madureira, Ozias, Khosravani, Ruano, and Ferreira, 2013]:

“A Support Vector Machine Seismic Detector for Early-Warning Applications”.

(keywords = early-warning systems, seismic detection, support vector machines)

Apresentado em, ***The 3rd IFAC International Conference on Intelligent Control and Automation Science (ICONS 2013), Chengdu, China, 2-4 September 2013.***

6.3 Limitações

Algumas limitações podem ser apontadas nas estratégias de testes dos classificadores, bem como nas experiências conduzidas.

Todo o processo de simulação do funcionamento em ambiente real é computacionalmente custoso, embora em princípio serem facilmente paralelizáveis e tirar proveito de tecnologias distribuídas. De forma a não haver amostras de dados não processadas, o processamento de informações registados mantém registos de dados processados anteriormente, para quando houver necessidade, fazer conexão no processo de extração de segmentos. Como a dimensão dos registos não é linear, paralelizar o funcionamento impõe perda de pequenos registos do sinal, o que não apresenta grande influência nos resultados alcançados de acordo com os testes feitos utilizando o esquema de processamento distribuído fornecido pela ferramenta utilizada MATLAB®.

Em relação as experiências realizadas, uma das limitações refere-se a escolha de valores para os parâmetros da função de kernel, a própria função de kernel, a melhor estratégia para utilizar os fechos convexos calculados e a normalização a utilizar.

Também no caso da utilização do catálogo sísmico, a utilização de um catálogo padrão onde são apresentados registos sísmicos de todas as estações, permitindo desenvolver classificadores que aproveitassem da capacidade de generalização das SVMs para desenvolver classificadores genéricos a utilizar para dados de qualquer estação sísmica.

Seria interessante analisar o funcionamento para um conjunto de treino com maior quantidade de sismos, permitindo quem sabe depois aplicar a SVM num esquema de classificação de apenas uma classe, onde são dados amostras de sismos e depois o sistema tentará determinar os eventos que não pertencem a classe (*“outlier”*). Dessa forma desde que sejam sismos, não havendo importância de diferenciar entre sismos locais regionais, e sismos con-

siderados distantes. Com essa abordagem a performance em termos de Especificidade (R), podia aumentar consideravelmente.

Um outro problema já referido anteriormente [Madureira, 2009], é o tratamento a dar a eventos que se encontram na região de transição ,entre, sismos locais e regionais, e sismos distantes.

A forma como é feito a seleção dos dados que constituíram o conjunto de treino, pode ser melhorado, utilizando metodologias determinísticas e aplicando aprendizagem ativa de forma a ter o mais pequeno conjunto de treino possível sem influenciar a performance do sistema.

Também o catálogo mantém registos de sismos locais e regionais assim como os distantes, convinha separar esses dados de forma a não influenciar a performance dos sistema.

Não foi possível ter acesso a mais dados para além das duas estações sísmicas apresentadas, seria interessante utilizar um número considerável de estações de modo a testar o desempenho da rede como um todo.

6.4 Trabalho Futuro

Tendo como ponto de partida as limitações apresentadas na secção anterior, um conjunto de trabalhos futuros podem ser apontados.

Uma pesquisa promissora seria aperfeiçoar o esquema de processamento em modo contínuo de forma a tirar partido das tecnologias distribuídas.

Aplicar treinos com diferentes features, e utilizar métodos determinísticos para escolha dos parâmetros da SVM. Também seria interessante utilizar outros métodos apresentados na revisão da literatura e comparar a performance dos métodos.

Seria importante procurar tornar o algoritmo mais completo, de forma escolher registos sísmicos que tornem possível criar um detetor que apresenta boa performance para diferentes estações.

Seria ainda de interesse estudar a possibilidade de implementação de uma estação sísmica de alerta antecipado, onde utiliza-se essa abordagem.

6.5 Resumo

Nesta dissertação de mestrado foi apresentado um trabalho em que explorou-se a utilização de um classificador do tipo SVM binário, para detetar de forma automática eventos de duas

classes, a primeira classe constituída por sismos locais e regionais, e a outra classe pelas restantes situações.

São descritos os aspetos mais importantes de forma a conseguir fazer as experiências pretendidas.

Por fim são apresentados os resultados e a sua respetiva análise.

O trabalho realizado abre ainda diversas possibilidades de pesquisas futuras. Levando em conta a importância do problema aqui tratado, e a qualidade dos resultados alcançados pode assim ser considerado um campo de investigação promissor o aqui analisado.

Apêndice A

Informações Adicionais

A.1 Rede de Vigilância Sísmica

[Madureira, 2009], descreve a rede de vigilância sísmica como tendo a função de detetar, registar e processar, os eventos sísmicos que são perceptíveis pelo equipamento numa determinada região do globo. Sendo que o elemento base é a estação sísmica.

Na Figura A.1 é apresentada a rede sismográfica de Portugal Continental. Para mais informações sobre a rede de vigilância sísmica em Portugal é recomendado a consulta do trabalho citado anteriormente e os trabalhos a ele associados.

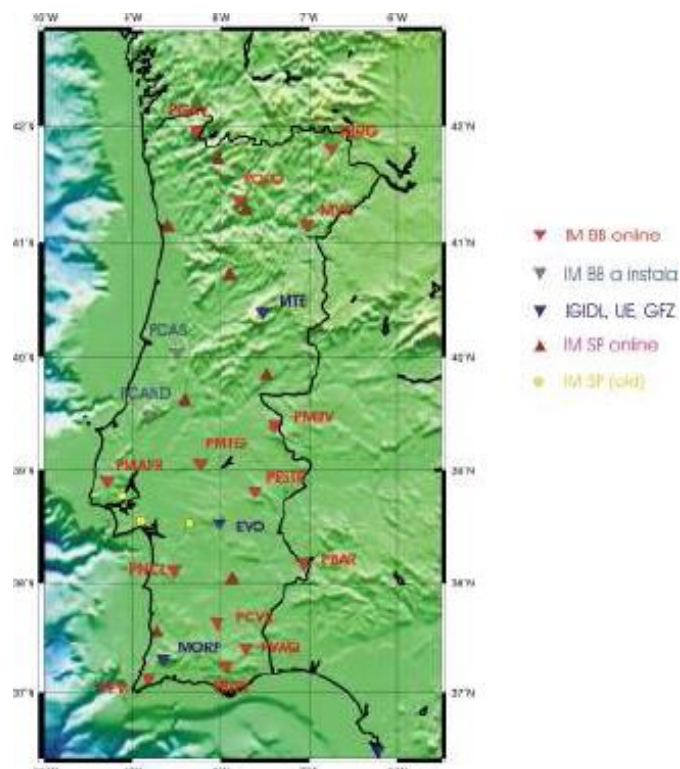


Figura A.1: Rede sismográfica de Portugal Continental.

Bibliografia

- M A Abu-ElSoud, F E Z Abou-Chadi, A M Amin, and M Mahana. Classification of seismic events in suez gulf area, egypt using artificial neural network. In *Electrical, Electronic and Computer Engineering, 2004. ICEEC '04. 2004 International Conference on*, pages 337–340, 2004. doi: 10.1109/ICEEC.2004.1374460.
- R V Allen. AUTOMATIC EARTHQUAKE RECOGNITION AND TIMING FROM SINGLE TRACES. *Bulletin of the Seismological Society of America*, 68(5), 1978. URL <GotoISI>://WOS:A1978FV07000017.
- J.K Anlauf and M BIEHL. The AdaTron: an Adaptive Perceptron Algorithm. 10(December): 687–692, 1989.
- S V Baranov. Application of the wavelet transform to automatic seismic signal detection. *Izvestiya-Physics of the Solid Earth*, 43(2):177–188, 2007. doi: 10.1134/s1069351307020085. URL <GotoISI>://WOS:000246402400008.
- M Benbrahim, A Daoudi, K Benjelloun, and A Ibenbrahim. Discrimination of Seismic Signals Using Artificial Neural Networks. In C Ardil, editor, *Proceedings of World Academy of Science, Engineering and Technology, Vol 4*, volume 4, pages 4–7. World Acad Sci, Eng & Tech-Waset, Canakkale, 2005. ISBN 1307-6884. URL <GotoISI>://WOS:000259996000002.
- Moritz Beyreuther and Joachim Wassermann. Continuous earthquake detection and classification using discrete Hidden Markov Models. *Geophysical Journal International*, 175(3): 1055–1066, December 2008a. ISSN 0956540X. doi: 10.1111/j.1365-246X.2008.03921.x. URL <http://gji.oxfordjournals.org/cgi/doi/10.1111/j.1365-246X.2008.03921.x>.
- Moritz Beyreuther and Joachim Wassermann. Continuous earthquake detection and classification using discrete Hidden Markov Models. *Geophysical Journal International*, 175(3):1055–1066, 2008b. doi: 10.1111/j.1365-246X.2008.03921.x. URL <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=35393799&site=ehost-live&scope=site>.

- Moritz Beyreuther, Roberto Carniel, and Joachim Wassermann. Continuous Hidden Markov Models: Application to automatic earthquake detection and classification at Las Cañadas caldera, Tenerife. *Journal of Volcanology & Geothermal Research*, 176(4):513–518, 2008. doi: 10.1016/j.jvolgeores.2008.04.021. URL <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=34533139&site=ehost-live&scope=site>.
- Moritz Beyreuther, Conny Hammer, Joachim Wassermann, Matthias Ohrnberger, and Tobias Megies. Constructing a Hidden Markov Model based earthquake detector: application to induced seismicity. *Geophysical Journal International*, 189(1):602–610, April 2012. ISSN 0956540X. doi: 10.1111/j.1365-246X.2012.05361.x. URL <http://gji.oxfordjournals.org/cgi/doi/10.1111/j.1365-246X.2012.05361.x>.
- F Botella, J Rosa-Herranz, J J Giner, S Molina, and J J Galiana-Merino. A real-time earthquake detector with prefiltering by wavelets. *Computers & Geosciences*, 29(7):911–919, 2003. doi: 10.1016/S0098-3004(03)00099-2. URL <GotoISI>://WOS:000184980200010.
- Christopher J C Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998a. ISSN 13845810. doi: 10.1023/A:1009715923555. URL <http://www.springerlink.com/index/Q87856173126771Q.pdf>.
- Christopher J C Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998b. ISSN 13845810. doi: 10.1023/A:1009715923555. URL <http://www.springerlink.com/index/Q87856173126771Q.pdf>.
- Olivier Chapelle and Nello Cristianini. Choosing multiple parameters for support vector machines. In *Machine Learning*, page 2002, 2002.
- Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet, and Sayan Mukherjee. Choosing multiple parameters for support vector machines. In *Machine Learning*, page 2002, 2000.
- David Cohn, Les Atlas, and Richard Ladner. Improving generalization with active learning. *Machine Learning*, 15(2):201–221, 1994. ISSN 0885-6125. doi: 10.1007/BF00993277. URL <http://dx.doi.org/10.1007/BF00993277>.
- Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. ISSN 08856125. doi: 10.1007/BF00994018. URL <http://www.springerlink.com/index/10.1007/BF00994018>.
- Gloria Curilem, Jorge Vergara, Gustavo Fuentealba, Gonzalo Acuña, and Max Chacón. Classification of seismic signals at Villarrica volcano (Chile) using neural networks and ge-

- netic algorithms. *Journal of Volcanology and Geothermal Research*, 180(1):1–8, 2009. doi: 10.1016/j.jvolgeores.2008.12.002. URL <http://www.sciencedirect.com/science/article/pii/S0377027308006355>.
- H C Dai and C MacBeth. Effects of learning parameters on learning procedure and performance of a BPNN. *Neural Networks*, 10(8):1505–1521, 1997. doi: 10.1016/s0893-6080(97)00014-2. URL <GotoISI>://WOS:A1997YJ81300013.
- E Del Pezzo, A Esposito, F Giudicepietro, M Marinaro, M Martini, and S Scarpetta. Discrimination of earthquakes and underwater explosions using neural networks. *Bulletin of the Seismological Society of America*, 93(1):215–223, 2003. doi: 10.1785/0120020005. URL <GotoISI>://WOS:000181971700015.
- S Diersen, E J Lee, D Spears, P Chen, and L Q Wang. Classification of Seismic Windows Using Artificial Neural Networks. In M Sato, S Matsuoka, P M A Soot, G D VanAlbada, and J Dongarra, editors, *Proceedings of the International Conference on Computational Science*, volume 4, pages 1572–1581. Elsevier Science Bv, Amsterdam, 2011. ISBN 1877-0509. doi: 10.1016/j.procs.2011.04.170. URL <GotoISI>://WOS:000299165200169.
- Farid U Dowla, Steven R Taylor, and Russell W Anderson. Seismic discrimination with artificial neural networks: Preliminary results with regional spectral data. *Bulletin of the Seismological Society of America*, 80(5):1346–1373, 1990. URL <http://www.bssaonline.org/content/80/5/1346.abstract>.
- Richard O Duda, Peter E Hart, and David G Stork. *Pattern Classification*, volume 2 of *Pattern Classification and Scene Analysis: Pattern Classification*. Wiley, 2001. ISBN 0471056693. doi: 10.1038/npp.2011.9. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.133.1318&rep=rep1&type=pdf>.
- A M Esposito, F Giudicepietro, S Scarpetta, L D'Auria, M Marinaro, and M Martini. Automatic Discrimination among Landslide, Explosion-Quake, and Microtremor Seismic Signals at Stromboli Volcano Using Neural Networks. *Bulletin of the Seismological Society of America*, 96(4A):1230–1240, 2006. doi: 10.1785/0120050097. URL <http://www.bssaonline.org/content/96/4A/1230.abstract>.
- TT Frie, N Cristianini, and Colin Campbell. The kernel-adatron algorithm: a fast and simple learning procedure for support vector machines. *Machine Learning: Proceedings of the ...*, 1998. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.42.2060&rep=rep1&type=pdf>.

- S Gentili and A Michelini. Automatic picking of P and S phases using a neural tree. *Journal of Seismology*, 10(1):39–63, 2006. doi: 10.1007/s10950-006-2296-6. URL <GotoISI>://WOS:000238205100003.
- Ferdinando Giacco, Antonietta Maria Esposito, Silvia Scarpetta, Flora Giudicepietro, and Maria Marinaro. Support Vector Machines and MLP for automatic classification of seismic signals at Stromboli volcano, 2009.
- Steven J Gibbons and Frode Ringdal. The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1):149–166, 2006. doi: 10.1111/j.1365-246X.2006.02865.x. URL <http://dx.doi.org/10.1111/j.1365-246X.2006.02865.x>.
- F Giudicepietro, A Esposito, M Martini, and S Scarpetta. REPORT.
- Olivia Grubert and Bridget Vuong. Earthquake Waveform Recognition. pages 1–5, 2012.
- S Haykin. *Neural Networks: A Comprehensive Foundation*, volume 13. Prentice Hall, 1999. ISBN 0132733501. doi: 10.1017/S0269888998214044. URL http://www.journals.cambridge.org/abstract_S0269888998214044.
- M A Hearst, B Schölkopf, S Dumais, E Osuna, and J Platt. Trends and Controversies - Support Vector Machines. *IEEE Intelligent Systems*, 13(4):18–28, 1998. URL <http://computer.org/intelligent/ex1998/pdf/x4018.pdf>.
- Tommi S. Jaakkola and David Haussler. Probabilistic kernel regression models. In *In Proceedings of the 1999 Conference on AI and Statistics*. Morgan Kaufmann, 1999.
- M Joswig. PATTERN-RECOGNITION FOR EARTHQUAKE DETECTION. *Bulletin of the Seismological Society of America*, 80(1), 1990. URL <GotoISI>://WOS:A1990CL67800011.
- K V Kislov and V V Gravirov. Earthquake arrival identification in a record with technogenic noise. *Seismic Instruments*, 47(1):66–79 LA – English, 2011. ISSN 0747-9239. doi: 10.3103/S0747923911010129. URL <http://dx.doi.org/10.3103/S0747923911010129>.
- L Kueperkoch, T Meier, J Lee, W Friederich, and Egelados Working Grp. Automated determination of P-phase arrival times at regional and local distances using higher order statistics. *Geophysical Journal International*, 181(2), 2010. doi: 10.1111/j.1365-246X.2010.04570.x. URL <GotoISI>://WOS:000276697000034.

- Satish Kumar, B K Sharma, Parkhi Sharma, and M A Shamshi. 24 Bit seismic processor for analyzing extra large dynamic range signals for early warnings. *Journal of Scientific & Industrial Research*, 68(5), 2009. URL <GotoISI>://WDS:000265899100005.
- H Langer, S Falsaperla, M Masotti, R Campanini, S Spampinato, and A Messina. Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna, Italy. *Geophysical Journal International*, 178(2):1132–1144, 2009. doi: 10.1111/j.1365-246X.2009.04179.x. URL <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=43227384&site=ehost-live&scope=site>.
- Ana Carolina Lorena. *Introdução as Máquinas de Vetores Suporte*. 2003.
- G Madureira, A E Ruano, and M G Ruano. On-line operation of an intelligent seismic detector. *Soft Computing Applications*, pages 531–542, 2013a.
- Guilherme Madureira. *Deteção de Eventos Sísmicos Utilizando Redes Neurais*. Master's thesis, 2009.
- Guilherme Madureira and A. E. Ruano. A neural network seismic detector. January 2009. ISSN 1789-6932. URL <http://sapientia.ualg.pt/handle/10400.1/2178>.
- Guilherme Madureira and António Ruano. *Aplicações de Técnicas de Inteligência Computacional à Deteção Sísmica Applications of Computational Intelligence Techniques for Seismic Detection*. pages 1–4, 2012.
- Guilherme Madureira, A. E. Ruano, and M. Graca Ruano. On-line operation of an intelligent seismic detector. January 2013b. URL <http://sapientia.ualg.pt/handle/10400.1/2134>.
- M Masotti, S Falsaperla, H Langer, S Spampinato, and R Campanini. Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy. *Geophys. Res. Lett.*, 33(20):L20304, 2006. doi: 10.1029/2006GL027441. URL <http://dx.doi.org/10.1029/2006GL027441>.
- M Massa, G Ferretti, D Spallarossa, and C Eva. Improving automatic location procedure by waveform similarity analysis: An application in the South Western Alps (Italy). *Physics of the Earth and Planetary Interiors*, 154(1):18–29, 2006. doi: 10.1016/j.pepi.2005.07.003. URL <http://www.sciencedirect.com/science/article/pii/S0031920105001603>.
- Bezzeghoud Mourad, Fernando Borges José, and Bento Caldeira. Contribuição da rede sísmica do Centro de Geofísica de Évora para a Investigação Sismológica em Por-

- tugal Mourad. *Vasa*, 80(lm), 2008. URL <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>.
- Niksa Orlic and Sven Loncaric. Earthquake—explosion discrimination using genetic algorithm-based boosting approach. *Computers & Geosciences*, 36(2):179–185, 2010. doi: 10.1016/j.cageo.2009.05.006. URL <http://www.sciencedirect.com/science/article/pii/S0098300409002714>.
- Domenico Patanè and Ferruccio Ferrari. ASDP: a PC-based program using a multi-algorithm approach for automatic detection and location of local earthquakes. *Physics of the Earth and Planetary Interiors*, 113(1–4):57–74, 1999. doi: 10.1016/S0031-9201(99)00030-8. URL <http://www.sciencedirect.com/science/article/pii/S0031920199000308>.
- V W Porto, L J Fogel, and D B Fogel. Using evolutionary computation for seismic signal detection: a homeland security application. In *Computational Intelligence for Homeland Security and Personal Safety, 2004. CIHSPS 2004. Proceedings of the 2004 IEEE International Conference on*, pages 62–66, 2004. doi: 10.1109/CIHSPS.2004.1360209.
- Pedro João Soares Rodrigues. Técnicas de segmentação e de classificação em imagens : estudo de um caso de aplicação, December 2008. URL <http://repositorium.sdum.uminho.pt/handle/1822/8909>.
- A. E. Ruano and Guilherme Madureira. Active learning in the detection of seismic events using artificial neural networks. January 2011. URL <http://sapiencia.ualg.pt/handle/10400.1/2275>.
- A.E. Ruano. *Artificial Neural Networks*. Center of Intelligent Systems, UALG, 1996.
- António E Ruano, G. Madureira, Barros Ozias, Hamid R. Khosravani, Maria G Ruano, and Pedro M. Ferreira. A Support Vector Machine Seismic Detector for Early-Warning Applications. 2013.
- Scarpetta. Automatic Classification of Seismic Signals at Mt. Vesuvius Volcano, Italy, Using Neural Networks, 2005. URL <http://hdl.handle.net/2122/436>.
- B K Sharma, S Kumar, S K Mittal, and M A Shamshi. Design improvements in digital seismograph for recording long duration seismic events and aftershocks. *Journal of Scientific & Industrial Research*, 65(1), 2006. URL <GotoISI>://WOS:000234443500004.
- Reinoud Sleeman and Torild van Eck. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and*

- Planetary Interiors*, 113(1–4):265–275, 1999. doi: 10.1016/S0031-9201(99)00007-2. URL <http://www.sciencedirect.com/science/article/pii/S0031920199000072>.
- A Smola, P Bartlett, Bernhard Schölkopf, and Dale Schuurmans. Introduction to large margin classifiers. 2000. ISSN 0262283972.
- Alex J Smola and Bernhard S C H Olkopf. A tutorial on support vector regression. pages 199–222, 2004.
- S W Stewart. REAL-TIME DETECTION AND LOCATION OF LOCAL SEISMIC EVENTS IN CENTRAL CALIFORNIA. *Bulletin of the Seismological Society of America*, 67(2):433–452, 1977. URL <GotoISI>://WOS:A1977DC76100016.
- Timo Tiira. Detecting teleseismic events using artificial neural networks. *Computers & Geosciences*, 25(8):929–938, 1999. doi: 10.1016/S0098-3004(99)00056-4. URL <http://www.sciencedirect.com/science/article/pii/S0098300499000564>.
- Simon Tong and Daphne Koller. with Applications to Text Classification. pages 45–66, 2001.
- V N Vapnik. *The Nature of Statistical Learning Theory*, volume 8. 1995. ISBN 0387945598. doi: 10.1109/TNN.1997.641482.
- Vladimir Vapnik. Estimation of Dependences Based on Empirical Data: Empirical Inference Science, 2006.
- Vladimir Vapnik and Olivier Chapelle. Bounds on error expectation for support vector machines, 1999.
- Vladimir N Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.
- Andreas Vlachos. *Active Learning with Support Vector Machines*. PhD thesis, 2004.
- J Wang and T L Teng. ARTIFICIAL NEURAL-NETWORK-BASED SEISMIC DETECTOR. *Bulletin of the Seismological Society of America*, 85(1):308–319, 1995. URL <GotoISI>://WOS:A1995QJ46900022.
- J Wang and T L Teng. Identification and picking of S phase using an artificial neural network. *Bulletin of the Seismological Society of America*, 87(5), 1997. URL <GotoISI>://WOS:A1997YC39200005.

P Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967. ISSN 00189278. doi: 10.1109/TAU.1967.1161901.

Mitchell Withers, Richard Aster, Christopher Young, Judy Beiriger, Susan Moore, and Julian Trujillo. A Comparison of Select Trigger Algorithms for Automated Global Seismic Phase and Event Detection. 88(1):95–106, 1998.

