

UNIVERSIDADE ABERTA



DEPARTAMENTO DE CIÊNCIAS E TECNOLOGIAS

Mestrado em Estatística, Matemática e Computação

**Aplicação de Técnicas de Aprendizagem Supervisionada a
Dados de Comportamento Motor**

Jailson Almeida Santos Lopes

Lisboa, 2017

UNIVERSIDADE ABERTA



DEPARTAMENTO DE CIÊNCIAS E TECNOLOGIAS

Mestrado em Estatística, Matemática e Computação

**Aplicação de Técnicas de Aprendizagem Supervisionada a
Dados de Comportamento Motor**

Dissertação apresentada para a obtenção do grau de Mestre em
Estatística, Matemática e Computação

Jailson Almeida Santos Lopes

Orientador: Doutor Pedro Miguel Picado de Carvalho Serranho, Universidade Aberta

Co-orientador: Doutor Luís Miguel Domingues Ferreira Silva, Laboratório de Comportamento Motor, Faculdade de Motricidade Humana, Universidade de Lisboa

Lisboa, 2017

Resumo

Neste trabalho pretende-se estabelecer um modelo de classificação multiclasse baseado em técnicas de aprendizagem supervisionada como Máquinas de Vetores Suporte (SVM – *Support Vector Machines*), em dados de comportamento motor. Para tal, consideramos dados clínicos de duas amostras: 161 sujeitos diagnosticados com a Doença de Parkinson (DP), submetidos a um estudo de evolução desta doença para a função motora, nos seus diferentes estadios; e 98 sujeitos com a DP, que foram submetidos a uma tarefa de visualizar um ponto móvel, onde foram usados 5 conjuntos de features relacionadas à Distância, Amplitude, Área, Velocidade e Frequência de oscilação do referido ponto. Para a primeira amostra foram consideradas as seguintes features de marcha: Comprimento, Largura, Velocidade e Simetria da passada. A partir destes dados o objetivo é assim entender quais destas features mais contribuem para a classificação dos sujeitos nos diferentes estadios da doença. Numa análise preliminar sobre a caracterização da amostra por grupo, foi aplicada a ANOVA simples para igualdade das médias de altura e idade por estadio. De seguida realizaram-se estudos de validação cruzada (*cross validation*) do modelo SVM para determinar os melhores parâmetros após uma pesquisa em grelha (*grid search*). Começamos por testar o SVM linear nos dados iniciais para a classificação, passando depois para o não-linear. Uma vez que o algoritmo tendia a classificar todos os elementos como pertencentes à classe com maior dimensão, de seguida testamos para novas amostras com as classes de dimensões equilibradas. Consideramos as features de maior diferença de médias entre as classes, a retirada das features correlacionadas e outliers da amostra, e ainda o caso binário com as classes agrupadas duas a duas, onde 1 e 2 formam a classe A, 3 e 4 a B, para tirar complexidade à classificação. Foi ainda realizada a classificação binária de discinesias do membro superior entre indivíduos diagnosticados com ou sem essa discinesia. Novamente equilibramos as dimensões das classes pelo mesmo motivo acima referido. Na maioria dos casos considerados, o modelo obtido não foi satisfatório para o caso multiclasse, para as duas amostras, pelas taxas de precisão baixa em relação aos dados. A precisão melhorou para o caso binário com as amostras equilibradas, para as features Distâncias, e Amplitudes com a extração das correlacionadas, com 80.77% de precisão com Kernel RBF, em relação aos dados. Concluímos que estas features são as que melhor distinguem os estadios de Parkinson.

Palavras-chave: Modelo de classificação, SVM, Dados de comportamento motor, *Features* discriminantes, Aprendizagem.

Abstract

In this work is intended to establish a classification model multiclasse based on supervised learning techniques such as Support Vector Machines (SVM), in motor behavior data. For this, we consider clinical data from two samples: 161 subjects diagnosed with Parkinson's disease (PD) subjected to a study of evolution of this disease to motor function, in its different status. 98 subjects with the DP, which have been submitted to a task to visualize a moving point, where were used five sets of features related to Distance, Amplitude, Area and Speed of oscillation frequency from that point. For the first sample were considered the following features: Length, Width, Speed and Symmetry of last. From these data the goal is to understand which of these features more contribute to the classification of the subjects in the different status of the disease. A preliminary analysis on the characterization of the sample group, was applied to ANOVA simple equal height and age averages for status. In row, were held cross-validation studies of the SVM model to determine the best parameters after a grid search. We start by testing the linear SVM in the initial data for the classification, then moving on to non-linear. Once the algorithm tended to classify all elements belonging to the class with larger dimension, then we test for new samples with balanced dimensions classes. We consider the features of higher average difference between the classes, the withdrawal of features correlated and outliers of the sample, and then the binary case with the classes combined two by two, where 1 and 2 constitute the class, 3 and 4 a B, to take complexity of the classification. It was further held the dyskinesia binary classification of the upper limb between individuals diagnosed with or without this dyskinesia. One more time we balance the size of classes for the same reason stated above. In most cases considered, the model obtained was not satisfactory for the multiclasse case, for the two sample, the rates of low precision in relation to data. Improved accuracy for the binary case with samples balanced to the features Distance, and Amplitudes with the extraction of correlated with 80.77% accuracy with RBF Kernel, in relation to the data. We conclude that these features are those that best distinguish the Parkinson's status.

Keywords: SVM classification model, motor behavior data, discriminant Features, learning.

Dedicatória

Dedico este trabalho à uma mulher muito corajosa que batalhou muito para que eu pudesse chegar até aqui onde cheguei. Admiro muito a coragem dela e espírito de luta que nunca lhe faltou, a minha amada e adorada Mãe querida, **Doroteia Isabel Santos Pires.**

À minha esposa Suzete Almeida e o nosso filho Gabriel Almeida Lopes, pela compreensão e contribuição neste longo caminho de aprendizagem.

Aos meus irmãos, professores, amigos colegas e todos aqueles que me amam e me querem bem.

Mensagem aos estudantes

Para aqueles que estudam, não desistam nunca, pois só assim poderão alcançar os vossos objetivos, assim como eu já alcancei mais um dos meus. Não se sintam cansados e nem pensam em desistir logo ao primeiro obstáculo, pois este é o pior inimigo dos estudos.

O segredo está precisamente na persistência, na força de boa vontade e no espírito de sacrifício.

Devemos apostar nos nossos estudos, pois é essa a única herança que as famílias carenciadas podem e com muito sacrifício deixar aos seus filhos – os estudos.

Estudar é arte de aprender e quanto mais estudarmos, maior é a possibilidade de aprofundar e sedimentar os nossos conhecimentos, e “quem cultiva o conhecimento colhe o sucesso”.

Façam isso para que todos possam ser homens e mulheres de amanhã.

Autor: Jailson Lopes

Reparo: Eneida Lopes

Agradecimentos

O meu mais sincero sentido de agradecimento ao meu orientador, Professor Dr. Pedro Serranho, primeiro por ter aceitado esse desafio em me orientar neste caminho de aprendizagem, e também pelo seu estênico e incondicional apoio que sempre me dispensou durante todo o andamento deste trabalho, pois, sem ele o objetivo não seria concretizado. Do fundo do meu coração, muito obrigado pelas preciosas recomendações e conselhos contribuindo assim para que o presente e singelo estudo chegasse a bom porto.

Ao meu co-orientador, Professor Dr. Luís Silva, também pela disponibilidade em aceitar, sem hesitar, esse mesmo desafio em colaborar com o Dr. Pedro rumo ao objetivo por mim traçado nesta caminhada.

Ao Laboratório de Comportamento Motor da Faculdade de Motricidade Humana pela cedência dos dados por intermédio do meu co-orientador referido no parágrafo antecedente, pelos quais todas as análises práticas deste estudo foram suportadas, pelos demais materiais cedidos e pelas sugestões durante toda a parte prática deste trabalho.

À Professora Dra. Catarina Godinho, por ser a responsável pela recolha dos mesmos.

Aos professores do Mestrado em Estatística, Matemática e Computação (MEMC), pelas orientações e pelo excelente trabalho neste caminho de aprendizagem, à coordenação do MEMC pela preocupação para com os mestrandos e a capacidade imediata em ultrapassar alguns obstáculos deparados ao longo desta jornada.

Aos serviços administrativos na Universidade Aberta pela prontidão e pelas respostas rápidas emitidas sempre que necessárias.

A todos aqueles que estando ao meu redor, me ajudaram no meu crescimento espiritual, mental, cultural, intelectual e social. Me refiro a minha mãe Doroteia Pires, meus irmãos,

colegas, e amigos e demais familiares. Um muito obrigado pela vossa força que constituiu um grande estímulo durante toda a elaboração deste trabalho.

Um especial agradecimento a minha esposa Suzete Almeida e o nosso filho Gabriel Almeida Lopes, pelo apoio de um modo geral, pela paciência e tolerância em todos os momentos em que a minha presença foi distante colocando em causa, a atenção que eles merecem por direito.

Agradecimentos extensivos ao meu colega Arciolindo Pinheiro pelo apoio incondicional aquando das inúmeras dicas relativamente ao octave, pelos conselhos e apoio moral, o mesmo sentido de agradecimento é também extensivo ao meu colega Gilberto Neves Alves, pelo apoio espiritual e pelas dicas, ao colega Adriano Arcanjo pela cedência dos materiais de apoio e também ao amigo/irmão Lenine Fortes Delgado pela disponibilidade imediata na compra dos livros, aquando da sua estadia em Portugal, e pelo envio atempadamente dos mesmos para que eu pudesse ter um bom acompanhamento curricular, contribuindo muito para o meu aproveitamento neste processo de aprendizagem.

Índice Geral

Resumo	i
Abstract.....	ii
Dedicatória.....	iii
Mensagem aos estudantes.....	iv
Agradecimentos	v
Índice Geral	vii
Índice de Figuras	xi
Índice de Tabelas	xii
Lista de Abreviaturas.....	xv
Definições Operacionais.....	xvi
Introdução	1
CAPÍTULO 1	4
1. Revisão da literatura	5
1.1. O contexto das Máquinas de Vetores Suporte	5
1.2. Teoria de Otimização Matemática	5
1.3. Teoria Estatística de Aprendizagem.....	6
1.4. Considerações sobre o Problema de Aprendizagem	8
1.5. Aprendizagem Automática Supervisionada	10
1.6. Máquinas de Vetores Suporte	13
1.6.1. SVM com dados separáveis ou margens rígidas	15
1.6.2. SVM com dados não separáveis ou de margens suaves	21
1.6.3. SVM não linear com “kernel trick”	25
1.6.3.1. Funções de Kernel	28
1.6.3.2. Propriedades de Kernel	29
1.6.4. SVM linear multiclasse	30
CAPÍTULO 2	32
2. Doença de Parkinson	33
2.1. Conceito	33
2.2. Testes de Posturografia Dinâmica Computorizada	33
2.2.1. Teste de marcha	34
2.2.2. Limites de Estabilidade	34
CAPÍTULO 3	36
3. Metodologia	37
3.1. Amostra e descrição dos dados	37
3.2. Objetivos do trabalho	37
3.3. Metodologia proposta.....	38
3.4. Problema do estudo	38
3.5. Instrumentos	38
3.6. Procedimentos	39
CAPÍTULO 4	41
4. Resultados	42
4.1. Análise Descritiva dos dados	43
4.1.1. Descrição dos dados de marcha.....	44
4.1.1.1. Distribuição dos doentes por Sexo.....	44
4.1.1.2. Distribuição e caracterização de Altura e Idade dos doentes por estadio..	44

4.1.1.3. ANOVA simples para a diferença de médias nas Alturas e Idades por estádios.....	45
4.1.2. Descrição dos dados visuais	45
4.1.2.1. Distribuição dos doentes por Sexo.....	46
4.1.2.2. Distribuição e caracterização de Altura e Idade dos doentes por estádio..	46
4.1.2.3. ANOVA simples para a diferença de médias nas Alturas e Idades por estádios.....	47
4.2. Classificação com as features de marcha	47
4.2.1. Caso Multiclasse.....	49
4.2.1.1. Kernel Linear	50
4.2.1.2. Kernel RBF	51
4.2.1.3. Equilíbrio das dimensões dos dados por classes.....	52
4.2.1.3.1. Kernel Linear.....	52
4.2.1.3.2. Kernel RBF	52
4.2.1.3.3. Caso binário.....	53
4.2.1.3.3.1. Kernel Linear.....	53
4.2.1.3.3.2. Kernel RBF.....	54
4.3. Classificação com as features Distância.....	55
4.3.1. Caso Multiclasse.....	55
4.3.1.1. Kernel Linear	55
4.3.1.2. Kernel RBF	56
4.3.1.3. Equilíbrio da dimensão dos dados por classes.....	57
4.3.1.3.1. Kernel Linear.....	57
4.3.1.3.2. Kernel RBF	57
4.3.1.3.3. Caso binário.....	58
4.3.1.3.3.1. Kernel Linear.....	58
4.3.1.3.3.2. Kernel RBF.....	59
4.4. Classificação com as features Amplitude.....	59
4.4.1. Caso Multiclasse.....	60
4.4.1.1. Kernel Linear	60
4.4.1.2. Kernel RBF	61
4.4.1.3. Equilíbrio das dimensões dos dados por classes.....	62
4.4.1.3.1. Kernel Linear.....	62
4.4.1.3.2. Kernel RBF	62
4.4.1.3.3. Caso binário.....	63
4.4.1.3.3.1. Kernel Linear.....	63
4.4.1.3.3.2. Kernel RBF.....	63
4.5. Classificação com as features com maior diferença de médias entre as classes ...	64
4.5.1. Caso Multiclasse com as features Distância.....	64
4.5.1.1. Kernel Linear	64
4.5.1.2. Kernel RBF	65
4.5.1.3. Equilíbrio das dimensões dos dados por classes.....	66
4.5.1.3.2. Kernel RBF	66
4.5.1.3.3. Caso binário.....	67
4.5.1.3.3.1. Kernel Linear.....	67
4.5.1.3.3.2. Kernel RBF.....	68
4.5.2. Classificação com as features Velocidade.....	68
4.5.2.1. Caso binário para as dimensões equilibradas por classes	69

4.5.2.1.1. Kernel Linear.....	69
4.5.2.1.2. Kernel RBF	70
4.6. Classificação com a exclusão das features correlacionadas entre si	70
4.6.1. Classificação com as features de marcha	70
4.6.1.1. Caso binário para as dimensões equilibradas por classes	71
4.6.1.1.1. Kernel Linear.....	71
4.6.1.1.2. Kernel RBF	72
4.6.2. Classificação com as features Amplitude.....	72
4.6.2.1. Caso binário para as dimensões equilibradas por classes	73
4.6.2.1.1. Kernel Linear.....	73
4.6.2.1.2. Kernel RBF	74
4.7. Classificação com a exclusão dos outliers em cada caso	75
4.7.1. Caso multiclasse para as features Distância com a exclusão dos outliers	75
4.7.1.1. Kernel Linear	78
4.7.1.2. Kernel RBF	78
4.7.1.3. Equilíbrio das dimensões dos dados por classes.....	79
4.7.1.3.1. Kernel Linear.....	79
4.7.1.3.2. Kernel RBF	80
4.7.1.2.3. Caso binário para as dimensões equilibradas por classes	81
4.7.1.2.3.1. Kernel Linear	81
4.7.1.2.3.2. Kernel RBF.....	81
4.8. Classificação de Discinesias do membro superior	82
4.8.1. Classificação com as features Distância.....	82
4.8.1.1. Kernel Linear	82
4.8.1.2. Kernel RBF.....	83
4.8.1.3. Equilíbrio das dimensões dos dados por classes.....	83
4.8.1.3.1. Kernel Linear.....	84
4.8.1.3.2. Kernel RBF	84
CAPÍTULO 5	85
5. Conclusões e Perspetivas futuras	86
Referências	89
Anexos.....	93
Escala de Hoehn & Yhar	94
Sistema Balance Master®.....	95
Outputs.....	96
Teste de K-S para a normalidade das amostras das de marcha	96
Teste de ANOVA para igualdade de médias de altura por estádio para as features de marcha	96
Teste de ANOVA para igualdade de médias de idade por estádio para as features de marcha	96
Teste de K-S de uma amostra para a normalidade das features Distância.....	97
Teste de K-S de uma amostra para a normalidade das features Amplitude.....	97
Teste de K-S de uma amostra para a normalidade das features Área	98
Teste de K-S de uma amostra para a normalidade das features Velocidade.....	98
Teste de K-S de uma amostra para a normalidade das features Frequência	99
Teste de ANOVA para igualdade de médias de altura por estádio para as features posturais	99

Teste de ANOVA para igualdade de médias de idade por estadio para as fetures posturais	99
Tabela de ANOVA para as features Distância.....	100
Tabela de ANOVA para as features Amplitude.....	101
Tabela de ANOVA para as features Velocidade.....	102
Tabela de ANOVA as features Área	104
Tabela de ANOVA para as features Frequência	105
Estaística de Resíduos das variáveis Distância	106
Gráficos de dispersão das features Distância	109
Boxplot para as features Amplitude	110
Boxplot para as features Área	111
Boxplot para as features Velocidade	112
Boxplot para as features Frequência	114
Função geral para a validação cruzada do modelo SVM para a escolha dos parâmetros	

Índice de Figuras

Figura 1.1 - Maximização das margens do hiperplano e o respetivo cálculo da distância d entre $H1$ e $H2$	17
Figura 1.2 - Maximização das margens do hiperplano com a introdução das variáveis de folga para o caso linear não separável.	23
Figura 2.1 - Representação da imagem motora característica do doente com DP	33
Figura 2.2 - Realização do teste WA	34
Figura 2.3 - Realização do teste LOS	35
Figura 4.1 - Visualização gráfica das features de marcha	49
Figura 4.2 - Visualização gráfica dos outliers das features Distância	76

Índice de Tabelas

Tabela 4.1 - Resumo dos melhores resultados obtidos.....	43
Tabela 4.2 - Distribuição por Sexo.....	44
Tabela 4.3 – Distribuição e caracterização dos doentes por estadios.....	45
Tabela 4.4 - Distribuição por Sexo.....	46
Tabela 4.5 – Distribuição e caracterização dos doentes por estadios.....	47
Tabela 4.6 - Estatísticas de grupo para as features de marcha, por classe.....	48
Tabela 4.7 - Matriz de confusão global para o caso multiclasse com features de marcha e kernel Linear.....	50
Tabela 4.8 - Matriz de confusão global para o caso multiclasse com features de marcha e Kernel RBF.....	51
Tabela 4.9 - Matriz de confusão global para o caso multiclasse com features de marcha e Kernel Linear para as dimensões equilibradas dos dados por classes.	52
Tabela 4.10 - Matriz de confusão global para o caso multiclasse com features de marcha e Kernel RBF para as dimensões equilibradas dos dados por classes.....	53
Tabela 4.11 - Matriz de confusão global para o caso binário com features de marcha e Kernel Linear para as dimensões equilibradas dos dados por classes.	54
Tabela 4.12 - Matriz de confusão global para o caso binário com features de marcha e Kernel RBF para as dimensões equilibradas dos dados por classes.....	54
Tabela 4.13 - Estatísticas de grupo para as features Distância, por classe.....	55
Tabela 4.14 - Matriz de confusão global para o caso multiclasse com as features Distância e Kernel Linear.....	56
Tabela 4.15 - Matriz de confusão global para o caso multiclasse com as features Distância e Kernel RBF.....	56
Tabela 4.16 - Matriz de confusão global para o caso multiclasse com as features Distância e Kernel Linear para as dimensões equilibradas dos dados por classes.	57
Tabela 4.17 - Matriz de confusão global para o caso multiclasse com as features Distância e Kernel RBF para as dimensões equilibradas dos dados por classes.....	58
Tabela 4.18 - Matriz de confusão global para o caso binário com as features Distâncias e Kernel Linear para as dimensões equilibradas dos dados por classes.	58
Tabela 4.19 - Matriz de confusão global para o caso binário com as features Distâncias e Kernel RBF para as dimensões equilibradas dos dados por classes.....	59
Tabela 4.20 - Estatísticas de grupo para as features Amplitude, por classes.	60
Tabela 4.21 - Matriz de confusão global para o caso multiclasse com as features Amplitude e Kernel Linear.	61
Tabela 4.22 - Matriz de confusão global para o caso multiclasse com as features Amplitude e Kernel RBF.....	61
Tabela 4.23 - Matriz de confusão global para o caso multiclasse com as features Amplitude e Kernel Linear para as dimensões equilibradas dos dados por classes.	62
Tabela 4.24 - Matriz de confusão global para o caso multiclasse com as features Amplitude e Kernel RBF para as dimensões equilibradas dos dados por classes.	63
Tabela 4.25 - Matriz de confusão global para o caso binário com as features Amplitude e Kernel Linear para as dimensões equilibradas dos dados por classes.	63
Tabela 4.26 - Matriz de confusão global para o caso binário com as features Amplitude e Kernel RBF para as dimensões equilibradas dos dados por classes.....	64

Tabela 4.27 - Matriz de confusão global para o caso multiclasse com as features Distância com maior diferença de médias entre as classes e Kernel Linear.	65
Tabela 4.28 - Matriz de confusão global para o caso multiclasse com as features Distância com maior diferença de médias entre as classes e Kernel RBF.	65
Tabela 4.29 - Matriz de confusão global para o caso multiclasse com as features Distância com maior diferença de médias entre as classes e Kernel Linear para as dimensões equilibradas dos dados por classes.	66
Tabela 4.30 - Matriz de confusão global para o caso multiclasse com as features Distância com maior diferença de médias entre as classes e Kernel RBF para as dimensões equilibradas dos dados por classes.	67
Tabela 4.31 - Matriz de confusão global para o caso binário com as features Distância com maior diferença de médias entre as classes e Kernel Linear para as dimensões equilibradas dos dados por classes.	67
Tabela 4.32 - Matriz de confusão global para o caso binário com as features Distância com maior diferença de médias entre as classes e Kernel RBF para as dimensões equilibradas dos dados por classes.	68
Tabela 4.33 - Estatísticas de grupo para as features Velocidade, por classe.	69
Tabela 4.34 - Matriz de confusão global para o caso binário com as features Velocidade com maior diferença de médias entre as classes e Kernel Linear para as dimensões equilibradas dos dados por classes.	69
Tabela 4.35 - Matriz de confusão global para o caso binário com as features Velocidade com maior diferença de médias entre as classes e Kernel RBF para as dimensões equilibradas dos dados por classes.	70
Tabela 4.36 - Matriz de correlação entre as features de Marcha.	71
Tabela 4.37 - Matriz de confusão global para o caso binário com as features de marcha com exclusão das features correlacionadas entre si e Kernel Linear para as dimensões equilibradas dos dados por classes.	71
Tabela 4.38 - Matriz de confusão global para o caso binário com as features de marcha com exclusão das features correlacionadas entre si e Kernel RBF para as dimensões equilibradas dos dados por classes.	72
Tabela 4.39 - Matriz de correlação entre as features Amplitude.	73
Tabela 4.40 - Matriz de confusão global para o caso binário com as features Amplitude com exclusão das features correlacionadas entre si e Kernel Linear para as dimensões equilibradas dos dados por classes.	74
Tabela 4.41 - Matriz de confusão global para o caso binário com as features Amplitude com exclusão das features correlacionadas entre si e Kernel RBF para as dimensões equilibradas dos dados por classes.	74
Tabela 4.42 - Matriz de confusão global para o caso multiclasse com as features Distância com a exclusão dos outliers e Kernel Linear.	78
Tabela 4.43 - Matriz de confusão global para o caso multiclasse com as features Distância com a exclusão dos outliers e Kernel RBF.	79
Tabela 4.44 - Matriz de confusão global para o caso multiclasse com as features Distância com a exclusão dos outliers e Kernel Linear para as dimensões equilibradas dos dados por classes.	80
Tabela 4.45 - Matriz de confusão global para o caso multiclasse com as features Distância com a exclusão dos outliers e Kernel RBF para as dimensões equilibradas dos dados por classes.	80

Tabela 4.46 - Matriz de confusão global para o caso binário com as features Distância com a exclusão dos outliers e Kernel Liner para as dimensões equilibradas dos dados por classes.	81
Tabela 4.47 - Matriz de confusão global para o caso binário com as features Distância com a exclusão dos outliers e Kernel RBF para as dimensões dos dados equilibrados por classes.	82
Tabela 4.48 - Matriz de confusão global para o caso binário de discinesias do membro superior com as features Distância e Kernel Linear.	83
Tabela 4.49 - Matriz de confusão global para o caso binário de discinesias do membro superior com as features Distância e kernel RBF.	83
Tabela 4.50 - Matriz de confusão global para o caso binário de discinesias do membro superior com as features Distância e Kernel Linear para as dimensões equilibradas dos dados por classes.	84
Tabela 4.51 - Matriz de confusão global para o caso binário de discinesias do membro superior com as features Distância e Kernel RBF para as dimensões equilibradas dos dados por classes.	84

Lista de Abreviaturas

- AA:** Aprendizagem Automática
- Ampl_AP:** Amplitude ântero-posterior;
- Ampl_ML:** Amplitude médio-lateral;
- ANOVA:** *Analysis of variance* (Análise de Variância)
- CG** – Centro de Gravidade;
- Dist_T:** Distância total;
- dp:** Desvio padrão
- DP:** Doença de Parkinson
- H&Y:** *Hoehn e Yahr*
- KKT:** *Karush-Kuhn-Tucker*
- K – S:** Kolmogorov – Smirnov
- LOS** – *Limits of Stability* (Limites de Estabilidade)
- WA** – *walk across* (Teste de Marcha)
- Mda:** Média
- PDC** – Posturografia Dinâmica Computorizada
- RBF:** *Radial Basis Function* (Função Base Radial)
- RNA:** Redes Neurais Artificiais
- SVM:** *Support Vector Machines* (Máquinas de Vetor Suporte)
- TEA:** Teoria Estatística de Aprendizagem
- TOM:** Teoria de Otimização Matemática
- Vel_M_AP:** Velocidade média ântero-posterior;
- Vel_M_ML:** Velocidade média médio-lateral;
- Vel_M_T:** Velocidade média total;
- WA – SIMETRIA:** *Symmetry stride* (Simetria da passada)
- WA – SL:** *Sride length* (Comprimento da passada)
- WA – SPEED:** *Average speed of stride* (Velocidade média da passada)
- WA – WS:** *Stride width* (Largura da passada)

Definições Operacionais

Features – termo atribuído as entradas características de um atributo tendo os componentes como catálogo um vetor de rótulos. Usualmente cada *feature* é uma coluna e o conjunto de *features* é uma matriz em que nas colunas dessa matriz são representados os valores de cada uma destas. (SILVA, 2014).

Pesquisa em grelha (*grid search*) – é a aplicação do método para vários pares de parâmetros do modelo (no caso de SVM com kernel RBF são os parâmetros C e γ), por forma a escolher o par que fornece melhores resultados (de precisão). Sendo que de uma forma geral os valores a considerar para os parâmetros C e γ são valores fixos, então o conjunto de pares é uma grelha no plano. Daí advém a designação de *grid search* e o que se pretende é procurar o ponto da grelha que fornece melhor resultado (HSU, CHANG, & LIN, 2010).

Validação cruzada (*cross validation*) – consiste em dividir a amostra disponível em k grupos e, à vez, retirar cada um desses grupos como conjunto de teste, considerar a restante amostra como grupo de treino para o modelo, e depois, testar o modelo no conjunto de teste. O procedimento é repetido, deixando cada um dos k grupos de fora do treino, como grupo de teste. Trata-se assim de validação (o grupo de treino e teste são disjuntos) cruzada (os grupos de teste vão sendo alterados) (HSU, CHANG, & LIN, 2010).

Introdução

Nas últimas décadas, a Aprendizagem Supervisionada tornou-se um dos pilares primordiais na tecnologia de informação. Dado o crescimento constante de dados, acreditou-se que com uma análise inteligente desses dados constituir-se-ia um ponto crucial para o progresso tecnológico. Nesta base, a aplicação das Técnicas de Aprendizagem Supervisionada tem tido um papel importante na extração de grande quantidade de informação armazenada e disponibilizada em documentos textuais, informações estas que o ser humano, por si só, não consegue obter e analisar em tempo útil (SMOLA & VISHWANATHAN, 2008; PELLUCCI, DE PAULA, SILVA, & LADEIRA, 2011).

Este trabalho emerge da realização de um estudo de avaliação clínica, onde se pretendia analisar a evolução da Doença de Parkinson (DP) para a função motora em diferentes estadios, de 1 a 4, isto é, do menos grave ao mais grave, através de técnicas de aprendizagem supervisionada, com base nos dados clínicos recolhidos para dois grupos de sujeitos diagnosticados com esta doença. No primeiro grupo, os dados foram recolhidos durante uma tarefa em que estes têm de andar ou marchar sobre uma passadeira e onde foram recolhidas as seguintes características: Largura da passada, Comprimento da passada, Velocidade da passada e Simetria da passada. No segundo grupo, os dados foram recolhidos durante uma tarefa em que os sujeitos tinham de acompanhar visualmente um ponto móvel num ecrã e dos quais foram recolhidas as seguintes características: Distância de oscilação, Amplitude de oscilação, Área de oscilação, Velocidade de oscilação e Frequência de oscilação.

Com base neste estudo e com os respetivos dados de comportamento motor referidos no parágrafo antecedente, procuramos conhecer quais as variáveis recolhidas que mais contribuem para a classificação dos sujeitos entre grupos, com recurso aos Métodos de Aprendizagem Supervisionada, no caso Máquinas de Vetor Suporte (SVM - *Support Vector Machine*).

Estes métodos estabelecem um modelo de classificação de objetos/variáveis em grupos de interesse, dado um chamado conjunto de treino (amostra de teste) e a respetiva classificação por grupos desse conjunto de treino. Assim, obtém-se um modelo com a capacidade de prever qual a classificação de um novo elemento cuja classificação é desconhecida. A importância destes modelos de classificação é grande em duas vertentes.

Por um lado e como já referido, fornece um modelo de classificação para novos elementos. Por outro, ao estabelecer o modelo, indica quais os atributos (ou conjunto deles) com maior capacidade discriminatória entre grupos, permitindo assim estabelecer a sua importância para os diferenciar.

Dos referidos métodos de aprendizagem, as Máquinas de Vetores Suporte têm uma atenção especial pelos bons resultados apresentados e pela sua fácil aplicação. Os resultados no reconhecimento de padrões em tarefas de classificação são, na maioria dos casos, superiores a outros obtidos por outros métodos, como é o caso das redes neurais artificiais (RNA) (HAYKIN, 1999; LORENA & CARVALHO, 2003). As SVM são algoritmos de aprendizagem supervisionada, tendo sido apresentadas na década de 90, como um algoritmo de treino cujo objetivo era maximizar, segundo padrões de treino, as margens entre classes distintas (BOSER, GUYON, & VAPNIK, 1992; SILVA, 2014). O conceito de aprendizagem supervisionada requer a intervenção do investigador a uma dada altura do processo. Este conceito é um campo da Teoria Estatística de Aprendizagem (TEA), que por sua vez é a base suporte das SVM. Esta teoria tem como ponto focal, minimizar ao máximo o risco empírico por forma a conseguir o melhor classificador possível, satisfazendo as restrições impostas. Posteriormente em 1995 as SVM foram alargadas para outros contextos de dados não separáveis e margens suaves, com a sua aplicação real no processamento e tratamento de dados, uma vez que, em casos reais, geralmente os dados não são separáveis linearmente (CORTES & VAPNIK, 1995; SILVA, 2014).

O objetivo deste trabalho é determinar um modelo de classificação multiclasse através de SVM entre grupos de sujeitos, com base nos dados de comportamento motor de sujeitos de diferentes grupos de interesse.

Desta forma propõe-se aplicar um método de aprendizagem supervisionada, neste caso as SVM, onde serão feitos estudos de validação cruzada para determinar os parâmetros do modelo escolhido que garantam a maior precisão possível. Para tal, começamos por definir atributos numéricos com base num conjunto de dados que foram recolhidos durante tarefas realizadas pelos sujeitos, de andar sobre uma passadeira e de acompanhar visualmente um ponto móvel num ecrã, conforme referido anteriormente. Assim será definido um modelo apropriado de classificação baseado em SVM entre grupos. Com base neste modelo, será possível determinar que *features* serão mais importantes para a classificação.

Após esta introdução, segue a apresentação do enquadramento teórico pelo qual esta dissertação é suportada. Esta dissertação encontra-se dividida em cinco capítulos chave: Revisão da literatura sobre o contexto das SVM, a Doença de Parkinson (DP), Metodologia, Resultados e Conclusões.

No primeiro capítulo, começamos por introduzir dois conceitos fundamentais para uma melhor compreensão das SVM: A TOM e a TEA. Esta última é introduzida seguida de uma breve retrospectiva histórica, com a sua passagem pelo conceito de aprendizagem supervisionada até às SVM. A explanação matemática encontra-se organizada em dois grandes grupos: SVM para dados separáveis linearmente e o seu alargamento para dados não separáveis.

No segundo capítulo é apresentado uma breve abordagem sobre a DP começando pelo seu conceito, as características motoras, passando de seguida para os testes de posturografia dinâmica computadorizada onde serão apresentados dois testes, sendo uma relacionada à marcha e a outra relacionada à postura.

No terceiro capítulo é apresentada a metodologia usada para aplicação prática das SVM, começando pela mostra e descrição dos dados, passando de seguida para os objetivos do trabalho, metodologia proposta, o problema do estudo, os instrumentos utilizados para caracterizar o nível de severidade da DP e para avaliar quantitativamente o comportamento postural e de marcha. Ainda neste capítulo são apresentados os procedimentos utilizados para a recolha dos dados para aplicação do *cross validation* no software *Octave*[®] 4.0.0 (*GUI*), para localizar os erros de classificação e a precisão de classificação global dada pelo modelo e as diversas tentativas utilizadas para a concretização dos objetivos propostos.

No quarto capítulo são apresentados os resultados do estudo, onde começamos por fazer a descrição dos dados, passando de seguida para a apresentação dos resultados obtidos relativamente aos casos usados: caso linear onde é otimizado o parâmetro custo C para a separação não linear; e Kernel RBF onde são otimizados dois parâmetros, o C e o γ , com base na pesquisa de grelha (*grid search*), para a seleção das features que mais contribuem para a classificação.

No quinto e último capítulo terminamos com as conclusões resultantes deste estudo e as perspetivas futuras para estudos sobre esta temática.



CAPÍTULO 1

REVISÃO DA LITERATURA

1. Revisão da literatura

1.1. O contexto das Máquinas de Vetores Suporte

Para introduzir o tema de Máquinas de Vetores Suporte (SVM, em inglês: *Support Vector Machines*), deve ter-se em conta duas teorias: a Teoria Estatística de Aprendizagem (TEA) e a Teoria de Otimização Matemática (TOM), pois essas são as bases para uma melhor compreensão da técnica de SVM.

1.2. Teoria de Otimização Matemática

A teoria de otimização matemática (TOM) é um ramo da matemática que visa minimizar ou maximizar uma determinada função custo, com o objetivo de encontrar uma melhor solução, ou seja, uma solução que seja ótima para um problema em estudo, sendo esta solução sujeita a certas restrições. Ao problema de otimização está geralmente associado um conjunto de variáveis independentes. Essas variáveis assumem valores em função das restrições que lhes são impostas. Se a partir dessas variáveis se conseguir obter um conjunto de valores que satisfaça as condições de restrições por forma a minimizar ou maximizar a função custo, diz-se que esse conjunto de valores é solução do problema de otimização. A otimização pode ser classificada da seguinte forma (SANTOS, 2002):

- **Otimização Irrestrita** – aqui as variáveis que a otimização engloba podem assumir qualquer valor, ou seja, não existem restrições;
- **Otimização Restrita** – ao contrário da anterior, aqui as variáveis estão sujeitas a algumas condições ou restrições e essas restrições podem ser de igualdade ou desigualdade. As restrições de igualdade, em determinada condição do sistema, faz a descrição de um conjunto de variáveis a serem consideradas, procurando encontrar uma solução para o problema em estudo. As condições de desigualdade, o seu objetivo é impor limites nas variáveis, ou seja, requerer que estas tenham limite inferior e/ou superior.

1.3. Teoria Estatística de Aprendizagem

A teoria estatística de aprendizagem (TEA) é um processo que permite gerar um classificador de um conjunto de classificadores possíveis, através do algoritmo de Aprendizagem Automática (AA) com base num conjunto de treino T . Seja p um classificador e P o conjunto de todos os classificadores que podem ser gerados pelo algoritmo de AA. Então o algoritmo de AA, vai gerar o classificador estimado \hat{p} com base no conjunto de treino T , classificador esse que deve pertencer ao conjunto de todos os classificadores P . Por sua vez, o conjunto de treino T deve ser composto por pares (x_i, y_i) em que x_i são os objetos de treino e y_i são os respetivos rótulos (ou classes). Rótulos ou classes são conjuntos de objetos que representam o fenómeno de interesse sobre o qual se deseja fazer previsões (LORENA & CARVALHO, 2007). Este processo de TEA tem como base tanto a classificação quanto a discriminação de padrões e funções discriminantes a que está associado (SILVA, 2014).

Ronald Aylmer Fisher deu um grande impulso na compreensão deste processo quando desenvolveu o trabalho intitulado “*The use of multiple measurements in taxonomic problems*” (O uso de medições múltiplas em problemas taxonómicos) onde o objetivo era classificar espécies de plantas (Iris setosa e Iris versicolor). Para tal classificação, Fisher incidiu o seu estudo nas pétalas e sépalas onde analisara o comprimento e largura destas em quatro medidas, através de procedimentos desenvolvidos por ele (FISHER, 1936). Generalizando o caso de quatro medidas para um universo de n medidas, é preciso conhecer uma função que permita maximizar a diferença de relação entre a média específica e os desvios padrão dentro das espécies. Assim, a função que se pretende conhecer para essas medidas, é uma função linear conforme se segue:

$$X = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

sendo a diferença das médias de X entre as duas espécies dada pela seguinte expressão:

$$D = \beta_1 d_1 + \beta_2 d_2 + \dots + \beta_n d_n$$

onde d_j é a diferença das médias das j – ésima medidas entre as duas espécies, e a variância de X dentro das espécies é proporcional a

$$S = \sum_{p=1}^n \sum_{q=1}^n \beta_p \beta_q S_{pq}$$

em que as n medidas são correspondidas uma a uma por p e q respetivamente, e a cada uma dessas medidas, o produto dos desvios padrão é representado por S_{pq} que também pode ser chamado de somatório dos quadrados entre as duas espécies. Ao construir a matriz S , os valores p e q são denotados por S_{pq} de 1 a n , nessa matriz. Conforme referido acima, a TEA está associada também às funções discriminantes. Nesse contexto é necessário que haja uma função discriminatória que maximize o rácio $\frac{D^2}{S}$. Esse rácio será maximizado em função dos parâmetros $\beta_1, \beta_2, \dots, \beta_n$ de forma independente para cada β conforme se segue (FISHER, 1936):

$$\frac{D}{S^2} \left\{ 2S \frac{\partial D}{\partial \beta} - D \frac{\partial S}{\partial \beta} \right\} = 0.$$

Uma vez maximizado o rácio, pode-se dizer então que a função de discriminação utilizada para o efeito é a melhor. A probabilidade dos desvios e os erros obtidos nos resultados finais deve ser quantificada. Isto faz-se com base nos resultados das diferenças das médias totais e o somatório dos quadrados obtidos através do rácio (BURGES, 1998; SANTOS, 2002).

Para além do grande contributo de Fisher, também um outro trabalho intitulado “*Theory of Reproducing kernels*” de Aronszajn foi desenvolvido para complementar e solidificar a compreensão da TEA. Este trabalho também foi de extrema importância para as SVM na classificação não-linear. Teve como base, um espaço linear F complexo (admite a multiplicação por constantes complexos) de funções $f(x)$ definidas num conjunto E onde E é um conjunto de classificadores (ARONSZAJN, 1950).

Há ainda mais dois termos que estão associados a este processo com o desenvolvimento e a evolução do conceito da TEA: a inferência e a extração/exploração de dados (em inglês: data mining) (SILVA, 2014).

A extração de dados aqui, pode ser definida tanto de uma perspetiva estatística como de uma perspetiva de aprendizagem automática (SÉRGIO, 2009; CAMILO & SILVA, 2009):

- De uma perspectiva estatística, a extração de dados consiste em tornar úteis e compreensíveis grandes conjuntos de dados através de um resumo destes;
- De uma perspectiva da aprendizagem automática, a extração de dados consiste em fazer uma análise destes com o propósito de descobrir novos padrões, aplicando algoritmos que auxiliam nessa descoberta. Apesar desses algoritmos terem limitações computacionais, o seu objetivo é extrair um conjunto de padrões presentes nos dados.

Mais contribuições para o desenvolvimento da TEA foram aparecendo, e no ano de 1965 surge o Teorema de Cover, onde a essência do seu estudo focaliza a separabilidade de grupos de dados de treino através de um hiperplano separador, com base num determinado conjunto de dados de treino num espaço euclidiano (COVER, 1965). Todas essas contribuições foram muito valiosas para a expansão da TEA, mas atingiu um dos seus picos de popularidade na década de 90 com a introdução das SVM, mais precisamente em 1992 para dados separáveis linearmente, alargando posteriormente para outros contextos de dados não separáveis e margens suaves em 1995 (BOSER, GUYON, & VAPNIK, 1992; CORTES & VAPNIK, 1995; SILVA, 2014).

1.4. Considerações sobre o Problema de Aprendizagem

O mapeamento de uma determinada amostra que é explicada pela função de aproximação, está relacionado com o conceito do problema de aprendizagem. Assim segue a definição do problema de aprendizagem:

O Problema de Aprendizagem consiste em encontrar entre os objetos $x \in \mathcal{X}$ e os rótulos $y \in \mathcal{Y}$ com $f: \mathbb{R}^n \rightarrow \{-1, 1\}$, uma relação funcional desconhecida com base numa única amostra $z = (x, y) = ((x_i, y_i), \dots, (x_n, y_n)) \in (\mathcal{X}, \mathcal{Y})^n$ de dimensão $n \in \mathbb{N}$ de dados independente e identicamente distribuídos, considerando uma distribuição $P(x, y)$ desconhecida. Está-se perante um problema de aprendizagem de classificação, se um número finito $|y|$ de elementos está contido no espaço de saída \mathcal{Y} (HERBRICH, 2002; SILVA, 2014).

Tendo em conta que a semelhança entre objetos deve ser modelada, isto é, com base numa função do produto interno, uma outra definição que deve estar presente é as entradas características (em inglês: *features*) e o seu espaço de entradas conforme se segue:

cada componente dos objetos $x \in \mathcal{X}$ mapeado com base numa função $f: \mathbb{R}^n \rightarrow \{-1,1\}$ que representa um classificador, é chamada entrada característica (*features*) a cada x . A representação do espaço euclidiano por $\mathcal{X} \subseteq \mathbb{R}^n$ deve-se ao espaço característico de dimensionalidade n construído por vários *features*. O mapeamento $\Phi: \mathcal{X} \rightarrow \mathcal{H}$ (*Kernel Trick*) num espaço dimensional superior resulta de n combinações, onde \mathcal{H} é denominado como espaço característico (HERBRICH, 2002; SILVA, 2014).

A descrição do modelo de aprendizagem na TEA baseia-se em três componentes principais conforme se segue (VAPNIK, 1999):

1. Com base numa probabilidade cumulativa $P(x)$, fazer com que os vetores $x \in \mathbb{R}^n$ sejam executadas de forma cuidadosa e independentes, com recurso a um Gerador, sendo essa probabilidade fixa e desconhecida;
2. Levando em consideração uma função distribuição condicional $P(y|x)$, é devolvido para cada vetor de entrada x , um vetor de saída y com recurso a um supervisor, tendo essa probabilidade as mesmas condições que a antecedente;
3. Dado um conjunto de funções $y = f(x, \alpha) \in \Lambda$, estas devem ser executadas através de uma máquina de aprendizagem com a capacidade para tal execução, onde y representa a classe de classificação, f representa a função classificadora e α os parâmetros da função que se pretende conhecer, x é o elemento no espaço dados, com Λ a representar o conjunto dos parâmetros possíveis.

A melhor resposta advém da boa escolha de uma função que é capaz de a prever, onde esta função é escolhida a partir de um conjunto de funções $y = f(x, \alpha) \in \Lambda$. Esta escolha ou seleção tem como base, um conjunto de treino de n observações aleatórias, independentes e identicamente distribuídas executada com recurso a $P(x, y) = P(x)P(y|x)$, onde o par (x, y) é uma amostra que não depende da distribuição P (SILVA, 2014; VAPNIK, 1999).

1.5. Aprendizagem Automática Supervisionada

No processo de aprendizagem supervisionada deve levar-se em consideração a relação ambiente e adaptação, ou seja, o ambiente deve ser conhecido com base num conjunto de exemplos que permitam, mediante a entrada destes exemplos, obter saídas desejadas, tendo a máquina uma sequência de instruções a percorrer para o efeito de transmissão desses exemplos. Quanto a adaptação, todos os exemplos considerados devem ser avaliados durante o processo (SILVA, 2014; LORENA & CARVALHO, 2007). Sendo esta aprendizagem considerada como um ramo da inteligência artificial que visa desenvolver algoritmos capazes de fazer com que as máquinas reconheçam padrões ou tarefas, sendo capazes de as replicar em testes futuros, deve-se ter em conta duas metodologias de aprendizagem no contexto da classificação de dados: aprendizagem não supervisionada e aprendizagem supervisionada (BURBIDGE & BUXTON, s.d.; LORENA & CARVALHO, 2003; SILVA, 2014).

Na **aprendizagem não supervisionada** não se consegue prever quantas classes é que o algoritmo não supervisionado de classificação irá produzir após ser utilizado. Os passos a serem seguidos pelo algoritmo na tomada de decisão são independentes, ou seja, não dependem do apoio do investigador. Exemplos de aprendizagem não supervisionada são classificação por clusters, análise fatorial, entre outros (BURBIDGE & BUXTON, s.d.; LORENA & CARVALHO, 2003; SILVA, 2014).

Na **aprendizagem supervisionada** ao contrário da anterior, há necessidade de intervenção do investigador. A aprendizagem advém de um processo de treino, baseado num conjunto de exemplos de entrada-saída, sendo esses exemplos provenientes de um ambiente qualquer. Neste caso, o ajustamento dos parâmetros para o classificador a construir vai depender da resposta que se pretende obter e do erro associado. Dentro da aprendizagem supervisionada destacam-se: a aprendizagem por reforço e aprendizagem corretiva. A aprendizagem por reforço dentro da aprendizagem supervisionada, independentemente do resultado obtido primeiro utiliza-se os exemplos de treino e só depois é utilizada a referida aprendizagem. Já na aprendizagem corretiva dentro da supervisionada, é considerado o erro em conjunto com o vetor de entrada, sendo a correção baseada neles (BURBIDGE & BUXTON, s.d.; LORENA & CARVALHO, 2003; SILVA, 2014).

O risco esperado $R(\alpha)$ que também é chamado de risco previsto, é dado pela expressão seguinte (BURGES, 1998):

$$R(\alpha) = \int \frac{1}{2} |y - f(x, \alpha)| dP(x, y) \quad (1)$$

sendo y a classe de classificação, f a função classificadora, x o elemento no espaço dados e α os parâmetros da função que se pretende conhecer (SILVA, 2014; VAPNIK, 1999). As entradas x e a classe y pertencem a dois espaços diferentes, isto é, \mathcal{X} e \mathcal{Y} respetivamente, onde y assume apenas dois valores possíveis, ou seja, $y \in \{-1, 1\}$. Neste caso pode-se dizer que se está perante um problema de classificação binária. A função classificadora que se pretende determinar $f(x, \alpha)$ é a procurada pela probabilidade condicional $P(x|y)$ com base numa descrição da relação existente entre os dados e as classes para efeito da sua procura, isto é, da função que se pretende conhecer (SILVA, 2014). Na expressão (1) se depara com um problema que impede com que esta seja calculada. Esse problema deve-se ao desconhecimento da probabilidade. Neste caso, uma solução plausível que geralmente se costuma utilizar para minimizar o risco esperado é o princípio indutivo para minimização do risco empírico, que visa deduzir uma determinada função $f(x, \alpha)$ que minimize o erro sobre os dados de treino (SILVA, 2014; VAPNIK, 1999). Focalizando no objetivo que é de minimizar o risco esperado, começa-se por minimizar o risco empírico $R_{emp}(\alpha)$. A expressão do risco empírico para o caso binário que se segue, é calculada com base num conjunto de treino com l observações aleatórias, independente e identicamente distribuídas:

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

onde l representa os exemplos de treino, y_i representa as classes, f representa o classificador, x_i os atributos e α os parâmetros da função de classificação f a determinar. Esse risco vai medir a taxa do erro médio de um conjunto de dados de treino de dimensão l para o efeito do desempenho do classificador. Volta-se a frisar que a distribuição de probabilidade condicional não é conhecida. De uma maneira particular, fixa-se $R_{emp}(\alpha)$ para escolher α com base no conjunto de treino escolhido $\{x_i, y_i\}$. A minimização do risco esperado tem o

seu custo, e esse custo é dado pela expressão ou quantidade $\frac{1}{2}|y_i - f(x_i, \alpha)|$. Como se trata de um caso binário onde y é rotulado por dois valores na saída, -1 e 1 conforme já se tinha referido anteriormente, a quantidade $\frac{1}{2}|y_i - f(x_i, \alpha)|$ toma esses valores (BURGES, 1998; SILVA, 2014). Tendo em conta que a minimização do risco empírico na realidade nem sempre proporciona um menor risco esperado devido a limitação dos conjuntos de treino, recorre-se então ao princípio indutivo da minimização do risco estrutural que é definido mais à frente. A expressão que se segue mostra que o risco esperado está limitado pela minimização do risco empírico e do risco estrutural. Esse limite é garantido com uma pequena probabilidade de no mínimo $1 - \beta$, onde o parâmetro β é escolhido por forma a satisfazer a condição $0 \leq \beta \leq 1$. Assim, minimizando os dois termos acima mencionados, têm-se a seguinte expressão (BURGES, 1998; LORENA & CARVALHO, 2007; SANTOS, 2002; SILVA, 2014; VAPNIK, 1998):

$$R(\alpha) \leq R_{emp}(\alpha) + \varphi\left(\frac{h}{l}, \frac{\log(\beta)}{l}\right) \quad (3)$$

onde φ é chamado termo de confiança e é definido da seguinte forma:

$$\varphi\left(\frac{h}{l}, \frac{\log(\beta)}{l}\right) = \sqrt{\left(\frac{h\left(\log\left(\frac{2l}{h}\right) + 1\right) - \log\left(\frac{\beta}{4}\right)}{l}\right)} \quad (4)$$

que substituído na expressão (3), traduz-se na seguinte expressão:

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\left(\frac{h\left(\log\left(\frac{2l}{h}\right) + 1\right) - \log\left(\frac{\beta}{4}\right)}{l}\right)} \quad (5)$$

em que h é conhecida como a dimensão de Vapnik Chervonenkis (VC) e é um número inteiro não negativo. A dimensão de VC é uma propriedade de um conjunto de funções $\{f(\alpha)\}$, podendo essa dimensão ser definida por várias funções classificadoras, onde o número máximo de pontos a serem usadas por estas funções classificadoras $\{f(\alpha)\}$, é determinado pela referida dimensão (BURGES, 1998). Posto isto, de seguida é apresentado um teorema e o seu respetivo corolário:

Teorema 1: Considere um conjunto de pontos de treino m tal que $m \subseteq \mathbb{R}^n$. Escolhe-se como origem, qualquer um desses pontos. Então esse conjunto m pode ser separado por hiperplanos orientados, em função das classes dadas, se e somente se, a posição vetorial dos demais pontos forem linearmente independentes.

Corolário 1: A dimensão VC do conjunto de hiperplanos orientados no espaço \mathbb{R}^n é $n + 1$, desde que possamos sempre escolher $n + 1$ pontos, escolhendo de seguida um desses pontos como origem, de forma que a posição vetorial dos n pontos remanescentes sejam linearmente independentes, mas nunca será possível encontrar $n + 2$ de tais pontos, uma vez que os demais $n + 1$ vetores em \mathbb{R}^n não são linearmente independentes.

(BURGES, 1998).

1.6. Máquinas de Vetores Suporte

O reconhecimento de padrões, a regressão e extração de características num determinado conjunto de números de entradas, em grande quantidade, tem custos computacionais elevados para o processo de aprendizagem. Nesta base, surge então as máquinas de vetores suportes (SVM – *Support Vector Machines*) como um poderoso reforço para este processo, com o objetivo de solucionar esses tipos de problemas por forma a reduzir esses custos, onde o seu primeiro alvo foi o caso para problemas de classificação binária, avançando mais tarde para os problemas de regressão (ZHANG, 2001). As SVMs são algoritmos de aprendizagem supervisionada.

Começa-se por introduzir um problema simples para explicar o funcionamento do modelo de SVM para dados separáveis com apenas duas classes. A aplicação deste algoritmo foi apresentada para maximizar a margem entre padrões de treino de classes diferenciadas.

O reconhecimento de padrões foi a prioridade na apresentação das SVMs, tendo como base a minimização do risco esperado através da dimensão VC (BOSER, GUYON, & VAPNIK, 1992; SILVA, 2014). Pode-se dizer que o russo Vapnik foi o pai das SVMs, pois foi ele o primeiro que sugeriu este algoritmo na década de 1960 como um método de aprendizagem de regras de classificação e regressão a partir dos dados de treino (BURBIDGE & BUXTON, s.d.; VAPNIK & LERNER, 1962). Conforme já se tinha referido anteriormente, as SVMs são um caso particular da TEA sendo utilizadas normalmente como técnicas de classificação para o reconhecimento de padrões para dados linearmente separáveis com duas classes. Esta técnica tem como suporte a minimização do risco estrutural. Risco estrutural é um princípio indutivo cujo objetivo é minimizar o erro do conjunto de treino juntamente com o erro do conjunto de teste para que o problema de aprendizagem possa realizar o melhor desempenho de generalização, por combinação entre a capacidade da máquina com a quantidade de dados do problema, tendo como variável de controle a dimensão VC (BURGES, 1998; SANTOS, 2002). Para que este risco (esperado) seja minimizado, deve-se primeiro minimizar o risco estrutural que vai depender de uma escolha adequada através do algoritmo de aprendizagem, minimizar também o risco empírico com recurso a um classificador qualquer que pertence a uma determinada classe de funções cuja dimensão VC seja baixa. Quanto mais pequena for a dimensão VC, menor será o risco empírico e assim pode-se conseguir o melhor classificador possível (SILVA, 2014).

Dada a diversidade dos problemas de otimização que se pode encontrar, as SVMs são abordadas consoante o tipo de problema (BURGES, 1998; LORENA & CARVALHO, 2003; 2007; SILVA, 2014):

- SVM com dados separáveis que também podem ser chamados de SVM com margens rígidas;
- SVM com dados não separáveis;

Dentro das SVM com dados não separáveis consideram-se ainda:

- SVM de margens suaves;
- SVM não linear com “*kernel trick*”;

1.6.1. SVM com dados separáveis ou margens rígidas

A aplicação de SVM começa por ser ilustrada no seu caso mais simples, e de seguida avança-se para contextos e versões mais complexas. Começa-se por considerar o caso de dados linearmente separáveis para duas classes. Assumindo que as classes são identificadas como A e B, o conjunto de dados que pertencem a elas podem ser rotulados por $y_i = 1$ para o elemento x_i que pertence à classe A e por $y_i = -1$ caso o elemento x_i pertença à classe B. Ao assumir que as classes são linearmente separáveis, existe um hiperplano que as separa e logo é possível obter um classificador sem erro de classificação (LORENA & CARVALHO, 2003; THOMÉ, 2012). Considera-se então um conjunto de treino T composto pelos objetos de treino e os respetivos rótulos de dados compostos pelos pares (x_i, y_i) . Considera-se ainda que $x_i \in \mathbb{R}^n$ com $i = 1, 2, \dots, n$ e tal como já referido $y_i \in \{-1, 1\}$ são os rótulos para as duas classes. Se se conseguir encontrar um hiperplano capaz de separar os dados das duas classes anteriores, diz-se então que o conjunto de treino T é linearmente separável. Como o conjunto de treino é formado por pares ou seja, $T = \{(x_i, y_i) : i = 1, \dots, n\}$, então com base nos pares $(x_1, y_1), \dots, (x_n, y_n)$, traduz-se a seguinte expressão para as duas classes (GUYON, BOSER, & VAPNIK, s.d.; SILVA, 2014):

$$\begin{cases} y = +1 \text{ se } x \in \text{classe } \mathcal{X}(A) \\ y = -1 \text{ se } x \in \text{classe } \mathcal{X}(B) \end{cases} \quad (6)$$

em que \mathcal{X} é o espaço caraterístico onde um conjunto de treino é realizado para as duas classes A e B. Se se conseguir ver no espaço geométrico do hiperplano, que as margens por ele separadas são máximas, pode-se dizer então que esse hiperplano é ótimo e é o que melhor separa linearmente as duas classes, que é exatamente aquilo que se pretende. Assim sendo, está-se perante um problema de maximização das margens. Dado um conjunto de exemplos de treino, a menor distância entre os exemplos do referido conjunto e o hiperplano separador dos dados por classe, define a margem de um classificador (LORENA & CARVALHO, 2003). Considerando um conjunto de exemplos positivos e negativos, num espaço caraterístico \mathcal{X} , e pretendendo separá-los uns dos outros no referido espaço, então os pontos x que se encontram sobre o hiperplano de separação, devem satisfazer a equação do plano seguinte:

$$f(x) = w \cdot x + b = 0 \quad (7)$$

onde $w \cdot x$ é o produto escalar entre os vetores w e x , w é o vetor normal ao hiperplano, $|b|/\|w\|$ é a distância perpendicular do hiperplano à origem, e $\|w\|$ é a norma Euclidiana de w . O algoritmo de AA escolhe o hiperplano com maior margem para o caso de dados linearmente separáveis, de forma que todos os dados de treino tenham a classificação correta, o que se traduz nas condições seguintes (BURGES, 1998; LORENA & CARVALHO, 2003):

$$\begin{cases} x_i \cdot w + b \geq +1 \text{ para } y_i = +1 \\ x_i \cdot w + b \leq -1 \text{ para } y_i = -1 \end{cases} \quad (8)$$

que podem ser escritas de forma única como:

$$y_i(x_i \cdot w + b) - 1 \geq 0, \quad \forall_i i = 1, \dots, n \quad (9)$$

pois o vetor w é ortogonal ao hiperplano definido por (w, b) que separa os dados treino.

Para encontrar o hiperplano com a margem máxima, considera-se os hiperplanos que definem as classes dados por:

$$H_1: w \cdot x_i + b = +1 \quad (10)$$

$$H_2: w \cdot x_i + b = -1$$

com o vetor normal w e as distâncias perpendiculares aos planos dadas por:

$$\frac{|1 - b|}{\|w\|} \text{ e } \frac{|-1 - b|}{\|w\|}.$$

Sejam x_1 e x_2 dois pontos pertencentes aos hiperplanos da expressão (10) respetivamente, ou seja, $x_1 \in H_1$ e $x_2 \in H_2$, então a distância entre os referidos hiperplanos que agora traduz-se no seguinte sistema:

$$\begin{cases} H_1: w \cdot x_1 + b = +1 \\ H_2: w \cdot x_2 + b = -1 \end{cases} \quad (11)$$

pode ser encontrada pela diferença entre os dois pontos fazendo uma projeção perpendicular ao hiperplano separador $w \cdot x + b = 0$ na direção de w . Segue a ilustração da distância entre os hiperplanos:

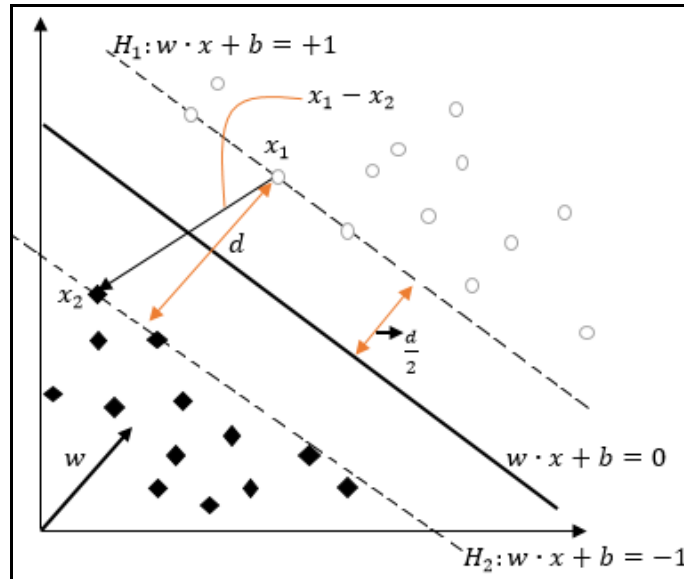


Figura 1.1 - Maximização das margens do hiperplano e o respectivo cálculo da distância d entre H_1 e H_2 .

A expressão seguinte é deduzida da expressão antecedente, isto é, da (11):

$$w \cdot (x_1 - x_2) = 2 \quad (12)$$

sendo que, ao hiperplano separador, são ortogonais os vetores w , $x_1 - x_2$, sendo esses vetores por sua vez paralelos entre si. Assim obtêm-se a seguinte expressão:

$$|w \cdot (x_1 - x_2)| = \|w\| \times \|x_1 - x_2\| \quad (13)$$

onde $\|\cdot\|$ representa a norma do vetor. Substituindo a expressão (11) na (12), obtêm-se uma nova expressão:

$$\|x_1 - x_2\| = \frac{2}{\|w\|} \quad (14)$$

onde a distância entre os hiperplanos em (10) é dada por $\frac{2}{\|w\|}$, pois, é medida pela norma $\|x_2 - x_1\|$. Posto isto, o hiperplano com margem máxima é obtido minimizando $\|w\|^2$ com a seguinte condição do problema de otimização:

$$\min_{w,b} \left(\frac{1}{2} \|w\|^2 \right) \quad (15)$$

sujeita às restrições da expressão (9) (BURGES, 1998; LORENA & CARVALHO, 2003; SILVA, 2014).

O objetivo é encontrar um hiperplano que seja ótimo, e só é possível o encontrar se se conseguir cumprir as condições dadas pelas desigualdades da expressão (9) de modo que a norma $\|w\|$ seja mínima conforme as condições da expressão (15). Ainda nas desigualdades mencionadas anteriormente, isto é, nas desigualdades da expressão (9), pode-se introduzir os multiplicadores de Lagrange (positivos) e transformá-las num problema de otimização quadrática. A introdução dos referidos multiplicadores deve-se a dois motivos fundamentais: 1º - para simplificar ao máximo possível o manuseamento computacional do algoritmo, substituindo as restrições anteriores nos multiplicadores de Lagrange; 2º - o produto interno entre os vetores que é formado pelos dados de treino, constitui uma propriedade fundamental que permite a generalização do procedimento, para o caso de separadores não lineares (BURGES, 1998).

Desta forma, as restrições devem ser lagrangeadas associando a elas os multiplicadores de Lagrange α_i . Assim segue-se a expressão:

$$\alpha_i (y_i (x_i \cdot w + b) - 1) > 0 \quad (16)$$

com $\alpha_i \geq 0$, $i = 1, \dots, n$, onde ao fazer o desdobramento completo do produto e tendo em consideração as restrições $\alpha_i \geq 0$, deve-se de seguida, fazer a soma dos pontos que pode ser escrita da seguinte forma:

$$\sum_{i=1}^n \alpha_i (y_i (x_i \cdot w + b) - 1) \quad (17)$$

pois está-se na presença de uma restrição com um conjunto de pontos (SILVA, 2014).

Assim, a expressão de Lagrange é obtida na sua forma primal pela agregação da função objetivo com a expressão (17) conforme se segue:

$$L_P(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i (x_i \cdot w + b) - 1) \quad (18)$$

que ainda pode ser estendida da seguinte forma:

$$L_P(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^n \alpha_i \quad (19).$$

Deve-se anular as derivadas parciais de $L_P(w, b, \alpha)$ minimizando a expressão (19) em relação a w e b , pois o problema que se pretende solucionar é um problema de minimização. Deve-se ainda, manter as restrições $\alpha_i \geq 0$. O aumento do parâmetro α_i pode provocar também o aumento da expressão (19), caso as restrições $y_i (x_i \cdot w + b) - 1 \geq 0$ sejam violadas. Uma vez que o objetivo é minimizar o problema com recurso à função lagrangeana, espera-se que o resultado final não seja um número muito elevado e nem negativo pois, as restrições anteriores devem ser respeitadas. Para isso, a escolha de w e b deve ser feita de forma muito cuidadosa pois, o objetivo do problema, referido anterior, depende da referida escolha (BURGES, 1998; THOMÉ, 2012).

Conforme já se tinha mencionado acima, deve-se anular as derivadas parciais em relação a w e b , obtendo assim as equações de “ponto de sela” no ponto ótimo (BURGES, 1998):

$$\frac{\partial}{\partial w} L_P(w, b, \alpha) = 0 \Rightarrow w = \sum_{i=1}^n \alpha_i x_i y_i \quad (20)$$

e

$$\frac{\partial}{\partial b} L_P(w, b, \alpha) = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \quad (21).$$

Assim, a formulação dual que se segue, é o resultado do problema de otimização que é obtido substituindo as expressões (20) e (21) na expressão (19)

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (22)$$

sujeita às restrições

$$\begin{cases} \alpha_i \geq 0 \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (23).$$

Se se pretender encontrar a expressão que minimize o problema de otimização, escreve-se novamente a mesma expressão (22) e troca-se os sinais, obtendo-se a seguinte expressão a minimizar:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j - \sum_{i=1}^n \alpha_i \quad (24).$$

Tendo em consideração as variáveis primais, recorre-se às condições de Karush-Kuhn-Tucker (KKT) para obter o ponto sela. Este ponto é obtido por minimização derivando L_P . Assim determina-se o valor de w , o subconjunto de treino expandir-se-á resultando no vetor suporte. De seguida, as margens do hiperplano, isto é, H_1 e H_2 serão determinadas pelos pontos α_i que satisfazem a condição $\alpha_i > 0$, pois os pontos respetivos encontram-se sob as referidas margens e são denominados os vetores suporte que dão nome ao método. Com recurso aos vetores suporte calcula-se o valor de b usando a expressão (16). Assim sendo, segue-se a expressão para o cálculo de b ,

$$b = \frac{1}{nSV} \sum_{x_j \in SV} \frac{1}{y_j} - w \cdot x_j \quad (25)$$

onde nSV corresponde ao número de vetores suporte e SV corresponde ao conjunto de vetores suporte (LORENA & CARVALHO, 2007). O vetor w pode ser substituída na equação anterior permitindo assim a expansão desta. Assim obtêm-se a seguinte equação expandida:

$$b = \frac{1}{nSV} \sum_{x_j \in SV} \left(\frac{1}{y_j} - \sum_{x_i \in SV} \alpha_i y_i x_i \cdot x_j \right) \quad (26).$$

Para os dados separados com as margens máximas pelo hiperplano, tendo em conta aquele com mais poder de generalização, segue-se a função linear classificadora:

$$g(x) = \text{sgn}(f(x)) = \text{sgn} \left(\sum_{x_i \in SV} \alpha_i y_i x_i \cdot x + b \right) \quad (27)$$

que assume os valores -1 ou 1, consoante o lado em que o ponto x_i se encontra do hiperplano de separação (LORENA & CARVALHO, 2007; SILVA, 2014).

1.6.2. SVM com dados não separáveis ou de margens suaves

Tendo já discutido anteriormente o algoritmo para dados separáveis, a compreensão das SVMs fica mais fácil, pois pode-se considerá-lo como algoritmo base para compreender outros casos que a seguir se apresenta. A sua aplicação a dados não separáveis torna-se útil, pois a realidade dos dados raramente estão próximas da ideal (SILVA, 2014). Nesta base surge então a necessidade de encontrar um hiperplano com recurso a uma folga aplicada às restrições onde o custo a introduzir será baseado na introdução de variáveis positivas, isto é, $\xi_i \geq 0$, $i = 1, 2, \dots, l$, permitindo assim a extensão do algoritmo aos dados não separáveis. Esse hiperplano é denominado por hiperplano de margens suaves para dados lineares não separáveis (BURGES, 1998). A restrição (15) será violada para permitir a introdução das variáveis folga, pois os dados são não separáveis, por forma a relaxar essa restrição

minimizando também o somatório de erros nos dados de treino. É com a aplicação deste processo que se consegue ter as margens suavizadas conforme as restrições que se seguem (BURGES, 1998; THOMÉ, 2012):

$$\begin{aligned}x_i \cdot w + b &\geq +1 - \xi_i \text{ para } y_i = +1 \\x_i \cdot w + b &\leq -1 + \xi_i \text{ para } y_i = -1\end{aligned}\tag{28}$$

com

$$\xi_i \geq 0 \quad \forall_i = 1, 2, \dots, n$$

que de forma resumida têm-se

$$y_i(x_i \cdot w + b) \geq 1 - \xi_i\tag{29}$$

onde ξ_i representa as variáveis de folga. A expressão (15) será reformulada novamente onde será adicionado uma constante de regularização C que tem por objetivo, quantificar a penalidade dos erros. Assim obtêm-se a seguinte expressão (BURGES, 1998):

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n \xi_i \right)^k\tag{30}$$

onde k representa os inteiros positivos para a expressão quadrática (30). Então para qualquer valor inteiro positivo de k , está-se perante um problema de programação quadrática. Quanto maior for o regulador C , maior é a penalidade atribuída aos erros ξ_i e logo maior separabilidade é exigida.

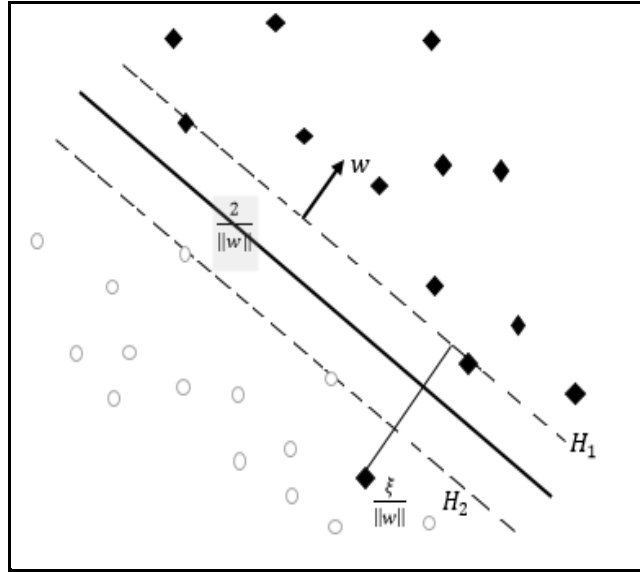


Figura 1.2 - Maximização das margens do hiperplano com a introdução das variáveis de folga para o caso linear não separável.

A escolha de $k = 1$ tem vantagem sobre os multiplicadores de Lagrange, e não permite que nem os ξ_i e nem os seus multiplicadores de Lagrange apareçam na expressão do problema dual que se segue:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (31)$$

com as restrições

$$\left\{ \begin{array}{l} 0 \leq \alpha_i \leq C, \quad \forall i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{array} \right. \quad (32)$$

com a seguinte solução do problema

$$w = \sum_{x_i \in SV} \alpha_i y_i x_i \quad (33)$$

onde SV corresponde ao número de vetores suporte. Como se pode ver, o que diferencia essa formulação da apresentada nas SVM de margens rígidas, isto é, na expressão (22) são apenas as restrições nos α_i a que estão sujeitas, em que há uma limitação imposta pela constante C aos valores dos α_i nas restrições anteriores (BURGES, 1998; SILVA, 2014).

Com base na função primal de Lagrange, o problema que a seguir se apresenta é obtido recorrendo novamente às condições de Karush-Kuhn-Tucker (KKT):

$$L_P(w, b, \alpha, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i - \sum_i \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} + \sum_i \mu_i \xi_i \quad (34)$$

onde α_i e μ_i são multiplicadores de Lagrange e portanto, são ambos não negativos e são parametros livres. A positividade de ξ_i é reforçada com a introdução destes multiplicadores sendo $\alpha_i, \mu_i \geq 0 \quad \forall_i \in \mathbb{N}$. (Nota: i vai de 1 ao número de pontos de treino). É de realçar que o que se pretende com a introdução dos ξ_i é medir a quantidade de violação causada por dados que podem ser classificados de forma errada, isto é, na restrição (9) e tentar encontrar um hiperplano para dados não separáveis que suaviza as margens desse hiperplano, com a extensão do algoritmo SVM para esses dados. Segue então as condições KKT para o problema primal, ou seja, as condições de primeira ordem em (w, b) para o processamento da função dual de $L_P(w, b, \alpha, \xi)$:

$$\frac{\partial}{\partial w} L_P = w - \sum_{i=1}^n \alpha_i y_i x_i = 0 \quad (35)$$

$$\frac{\partial}{\partial b} L_P = - \sum_{i=1}^n \alpha_i y_i = 0 \quad (36)$$

$$\frac{\partial}{\partial \xi_i} L_P = C - \alpha_i + \mu_i \quad (37)$$

$$y_i(x_i \cdot w + b) - 1 + \xi_i = 0 \quad (38)$$

$$\alpha_i(y_i(x_i \cdot w + b) - 1 + \xi_i) = 0 \quad (39)$$

$$\mu_i \xi_i = 0 \quad (40)$$

$$0 \leq \alpha_i \leq C, \quad \forall_i \in \mathbb{N}.$$

O valor de b é calculado tendo em conta as condições de KKT, considerando a condição $\alpha_i(y_i(x_i \cdot w + b) - 1 + \xi_i) = 0$ em que $\xi_i = 0$, que deve ser satisfeita para $0 < \alpha_i < C$ (BURGES, 1998).

1.6.3. SVM não linear com “kernel trick”

Como já foi visto anteriormente, o caso SVM linear com dados separáveis e não separáveis, tem como base o hiperplano separador. No entanto podemos deparar-nos com problemas em que os dados a serem discriminados podem não ser linearmente separáveis no seu espaço de origem. Problemas desse tipo podem ser, por exemplo, resolvidos com a introdução do “kernel trick”, baseado em funções adequadas denominadas por “Kernels”. Estas funções permitem obter um separador não-linear dos dados de origem, com base no mapeamento não-linear (artificial) desses dados num espaço de dimensão superior, espaço no qual são usados classificadores lineares (HOFMANN, 2006). Assim, considerando um mapeamento Φ dos dados num espaço Euclidiano \mathcal{H} tal que

$$\begin{aligned} \Phi: \mathbb{R}^d &\rightarrow \mathcal{H} \\ x &\mapsto \Phi(x) \end{aligned}$$

onde d é a dimensão do espaço de origem. Considera-se agora SVM para dados separáveis em \mathcal{H} com o conjunto de treino definido pelos pontos $(\Phi(x_1), y_1), \dots, (\Phi(x_d), y_d) \in \mathcal{H} \times Y$ (BURGES, 1998; MÜLLER, MIKA, RÄTSCH, TSUDA, & SCHÖLKOPF, 2001). Uma vez que determinar um mapeamento Φ adequado aos dados do problema pode ter custos

computacionais enormes, surge então um truque denominado “*kernel trick*”. Este tem como objetivo substituir o produto interno no espaço de origem nas expressões (27) e (28), pelo produto interno $\Phi(x_i) \cdot \Phi(x_j)$ no espaço \mathcal{H} , sem ter de especificar o mapeamento Φ . Este processo passa por considerar um kernel K adequado tal que

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$$

que tem como base, as condições de separabilidade apresentadas por Thomas Cover (COVER, 1965). Para isso é necessário que duas condições sejam satisfeitas: a não linearidade dos dados a serem transferidos e a dimensionalidade do espaço característico, ou seja, o espaço característico deve ser extenso (LORENA & CARVALHO, 2007). Para o mapeamento do espaço característico, começa-se com um exemplo partindo de um espaço \mathbb{R}^2 onde os dados não são linearmente separáveis devido à dimensão do espaço de origem que é reduzido, para um espaço dimensional isto é, para um espaço \mathbb{R}^3 , onde serão linearmente separáveis, com a notação $\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^3$, conforme se segue (MÜLLER, MIKA, RÄTSCH, TSUDA, & SCHÖLKOPF, 2001):

$$\Phi(y) = (y_1^2, \sqrt{2}y_1y_2, y_2^2)^T \quad (41)$$

com a seguinte função linear em \mathbb{R}^3

$$w^T \cdot \Phi(y) = w_1y_1^2 + w_2\sqrt{2}y_1y_2 + w_3y_2^2 = 0 \quad (42).$$

Para que esses dados em \mathbb{R}^3 sejam linearmente separáveis, é necessário que lhes sejam aplicado um hiperplano linear. Na expressão (31) aplica-se o mapeamento Φ , e novamente obtém-se o problema de otimização com a seguinte forma (BURGES, 1998; LORENA & CARVALHO, 2007; SILVA, 2014):

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\Phi(x_i) \cdot \Phi(x_j)) \quad (43)$$

sujeito às restrições (32), com o classificador

$$g(x) = \text{sgn}(f(x)) = \text{sgn}\left(\sum_{x_i \in SV} \alpha_i y_i \Phi(x_i) \cdot \Phi(x_j) + b\right) \quad (44).$$

A expressão (43) pode ser reformulada com a introdução de uma função kernel, que vai permitir o cálculo do produto interno entre dois espaços característicos x e y , pois, devido à dimensão do espaço que pode ser muito elevada, pode-se deparar com algumas dificuldades em controlar o problema conforme já se tinha referido anteriormente. Assim obtém-se a seguinte expressão (MÜLLER, MIKA, RÄTSCH, TSUDA, & SCHÖLKOPF, 2001):

$$\begin{aligned} (\Phi(x) \cdot \Phi(y)) &= (x_1^2, \sqrt{2}x_1x_2, x_2^2)(y_1^2, \sqrt{2}y_1y_2, y_2^2)^T \\ &= ((x_1, x_2)(y_1, y_2)^T)^2 \\ &= (x \cdot y)^2 \\ &=: k(x, y). \end{aligned}$$

Esta função pode ser generalizada da seguinte forma (MÜLLER, MIKA, RÄTSCH, TSUDA, & SCHÖLKOPF, 2001):

$$k(x, y) = (x \cdot y)^d \quad (45).$$

Substituindo $(\Phi(x_i) \cdot \Phi(x_j))$ com as funções *Kernel* $K(x_i, x_j)$ ao invés de $K(x, y)$, na expressão (43), tem-se o seguinte problema de otimização quadrática na sua forma dual (MÜLLER, MIKA, RÄTSCH, TSUDA, & SCHÖLKOPF, 2001; BEGG, PALANISWAMI, MEMBER, & OWEN, 2005; SILVA, 2014):

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (46)$$

sujeito às restrições

$$\begin{aligned} \alpha_i &\geq 0, & i &= 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i &= 0 \end{aligned} \tag{47}$$

com a seguinte solução do problema

$$g(x) = \text{sgn}(f(x)) = \text{sgn}\left(\sum_{x_i \in SV} \alpha_i y_i k(x_i, x_j) + b\right) \tag{48}.$$

1.6.3.1. Funções de Kernel

As funções básicas de *Kernel* mais utilizadas em SVM são quatro, conforme ilustradas abaixo (HSU, CHANG, & LIN, 2010):

Linear:

$$K(X_i, X_j) = X_i^T X_j \tag{49}$$

Polynomial:

$$K(X_i, X_j) = (\gamma X_i^T X_j + r)^d, \quad \gamma > 0, \quad d \geq 0 \tag{50}$$

Radial Basis Function (RBF):

$$K(X_i, X_j) = \exp\left(-\gamma \|X_i - X_j\|^2\right), \gamma > 0 \tag{51}$$

Sigmoid:

$$K(X_i, X_j) = \tanh(\gamma X_i^T X_j + r) \quad (52)$$

onde γ, r e d são parâmetros de Kernel.

1.6.3.2. Propriedades de Kernel

Normalmente as funções de *Kernel* devem satisfazer as três propriedades seguintes (GENTON, 2001; HOFMANN, 2006; WEHBE, 2013):

1. Simetria,

$$K(x, z) = \Phi(x) \cdot \Phi(z) = K(z, x);$$

2. Semi-definida positiva,

$$\alpha K \alpha \geq 0, \quad \forall \alpha \in \mathbb{R}^m$$

e $k \in \mathbb{R}^{m \times m}$ para todas as matrizes Kernel

3. Desigualdade de Cauchy-Schwarz,

$$\begin{aligned} K(x, z)^2 &= (\Phi(x) \cdot \Phi(z))^2 \leq \|\Phi(x)\|^2 \cdot \|\Phi(z)\|^2 \\ &\leq (\Phi(x) \cdot \Phi(x)) \cdot (\Phi(z) \cdot \Phi(z)) \\ &K(x, z)^2 \leq K(x, x)K(z, z); \end{aligned}$$

em que a função simétrica $K(x, z)$ deve ser positiva semi-definida conforme é ilustrada pela expressão seguinte (RAPAPORT, s.d.):

$$\forall x_1, \dots, x_n \text{ e } \forall c_1, \dots, c_n \in \mathbb{R}^n, \sum_{i=1}^n \sum_{j=1}^n K(x_i, x_j) c_i c_j \geq 0 \quad (53)$$

onde x_1, \dots, x_n representa o conjunto de exemplos de treino e c_1, \dots, c_n representa qualquer conjunto de números reais. Novas *Kernels* podem ser criadas a partir das *Kernels* existentes. Assim assumindo que \tilde{K} e \hat{K} são *Kernels* pertencentes ao espaço $\mathcal{X} \times \mathcal{X}$, $\mathcal{X} \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^+$, $f(\cdot)$ uma função real, o mapeamento $\Phi: \mathcal{X} \rightarrow \mathbb{R}^m$ com \bar{K} um *Kernel* pertencente ao espaço $\mathbb{R}^m \times \mathbb{R}^m$ e B uma matriz simétrica semi-definida $n \times n$, segue algumas propriedades para novos *kernels*, propriedades estas que são chamadas propriedades de fecho (HOFMANN, 2006):

4. $K(x, z) = c \cdot \tilde{K}(x, z)$,
5. $K(x, z) = c + \tilde{K}(x, z)$,
6. $K(x, z) = \tilde{K}(x, z) + \hat{K}(x, z)$,
7. $K(x, z) = \tilde{K}(x, z) \cdot \hat{K}(x, z)$,
8. $K(x, z) = f(x) \cdot f(z)$,
9. $K(x, z) = \bar{K}(\Phi(x), \Phi(z))$,
10. $K(x, z) = x^T B z$

com $c > 0$.

1.6.4. SVM linear multiclasse

Conforme exposto até este momento, as SVMs visam solucionar problemas de classificação binária, ou seja, a classificação dos dados em duas classes distintas. No entanto, pode-se estender o mesmo princípio para a classificação de mais de duas classes. Neste caso está-se perante um problema de classificação multiclasse. A classificação multiclasse pode ser decomposta em vários problemas de classificação binária, que podem ser resolvidos utilizando classificadores binários, permitindo assim o uso das SVMs (FRANC & HLAVÁČ, 2002). O conjunto de treino para um problema multiclasse é formado por pares (x_i, y_i) , onde y_i assume valores de 1 a k , isto é, $y_i \in \{1, \dots, k\}$. Para tal decomposição várias abordagens podem ser consideradas, como a decomposição “um – contra – todos” (“one – against – all”), “um – contra – um” (one – against – one”), entre outros (LORENA & CARVALHO, 2003; SILVA, 2014).

Na decomposição “um – contra – todos” é construído um modelo de classificação SVM para cada classe contra todas as outras, ou seja, para M classes, o objetivo é separar cada classe das restantes $(M - 1)$ classes. Desta forma temos M modelos SVM. A decisão utilizada para classificar um novo elemento x é o winner – takes – all, ou seja, a classe da função discriminante $f(x)$ que possui o valor máximo na saída é atribuído a este. Dado um conjunto de N elementos de treino, todos esses exemplos são utilizados para construir cada modelo SVM para uma classe (LORENA & CARVALHO, 2003; SEO, 2007).

Na decomposição “um – contra – um”, duas classes são discriminadas em cada treino realizado, com a construção de um modelo SVM para a discriminação entre cada par de classes, onde $M(M - 1)/2$ corresponde ao número de modelos SVM obtidos. Assim é possível construir um modelo SVM para um par de classes (k, m) através de exemplos de treino pertencentes apenas a duas classes. Uma estratégia utilizada é baseada na escolha da classe com maior número de classificação do elemento x , usando todas as funções discriminatórias entre pares. Se para a função $f_{km}(x)$, o valor da função discriminante para o par de classes (k, m) for positivo, então a classe k ganha maior número de pontos, caso contrário, a classe m ganha mais. O número de classificação do elemento x obtido em cada classe, é determinado pelas saídas das SVMs. Um padrão de teste recebe como atribuição, a classe com o máximo número de vezes que o elemento x é classificado. À função discriminante, é atribuída a classe com o máximo valor de amplitude total, caso o número máximo de classificação do elemento x for observado em várias classes. Se o elemento x estiver bem classificado, a amplitude total para a função discriminante (ATFD) para k é ilustrada abaixo:

$$ATFD_k = \sum_m |f_{km}(x)| \quad (54)$$

onde o somatório que está sobre todos os m com k classes é emparelhado (LORENA & CARVALHO, 2003; SEO, 2007).



CAPÍTULO 2

DOENÇA DE PARKINSON.

2. Doença de Parkinson

Este capítulo é feita uma breve abordagem sobre o conceito da doença de Parkinson (DP) e algumas características motoras dos doentes de Parkinson pois, os dados recolhidos para a elaboração deste trabalho são dados que foram utilizados num estudo para a avaliação clínica da evolução desta doença para a função motora. Uma outra abordagem também relevante são os testes de posturografia dinâmica computadorizada utilizados para avaliar a marcha e a postura dos doentes. Para tal são apresentados dois dos referidos testes.

2.1. Conceito

A DP é considerada uma doença neurológica, degenerativa, crónica e progressiva, do sistema nervoso central que ataca principalmente o cérebro. Este é um dos principais e mais comuns distúrbios nervosos da terceira idade e é caracterizado, principalmente, por prejudicar a coordenação motora e provocar tremores e dificuldades para caminhar e se movimentar. Os DP apresentam algumas características motoras conforme apresentado na figura 2.3 (GODINHO, 2013; LIMA, s.d.; PEREIRA & GARRETT, 2010).

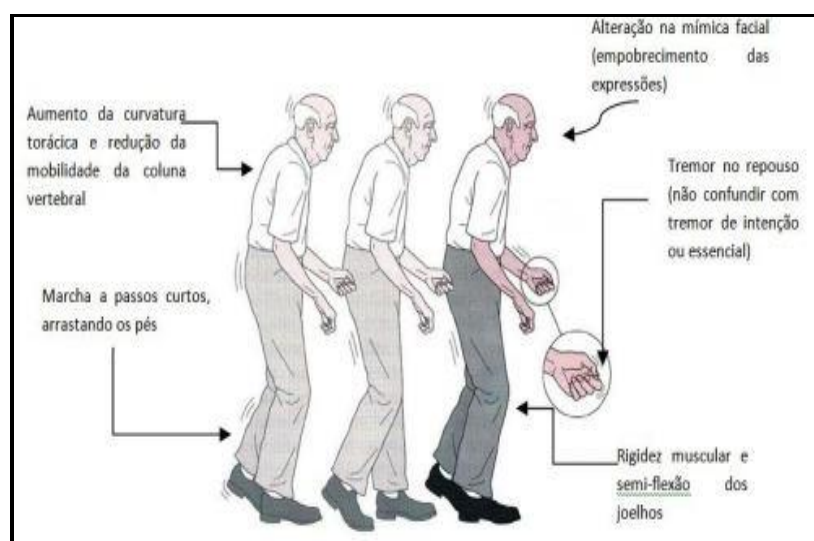


Figura 2.3 - Representação da imagem motora característica do doente com DP (GODINHO, 2013).

2.2. Testes de Posturografia Dinâmica Computorizada

Os testes de posturografia dinâmica computadorizada são testes realizados aos doentes através de um sistema chamado *Balance Mester*. O *Balance Mester* é um instrumento pelo

qual, as condições relacionadas com o equilíbrio são avaliadas e reabilitadas, sendo a medição da oscilação postural através de uma plataforma de forças com 1,52 m e de um software específico para o processamento dos dados, a base deste instrumento (GODINHO, 2013).

A seguir são apresentados dois testes, sendo o primeiro relacionado com a marcha e o segundo relacionado com a postura dos doentes.

2.2.1. Teste de marcha

Este teste (WA - em inglês: *walk across*), é utilizado para quantificar as características de marcha do doente durante o período em que este se desloca ao longo da plataforma de forças, onde os parâmetros a serem avaliados são: Largura da passada (cm); Comprimento da passada (cm); Velocidade média da marcha (cm/seg) e Simetria da marcha (%). A figura 2.5 mostra a realização do teste WA, onde a marcha é iniciada antes do limite inicial da área de registo da plataforma e só termina depois de ultrapassar o limite final da mesma. A tarefa é realizada 3 vezes (GODINHO, 2013).



Figura 2.4 - Realização do teste WA (GODINHO, 2013).

2.2.2. Limites de Estabilidade

Este teste (LOS - em inglês: *limits of stability*), tem por objetivo quantificar a distância máxima a que um indivíduo pode deslocar-se voluntariamente o CG, mantendo dentro da base de sustentação a projeção vertical, isto é, sem perder o equilíbrio, dar um passo ou apoiar-se, manter o seu corpo inclinado numa determinada direção. Segue a ilustração do teste na figura 2.7 (GODINHO, 2013).



Figura 2.5 - Realização do teste LOS (GODINHO, 2013).

Um conjunto de alvos é facultado ao indivíduo, no ecrã do monitor, para serem atingidos, mantendo o seu CG no interior de um desses alvos que serão indicados no ecrã do monitor, onde o CG é representado através de um cursor. A medida em que o programa vai fornecendo os comandos, em cada repetição realizada, o cursor do CG deverá ser movido o mais rápido possível, pelo indivíduo, para o alvo seguinte, assinalado no ecrã do instrumento, localizado no perímetro dos limites de estabilidade (100% dos limites de estabilidade teóricos), tentando manter o mais próximo possível, a sua posição, encima deste novo alvo. O percurso é realizado apenas uma vez e o tempo estabelecido ao indivíduo para tal é igual ao número de alvos, em segundos. Por exemplo se lhe for fornecido 8 alvos, então o tempo para realizar cada percurso é de 8 segundos (GODINHO, 2013).



CAPÍTULO 3

METODOLOGIA DO TRABALHO PROPOSTO.

3. Metodologia

3.1. Amostra e descrição dos dados

Os dados utilizados para este estudo foram obtidos na Faculdade de Motricidade Humana, Universidade de Lisboa, facultados pelo meu co-orientador Professor Doutor Luís Silva, Universidade onde ele trabalha, num ficheiro de dados em Excel. Estes dados são provenientes de um estudo realizado a grupos de indivíduos para a avaliação clínica sobre a evolução da DP para a função motora, baseada na escala de Hoehn & Yahr (H&Y). Escala de H & Y é uma escala de avaliação clínica que define grandes categorias da função motora na Doença de Parkinson (HOEHN & YAHR, 2012).

Como dito anteriormente, consideramos dois conjuntos de dados. O primeiro é composto por 161 elementos com idade e altura média de 69.43 anos e de 162.97 cm, que consistiu na marcha dos sujeitos sobre uma passadeira, recolhendo-se valores relativamente às características da marcha. Estas características estão relacionadas com o Comprimento da passada (cm), Largura da passada (cm), Velocidade da passada (cm/seg) e Simetria da passada (%). O segundo é composto por 98 elementos com idade e altura média de 70.67 anos e 190.96 cm e consistiu na observação de um ponto móvel num ecrã, recolhendo-se valores relacionadas com as características visuais. Estes estão relacionados com os seguintes conjuntos de *features*: Distância (mm), Amplitude (mm), Área (mm), Velocidade (%seg) e Frequência de oscilação (Hz).

Para ambas as amostras, os indivíduos foram diagnosticados com a doença de Parkinson cuja evolução desta doença é avaliada em diferentes estadios, codificados de 1 a 4 em que 1 corresponde a evolução menos grave da doença e 4 a mais grave. Estes doentes foram todos seguidos na consulta de doenças do movimento do Hospital Santo António dos Capuchos, (Centro Hospitalar de Lisboa Central – EPE) entre fevereiro de 2008 e julho de 2010 (GODINHO, 2013).

3.2. Objetivos do trabalho

De um modo geral, pretende-se com este estudo, rever a técnica SVM, aplicá-la a dados de comportamento motor para a classificação de Parkinson e retirar conclusões.

Este estudo tem como objetivo específico, determinar um modelo de classificação multiclasse através de SVM entre grupos de sujeitos, com base em dados de comportamento motor de sujeitos de diferentes grupos de interesse. Assim, pretende-se entender quais as variáveis que mais contribuem para a classificação dos sujeitos entre grupos, ou seja, quais

as variáveis com maior poder discriminatório dos sujeitos entre grupos e a sua precisão na classificação.

3.3. Metodologia proposta

Começamos com uma análise preliminar dos dados da amostra por grupo, onde aplicamos a ANOVA simples para averiguar a igualdade das médias de altura e idade por estádio.

Considerando os dados de marcha, começamos por aplicar em primeiro lugar o SVM linear, pois é o mais direto, passando de seguida para o RBF, pois permite separar dados não separáveis linearmente.

Finalmente consideramos os dados obtidos por visualização de um ponto móvel num ecrã. Neste caso propomos aplicar SVM linear e RBF para cada caso, com o objetivo de verificar quais as features que garantam maior poder discriminatório entre grupos de sujeitos.

3.4. Problema do estudo

Tendo iniciada a classificação com as 4 *features* de marcha, Comprimento da passada (WA – WS), Largura da passada (WA – SL), Velocidade média da passada (WA – SPEED) e Simetria da passada (WA – SIMETRIA), partindo de seguida para a classificação com 5 conjuntos de *features* relacionadas à visualização de um ponto móvel num ecrã, sendo estes por sua vez, relacionados com a Distância, a Amplitude, a Área, a Velocidade e a Frequência, o problema que sustenta o presente estudo é então discriminar os estádios de Parkinson através das variáveis de marcha e das variáveis de visualização.

3.5. Instrumentos

Para a caracterização do nível de severidade da doença foi utilizada a escala de H & Y que se encontra em anexo, durante a observação clínica dos doentes. De acordo com esta escala os 4 estádios utilizados para o presente estudo foram formados da seguinte forma: 0 e 1 formam o primeiro estádio, 1.5 e 2 o segundo, 2.5 e 3 o terceiro, sendo 4 e 5 o quarto estádio.

Para a realização dos testes posturográficos para a avaliação quantitativa do comportamento postural e da marcha dos doentes foi utilizada a plataforma Balance Master® (Neurocom® International Inc., USA), com 1,52 m de comprimento (GODINHO, 2013).

3.6. Procedimentos

Os dados foram recolhidos com base nos testes posturográficos submetidos aos doentes. Estes posicionavam-se sobre uma plataforma de forças, com os pés descalços, e após um período de habituação à posição os doentes faziam sinal para que se pudesse dar início ao registo dos dados. Cada sessão de avaliação teve uma duração de 30 a 40 minutos (GODINHO, 2013).

Os dados sócio-demográficos, a história clínica, o estadió da doença e as AVDs foram registados no dia da consulta, pelo clínico, com o consentimento livre e esclarecido dos doentes após esclarecimento prestado aos doentes sobre o estudo (GODINHO, 2013).

A avaliação posturográfica do comportamento postural e da marcha só foi efetuada num período máximo de 2 horas após a toma da medicação anti-parkinsoniana (fase on), e sempre da parte da manhã, seguindo as indicações de Jorgensen (GODINHO, 2013; JORGENSEN, et al., 2012).

Foram feitos estudos de validação cruzada (*cross validation*) do modelo, SVM, para determinar os parâmetros do modelo escolhido que garantam a maior precisão possível. Face a fraca precisão dos resultados com a aplicação do *cross validation*, foi necessário criar rotinas automáticas para gerar grupos aleatórios a partir das amostras em estudo. Foi também criada uma rotina que garantisse a existência de pelo menos um elemento em cada grupo de treino, pois se nestes não houver elementos de uma das classes, o algoritmo nunca há-de classificá-los como sendo desta classe porque ele a desconhecerá pela falta de elementos. Para os dois casos, Kernel linear e RBF, foi criada k matrizes de confusão 4×4 para cada grupo de teste, grupo que fica de fora, para poder visualizar os elementos que foram mal classificados e como é que estes estão mal classificados. No final foi feito o somatório das k matrizes para obter uma matriz global. Esta matriz dá a precisão global que é calculada pela percentagem do quociente da soma dos elementos da diagonal da matriz pelo total de elementos. É esta a precisão de interesse para a classificação. Após a implementação da rotina exposta anterior, no software *Octave*[®] VR4.0.0 (*GUI*), deu-se início aos testes, onde todas as tarefas desempenhadas pelo algoritmo são suportadas por esta rotina.

Para que o objetivo fosse concretizado, diversas tentativas foram realizadas, com os dados em estudo, em busca de resultados ótimos que garantem uma boa precisão para a classificação. Assim, começamos com a classificação multiclasse para o primeiro grupo de

dados. De seguida equilibramos as dimensões dos dados por classes, pois o algoritmo classificava todos os elementos como sendo da classe 3 pelo facto desta ter mais elementos nos dados iniciais. Consideramos grupos de features com maior diferença de médias entre as classes, com a retirada dos outliers e das features correlacionadas. De seguida passamos para a classificação dos dados em apenas duas classes onde as 4 classes foram agrupadas duas a duas. Assim as classes 1 e 2 formam a classe A, sendo que 3 e 4 formam a B.

Face a fraca precisão dos resultados obtidos na classificação multiclasse, partimos para o segundo grupo de dados. Foram replicadas as mesmas tentativas descritas anteriores para o primeiro grupo. Novamente estas variáveis não apresentaram poder discriminatório para o caso multiclasse. Foi ainda realizada, no segundo grupo, a classificação binária de discinesias do membro superior para os indivíduos diagnosticados com ou sem essa discinesia.



CAPÍTULO 4

RESULTADOS

4. Resultados

Neste capítulo apresentamos os resultados obtidos pela metodologia proposta aos dados considerados.

Começamos com a descrição estatística dos dados e de seguida apresentamos uma análise exaustiva dos resultados onde os primeiros resultados de classificação no ponto 4.2. Os segundos resultados são apresentados nos pontos 4.3 e 4.4, os terceiros, os quartos e os quintos são apresentados nos pontos 4.5, 4.6 e 4.7, sendo o sexto resultado apresentado no ponto 4.8.

É de salientar que o algoritmo foi aplicado em todos os grupos de variáveis mas serão apresentados apenas os melhores resultados obtidos pela referida metodologia dado o objetivo do estudo. Estes resultados foram obtidos utilizando o *cross validation* com 5 grupos, ou seja, o *cross validation* divide os dados das amostras em 5 grupos e cada vez deixa um grupo de fora para o teste e treina os outros quatro. Ao treinar estes 4 grupos cria o modelo e vai testá-lo no grupo que ficou de fora (de teste) por forma a obter a classificação correta de cada elemento de entrada inicial.

Para os dados de marcha o SVM não consegue bons resultados na maioria dos casos. O caso de maior sucesso é obtido por aplicação ao modelo RBF no caso binário apresentado na secção 4.6.1.1.2.

Para os dados visuais o SVM comporta-se mal na classificação multiclasse. Quando são classificados apenas em duas classes (caso binário resultante da junção das classes 1 e 2 na classe A e 3 e 4 na classe B) os resultados melhoram significativamente. Este caso é também o de maior sucesso por aplicação do mesmo modelo anterior. Os resultados são apresentados nas secções 4.3.1.3.3.2, 4.4.1.3.3.2, 4.5.1.3.3.2, 4.5.2.1.2, 4.6.2.1.2, 4.7.1.2.3.2 e 4.8.1.3.2.

Assim sendo, segue a tabela 4.1 onde é apresentado um resumo dos melhores resultados, referidos anteriores, com as seguintes características: Secções, Caso, Função *Kernel*, Valores dos parâmetros, a Precisão global e a Página em que se encontra cada secção.

Tabela 4.1 - Resumo dos melhores resultados obtidos.

Secções	Caso	Função <i>Kernel</i>	Valores dos parâmetros		Precisão global	Páginas
			C	γ		
4.3.1.3.3.2	Binário	RBF	10^4	10^{-4}	80.8%	59
4.4.1.3.3.2	Binário	RBF	10^3	10^{-4}	78.9%	63
4.5.1.3.3.2	Binário	RBF	10^3	10^{-3}	73.1%	68
4.5.2.1.2	Binário	RBF	10^4	10^{-3}	73.1%	70
4.6.1.1.2	Binário	RBF	10^2	10^{-1}	63.2%	72
4.6.2.1.2	Binário	RBF	10^2	10^{-3}	80.8%	74
4.7.1.2.3.2	Binário	RBF	10^4	10^{-3}	70.0%	81
4.8.1.3.2	Binário	RBF	10^3	10^{-4}	75.0%	84

4.1. Análise Descritiva dos dados

Para a descrição dos dados em estudo foram utilizados medidas de tendência central (médias) e de dispersão (desvio padrão, máximo, mínimo), e a percentagem para variáveis nominais.

Este tratamento foi ainda complementado com a inferência estatística, isto é, foram confirmados os pressupostos da normalidade dos dados de cada uma das *features* com recurso ao teste de Kolmogorov- Smirnov (K – S) e a ANOVA a um fator para a diferença entre as médias de cada uma das variáveis, Altura e idade por estadio. Na Altura o objetivo é averiguar qual é a média mais afetada pela DP, sendo que na Idade, que é considerada um fator de risco, o objetivo é também averiguar a média mais afaetada pela mesma.

O tratamento estatístico a que esses dados foram submetidos foi um tratamento descritivo com recurso ao programa SPSS® VR24.0 da IBM, permitindo assim estabelecer todas as análises convenientes das amostras utilizadas neste estudo.

4.1.1. Descrição dos dados de marcha

4.1.1.1. Distribuição dos doentes por Sexo

Começamos por fazer a distribuição dos doentes por sexo, isto é, masculino (M) e feminino (F), para averiguar qual destes é que são mais afetados pela DP de acordo com as amostras consideradas para o estudo.

Podemos verificar na tabela 4.2 que, dos 161 indivíduos diagnosticados com a DP a maior parte é do sexo masculino com um total de 102 indivíduos que corresponde a 63.4% do total, sendo 59 indivíduos do sexo feminino correspondendo a 36.6% do total.

Tabela 4.2 - Distribuição por Sexo

	Frequência	Percentagem (%)
M	102	63.4%
Válido F	59	36.6%
Total	161	100.0%

4.1.1.2. Distribuição e caracterização de Altura e Idade dos doentes por estadio

Na tabela 4.3 os doentes encontram-se distribuídos por frequência em cada estadio ou por classe e as respetivas percentagens. Daí é extraída a seguinte interpretação: a frequência de indivíduos com a DP no estadio 1 é de 16 que corresponde a 9.9% do total, 44 no estadio 2 que corresponde a 27.3% do total, 73 no estadio 3 que correspondendo a 45.3% e 28 no estadio 4 correspondendo a 17.4% do total. Quanto a caracterização estes foram caracterizados em termos de estadios, para verificar se existem diferenças significativas nas Alturas e Idade, por estadio. Na mesma tabela, isto é, 4.3 são apresentadas as variáveis Altura e Idade, a média amostral e o desvio padrão destas em cada estadio. As respostas são dadas numa escala de H&Y de estadios. É sobre esses valores que são calculadas as estatísticas.

Tabela 4.3 – Distribuição e caracterização dos doentes por estadios

Estadios	Frequência	Percentagem (%)	Variáveis	média	dp
1	16	9.9%	Altura	163.5	9.5
			Idade	72.8	5.2
2	44	27.3%	Altura	162.8	9.9
			Idade	67.6	10.6
3	73	45.3%	Altura	163.3	10.8
			Idade	69.2	9.5
4	28	17.4%	Altura	162.1	9.1
			Idade	71.0	7.8
Total	161	100.0%	Altura	163.0	10.1
			Idade	69.4	9.3

4.1.1.3. ANOVA simples para a diferença de médias nas Alturas e Idades por estadios

Aplicamos a ANOVA simples a um fator, para estudar se existem diferenças significativas nas Alturas e Idades por estadio, ou seja, se há diferenças entre médias de cada uma dessas variáveis por estadio. Verificamos que a um nível de significância $\alpha = 0.05$, não existe evidência estatística de diferenças entre as médias de Altura em cada estadio, pois o $p - value$ é 0.951 para o teste de ANOVA.

A mesma interpretação anterior é também válida para a Idade, ou seja, não existe evidência estatística de diferenças entre as médias de Idade em cada estadio, a um mesmo nível de significância ($\alpha = 0.05$), pois o $p - value$ é 0.200 para o teste de ANOVA.

4.1.2. Descrição dos dados visuais

Nesta secção começamos também por fazer a distribuição dos dados por sexo pelo mesmo objetivo apontado na secção 4.1.1.1.

4.1.2.1. Distribuição dos doentes por Sexo

A tabela 4.5 mostra que, dos 98 indivíduos diagnosticados com a DP, a maior frequência foi do sexo masculino com um total de 56 indivíduos que correspondem a 57.1% do total, sendo os 42 do sexo feminino correspondendo a 42.9% do total.

Tabela 4.4 - Distribuição por Sexo

		Frequência	Percentagem (%)
Válido	M	56	57.1%
	F	42	42.9%
	Total	98	100.0%

4.1.2.2. Distribuição e caracterização de Altura e Idade dos doentes por estadio

Na tabela 4.5 é realizada a mesma análise da tabela 4.3 sendo que a frequência de indivíduos com a DP no estadio 1 é de 15 que corresponde a 15.3% do total, 29 no estadio 2 que corresponde a 29.6% do total, 47 no estadio 3 correspondendo a 48.0% e 28 no estadio 4 correspondendo a 7.1% do total. Quanto a caracterização é ainda válida, para este ponto, a mesma análise da referida tabela.

Tabela 4.5 – Distribuição e caracterização dos doentes por estadios

Estadios	Frequência	Percentagem (%)	Variáveis	média	dp
1	15	15.3%	Altura	72.9	5.3
			Idade	162.8	9.4
2	29	29.6%	Altura	70.8	9.6
			Idade	162.1	8.0
3	47	48.0%	Altura	70.4	7.8
			Idade	159.4	9.2
4	7	7.1%	Altura	67.1	6.9
			Idade	162.7	8.2
Total	98	100.0%	Altura	70.7	8.0
			Idade	161.0	8.8

4.1.2.3. ANOVA simples para a diferença de médias nas Alturas e Idades por estadios

Novamente verificamos que a um nível de significância $\alpha = 0.05$, também não existe evidência estatística de diferenças entre as médias de Altura em cada estadio, conforme o ponto 4.1.1.3.1, pois o $p - value$ é 0.404 para o teste de ANOVA. Esta mesma interpretação é também válida para a Idade a um nível de significância $\alpha = 0.05$, pois o $p - value$ é 0.475 para o teste de ANOVA.

4.2. Classificação com as *features* de marcha

Começamos por considerar as seguintes *features* de marcha: WA – WS, WA – SL, WA – SPEED e WA – SIMETRIA. O objetivo é verificar se têm poder discriminatório entre as classes. Foi calculada a média e o desvio padrão para cada classe e o valor total. Também foi ainda calculado o número de elementos por classe.

Para a referida análise foi feita ainda representações gráficas. A tabela 4.6 apresenta a média e o desvio padrão para cada *feature* por classe. É notório que a *feature* é distinta das restantes.

Tabela 4.6 - Estatísticas de grupo para as *features* de marcha, por classe.

	Estádios									
	1		2		3		4		Total	
	média	dp ¹	média	dp	média	dp	média	dp	média	dp
WA_SW	16.3	2.6	16.9	3.4	16.9	3.8	15.5	4.7	16.6	3.8
WA_SL	42.6	13.7	43.0	14.0	40.9	16.9	35.8	19.9	40.7	16.5
WA_SPEED	46.3	14.8	49.1	17.8	46.3	21.8	38.8	22.2	45.8	20.3
WA_SIMETRIA	20	17	16	13	22	20	21	20	20	18

Na figura 4.1 são apresentadas as *features* 3-a-3, com 3 dimensões, para averiguar a relação existente entre os valores de cada uma delas consoante as classes. Verificamos que as classes estão todas associadas. Podemos verificar ainda, na referida figura, que aparentemente estas *features* não servem para separar as classes, pelo fato referido acima.

¹ desvio padrão

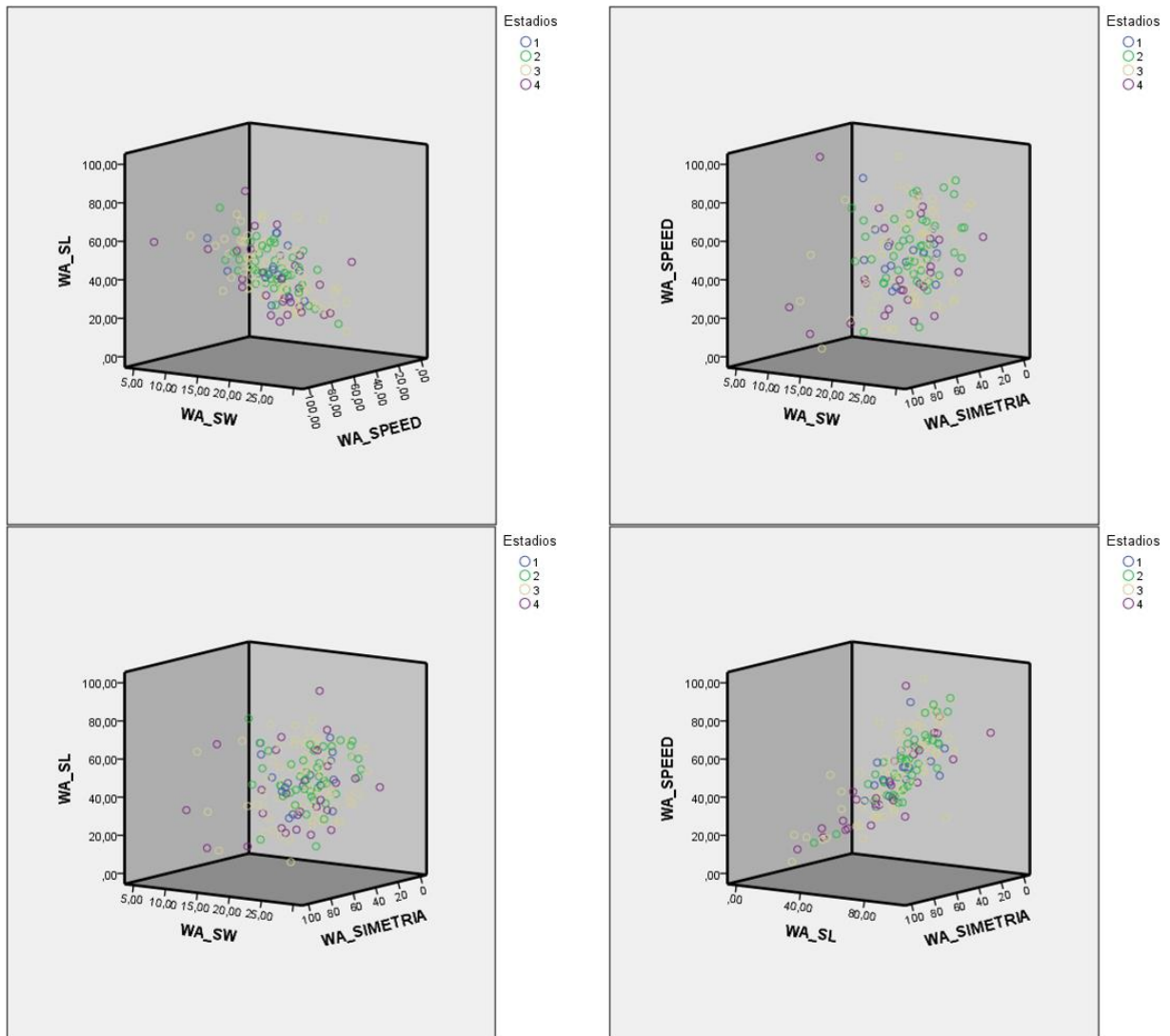


Figura 6.1 - Visualização gráfica das *features* de marcha

4.2.1. Caso Multiclasse

Tendo sido já realizada uma revisão da literatura sobre as SVM, começamos por indicar os parâmetros específicos deste algoritmo usados para a classificação, com base nas 4 *features* referidas anteriormente, para testar se têm poder discriminatório entre as classes. Para tal, foram considerados o Kernel Linear e o Kernel RBF.

Os melhores parâmetros serão escolhidos por *cross validation* com 5 grupos, com suporte num *grid search*.

4.2.1.1. Kernel Linear

Começamos com a separação linear, pois é o mais direto e permite saber os pesos de cada feature e a importância destes para a classificação. A fórmula pela qual esta aplicação é suportada, encontra-se ilustrada no (49). Foram atribuídos os seguintes valores ao parâmetro C onde é realizado o grid search,

$$C = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 0.75, 1, 10, 50, 10^2, 10^3, 10^4.$$

Após ter realizado o *grid search*, a melhor precisão global obtida foi de 45.3% para os valores de $C = 10^{-4}, 10^{-1}, 0.5, 1, 50$ selecionados por *cross validation* como sendo os melhores. Ainda assim não é suficiente para garantir uma boa classificação dos dados para este caso. A matriz de confusão global é apresentada na tabela 4.7 para verificar como é que os erros estão distribuídos. É notório que este modelo é inútil pois classifica todos os elementos como classe 3.

Tabela 4.7 - Matriz de confusão global para o caso multiclasse com *features* de marcha e kernel Linear.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	3	16 (9.9%)	44 (27.3%)	73 (45.3%)	28 (17.4%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

De notar que isto pode ser influenciado pelo facto de a classe ter uma dimensão muito superior às restantes nos dados de treino (45%). Tentámos então outros modelos para averiguar se o modelo obtido é utilizável. Antes de testar o modelo com classes de dimensões equilibradas, vamos testar um modelo não linear, para verificar se há melhoras significativas na classificação.

4.2.1.2. Kernel RBF

Sendo o classificador linear mais limitado na separação dos dados em relação ao RBF, partimos para a separação RBF pois, este separador permite realizar separações não-lineares, enquanto que o Kernel Linear permite apenas a separação de dados linearmente separáveis, ficando restrito a poucas aplicações práticas, (HSU, CHANG, & LIN, 2010; SANTOS, 2002). No RBF são otimizados dois parâmetros, o parâmetro custo (C) e o gamma (γ), tendo a sua aplicação como suporte, a equação ilustrada no ponto (51). Os valores atribuídos ao C no 4.2.1.1 são os mesmos atribuídos ao γ , para o mesmo efeito,

$$\gamma = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.5, 0.75, 1, 10, 50, 10^2, 10^3, 10^4.$$

Assim, após ter realizado o *grid search*, a melhor precisão global obtida foi de 47.2% para os valores de $C = 10^{-1}$, $\gamma = 10$ levando a concluir que, para estes valores e com os dados em estudo o modelo RBF continua a não conseguir uma boa classificação, embora há uma ligeira melhoria na precisão em relação ao caso linear. A matriz de confusão global na tabela 4.8 para estes valores, mostra que o algoritmo classificou a maioria dos elementos como classe 3, mas consegue também acertar a classificação de elementos residuais das classes 2 e 4.

Tabela 4.8 - Matriz de confusão global para o caso multiclasse com *features* de marcha e Kernel RBF.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (0.0%)	3 (1.9%)	1 (0.6%)	1 (0.6%)
	3	16 (9.9%)	41 (25.5%)	72 (44.7%)	25 (15.5%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.2%)

Com base nos resultados obtidos podemos afirmar que os valores atribuídos aos parâmetros C e γ , não garantiram uma boa classificação dos dados em estudo, nem para o caso linear e nem para o RBF.

4.2.1.3. Equilíbrio das dimensões dos dados por classes

Como o facto da dimensão das classes não estar equilibrada nos dados influenciou muito os resultados, uma vez que a maioria dos elementos são classificados como classe 3 tanto com Kernel Linear ou RBF, tentamos de seguida utilizar um conjunto de treino equilibrado por classes. Assim, mantendo todos os elementos da classe 1, foi extraída uma amostra aleatória de 20 elementos em cada uma das classes 2, 3 e 4. Posto isto, foi considerada uma nova amostra de 76 elementos para novos experimentos.

4.2.1.3.1. Kernel Linear

Para este caso a melhor precisão global obtida foi de 31.6% para $C = 0.5$. Por um lado o algoritmo consegue variar a tendência de classificar tudo de uma só classe. Por outro lado, o resultado obtido tende a piorar em relação ao obtido no 4.2.1.1. A tabela 4.9 mostra que o algoritmo classificou a maioria dos elementos como classe 4, conseguindo também acertar a classificação de elementos residuais das restantes classes.

Tabela 4.9 - Matriz de confusão global para o caso multiclasse com *features* de marcha e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	1 (1.3%)	4 (5.3%)	3 (4.0%)	0 (0.0%)
	2	1 (1.3%)	7 (9.2%)	4 (5.3%)	4 (5.3%)
	3	9 (11.8%)	2 (2.6%)	5 (6.6%)	5 (6.6%)
	4	5 (6.6%)	7 (9.2%)	8 (10.5%)	11 (14.5%)

4.2.1.3.2. Kernel RBF

Para o Kernel RBF a melhor precisão obtida foi de 36.8% com $C = 10^4$ e $\gamma = 10^{-4}$. A mesma análise do ponto antecedente é também válida para este, sendo o resultado comparado com o obtido no 4.2.1.2. De forma análoga, a mesma interpretação da tabela 4.9 é também válida para a tabela 4.10.

Tabela 4.10 - Matriz de confusão global para o caso multiclasse com *features* de marcha e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	1 (1.3%)	2 (2.6%)	1 (1.3%)
	2	6 (7.9%)	9 (11.8%)	6 (7.9%)	3 (4.0%)
	3	3 (4.0%)	2 (2.6%)	5 (6.6%)	2 (2.6%)
	4	7 (9.2%)	8 (10.5%)	7 (9.2%)	14 (18.4%)

Desta secção podemos concluir, através das matrizes de confusão global, que o facto da classe 3 ter mais elementos de entradas iniciais fez com que o algoritmo classificasse todos os elementos como sendo desta classe. Mesmo assim as precisões globais obtidas, quer para o Kernel Linear quer para o RBF, não garantiram uma boa classificação dos dados. Face às dificuldades do algoritmo na classificação multiclasse, foi realizada uma nova tentativa considerando o caso binário com as dimensões das classes equilibradas.

4.2.1.3.3. Caso binário

Com base nos resultados obtidos anterior, o objetivo para o novo teste é retirar complexidade à classificação. Assim as 4 classes foram agrupadas duas a duas em que 1 e 2 formam a classe A sendo que a 3 e 4 formam a B tendo ficado com apenas duas classes, considerando a amostra referida no ponto 4.2.2. Assim sendo, a classe A ficou com 36 elementos que corresponde a uma proporção de 47.4% do total, tendo a B ficado com 40 correspondendo a uma proporção de 52.6% do total. Replicando os mesmos passos anteriores passamos a apresentar os resultados obtidos.

4.2.1.3.3.1. Kernel Linear

A melhor precisão global obtida para o caso linear foi de 57.9% para $C = 10^3$. Embora este resultado não seja assim tão relevante para a classificação, este algoritmo é utilizável tendo em conta a relação de proporção entre as duas classes no 4.2.1.3.3. Verificamos na matriz de confusão global apresentada na tabela 4.11 que em todas as classes

o algoritmo erra na classificação. Classifica elementos da classe A como sendo da B e vice versa. Porém, quer para a classe A quer para a classe B o algoritmo classifica bem a maiorias dos elementos nestas classes.

Tabela 4.11 - Matriz de confusão global para o caso binário com *features* de marcha e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	20 (26.3%)	16 (21.1%)
	B	16 (21.1%)	24 (31.6%)

4.2.1.3.3.2. Kernel RBF

Para o Kernel RBF a melhor precisão global obtida foi de 60.5% para os valores de $C = 10^3$ e $\gamma = 10^{-3}$, selecionada por *cross validation*. Este resultado é melhor que o obtido no ponto 4.2.1.3.3.1, portanto é de notar que, embora não seja um resultado desejável para a classificação, o algoritmo melhorou o seu performance, para este caso, atingindo uma precisão de classificação acima dos 60.0%. A mesma análise do resultado obtido no 4.2.1.3.3.1 é também válida para este ponto. A mesma interpretação da tabela 4.11 é também válida para a tabela 4.12.

Tabela 4.12 - Matriz de confusão global para o caso binário com *features* de marcha e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	23 (30.3%)	17 (22.4%)
	B	13 (17.1%)	23 (30.3%)

4.3. Classificação com as *features* Distância

Nesta secção testamos novos conjuntos de *features*, isto é, as relacionadas com a tarefa visual pelo qual os sujeitos foram submetidos, descritas na secção 4.1.2, para averiguar se melhoram os resultados. Foram consideradas as *features* em título porque são as que apresentaram melhores resultados para a classificação multiclasse. Este conjunto é constituído pelas seguintes *features*: Dist_T_11, Dist_T_12, Dist_T_13, Dist_T_21, Dist_T_22, Dist_T_23², apresentadas na tabela 4.13.

Tabela 4.13 - Estatísticas de grupo para as *features* Distância, por classe.

	Estádios H&Y									
	1		2		3		4		Total	
	média	dp	média	dp	média	dp	média	dp	média	dp
Dist_T_11	8.9	5.1	8.5	6.8	6.8	5.7	12.2	7.2	8.0	6.1
Dist_T_12	4.9	2.7	5.3	4.3	6.8	5.2	17.1	14.7	6.8	6.4
Dist_T_13	6.0	6.4	5.7	4.1	6.7	8.0	10.2	8.0	6.6	6.8
Dist_T_21	5.3	3.6	4.9	5.1	6.8	7.0	12.4	10.4	6.4	6.5
Dist_T_22	5.1	3.8	4.6	4.1	6.0	5.6	8.8	7.4	5.6	5.1
Dist_T_23	13.6	6.6	15.5	9.8	12.9	9.2	18.3	20.5	14.1	10.1

4.3.1. Caso Multiclasse

4.3.1.1. Kernel Linear

A melhor precisão global obtida para o Kernel Linear foi de 48.0% para $C = 10^{-4}$. Este resultado é inútil para a classificação uma vez que o algoritmo classifica tudo como classe 3. A matriz de confusão global apresentada na tabela 4.14 tem a mesma interpretação da tabela 4.7.

² Distância Total

Tabela 4.14 - Matriz de confusão global para o caso multiclasse com as *features* Distância e Kernel Linear.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	3	15 (15.3%)	29 (29.6%)	47 (48.0%)	7 (7.1%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

4.3.1.2. Kernel RBF

Neste caso, o RBF apresenta uma ligeira melhoria em relação ao caso linear, com um precisão global de 49.0% para $C = 10^{-4}$ e $\gamma = 50$. Mesmo assim este resultado continua sendo inútil para a classificação pelo mesmo motivo reportado no 4.3.1.1. Para a tabela 4.15 é válida a mesma interpretação da tabela 4.18.

Tabela 4.15 - Matriz de confusão global para o caso multiclasse com as *features* Distância e Kernel RBF.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	4 (4.1%)	5 (5.1%)	4 (4.1%)	0 (0.0%)
	3	11 (11.2%)	24 (24.5%)	42 (42.9%)	6 (6.1%)
	4	0 (0.0%)	0 (0.0%)	1 (1.0%)	1 (1.0%)

4.3.1.3. Equilíbrio da dimensão dos dados por classes

Para este caso foi considerada a mesma análise do 4.2.1.3 replicando também o mesmo processo. Neste caso foram mantidos os elementos das classes 1 e 4, tendo extraída uma amostra aleatória de 15 elementos na classe 2 e 15 na classe 3. Assim, para os novos experimentos consideramos uma nova amostra com um total de 52 elementos.

4.3.1.3.1. Kernel Linear

Para o Kernel Linear a melhor precisão global obtida foi de 32.7% com $C = 1$. Embora com as dimensões das classes equilibradas a precisão é no mínimo 25.0% e o algoritmo conseguiu uma precisão de 32.7%, este resultado é péssimo para a classificação. A tabela 4.16 mostra que o algoritmo classificou a maioria dos elementos como sendo da classe 2. As restantes classes são interpretadas de forma análoga às da tabela 4.8.

Tabela 4.16 - Matriz de confusão global para o caso multiclasse com as *features* Distância e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	4 (7.7%)	3 (5.8%)	5 (9.6%)	0 (0.0%)
	2	6 (11.5%)	8 (15.4%)	4 (7.7%)	3 (5.8%)
	3	5 (9.6%)	3 (3.8%)	4 (7.7%)	3 (5.8%)
	4	0 (0.0%)	1 (1.9%)	2 (3.9%)	1 (1.9%)

4.3.1.3.2. Kernel RBF

A melhor precisão global obtida para o Kernel RBF foi de 50.0% para os valores de $C = 10^2$ e $\gamma = 10^{-3}$. Uma vez que o algoritmo conseguiu o dobro da precisão mínima esperada, conforme referido no ponto 4.3.2.1, este tem uma certa relevância para a classificação mas não é um resultado ótimo para o efeito. A tabela 4.17 é interpretada de forma análoga à 4.9.

Tabela 4.17 - Matriz de confusão global para o caso multiclasse com as *features* Distância e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	9 (7.7%)	2 (3.9%)	2 (3.9%)	0 (0.0%)
	2	3 (5.8%)	8 (15.4%)	4 (7.7%)	1 (1.9%)
	3	3 (5.8%)	3 (5.8%)	7 (13.5%)	4 (7.7%)
	4	0 (0.0%)	2 (3.9%)	2 (3.9%)	2 (3.9%)

4.3.1.3.3. Caso binário

O mesmo objetivo referido no ponto 4.2.1.3.3 é também válida para este caso. A nova amostra obtida no 4.3.2 encontra-se distribuída da seguinte forma para as duas classes, A e B: 30 elementos para a classe A que corresponde a uma proporção de 57.7% do total, 22 para a classe B que corresponde a uma proporção de 42.3% do total.

4.3.1.3.3.1. Kernel Linear

Assim sendo, a melhor precisão global encontrada para o caso linear foi de 69.2% para os valores de $C = 10^4$. Este resultado tem alguma relevância uma vez que a precisão mínima esperada é de 50.0%. A mesma interpretação da tabela 4.11 é válida para a 4.18.

Tabela 4.18 - Matriz de confusão global para o caso binário com as *features* Distâncias e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	24 (46.2%)	10 (19.2%)
	B	6 (11.5%)	12 (23.1%)

4.3.1.3.3.2. Kernel RBF

A melhor precisão global obtida para o Kernel RBF foi de 80.8% para os valores de $C = 10^4$ e $\gamma = 10^{-4}$, escolhidos por *cross validation*. É um resultado ótimo para a classificação considerando a relação de proporção no ponto 4.3.1.3.3 e a proporção mínima esperada referida no ponto 4.3.1.3.3.1. A interpretação da tabela 4.11 é também válida para a tabela 4.19.

Tabela 4.19 - Matriz de confusão global para o caso binário com as *features* Distâncias e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	25 (48.1%)	5 (9.6%)
	B	5 (9.6%)	17 (32.7%)

4.4. Classificação com as *features* Amplitude

O conjunto de *features* Amplitude tem um total de 12 *features* que passam a ser apresentadas: Ampl_AP_11, Ampl_AP_12, Ampl_AP_13, Ampl_AP_21, Ampl_AP_22, Ampl_AP_23³, Ampl_ML_11, Ampl_ML_12, Ampl_ML_13, Ampl_ML_21, Ampl_ML_22, Ampl_ML_23⁴. Na tabela 4.20 podemos verificar que a *feature* Ampl_ML_23 é distinta das restantes.

³ Amplitude Ântero Posterior

⁴ Amplitude Médio Lateral

Tabela 4.20 - Estatísticas de grupo para as *features* Amplitude, por classes.

	Estádios H&Y									
	1		2		3		4		Total	
	média	dp	média	dp	média	dp	média	dp	média	dp
Ampl_AP_11	12.8	5.0	12.1	8.2	10.8	6.4	19.4	9.7	12.1	7.3
Ampl_AP_12	10.8	4.4	9.2	5.3	10.5	6.4	21.3	13.1	11.0	7.0
Ampl_AP_13	10.9	7.5	9.2	5.3	11.2	8.3	19.3	11.7	11.1	7.9
Ampl_AP_21	13.7	11.7	10.2	6.8	13.0	8.5	21.7	14.4	12.9	9.4
Ampl_AP_22	10.9	7.5	8.8	5.2	11.2	7.1	17.4	8.7	10.9	7.0
Ampl_AP_23	15.2	4.8	14.7	8.3	15.6	8.6	22.4	16.1	15.8	8.8
Ampl_ML_11	5.6	3.8	5.0	5.0	5.5	7.6	18.5	13.4	6.3	7.7
Ampl_ML_12	4.1	2.7	3.7	3.0	5.0	4.3	17.3	17.6	5.4	6.6
Ampl_ML_13	4.1	2.0	3.8	3.3	5.5	7.8	17.1	23.8	5.6	8.9
Ampl_ML_21	5.3	5.5	4.0	3.3	7.3	10.0	13.3	8.4	6.5	8.1
Ampl_ML_22	4.9	7.1	2.8	2.3	5.6	10.9	13.2	10.8	5.2	8.9
Ampl_ML_23	11.9	6.3	10.9	8.9	10.8	9.3	26.7	18.8	12.1	10.4

4.4.1. Caso Multiclasse

4.4.1.1. Kernel Linear

A melhor precisão global obtida para este ponto é a mesma obtida no ponto 4.3.1.1 para $C = 10^{-4}$, sendo também válida a mesma análise. A matriz de confusão global apresentada na tabela 4.21 tem a mesma interpretação da tabela 4.7.

Tabela 4.21 - Matriz de confusão global para o caso multiclasse com as *features* Amplitude e Kernel Linear.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	3	15 (15.3%)	29 (29.6%)	47 (48.0%)	7 (7.1%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

4.4.1.2. Kernel RBF

Para o RBF a melhor precisão global obtida também foi a mesma obtida no ponto 4.3.1.2 para os valores de $C = 10, \gamma = 10^{-4}$, sendo válida a mesma análise. Para a tabela 4.22 onde é ilustrada a matriz de confusão global para este ponto, é válida a mesma interpretação da tabela 4.9.

Tabela 4.22 - Matriz de confusão global para o caso multiclasse com as *features* Amplitude e Kernel RBF.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (4.1%)	1 (1.0%)	0 (0.0%)	1 (1.0%)
	3	14 (14.3%)	28 (28.6%)	46 (46.9%)	5 (5.1%)
	4	1 (1.0%)	0 (0.0%)	1 (1.0%)	1 (1.0%)

4.4.1.3. Equilíbrio das dimensões dos dados por classes

Para este caso também foi extraída uma nova amostra mantendo as duas mesmas classes do ponto 4.3.2, escolhendo 15 elementos nas outras restantes, isto é, em cada uma.

4.4.1.3.1. Kernel Linear

A melhor precisão global obtida foi de 30.8% para $C = 0.5$. Para este resultado é válida a mesma interpretação do resultado obtido no ponto 4.3.1.3.3.1. A interpretação da tabela 4.23 onde o algoritmo classifica a maioria do elementos como classe 3, é análoga a da tabela 4.7.

Tabela 4.23 - Matriz de confusão global para o caso multiclasse com as *features* Amplitude e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	3 (5.8%)	4 (7.7%)	3 (5.8%)	0 (0.0%)
	2	7 (13.5%)	5 (9.6%)	4 (7.7%)	2 (3.9%)
	3	4 (7.7%)	6 (11.5%)	7 (13.5%)	4 (7.7%)
	4	1 (1.9%)	0 (0.0%)	1 (1.9%)	1 (1.9%)

4.4.1.3.2. Kernel RBF

A melhor precisão obtida para o RBF foi de 36.5% para $C = 10^3$ e $\gamma = 10^{-3}$. Também é ainda válida para este resultado a mesma interpretação do obtido no ponto 4.3.2.1. A interpretação da tabela 4.24 é também análoga à da tabela 4.9, sendo que nesta o algoritmo classifica a maioria como classe 3.

Tabela 4.24 - Matriz de confusão global para o caso multiclasse com as *features* Amplitude e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	6 (11.5%)	1 (1.9%)	3 (5.8%)	0 (0.0%)
	2	4 (7.7%)	7 (13.5%)	5 (9.6%)	2 (3.9%)
	3	4 (7.7%)	4 (7.7%)	6 (11.5%)	5 (9.6%)
	4	1 (1.9%)	3 (5.8%)	1 (1.9%)	1 (1.9%)

4.4.1.3.3. Caso binário

4.4.1.3.3.1. Kernel Linear

Foi obtida uma precisão global de 65.4%, como sendo a melhor, com $C = 10^4$ para o Kernel Linear. Este resultado é interpretado de forma análoga ao obtido no 4.3.1.3.3.1. Para a tabela 4.25 é válida a mesma interpretação da tabela 4.11.

Tabela 4.25 - Matriz de confusão global para o caso binário com as *features* Amplitude e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	22 (42.3%)	10 (19.2%)
	B	8 (15.4%)	12 (23.1%)

4.4.1.3.3.2. Kernel RBF

A melhor precisão global obtida foi de 78.9% para $C = 10^3$ e $\gamma = 10^{-4}$. Considerando a proporção das duas classes, A e B, no ponto 4.3.1.3.3. e a interpretação do

resultado obtido no ponto 4.3.1.3.3.2, este resultado é ótimo para a classificação. É ainda válida para a tabela 4.26 a mesma interpretação da 4.11.

Tabela 4.26 - Matriz de confusão global para o caso binário com as *features* Amplitude e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Previsita	A	24 (46.2%)	5 (9.6%)
	B	6 (11.5%)	17 (32.7%)

Concluimos que com estas novas *features* o algoritmo não conseguiu bom performance para a classificação multiclasse. Os resultados melhoraram significativamente com as *features* Distâncias e Amplitude para o caso binário com Kernel RBF, com as dimensões dos dados equilibrados por classes.

4.5. Classificação com as *features* com maior diferença de médias entre as classes

Para tentar melhorar os resultados de classificação tentamos reduzir as *features* utilizadas, considerando em cada caso apenas as variáveis com maior diferença de médias entre as classes. Para o efeito foram considerados gráficos de extremos e quartis e teste de anova a um fator.

4.5.1. Caso Multiclasse com as *features* Distância

Para este caso, as *features* que apresentaram maior diferença de médias entre as classes são Dist_T_12 e Dist_T_21.

4.5.1.1. Kernel Linear

Para o Kernel Linear a melhor precisão global obtida foi de 49.0% para $C = 10^4$. De notar que este resultado é ligeiramente superior ao obtido no 4.3.1.1, mas é inútil para a

classificação uma vez que o algoritmo classifica quase todos os elementos como sendo da classe 3 conforme a apresentação da tabela 4.27.

Tabela 4.27 - Matriz de confusão global para o caso multiclasse com as *features* Distância com maior diferença de médias entre as classes e Kernel Linear.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	3	15 (15.3%)	29 (29.6%)	47 (48.0%)	6 (6.1%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.0%)

4.5.1.2. Kernel RBF

Para o Kernel RBF a melhor precisão global obtida foi de 50.0% com $C = 0.75$ e $\gamma = 10^{-1}$. Também este resultado é ligeiramente melhor que o obtido no ponto 4.3.1.2 mas é válida a mesma interpretação. De forma análoga, a mesma interpretação da tabela 4.8 é também válida para a tabela 4.28.

Tabela 4.28 - Matriz de confusão global para o caso multiclasse com as *features* Distância com maior diferença de médias entre as classes e Kernel RBF.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	3 (3.1%)	3 (3.1%)	1 (1.0%)	0 (0.0%)
	3	12 (12.2%)	26 (26.5%)	46 (46.9%)	7 (7.1%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

4.5.1.3. Equilíbrio das dimensões dos dados por classes

Para este ponto continua sendo válido o mesmo processo do ponto agora com as duas *features* do ponto 4.5.1.

4.5.1.3.1. Kernel Linear

Para o Kernel Linear foi obtida uma precisão global de 32.7%, como sendo a melhor, com $C = 10^4$. Este resultado tem a mesma interpretação do obtido no 4.3.1.3.1. Na tabela 4.29 verificamos que o algoritmo classifica a maioria dos elementos como classe 2, sendo válida a mesma analogia de interpretação da tabela 4.9.

Tabela 4.29 - Matriz de confusão global para o caso multiclasse com as *features* Distância com maior diferença de médias entre as classes e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	3 (5.8%)	5 (9.6%)	3 (5.8%)	0 (0.0%)
	2	8 (15.4%)	7 (13.5%)	8 (15.4%)	1 (1.9%)
	3	4 (7.7%)	3 (5.8%)	6 (11.5%)	4 (7.7%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.9%)

4.5.1.3.2. Kernel RBF

A melhor precisão global obtida para o RBF foi de 44.2%, com $C = 10$ e $\gamma = 10^{-1}$. Comparando este resultado com o obtido no 4.3.1.3.2, este tende a piorar. A mesma análise da tabela anterior é também válida para a tabela 4.30 sendo ainda válida a mesma interpretação da tabela 4.9.

Tabela 4.30 - Matriz de confusão global para o caso multiclasse com as *features* Distância com maior diferença de médias entre as classes e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Previsita	1	7 (13.5%)	3 (5.8%)	1 (1.9%)	0 (0.0%)
	2	7 (13.5%)	8 (15.4%)	5 (9.6%)	3 (5.8%)
	3	1 (1.9%)	4 (7.7%)	8 (15.4%)	4 (7.7%)
	4	0 (0.0%)	0 (0.0%)	1 (1.9%)	0 (0.0%)

4.5.1.3.3. Caso binário

4.5.1.3.3.1. Kernel Linear

A melhor precisão global obtida foi de 71.2% com $C = 10^2, 10^3, 10^4$. Este resultado comparado com o obtido no ponto 4.3.1.3.3.1 é ligeiramente melhor, sendo relevante para a classificação considerando o ponto 4.3.1.3.3. Pela apresentação da tabla 4.31 podemos verificar que o algoritmo classifica bem a maioria dos elementos na classe A, quando passa para B erra ao classificar a mior parte dos elementos desta classe como sendo da classe A.

Tabela 4.31 - Matriz de confusão global para o caso binário com as *features* Distância com maior diferença de médias entre as classes e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Previsita	A	27 (51.9%)	12 (23.1%)
	B	3 (5.8%)	10 (19.2%)

4.5.1.3.3.2. Kernel RBF

Para o referido modelo a melhor precisão global obtida foi de 73.1% para os valores de $C = 10^3$ e $\gamma = 10^{-3}$. Ao contrário da análise do resultado obtido no ponto antecedente, este resultado é menor que o obtido no ponto 4.3.1.3.3.2. Ainda assim é um resultado relevante para a classificação pelas mesmas razões apontadas no ponto 4.3.1.3. Para a tabela 4.32 é válida a mesma interpretação da tabela anterior.

Tabela 4.32 - Matriz de confusão global para o caso binário com as *features* Distância com maior diferença de médias entre as classes e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	28 (53.85%)	12 (23.08%)
	B	2 (3.85%)	10 (19.23%)

4.5.2. Classificação com as *features* Velocidade

As *features* em título encontram-se discriminadas na tabela 4.33. Neste caso consideramos as *features* com maior diferença de médias entre as classes são, Vel_M_T_11, Vel_M_T_12, Vel_M_T_13, Vel_M_T_23⁵, Vel_M_AP_13, Vel_M_AP_23⁶, Vel_M_ML_11, Vel_M_ML_12, Vel_M_ML_13 e Vel_M_ML_23⁷.

⁵ Velocidade Média Total

⁶ Melocidade Média Ântero Posterior

⁷ Velocidade Média Médio-Lateral

Tabela 4.33 - Estatísticas de grupo para as *features* Velocidade, por classe.

	Estádios H&Y									
	1		2		3		4		Total	
	média	dp	média	dp	média	dp	média	dp	média	dp
Vel_M_T_11	3.8	2.1	3.6	2.6	3.8	2.8	9.4	4.7	4.1	3.1
Vel_M_T_12	3.3	1.9	2.8	2.0	3.8	2.6	9.2	5.6	3.8	3.0
Vel_M_T_13	3.7	2.0	3.2	1.9	3.8	3.1	10.9	10.3	4.1	4.0
Vel_M_T_23	5.8	3.2	4.1	1.9	5.7	4.1	14.7	12.1	5.9	5.1
Vel_M_AP_13	3.3	1.8	2.9	1.6	3.3	2.3	7.4	5.0	3.4	2.6
Vel_M_AP_23	4.8	2.4	3.3	1.8	4.6	3.1	9.8	7.8	4.6	3.5
Vel_M_ML_11	1.3	.9	1.3	1.2	1.5	1.9	5.3	3.4	1.6	2.0
Vel_M_ML_12	1.0	.8	.9	.7	1.4	1.3	5.4	4.9	1.5	1.9
Vel_M_ML_13	1.2	.7	1.0	.7	1.5	1.8	6.6	8.5	1.7	2.9
Vel_M_ML_23	2.4	1.7	1.8	1.1	2.6	2.4	9.2	8.6	2.8	3.4

4.5.2.1. Caso binário para as dimensões equilibradas por classes

4.5.2.1.1. Kernel Linear

A melhor precisão global obtida para o Kernel Linear foi de 67.3% com $C = 50$. Replicamos a mesma interpretação do ponto 4.3.1.3.3.1 para este ponto. Para a interpretação da tabela 4.34 replicamos a mesma análise da tabela 4.31.

Tabela 4.34 - Matriz de confusão global para o caso binário com as *features* Velocidade com maior diferença de médias entre as classes e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	25 (48.1%)	12 (23.1%)
	B	5 (9.6%)	10 (19.2%)

4.5.2.1.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi a mesma obtida no ponto 4.5.1.3.3.2, ou seja, uma precisão global de 73.1% para os valores de $C = 10^4$ e $\gamma = 10^{-3}$, sendo válida a mesma interpretação do resultado obtido no referido ponto. A mesma interpretação da tabela 4.11 é também válida para a tabela 4.35.

Tabela 4.35 - Matriz de confusão global para o caso binário com as *features* Velocidade com maior diferença de médias entre as classes e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	25 (48.1%)	9 (17.3%)
	B	5 (9.6%)	13 (25.0%)

Destes novos experimentos podemos concluir que, de um modo geral não houve melhorias significativas nos resultados para a classificação multiclasse. Para o caso binário, houve melhoria significativa nos resultados com o Kernel RBF.

4.6. Classificação com a exclusão das *features* correlacionadas entre si

Averiguamos a correlação entre as *features*. O objetivo é verificar as fortemente correlacionadas entre si, com valor acima de 0.7, e excluí-las considerando apenas uma, por forma a verificar se melhora os resultados de classificação. Neste caso foram consideradas as *features* de marcha e as *features* Amplitude que apresentaram melhores resultados na classificação com duas classes.

4.6.1. Classificação com as *features* de marcha

Na matriz de correlação apresentada na tabela 4.36, verificamos que as *features* WA-SL e WA-SPEED apresentam uma correlação significativa.

Tabela 4.36 - Matriz de correlação entre as *features* de Marcha.

	WA_SW	WA_SL	WA_SPEED	WA_SIMETRIA
WA_SW	1.0	0.1	0.1	-0.2
WA_SL	0.1	1.0	0.8	-0.1
WA_SPEED	0.1	0.8	1.0	-0.2
WA_SIMETRIA	-0.2	-0.1	-0.2	1.0

4.6.1.1. Caso binário para as dimensões equilibradas por classes

Para novos experimentos foi excluída a *feature* WA-SPEED e as restantes foram utilizadas para a classificação. Assim, replicando a mesma análise do 4.2.1.3.3 seguem os resultados obtidos pelos modelos.

4.6.1.1.1. Kernel Linear

A melhor precisão global obtida para o caso linear foi de 55.3% para $C = 10^4$. Este resultado tende a piorar comparado com o obtido no ponto 4.2.1.3.3.1. Na tabela 4.37 o número de elementos classificados como classe B é o mesmo que o algoritmo erra ao classificar elementos desta classe como classe A. Na classe A este classifica bem a maioria dos elementos.

Tabela 4.37 - Matriz de confusão global para o caso binário com as *features* de marcha com exclusão das *features* correlacionadas entre si e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	22 (29.0%)	20 (26.3%)
	B	14 (18.4%)	20 (26.3%)

4.6.1.1.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi de 63.2% com $C = 10^2$ e $\gamma = 10^{-1}$. Este resultado é melhor que o obtido no ponto 4.2.1.3.3.2 sendo ainda válida a mesma interpretação. Na tabela 4.38 podemos verificar que o algoritmo classifica a maioria dos elementos da classe A como classe B. Na classe B acontece o contrário, ou seja, classifica bem a maioria dos elementos nesta classe.

Tabela 4.38 - Matriz de confusão global para o caso binário com as *features* de marcha com exclusão das *features* correlacionadas entre si e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	15 (19.7%)	7 (9.2%)
	B	21 (27.6%)	33 (43.4%)

4.6.2. Classificação com as *features* Amplitude

Na matriz de correlação apresentada na tabela 4.39 as *features* foram codificadas com letras alfabéticas de A à L, respetivamente, por forma a simplificar a sua apresentação. Verificamos que as *features* C, D, E, G, H, I e K (Ampl_AP_13, Ampl_AP_21, Ampl_AP_22, Ampl_AP_11, Ampl_AP_12, Ampl_AP_13 e Ampl_AP_22) estão correlacionadas entre si. Assim sendo, foram excluídas as D, E, G, H, I e K, tendo ficado com a C que é associada às restantes para a classificação. Assim foram consideradas as *features* A, B, C, F, J, e L (Ampl_AP_11, Ampl_AP_12, Ampl_AP_13, Ampl_AP_23, Ampl_AP_11 e Ampl_AP_23) para o efeito.

Tabela 4.39 - Matriz de correlação entre as *features* Amplitude.

	A	B	C	D	E	F	G	H	I	J	K	L
A	1.0	0.6	0.5	0.5	0.5	0.3	0.6	0.4	0.3	0.2	0.3	0.3
B	0.6	1.0	0.6	0.5	0.6	0.3	0.5	0.6	0.4	0.4	0.4	0.3
C	0.5	0.6	1.0	0.7	0.7	0.3	0.7	0.6	0.7	0.6	0.7	0.3
D	0.5	0.5	0.7	1.0	0.7	0.3	0.6	0.5	0.4	0.7	0.6	0.3
E	0.5	0.6	0.7	0.7	1.0	0.3	0.6	0.5	0.5	0.6	0.7	0.3
F	0.3	0.3	0.3	0.3	0.3	1.0	0.4	0.5	0.4	0.3	0.3	0.4
G	0.6	0.5	0.7	0.6	0.6	0.4	1.0	0.7	0.8	0.6	0.8	0.4
H	0.4	0.6	0.6	0.5	0.5	0.5	0.7	1.0	0.8	0.6	0.5	0.4
I	0.3	0.4	0.7	0.4	0.5	0.4	0.8	0.8	1.0	0.7	0.7	0.4
J	0.2	0.4	0.6	0.7	0.6	0.3	0.6	0.6	0.7	1.0	0.7	0.3
K	0.3	0.4	0.7	0.6	0.7	0.3	0.8	0.5	0.7	0.7	1.0	0.4
L	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.3	0.4	1.0

4.6.2.1. Caso binário para as dimensões equilibradas por classes

Tendo já sido realizado o equilíbrio das dimensões dos dados por classes no ponto 4.3.1.3, partimos para a aplicação dos modelos.

4.6.2.1.1. Kernel Linear

Para o Kernel Linear a melhor precisão global obtida foi de 67.3% com $C = 10^4$. Comparando este resultado com o obtido no 4.4.1.3.3.2, este tende a melhorar. Este resultado tem alguma relevancia para a classificação. Para a tabela 4.40 é válida a mesma interpretação da tabela 4.37.

Tabela 4.40 - Matriz de confusão global para o caso binário com as *features* Amplitude com exclusão das *features* correlacionadas entre si e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	24 (46.2%)	11 (21.2%)
	B	6 (11.5%)	11 (21.2%)

4.6.2.1.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi de 80.8% com $C = 10^2$ e $\gamma = 10^{-3}$. Houve uma melhoria significativa em relação ao resultado obtido no ponto 4.4.1.3.3.2. Este resultado é ótimo para a classificação considerando as mesmas razões apontadas no ponto 4.3.1.3.3.2. Para a tabela 4.41 é válida a mesma interpretação da tabela 4.11.

Tabela 4.41 - Matriz de confusão global para o caso binário com as *features* Amplitude com exclusão das *features* correlacionadas entre si e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	24 (46.2%)	4 (7.7%)
	B	6 (11.5%)	18 (34.6%)

Desta seção concluímos que os resultados obtidos com as *features* de marcha melhoraram para o caso binário com Kernel RBF, embora a precisão de classificação não é de muita relevância para a classificação. As *features* Amplitude no 4.6.2 apresentaram melhor performance com o Kernel RBF, tendo obtido uma precisão de classificação de 80.77%.

4.7. Classificação com a exclusão dos *outliers* em cada caso

Para testar a influência de *outliers* no modelo e testar se a sua exclusão melhora os resultados, foram realizados novos experimentos. Para tal foram excluídos, em cada caso, os valores mais extremos. Estes valores foram observados através dos gráficos de extremos e quartis (*Boxplot*). Foi ainda realizada uma análise de regressão linear multivariada onde foram analisadas as estatísticas de resíduos para cada caso.

4.7.1. Caso multiclasse para as *features* Distância com a exclusão dos *outliers*

Foram consideradas as *features* em título pelo mesmo motivo frisado no ponto 4.6. Foram excluídos 5 valores mais extremos tendo ficado com uma amostra de 93 elementos com 15,28,45,5 para as classes 1,2,3 e 4. Na figura 4.6 são apresentados os valores mais extremos em cada *feature* que foram excluídos.

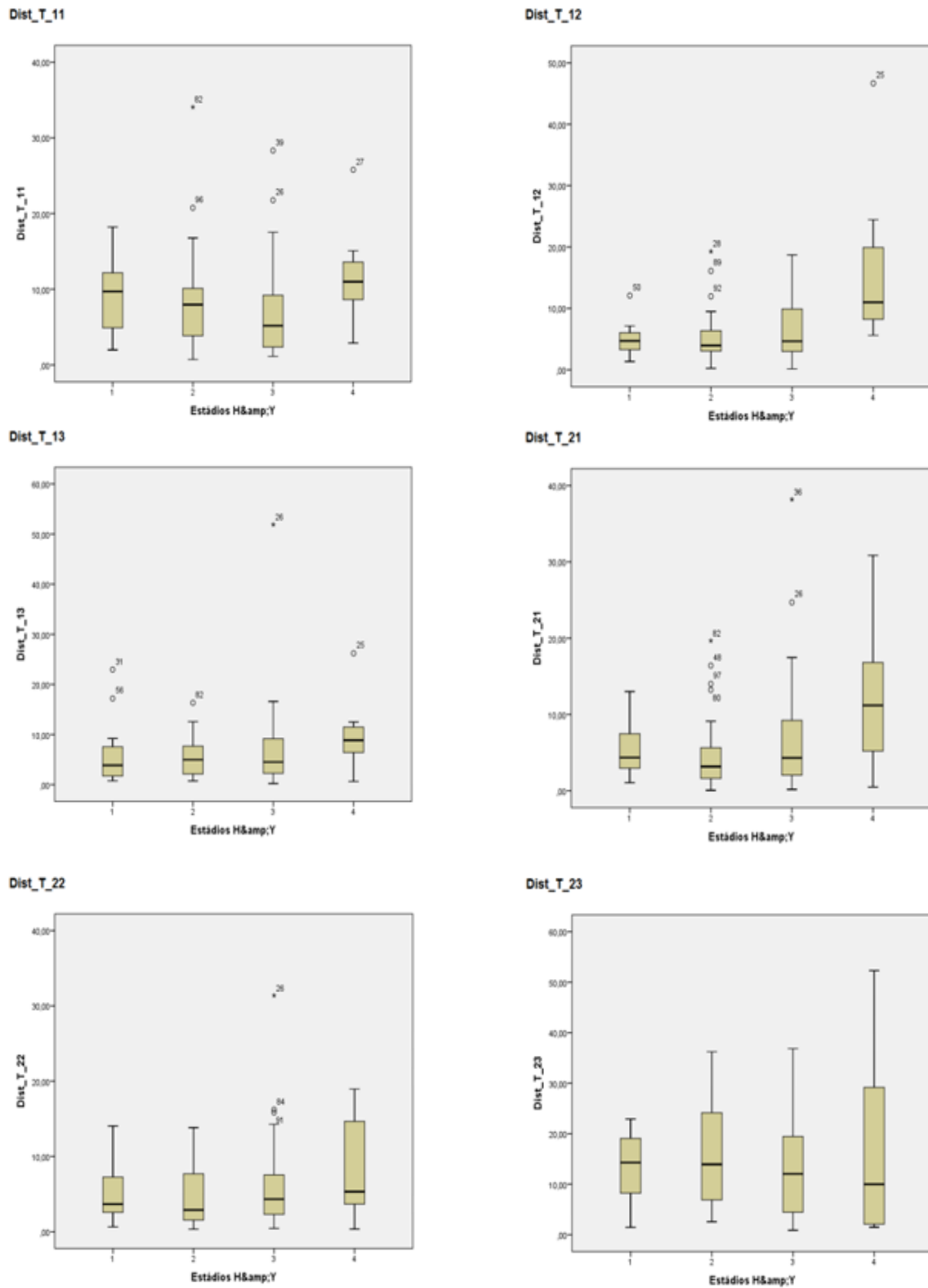


Figura 4.2 - Visualização gráfica dos *outliers* das *features* Distância

4.7.1.1. Kernel Linear

A melhor precisão global obtida para o Kernel Linear foi de 48.4% para os valores de $C = 10^{-4}, 10^{-3}$. Embora este resultado tenha apresentado uma ligeira melhoria em relação ao obtido no ponto 4.3.1.1, esta melhoria não é significativa para a classificação. Consideramos ainda válida, para este, a mesma interpretação do ponto 4.3.1.1. Para a tabela 4.42 é válida a mesma interpretação da tabela 4.7.

Tabela 4.42 - Matriz de confusão global para o caso multiclasse com as *features* Distância com a exclusão dos *outliers* e Kernel Linear.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	3	15 (16.1%)	28 (30.1%)	45 (48.4%)	5 (5.4%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

4.7.1.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi de 50.5% com $C = 10^4$ e $\gamma = 10^{-4}$. Este resultado, comparado com o obtido no ponto 4.3.1.2, é interpretado de forma análoga ao 4.7.1.1. Embora não seja um resultado tão relevante para a classificação, o algoritmo consegue acertar em todas as classes, conforme a tabela 4.43, cuja a interpretação é análoga a da tabela 4.9 sendo que nesta o algoritmo classifica a maioria dos elementos como classe 3.

Tabela 4.43 - Matriz de confusão global para o caso multiclasse com as *features* Distância com a exclusão dos *outliers* e Kernel RBF.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	5 (5.4%)	2 (2.2%)	3 (3.2%)	0 (0.0%)
	2	2 (2.2%)	7 (7.5%)	6 (6.5%)	1 (1.1%)
	3	7 (7.5%)	18 (19.4%)	35 (37.6%)	4 (4.3%)
	4	1 (1.1%)	1 (1.1%)	1 (1.1%)	0 (0.0%)

4.7.1.3. Equilíbrio das dimensões dos dados por classes

Para equilibrar as dimensões dos dados por classes foram mantidos os elementos das classes 1 e 4. Nas classes 2 e 3 foi extraída uma amostra aleatória de 15 elementos em cada uma. Assim, para novos experimentos foi considerada uma nova amostra com um total de 50 elementos.

4.7.1.3.1. Kernel Linear

A melhor precisão global obtida para o Kernel Linear foi de 40.0% com $C = 10$. Este resultado não é ótimo para a classificação. Para a tabela 4.44 são válidas as mesmas interpretações das tabelas 4.9 e 4.43.

Tabela 4.44 - Matriz de confusão global para o caso multiclasse com as *features* Distância com a exclusão dos *outliers* e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	6 (12.0%)	5 (10.0%)	2 (4.0%)	1 (2.0%)
	2	4 (8.0%)	5 (10.0%)	6 (12.0%)	0 (0.0%)
	3	5 (10.0%)	4 (8.00%)	7 (14.%)	2 (4.0%)
	4	0 (0.0%)	1 (2.0%)	0 (0.0%)	2 (4.0%)

4.7.1.3.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi 50.0% com $C = 1$ e $\gamma = 10^{-2}$. Embora este resultado seja melhor com o obtido no ponto 4.5.1.3.2, esta melhoria não satisfaz a classificação. A interpretação da tabela 4.45 é análoga a da tabela 4.9, sendo que o algoritmo classifica a maioria dos elementos como sendo da classe 3.

Tabela 4.45 - Matriz de confusão global para o caso multiclasse com as *features* Distância com a exclusão dos *outliers* e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real			
		1	2	3	4
Classificação Prevista	1	8 (16.0%)	2 (4.0%)	2 (4.0%)	0 (0.0%)
	2	2 (4.0%)	7 (14.0%)	4 (8.0%)	4 (8.0%)
	3	5 (10.0%)	6 (12.0%)	9 (18.0%)	1 (2.0%)
	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

4.7.1.2.3. Caso binário para as dimensões equilibradas por classes

Da amostra considerada no ponto 4.7.1.3, 30 elementos compõem a classe A que corresponde a uma proporção de 60.00% do total, sendo 20 a B que corresponde a uma proporção de 40.0% do total.

4.7.1.2.3.1. Kernel Linear

A melhor precisão global obtida para o Kernel Linear foi de 64.0% com $C = 10^4$. A precisão mínima esperada é de 50.0%. Assim sendo, o resultado obtido tem a mesma interpretação do 4.3.1.3.3.1. A mesma interpretação da tabela 4.37 é também válida para a tabela 4.46.

Tabela 4.46 - Matriz de confusão global para o caso binário com as *features* Distância com a exclusão dos outliers e Kernel Liner para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	22 (44.0%)	10 (20.0%)
	B	8 (16.0%)	10 (20.0%)

4.7.1.2.3.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi de 70.0% com $C = 10^4$ e $\gamma = 10^{-3}$. Com base no 4.7.1.2.3 e 4.7.1.2.3.1 este resultado é considerado relevante para a classificação. Para a tabela 4.47 é válida a mesma da tabela 4.11.

Tabela 4.47 - Matriz de confusão global para o caso binário com as *features* Distância com a exclusão dos *outliers* e Kernel RBF para as dimensões dos dados equilibrados por classes.

		Classificação Real	
		A	B
Classificação Prevista	A	22 (44.0%)	7 (14.0%)
	B	8 (16.0%)	13 (26.0%)

Neste contexto concluímos que os resultados obtidos para o caso multiclasse não são ótimos para a classificação. O caso binário para as dimensões equilibradas apresenta melhorias com o Kernel RBF.

4.8. Classificação de Discinesias do membro superior

Esta classificação foca o seu estudo em indivíduos diagnosticados com (C) ou sem (S) essa discinesia, onde C e S são os *labels* de classificação. Dos indivíduos considerados para o estudo conforme o ponto 4.1.2.1, 26 compõe a classe C e 72 compõe a S, que correspondem a uma relação de proporção de 26.5% e 73.5% do total, respetivamente.

4.8.1. Classificação com as *features* Distância

4.8.1.1. Kernel Linear

Para o Kernel Linear a melhor precisão global obtida foi de 76.5% com $C = 10^{-1}$. Este resultado não é ótimo para a classificação pois, à partida 70.0% dos dados já são classificados como classe S. Isto deve-se ao fato da classe S ter uma dimensão muito superior a classe C. Na tabela 4.48 verificamos que o modelo classifica mal a maioria dos elementos da classe C como sendo da classe S. Quando passa para a classe S classifica bem todos os elementos desta classe.

Tabela 4.48 - Matriz de confusão global para o caso binário de discinesias do membro superior com as *features* Distância e Kernel Linear.

		Classificação Real	
		C	S
Classificação Prevista	C	3 (3.1%)	0 (0.0%)
	S	23 (23.5%)	72 (73.5%)

4.8.1.2. Kernel RBF

Para o RBF a melhor precisão global obtida foi de 79.6% com $C = 10^3$ e $\gamma = 10^{-4}$. As mesmas interpretações do resultado e da tabela 4.48 no 4.8.1.1 são também válidas para este resultado e para a tabela 4.49.

Tabela 4.49 - Matriz de confusão global para o caso binário de discinesias do membro superior com as *features* Distância e kernel RBF.

		Classificação Real	
		C	S
Classificação Prevista	C	6 (5.1%)	0 (3.1%)
	S	20 (21.4%)	72 (70.4%)

4.8.1.3. Equilíbrio das dimensões dos dados por classes

Equilibramos as dimensões dos dados por classes sendo que neste caso foram consideradas apenas as duas classes referidas no ponto 4.8. Assim, mantendo todos os elementos da classe C foi extraída uma amostra de 26 elementos na classe S de forma aleatória. Assim, ficamos com uma nova amostra para os estudos, composto por um total 52 elementos e com a mesma proporção para as duas classes, ou seja, 50.0% para a classe C e 50.0% para a S.

4.8.1.3.1. Kernel Linear

A melhor precisão global obtida foi de 67.3% com $C = 10^{-1}$. Este resultado tem relevância para a classificação considerando a proporção entre as classes no 4.8.1.3. A tabela 4.50 é interpretada de forma análoga a tabela 4.37.

Tabela 4.50 - Matriz de confusão global para o caso binário de discinesias do membro superior com as *features* Distância e Kernel Linear para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		C	S
Classificação Prevista	C	13 (25.0%)	4 (7.7%)
	S	13 (25.0%)	22 (42.3%)

4.8.1.3.2. Kernel RBF

A melhor precisão global obtida para o RBF foi de 75.0% com $C = 10^3$ e $\gamma = 10^{-4}$. É um resultado ótimo para a classificação considerando as mesmas razões referidas no ponto 4.8.1.3.1. Para a tabela 4.51 é válida a mesma interpretação da tabela 4.11.

Tabela 4.51 - Matriz de confusão global para o caso binário de discinesias do membro superior com as *features* Distância e Kernel RBF para as dimensões equilibradas dos dados por classes.

		Classificação Real	
		C	S
Classificação Prevista	C	17 (32.7%)	4 (7.7%)
	S	9 (17.3%)	22 (42.3%)

Concluimos que com as dimensões equilibradas por classes o algoritmo consegue uma boa performance com o Kernel RBF.



CAPÍTULO 5

CONCLUSÕES RESULTANTES DESTE TRABALHO E PERSPETIVAS
FUTURAS.

5. Conclusões e Perspetivas futuras

Neste trabalho, o objetivo principal era determinar um modelo de classificação multiclasse através de SVM entre grupos de sujeitos com base em dados de comportamento motor de sujeitos de diferentes estadios da doença de Parkinson. Nesse sentido foram obtidos vários modelos SVM para vários conjuntos de dados como, o equilíbrio das dimensões das classes, as *features* com maior diferença de médias entre as classes, a exclusão dos *outliers* da amostra, a exclusão das *features* correlacionadas, o caso binário para as classes agrupadas duas a duas, e a classificação binária de descinesias do membro superior para os indivíduos diagnosticados com ou sem essa descinesia.

Sumarizando os resultados obtidos, a utilização das 4 *features* de marcha assim como os 5 conjuntos de *features* relacionados com a tarefa visual para a classificação nos estadios de Parkinson não têm poder discriminatório entre estadios. Em outras palavras, não foram suficientes para garantir uma boa classificação dos dados de comportamento motor nestes estadios, nem para o caso multiclasse nem para o caso binário para os dados de entradas iniciais, em que as dimensões entre classes não estavam equilibradas.

Os dados de entradas iniciais traduziram-se num grau de dificuldade para a classificação, pois a classe que apresentou maior número de elementos conduziu o algoritmo para uma performance tendenciosa. Este tende a classificar todos os elementos como sendo da referida classe, aumentando assim a dificuldade de deteção dos erros de classificação. Desta forma, uma sugestão aqui evidenciada foca-se no equilíbrio das dimensões dos dados por classes para tentar melhorar os resultados de classificação de modo a evitar que o algoritmo seja sensível às classes com maior número de elementos.

Esta sugestão ajudou o algoritmo a diminuir a referida tendência, ainda que esta tenha sido pouco atenuada na classificação multiclasse. Mesmo considerando o caso das *features* com maior diferença de médias entre as classes, a exclusão de *features* correlacionadas e a exclusão de *outliers* da amostra os resultados não tiveram melhorias significativas. Logo o objetivo inicial não foi concretizado.

Neste caso, a utilização de SVM multiclasse para a detecção de *features* com maior poder discriminatório entre grupos acaba por apresentar maior grau de dificuldade na construção de algoritmos que especificam quais é que são essas *features* para o efeito, pois classifica tudo como classe 3.

Os parâmetros utilizados para o kernel Linear e o RBF, identificados na grid search, por cross validation, apresentaram diferenças nas precisões dadas pelo modelo, sendo que, no geral, as apresentadas pelo RBF são melhores em quase todos os casos. Uma vez que essas precisões apresentadas pelo modelo não foram ótimas para a classificação, este requer maior número de vectores suporte, o que dificulta muito a classificação.

No entanto para o caso binário, com as classes iniciais agrupadas duas a duas e com as dimensões equilibradas por classes, o algoritmo apresenta melhor performance. De igual modo acontece com o caso binário de descinesias do membro superior. Concluímos que para o caso binário das dimensões equilibradas por classes, as features Distância, e Amplitude com a exclusão das features correlacionadas têm maior poder discriminatório entre grupos nos estadios de Parkinson, pois apresentaram melhor precisão de classificação com 80.8% com Kernel RBF.

Espera-se que esta dissertação seja uma ferramenta de estudo no sentido de proporcionar uma visão técnica dos SVM sobre esta temática.

De igual modo espera-se que a sua contribuição proporcione uma boa compreensão da referida técnica, servindo também de um estímulo aos futuros estudantes de ciência exata no sentido de, neles despertar o interesse em realizar pesquisas mais avançadas que permitam gerar resultados que retificam a sua aplicabilidade.

Este trabalho abre vertentes para ser complementado com trabalhos futuros, como por exemplo:

Utilizar dados de comportamento motor cujos rótulos de classificação não estejam muito associados entre si, por forma a facilitar a separabilidade destes rótulos e assim verificar se se consegue melhorar as precisões de classificação. Como visto nesta

dissertação, o fato de os rótulos estarem muito associados entre si traduz-se num problema de separabilidade destes bem como a obtenção de precisões ótimas para a classificação.

Investigar outros atributos relacionados com o tratamento de outros dados de comportamento motor normalizados, por forma a verificar se aumenta a precisão de classificação multiclasse.

Investigar técnicas SVM mais avançadas por forma a aprimorar as utilizadas nesta dissertação. Neste caso o objetivo era aplicar as técnicas SVM linear e não-linear para obter um classificador multiclasse. Mas pelos resultados obtidos estas técnicas não são as mais adequadas para o problema abordado.

Referências

- ARONSZAJN, N. (1950). Theory of Reproducing Kernels. *Transactions of the American Mathematical Society*, Volume 68, 337-404.
- BEGG, R. K., PALANISWAMI, M., MEMBER, S., & OWEN, B. (MAY de 2005). *Support Vector Machines for Automated Gait Classification*. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, VOL. 52, NO. 5. Fonte: <http://www.google.com.br/http://cbio.mines-paristech.fr/~jvert/svn/bibli/local/Begg2005Support.pdf>
- BOSER, E. B., GUYON, I. M., & VAPNIK, V. N. (1992). *A Training Algorithm for Optimal Margin Classifiers*. Fonte: <http://www.google.com.br/http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.3818>
- BURBIDGE, R., & BUXTON, B. (s.d.). *An Introduction to Support Vector Machines for Data Mining*. Fonte: <https://www.google.com.br/http://svms.org/tutorials/BurbidgeBuxton2001.pdf>
- BURGES, C. J. (1998). A Tutorial on Support Vector Machines for Pattern. *Kluwer Academic Publishers, Boston*, 121-167.
- CAMILO, C. O., & SILVA, J. C. (2009). *Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas*. Brasil: Instituto de Informática - Universidade Federal de Goiás.
- CORTES, C., & VAPNIK, V. (1995). *Support-Vector Networks*. *Machine Learning*, 20, 273-297. Editor: Lorenza Saitta. Fonte: www.google.com.br/http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf
- COVER, T. M. (1 de March de 1965). *GEOMETRICAL END STATISTICAL PROPERTIES OF SYSTEMS OF LINEAR INEQUALITIES WITH APPLICATIONS IN PATTERN RECOGNITION*. *IEEE TRANSACTIONS ON ELETRONIC COMPUTERS*. Fonte: <http://www.google.com.br/http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.366.5645&rep=rep1&type=pdf>
- FISHER, R. A. (1936). *THE USE OF THE MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS*. Fonte: <https://www.google.com.br/http://rcs.chemometrics.ru/Tutorials/classification/Fisher.pdf>
- FRANC, V., & HLAVÁČ, V. (August de 2002). *Multi-class Support Vector Machine*. Fonte: <http://www.google.com.br/http://cmp.felk.cvut.cz/ftp/articles/franc/franc-multiSVM02.pdf>
- GENTON, M. G. (2001). *Classes of Kernels for Machine Learning: A Statistics Perspective*. Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA . *Journal of Machine Learning Research* 2. Submitted 3/01;Published 12/01, 299-312.
- GODINHO, C. A. (2013). *ESTUDO DO COMPORTAMENTO POSTURAL EM DOENTES COM DOENÇA DE PARKINSON*. *Caracterização e Análise segundo Metodologias Tradicionais e Não Lineares*. *Dissertação apresentada com vista à obtenção do grau*

de Doutor no Ramo de Motricidade Humana. . Lisboa: Universidade Técnica de Lisboa. Faculdade de Motricidade Humana.

- GÖNEN, G., TANUĞUR, A., & ALPAYDIN, E. (JANUARY de 2008). *Multiclass Posterior Probability Support Vector Machines.* *IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 19, NO. 1. Senior Member, IEEE.* Fonte: <http://www.google.com.br>: http://www.cmpe.boun.edu.tr/~ethem/files/papers/mehmet_tnn.pdf
- GUYON, I., BOSER, B., & VAPNIK, V. (s.d.). *Automatic Capacity Tuning of Very Large VC-dimension Classifiers.* Fonte: <https://www.google.com.br>: <https://papers.nips.cc/paper/653-automatic-capacity-tuning-of-very-large-vc-dimension-classifiers.pdf>
- HAYKIN, S. (1999). *Neural Networks - A Comprehensive Foundation.* Prentice-Hall, New Jersey, Second Edition. An imprint of Pearson Education. Fonte: www.google.com.br: [https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural Networks](https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural_Networks)
- HERBRICH, R. (2002). *LEARNING KERNEL CLASSIFIERS. THEORY AND ALGORITHMS.* Massachusetts: Massachusetts Institute of Technology.
- HOEHN, M. M., & YAHR, M. D. (2012). *Chapter 2. Parkinson's Disease: Hoehn and Yahr Scale. Background.* Fonte: <http://www.google.com.br>: [file:///C:/Users/Djay/Downloads/9781603274258-c1%20\(1\).pdf](file:///C:/Users/Djay/Downloads/9781603274258-c1%20(1).pdf)
- HOFMANN, M. (26 de June de 2006). *Support Vector Machines — Kernels and the Kernel Trick.* Fonte: www.google.com.br: www.cogsys.wiai.uni-bamberg.de/teaching/ss06/hs_svm/slides/SVM_Seminarbericht_Hofmann.pdf
- HSU, C.-W., CHANG, C.-C., & LIN, C.-J. (15 de April de 2010). *A Practical Guide to Support Vector Classification.* Fonte: <https://www.google.com.br>: <https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>
- JORGENSEN, M., RATHLEFF, M., LAESSOE, U., CASEROTTI, P., NIELSEN, O., & AAGAARD, P. (2012). *Time-of-day influences postural balance in older adults.* 35 (2012) 653–657. Fonte: www.google.com.br: https://www.researchgate.net/publication/221888615_Time-of-day_influences_postural_balance_in_older_adults
- LIMA, B. M. (s.d.). *Parkinson: sintomas, tratamentos e causas.* Fonte: www.google.com.br: <http://www.minhavidade.com.br/saude/temas/parkinson>
- LORENA, A. C., & CARVALHO, A. C. (2003). *Introdução às Máquinas de Vetores Suporte (Support Vector Machines).* ISSN - 0103-2569.No. 192. São Carlos: Instituto de Ciências Matemáticas e de Computação.
- LORENA, A., & CARVALHO, A. (2007). *Uma Introdução às Support Vector Machines.* RITA XIV (2). Fonte: <https://www.google.com.br>: [file:///C:/Users/Djay/Downloads/5690-18885-1-PB%20\(1\).pdf](file:///C:/Users/Djay/Downloads/5690-18885-1-PB%20(1).pdf)
- MÜLLER, K.-R., MIKA, S., RÄTSCH, G., TSUDA, K., & SCHÖLKOPF, B. (MARCH de 2001). *An Introduction to Kernel-Based Learning Algorithms.* *IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 12.* Fonte:

- <https://www.google.cv>:
<http://media.cs.tsinghua.edu.cn/~taopin/ML2005/intro2Kernel-basedLearn-TNN-2001.pdf>
- PAL, M. (s.d.). *Multiclass Approaches for Support Vector Machine Based Land Cover Classification. Lecturer, Department of Civil engineering, National Institute of Technology. Kurukshetra, 136119. Haryana (INDIA). mpce_pal@yahoo.co.uk.*
Fonte: <http://www.google.com.br>:
<https://arxiv.org/ftp/arxiv/papers/0802/0802.2411.pdf>
- PELLUCCI, P. R., DE PAULA, R. R., SILVA, W. B., & LADEIRA, A. P. (2011). *UTILIZAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NO RECONHECIMENTO DE ENTIDADES NOMEADAS NO PORTUGUÊS. e-xacta, Belo Horizonte, v. 4, n. 1, p. 73-81. Editora UniBH. Disponível em: www.unibh.br/revistas/exacta/.* Fonte: www.google.com.br:
<http://revistas2.unibh.br/index.php/dcet/article/viewFile/305/164>
- PEREIRA, D., & GARRETT, C. (2010). *FACTORES DE RISCO DA DOENÇA DE PARKINSON Um Estudo Epidemiológico*. ARTIGO ORIGINAL. Acta Med Port; 23: 015-024.* Fonte: www.google.com.pt:
<http://actamedicaportuguesa.com/revista/index.php/amp/article/viewFile/599/283>
- RAPAPORT, M. F. (s.d.). *Introduction de la connaissance a priori dans l'étude des puces à ADN. THESE DE DOCTORAT DE L'UNIVERSITE PIERRE ET MARIE CURIE. Spécialité Biochimie et Biologie Moléculaire (Ecole doctorale). Institut Formation Doctorale. Paris - 6: Université Pierre & Marie Curie. Bureau d'accueil, inscription des doctorants et base de données. Esc G, 2ème étage, 15 rue de l'école de médecine.*
- SANTOS, E. M. (junho de 2002). *Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência.* Campina Grande, Paraíba, Brasil, Brasil: Universidade Federal da Paraíba, Campus II.
- SEO, N. (6 de December de 2007). *A Comparison of Multi-class Support Vector Machine Methods for Face Recognition. Department of Electrical and Computer Engineering. The University of Maryland. sonots@umd.edu.* Fonte: <http://www.google.com.br>:
<http://note.sonots.com/?plugin=attach&refer=SciSoftware%2FMSVM&openfile=report.pdf>
- SÉRGIO, M. G. (2009). *Adaptação Preditiva de Conteúdo Multimédia.* Porto-Portugal: FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO.
- SEWELL, M. (2008). *Structural Risk Minimization. University College London.*
- SILVA, L. M. (2014). *Máquinas de Vetores Suporte para Classificação do Onset em dados Temporais de Eletromiografia.* Lisboa: UNIVERSIDADE ABERTA - DEPARTAMENTO DE CIÊNCIAS E TECNOLOGIA.
- SMOLA, A., & VISHWANATHAN, S. V. (2008). *INTRODUCTION TO MACHINE LEARNING. Cambridge University Press.* Fonte: www.google.com.br:
<http://alex.smola.org/drafts/thebook.pdf>
- STEINWART, I., & CHRISTMANN, A. (2008). *Information Science and Statistics - Support Vector Machines.* USA: Springer.

- THOMÉ, A. C. (2012). *SVM Classifiers – Concepts and Applications to Character Recognition*. Fonte: <https://www.google.com.br>: <http://dx.doi.org/10.5772/52009>
- VAPNIK, V. N. (1998). *Statistical Learning Theory*. Read Bank, New Jersey: Wiley.
- VAPNIK, V. N. (SEPTEMBER de 1999). *An Overview of Statistical Learning Theory*. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 10(5), 988-999. Fonte: <http://www.google.com.br/>: <http://math.arizona.edu/~hzhang/math574m/vapnik.pdf>
- VAPNIK, V. N. (August de 1999). *The Nature of Statistical Learning Theory. Second Edition with 50 Illustrations*. New York. Red Bank, New Jersey: Springer-Verlag. Fonte: <http://down.cenet.org.cn/upfile/28/20081111224649178.pdf>
- VAPNIK, V. N., & LERNER, A. Y. (26 de December de 1962). *PATTERN RECOGNITION USING GENERALIZED PORTRAITS*. *Translated from Avtamarika I Telerckllanjka. Vol. 21. No. 6. pp. 774-780, Jun, 1963*. Fonte: <http://www.google.com.br/>: <http://web.cs.iastate.edu/~cs573x/vapnik-portraits1963.pdf>
- WEHBE, L. (1 de October de 2013). *Kernel Properties - Convexity*. Fonte: www.google.com.br: http://alex.smola.org/teaching/cmu2013-10-701x/slides/recitation_3_kernel_properties_convexity.pdf
- ZHANG, T. (2001). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. A Review. Book Reviews. Volume 22 Number 2 . *AI Magazine*, 183-184.



Anexos

Escala de Hoehn & Yhar

ESCALA DE HOEHN E YAHR MODIFICADA		
NOME:	Sexo:	Prontuário:
	Idade:	Data da Lesão:
Lado Dominante ou parético: (D) (E)		Data da Avaliação:
Diagnóstico:		Avaliador:

Estágio	Descrição
0	Nenhum sinal da doença
1	Doença unilateral
1,5	Envolvimento unilateral e axial.
2	Doença bilateral sem déficit de equilíbrio (recupera o equilíbrio dando três passos para trás ou menos).
2,5	Doença bilateral leve, com recuperação no "teste do empurrão"(empurra-se bruscamente o paciente para trás a partir dos ombros, o paciente dá mais que três passos, mas recupera o equilíbrio sem ajuda).
3	Doença bilateral leve a moderada; alguma instabilidade postural; capacidade para viver independente.
4	Incapacidade grave, ainda capaz de caminhar ou permanecer de pé sem ajuda.
5	Confinado à cama ou cadeira de rodas a não ser que receba ajuda.

* Sugere-se iniciar o teste do item 5 para o item 1.

Classificação

Estágios 1 a 3 = incapacidades leve a moderada

Estágios 4 e 5 = incapacidade grave.

FONTE: SHENKMAN M. L.; CLARK K.; XIE T.; KUCHIBHATLA M.; SHINBERG M.; RAY L.; Spinal movement and performance of standing reach task in participants with and without Parkinson disease. Phys Ther, vol. 81, p. 1400-1411, 2001.

Sistema Balance Master®



Componentes do Sistema

O sistema Balance Master® é fabricado pela Neurocom® International Inc., Clackamas, OR (E.U.A.) e é constituído pelos seguintes componentes:

- Dupla plataforma de forças (46 cm X 152 cm)
- Suportes circulares para plataforma
- Computador Pentium com 16 MB RAM e disco rígido de 1 GB
- Placa gráfica SVGA
- Projector de parede Samsung
- Software Balance Master System Versão 6.1 (para Windows)
- Impressora de jacto tinta
- Rato sem fios, tipo “trackball”
- Transformador eléctrico

Bateria de testes

- WBS – Weight Bearing Squat
- mCTSIB – modified Clinical Test of Sensory Interaction on Balance
- US – Unilateral Stance
- LOS – Limits of Stability
- RWS – Rhythmic Weight Shift
- STS – Sit-to-Stand
- WA – Walk Across
- TW – Tandem Walk
- SQT – Step/Quick Turn
- SUO – Step Up / Over
- FL – Forward Lunge

(GODINHO, 2013)

Outputs

Teste de K-S para a normalidade das amostras das de marcha

	N	Parâmetros normais ^{a,b}		Diferenças Mais Extremas			Estatística do teste	Significância Assint. (Bilateral)	Sig exata (bilateral)	Probabilidade de ponto
		Média	Desvio Padrão	Absoluto	Positivo	Negativo				
WA_SW	161	16,5863	3,79225	,048	,040	-,048	,048	,200 ^{c,d}	,834	,000
WA_SL	161	40,7329	16,46898	,055	,046	-,055	,055	,200 ^{c,d}	,696	,000
WA_SPEED	161	45,7845	20,31621	,065	,065	-,040	,065	,091 ^c	,480	,000
WA_SIMETRIA	161	19,75	17,913	,144	,144	-,135	,144	,000 ^c	,002	,000

a. A distribuição do teste é Normal.

b. Calculado dos dados.

c. Correção de Significância de Lilliefors.

d. Este é um limite inferior da significância verdadeira.

Teste de ANOVA para igualdade de médias de altura por estadio para as features de marcha

	Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Entre Grupos	35,483	3	11,828	,115	,951
Nos grupos	16115,361	157	102,646		
Total	16150,845	160			

Teste de ANOVA para igualdade de médias de idade por estadio para as features de marcha

	Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Entre Grupos	401,954	3	133,985	1,567	,200
Nos grupos	13423,611	157	85,501		
Total	13825,565	160			

Teste de K-S de uma amostra para a normalidade das features Distância

	N	Parâmetros normais ^{a,b}		Diferenças Mais Extremas			Estatística do teste	Significância Assint. (Bilateral)	Sig exata (bilateral)	Probabilidade de ponto
		Média	Desvio Padrão	Absoluto	Positivo	Negativo				
		Dist_T_11	98	7,9810	6,12691	,133				
Dist_T_12	98	6,7547	6,41846	,185	,185	-,151	,185	,000 ^c	,002	,000
Dist_T_13	98	6,5570	6,83152	,177	,139	-,177	,177	,000 ^c	,004	,000
Dist_T_21	98	6,4045	6,51169	,176	,176	-,165	,176	,000 ^c	,004	,000
Dist_T_22	98	5,6290	5,11229	,182	,182	-,150	,182	,000 ^c	,003	,000
Dist_T_23	98	14,1174	10,14245	,107	,107	-,096	,107	,007 ^c	,194	,000

a. A distribuição do teste é Normal.

b. Calculado dos dados.

c. Correção de Significância de Lilliefors.

Teste de K-S de uma amostra para a normalidade das features Amplitude

	N	Parâmetros normais ^{a,b}		Diferenças Mais Extremas			Estatística do teste	Significância Assint. (Bilateral)	Sig exata (bilateral)	Probabilidade de ponto
		Média	Desvio Padrão	Absoluto	Positivo	Negativo				
		Ampl_AP_11	98	12,0715	7,26367	,118				
Ampl_AP_12	98	10,9566	7,03414	,158	,158	-,088	,158	,000 ^c	,013	,000
Ampl_AP_13	98	11,1205	7,92297	,173	,173	-,118	,173	,000 ^c	,005	,000
Ampl_AP_21	98	12,9023	9,37703	,181	,181	-,126	,181	,000 ^c	,003	,000
Ampl_AP_22	98	10,8926	6,99219	,087	,087	-,076	,087	,067 ^c	,429	,000
Ampl_AP_23	98	15,7452	8,82735	,078	,078	-,061	,078	,164 ^c	,570	,000
Ampl_ML_11	98	6,2940	7,67646	,223	,222	-,223	,223	,000 ^c	,000	,000
Ampl_ML_12	98	5,3611	6,56082	,220	,220	-,210	,220	,000 ^c	,000	,000
Ampl_ML_13	98	5,6332	8,84978	,274	,272	-,274	,274	,000 ^c	,000	,000
Ampl_ML_21	98	6,4498	8,05119	,246	,246	-,228	,246	,000 ^c	,000	,000
Ampl_ML_22	98	5,2002	8,89069	,288	,272	-,288	,288	,000 ^c	,000	,000
Ampl_ML_23	98	12,1383	10,41306	,141	,138	-,141	,141	,000 ^c	,036	,000

a. A distribuição do teste é Normal.

b. Calculado dos dados.

c. Correção de Significância de Lilliefors.

Teste de K-S de uma amostra para a normalidade das features Área

	N	Parâmetros normais ^{a,b}		Diferenças Mais Extremas			Estatística do teste	Significância Assint. (Bilateral)	Sig exata (bilateral)	Probabilidade de ponto
		Média	Desvio Padrão	Absoluto	Positivo	Negativo				
Area_Elip_11	98	100,7275	196,56363	,304	,293	-,304	,304	,000 ^c	,000	,000
Area_Elip_12	98	80,3663	174,31252	,322	,305	-,322	,322	,000 ^c	,000	,000
Area_Elip_13	98	90,4236	331,13513	,392	,390	-,392	,392	,000 ^c	,000	,000
Area_Elip_21	98	123,9506	285,96422	,332	,309	-,332	,332	,000 ^c	,000	,000
Area_Elip_22	98	98,8633	323,39946	,380	,346	-,380	,380	,000 ^c	,000	,000
Area_Elip_23	98	107,6775	249,20485	,338	,338	-,333	,338	,000 ^c	,000	,000

a. A distribuição do teste é Normal.

b. Calculado dos dados.

c. Correção de Significância de Lilliefors.

Teste de K - S de uma amostra para a normalidade das features Velocidade

	N	Parâmetros normais ^{a,b}		Diferenças Mais Extremas			Estatística do teste	Significância Assint. (Bilateral)	Sig exata (bilateral)	Probabilidade de ponto
		Média	Desvio Padrão	Absoluto	Positivo	Negativo				
Vel_M_T_11	98	4,1178	3,14057	,179	,179	-,112	,179	,000 ^c	,003	,000
Vel_M_T_12	98	3,7907	3,03056	,172	,172	-,114	,172	,000 ^c	,005	,000
Vel_M_T_13	98	4,1216	4,03608	,215	,215	-,178	,215	,000 ^c	,000	,000
Vel_M_T_21	98	5,3835	4,45142	,161	,161	-,126	,161	,000 ^c	,011	,000
Vel_M_T_22	98	4,6013	4,20400	,176	,176	-,145	,176	,000 ^c	,004	,000
Vel_M_T_23	98	5,9065	5,10369	,200	,200	-,154	,200	,000 ^c	,001	,000
Vel_M_AP_11	98	3,4781	2,22705	,158	,158	-,076	,158	,000 ^c	,013	,000
Vel_M_AP_12	98	3,2025	2,15476	,129	,129	-,088	,129	,000 ^c	,071	,000
Vel_M_AP_13	98	3,4374	2,55792	,159	,159	-,115	,159	,000 ^c	,013	,000
Vel_M_AP_21	98	4,6036	3,57858	,155	,155	-,113	,155	,000 ^c	,016	,000
Vel_M_AP_22	98	3,9411	3,13351	,119	,119	-,110	,119	,002 ^c	,115	,000
Vel_M_AP_23	98	4,6453	3,53976	,129	,129	-,120	,129	,000 ^c	,069	,000
Vel_M_ML_11	98	1,6509	1,99749	,241	,241	-,214	,241	,000 ^c	,000	,000
Vel_M_ML_12	98	1,5017	1,93002	,246	,246	-,220	,246	,000 ^c	,000	,000
Vel_M_ML_13	98	1,6878	2,85676	,283	,268	-,283	,283	,000 ^c	,000	,000
Vel_M_ML_21	98	2,0325	2,23428	,229	,229	-,188	,229	,000 ^c	,000	,000
Vel_M_ML_22	98	1,7276	2,40522	,264	,264	-,242	,264	,000 ^c	,000	,000
Vel_M_ML_23	98	2,7939	3,37278	,238	,238	-,215	,238	,000 ^c	,000	,000

a. A distribuição do teste é Normal.

b. Calculado dos dados.

c. Correção de Significância de Lilliefors.

Teste de K-S de uma amostra para a normalidade das features Frequência

	N	Parâmetros normais ^{a,b}		Diferenças Mais Extremas			Estatística do teste	Significância Assint. (Bilateral)	Sig exata (bilateral)	Probabilidade de ponto
		Média	Desvio Padrão	Absoluto	Positivo	Negativo				
Freq_Osc_AP_11	98	,3131	,06956	,084	,084	-,073	,084	,083 ^c	,465	,000
Freq_Osc_AP_12	98	,2987	,07081	,081	,081	-,069	,081	,110 ^c	,509	,000
Freq_Osc_AP_13	98	,3094	,08284	,091	,091	-,061	,091	,046 ^c	,376	,000
Freq_Osc_AP_21	98	,3378	,09615	,152	,152	-,064	,152	,000 ^c	,019	,000
Freq_Osc_AP_22	98	,3264	,09281	,064	,060	-,064	,064	,200 ^{c,d}	,786	,000
Freq_Osc_AP_23	98	,3314	,07995	,076	,076	-,047	,076	,182 ^c	,588	,000
Freq_Osc_ML_11	98	,3317	,10261	,138	,118	-,138	,138	,000 ^c	,043	,000
Freq_Osc_ML_12	98	,3386	,09739	,121	,121	-,084	,121	,001 ^c	,105	,000
Freq_Osc_ML_13	98	,3471	,09442	,066	,066	-,050	,066	,200 ^{c,d}	,766	,000
Freq_Osc_ML_21	98	,3524	,09724	,078	,078	-,054	,078	,161 ^c	,567	,000
Freq_Osc_ML_22	98	,3568	,10736	,048	,037	-,048	,048	,200 ^{c,d}	,970	,000
Freq_Osc_ML_23	98	,3495	,10046	,093	,093	-,071	,093	,036 ^c	,342	,000

a. A distribuição do teste é Normal.

b. Calculado dos dados.

c. Correção de Significância de Lilliefors.

d. Este é um limite inferior da significância verdadeira.

Teste de ANOVA para igualdade de médias de altura por estadio para as features posturais

	Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Entre Grupos	229,454	3	76,485	,983	,404
Nos grupos	7314,383	94	77,813		
Total	7543,837	97			

Teste de ANOVA para igualdade de médias de idade por estadio para as fetures posturais

	Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Entre Grupos	162,713	3	54,238	,841	,475
Nos grupos	6062,838	94	64,498		
Total	6225,551	97			

Tabela de ANOVA para as features Distância

				Soma dos	gl	Quadrado	F	Sig.	
				Quadrados		Médio			
Dist_T_11	*	Estádios	Entre Grupos	(Combinado)	196,972	3	65,657	1,792	,154
			Nos grupos		3444,319	94	36,642		
			Total		3641,291	97			
H&Y			Entre Grupos	(Combinado)	867,017	3	289,006	8,682	,000
			Nos grupos		3129,062	94	33,288		
			Total		3996,079	97			
Dist_T_12	*	Estádios	Entre Grupos	(Combinado)	119,859	3	39,953	,852	,469
			Nos grupos		4407,098	94	46,884		
			Total		4526,958	97			
H&Y			Entre Grupos	(Combinado)	334,128	3	111,376	2,770	,046
			Nos grupos		3778,871	94	40,201		
			Total		4112,999	97			
Dist_T_21	*	Estádios	Entre Grupos	(Combinado)	107,716	3	35,905	1,390	,251
			Nos grupos		2427,425	94	25,824		
			Total		2535,140	97			
H&Y			Entre Grupos	(Combinado)	243,535	3	81,178	,784	,506
			Nos grupos		9734,784	94	103,562		
			Total		9978,320	97			
Dist_T_22	*	Estádios	Entre Grupos	(Combinado)	243,535	3	81,178	,784	,506
			Nos grupos		9734,784	94	103,562		
			Total		9978,320	97			
H&Y			Entre Grupos	(Combinado)	243,535	3	81,178	,784	,506
			Nos grupos		9734,784	94	103,562		
			Total		9978,320	97			
Dist_T_23	*	Estádios	Entre Grupos	(Combinado)	243,535	3	81,178	,784	,506
			Nos grupos		9734,784	94	103,562		
			Total		9978,320	97			
H&Y			Entre Grupos	(Combinado)	243,535	3	81,178	,784	,506
			Nos grupos		9734,784	94	103,562		
			Total		9978,320	97			

Tabela de ANOVA para as features Amplitude

			Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Ampl_AP_11 Estádios H&Y	*	Entre Grupos (Combinado)	460,152	3	153,384	3,096	,031
		Nos grupos	4657,659	94	49,550		
		Total	5117,812	97			
Ampl_AP_12 Estádios H&Y	*	Entre Grupos (Combinado)	843,530	3	281,177	6,681	,000
		Nos grupos	3955,940	94	42,084		
		Total	4799,470	97			
Ampl_AP_13 Estádios H&Y	*	Entre Grupos (Combinado)	570,418	3	190,139	3,239	,026
		Nos grupos	5518,607	94	58,709		
		Total	6089,024	97			
Ampl_AP_21 Estádios H&Y	*	Entre Grupos (Combinado)	765,653	3	255,218	3,090	,031
		Nos grupos	7763,422	94	82,590		
		Total	8529,076	97			
Ampl_AP_22 Estádios H&Y	*	Entre Grupos (Combinado)	431,343	3	143,781	3,135	,029
		Nos grupos	4311,053	94	45,862		
		Total	4742,396	97			
Ampl_AP_23 Estádios H&Y	*	Entre Grupos (Combinado)	342,618	3	114,206	1,488	,223
		Nos grupos	7215,821	94	76,764		
		Total	7558,439	97			
Ampl_ML_11 Estádios H&Y	*	Entre Grupos (Combinado)	1135,222	3	378,407	7,765	,000
		Nos grupos	4580,804	94	48,732		
		Total	5716,026	97			
Ampl_ML_12 Estádios H&Y	*	Entre Grupos (Combinado)	1098,971	3	366,324	11,193	,000
		Nos grupos	3076,336	94	32,727		
		Total	4175,307	97			
Ampl_ML_13 Estádios H&Y	*	Entre Grupos (Combinado)	1047,806	3	349,269	5,013	,003
		Nos grupos	6549,091	94	69,671		
		Total	7596,897	97			
Ampl_ML_21 Estádios H&Y	*	Entre Grupos (Combinado)	564,081	3	188,027	3,088	,031
		Nos grupos	5723,621	94	60,890		
		Total	6287,703	97			
Ampl_ML_22 Estádios H&Y	*	Entre Grupos (Combinado)	622,977	3	207,659	2,771	,046
		Nos grupos	7044,321	94	74,940		
		Total	7667,298	97			
Ampl_ML_23 Estádios H&Y	*	Entre Grupos (Combinado)	1606,860	3	535,620	5,650	,001
		Nos grupos	8911,035	94	94,798		
		Total	10517,896	97			

Tabela de ANOVA para as features Velocidade

			Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Vel_M_T_11 * Estádios	Entre Grupos	(Combinado)	210,474	3	70,158	8,837	,000
H&Y	Nos grupos		746,253	94	7,939		
	Total		956,727	97			
Vel_M_T_12 * Estádios	Entre Grupos	(Combinado)	235,660	3	78,553	11,270	,000
H&Y	Nos grupos		655,218	94	6,970		
	Total		890,878	97			
Vel_M_T_13 * Estádios	Entre Grupos	(Combinado)	358,166	3	119,389	9,184	,000
H&Y	Nos grupos		1221,961	94	13,000		
	Total		1580,126	97			
Vel_M_T_21 * Estádios	Entre Grupos	(Combinado)	208,873	3	69,624	3,820	,012
H&Y	Nos grupos		1713,200	94	18,226		
	Total		1922,072	97			
Vel_M_T_22 * Estádios	Entre Grupos	(Combinado)	216,862	3	72,287	4,538	,005
H&Y	Nos grupos		1497,481	94	15,931		
	Total		1714,342	97			
Vel_M_T_23 * Estádios	Entre Grupos	(Combinado)	635,527	3	211,842	10,530	,000
H&Y	Nos grupos		1891,095	94	20,118		
	Total		2526,621	97			
Vel_M_AP_11 * Estádios	Entre Grupos	(Combinado)	80,558	3	26,853	6,302	,001
H&Y	Nos grupos		400,536	94	4,261		
	Total		481,094	97			
Vel_M_AP_12 * Estádios	Entre Grupos	(Combinado)	72,272	3	24,091	5,989	,001
H&Y	Nos grupos		378,097	94	4,022		
	Total		450,368	97			
Vel_M_AP_13 * Estádios	Entre Grupos	(Combinado)	118,599	3	39,533	7,201	,000
H&Y	Nos grupos		516,067	94	5,490		
	Total		634,666	97			
Vel_M_AP_21 * Estádios	Entre Grupos	(Combinado)	115,639	3	38,546	3,216	,026
H&Y	Nos grupos		1126,566	94	11,985		
	Total		1242,205	97			
Vel_M_AP_22 * Estádios	Entre Grupos	(Combinado)	112,192	3	37,397	4,184	,008
H&Y	Nos grupos		840,242	94	8,939		
	Total		952,434	97			
Vel_M_AP_23 * Estádios	Entre Grupos	(Combinado)	232,460	3	77,487	7,410	,000
H&Y	Nos grupos		982,938	94	10,457		

	Total		1215,398	97			
Vel_M_ML_11 * Estádios	Entre Grupos (Combinado)		101,521	3	33,840	11,142	,000
H&Y	Nos grupos		285,507	94	3,037		
	Total		387,028	97			
Vel_M_ML_12 * Estádios	Entre Grupos (Combinado)		121,931	3	40,644	15,959	,000
H&Y	Nos grupos		239,392	94	2,547		
	Total		361,323	97			
Vel_M_ML_13 * Estádios	Entre Grupos (Combinado)		185,078	3	61,693	9,561	,000
H&Y	Nos grupos		606,547	94	6,453		
	Total		791,625	97			
Vel_M_ML_21 * Estádios	Entre Grupos (Combinado)		68,476	3	22,825	5,161	,002
H&Y	Nos grupos		415,749	94	4,423		
	Total		484,225	97			
Vel_M_ML_22 * Estádios	Entre Grupos (Combinado)		72,199	3	24,066	4,627	,005
H&Y	Nos grupos		488,952	94	5,202		
	Total		561,152	97			
Vel_M_ML_23 * Estádios	Entre Grupos (Combinado)		324,471	3	108,157	13,052	,000
H&Y	Nos grupos		778,968	94	8,287		
	Total		1103,439	97			

Tabela de ANOVA as features Área

			Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Area_Elip_11 * Estádios	Entre Grupos	(Combinado)	720840,731	3	240280,244	7,462	,000
H&Y	Nos grupos		3026973,716	94	32201,848		
	Total		3747814,447	97			
Area_Elip_12 * Estádios	Entre Grupos	(Combinado)	954524,701	3	318174,900	15,008	,000
H&Y	Nos grupos		1992806,139	94	21200,065		
	Total		2947330,840	97			
Area_Elip_13 * Estádios	Entre Grupos	(Combinado)	972185,799	3	324061,933	3,152	,029
H&Y	Nos grupos		9663910,421	94	102807,558		
	Total		10636096,220	97			
Area_Elip_21 * Estádios	Entre Grupos	(Combinado)	592158,203	3	197386,068	2,528	,062
H&Y	Nos grupos		7340068,541	94	78085,836		
	Total		7932226,744	97			
Area_Elip_22 * Estádios	Entre Grupos	(Combinado)	612194,583	3	204064,861	2,012	,118
H&Y	Nos grupos		9532764,926	94	101412,393		
	Total		10144959,510	97			
Area_Elip_23 * Estádios	Entre Grupos	(Combinado)	1628523,522	3	542841,174	11,609	,000
H&Y	Nos grupos		4395473,272	94	46760,354		
	Total		6023996,794	97			

Tabela de ANOVA para as features Frequência

			Soma dos Quadrados	gl	Quadrado Médio	F	Sig.
Freq_Osc_AP_11 * Estádios H&Y	Entre Grupos (Combinado)		,019	3	,006	1,301	,279
	Nos grupos		,451	94	,005		
	Total		,469	97			
Freq_Osc_AP_12 * Estádios H&Y	Entre Grupos (Combinado)		,028	3	,009	1,897	,135
	Nos grupos		,459	94	,005		
	Total		,486	97			
Freq_Osc_AP_13 * Estádios H&Y	Entre Grupos (Combinado)		,032	3	,011	1,570	,202
	Nos grupos		,634	94	,007		
	Total		,666	97			
Freq_Osc_AP_21 * Estádios H&Y	Entre Grupos (Combinado)		,013	3	,004	,445	,721
	Nos grupos		,884	94	,009		
	Total		,897	97			
Freq_Osc_AP_22 * Estádios H&Y	Entre Grupos (Combinado)		,028	3	,009	1,073	,365
	Nos grupos		,808	94	,009		
	Total		,836	97			
Freq_Osc_AP_23 * Estádios H&Y	Entre Grupos (Combinado)		,058	3	,019	3,219	,026
	Nos grupos		,562	94	,006		
	Total		,620	97			
Freq_Osc_ML_11 * Estádios H&Y	Entre Grupos (Combinado)		,028	3	,009	,870	,459
	Nos grupos		,994	94	,011		
	Total		1,021	97			
Freq_Osc_ML_12 * Estádios H&Y	Entre Grupos (Combinado)		,039	3	,013	1,370	,257
	Nos grupos		,882	94	,009		
	Total		,920	97			
Freq_Osc_ML_13 * Estádios H&Y	Entre Grupos (Combinado)		,011	3	,004	,393	,758
	Nos grupos		,854	94	,009		
	Total		,865	97			
Freq_Osc_ML_21 * Estádios H&Y	Entre Grupos (Combinado)		,006	3	,002	,210	,890
	Nos grupos		,911	94	,010		
	Total		,917	97			
Freq_Osc_ML_22 * Estádios H&Y	Entre Grupos (Combinado)		,019	3	,006	,544	,653
	Nos grupos		1,099	94	,012		
	Total		1,118	97			
Freq_Osc_ML_23 * Estádios H&Y	Entre Grupos (Combinado)		,013	3	,004	,421	,738
	Nos grupos		,966	94	,010		
	Total		,979	97			

Estatística de Resíduos das variáveis Distância

Estatísticas de resíduos³ para a variável Dist_T_11

	Mínimo	Máximo	Média	Desvio Padrão	N
Valor predito	2,38	2,49	2,47	,020	98
Valor Predito Padrão	-4,256	1,183	,000	1,000	98
Erro padrão do valor predito	,085	,375	,113	,043	98
Valor predito ajustado	2,24	2,52	2,47	,035	98
Resíduo	-1,489	1,589	,000	,840	98
Resíduo Padrão	-1,764	1,883	,000	,995	98
Resíduo studentizado	-1,782	1,982	,001	1,005	98
Resíduo Excluído	-1,520	1,760	,001	,858	98
Resíduo Studentizado Excluído	-1,803	2,013	-,001	1,013	98
Mahal. Distância	,000	18,114	,990	2,330	98
Distância de Cook	,002	,211	,011	,024	98
Valor Centralizado de Alavancagem	,000	,187	,010	,024	98

a. Variável Dependente: Estádios H&Y

Estatísticas de resíduos³ para a variável Dist_T_12

	Mínimo	Máximo	Média	Desvio Padrão	N
Valor predito	2,18	4,20	2,47	,278	98
Valor Predito Padrão	-1,031	6,220	,000	1,000	98
Erro padrão do valor predito	,080	,509	,103	,048	98
Valor predito ajustado	2,16	4,34	2,47	,286	98
Resíduo	-1,700	1,580	,000	,792	98
Resíduo Padrão	-2,135	1,983	,000	,995	98
Resíduo studentizado	-2,154	1,994	,000	1,003	98
Resíduo Excluído	-1,730	1,597	-,001	,805	98
Resíduo Studentizado Excluído	-2,196	2,026	-,002	1,010	98
Mahal. Distância	,000	38,690	,990	3,980	98
Distância de Cook	,000	,049	,008	,010	98
Valor Centralizado de Alavancagem	,000	,399	,010	,041	98

a. Variável Dependente: Estádios H&Y

Estatísticas de resíduos³ para variável Dist_T_13

	Mínimo	Máximo	Média	Desvio Padrão	N
Valor predito	2,38	3,14	2,47	,101	98
Valor Predito Padrão	-,927	6,636	,000	1,000	98
Erro padrão do valor predito	,085	,571	,107	,053	98
Valor predito ajustado	2,35	3,25	2,47	,111	98
Resíduo	-1,711	1,618	,000	,834	98
Resíduo Padrão	-2,041	1,930	,000	,995	98
Resíduo studentizado	-2,116	1,948	-,001	1,005	98
Resíduo Excluído	-1,839	1,647	-,001	,850	98
Resíduo Studentizado Excluído	-2,156	1,977	-,003	1,012	98
Mahal. Distância	,000	44,037	,990	4,519	98
Distância de Cook	,002	,168	,010	,023	98
Valor Centralizado de Alavancagem	,000	,454	,010	,047	98

a. Variável Dependente: Estádios H&Y

Estatísticas de resíduos³ para a variável Dist_T_21

	Mínimo	Máximo	Média	Desvio Padrão	N
Valor predito	2,29	3,37	2,47	,184	98
Valor Predito Padrão	-,973	4,881	,000	1,000	98
Erro padrão do valor predito	,083	,417	,108	,046	98
Valor predito ajustado	2,27	3,50	2,47	,186	98
Resíduo	-1,656	1,698	,000	,819	98
Resíduo Padrão	-2,011	2,062	,000	,995	98
Resíduo studentizado	-2,032	2,081	,000	1,004	98
Resíduo Excluído	-1,691	1,730	,001	,835	98
Resíduo Studentizado Excluído	-2,067	2,119	-,002	1,011	98
Mahal. Distância	,000	23,820	,990	2,893	98
Distância de Cook	,000	,113	,009	,015	98
Valor Centralizado de Alavancagem	,000	,246	,010	,030	98

a. Variável Dependente: Estádios H&Y

Estatísticas de resíduos³ para a variável Dist_T_22

	Mínimo	Máximo	Média	Desvio Padrão	N
Valor predito	2,33	3,15	2,47	,135	98
Valor Predito Padrão	-1,037	5,037	,000	1,000	98
Erro padrão do valor predito	,084	,434	,111	,043	98
Valor predito ajustado	2,29	3,20	2,47	,135	98
Resíduo	-1,691	1,670	,000	,829	98
Resíduo Padrão	-2,029	2,004	,000	,995	98
Resíduo studentizado	-2,069	2,026	,000	1,004	98
Resíduo Excluído	-1,758	1,706	,001	,845	98
Resíduo Studentizado Excluído	-2,106	2,059	-,001	1,012	98
Mahal. Distância	,003	25,375	,990	2,720	98
Distância de Cook	,002	,095	,010	,016	98
Valor Centralizado de Alavancagem	,000	,262	,010	,028	98

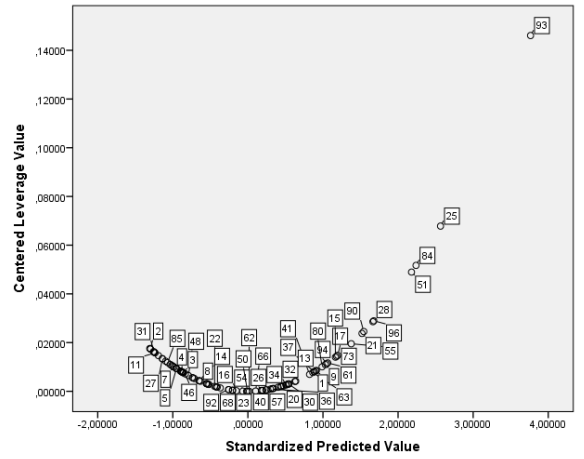
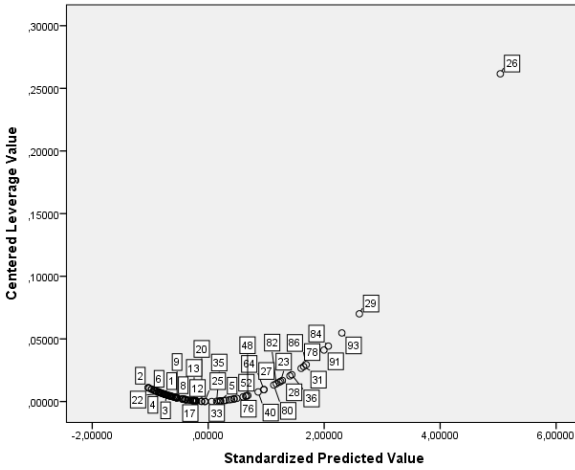
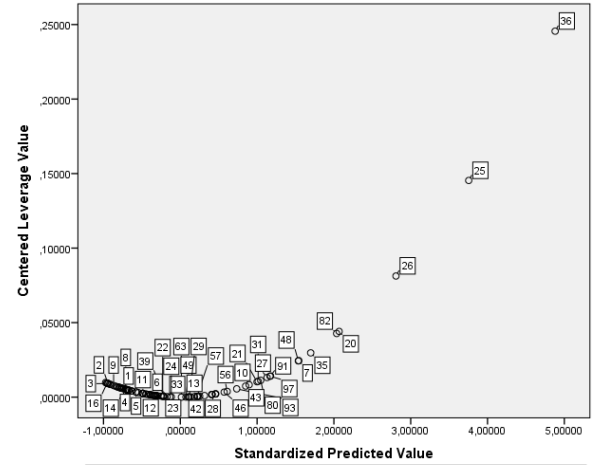
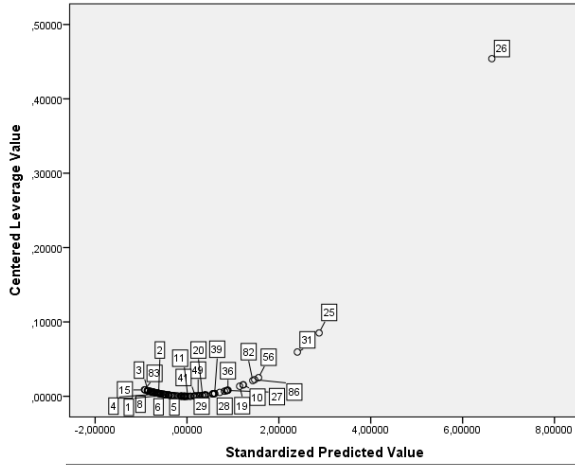
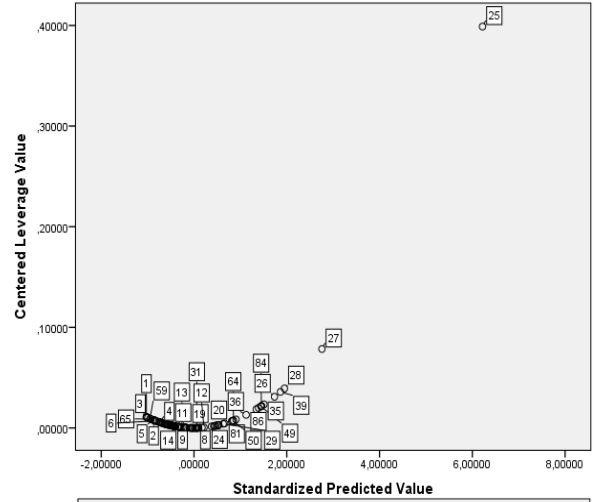
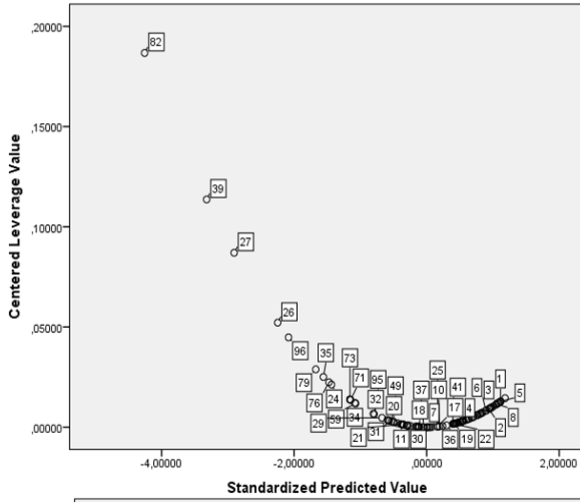
a. Variável Dependente: Estádios H&Y

Estatísticas de resíduos³ para a variável Dist_T_23

	Mínimo	Máximo	Média	Desvio Padrão	N
Valor predito	2,46	2,48	2,47	,004	98
Valor Predito Padrão	-1,306	3,764	,000	1,000	98
Erro padrão do valor predito	,085	,334	,115	,035	98
Valor predito ajustado	2,20	2,51	2,46	,033	98
Resíduo	-1,473	1,536	,000	,840	98
Resíduo Padrão	-1,745	1,819	,000	,995	98
Resíduo studentizado	-1,761	1,955	,003	1,007	98
Resíduo Excluído	-1,504	1,796	,005	,861	98
Resíduo Studentizado Excluído	-1,780	1,984	,001	1,015	98
Mahal. Distância	,000	14,168	,990	1,703	98
Distância de Cook	,002	,354	,013	,039	98
Valor Centralizado de Alavancagem	,000	,146	,010	,018	98

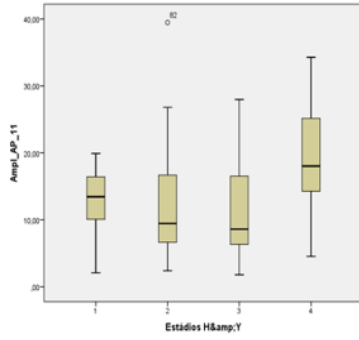
a. Variável Dependente: Estádios H&Y

Gráficos de dispersão das features Distância

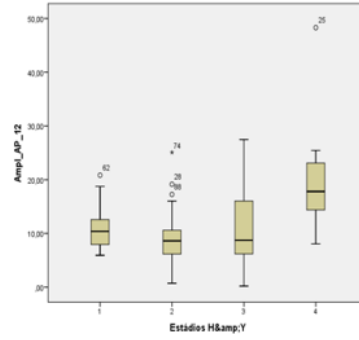


Boxplot para as features Amplitude

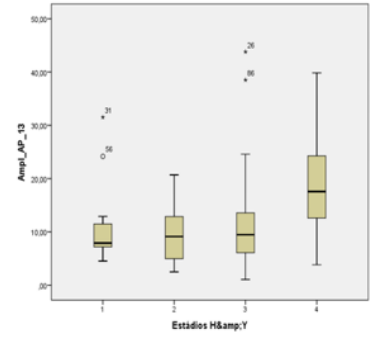
Ampl_AP_11



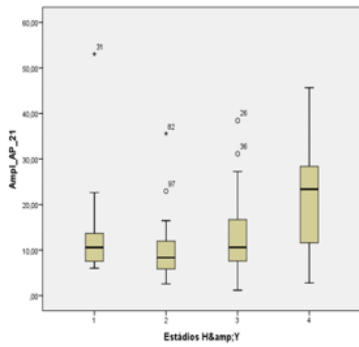
Ampl_AP_12



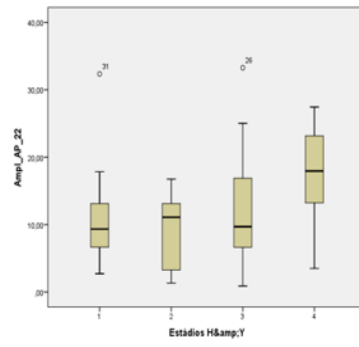
Ampl_AP_13



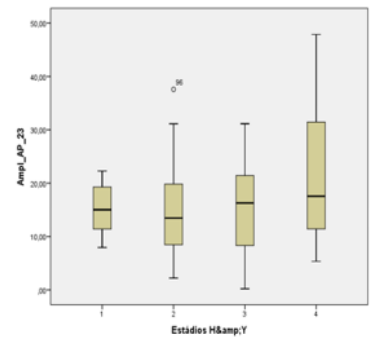
Ampl_AP_21



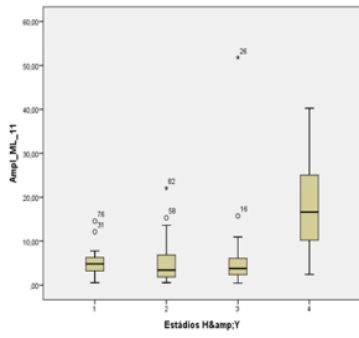
Ampl_AP_22



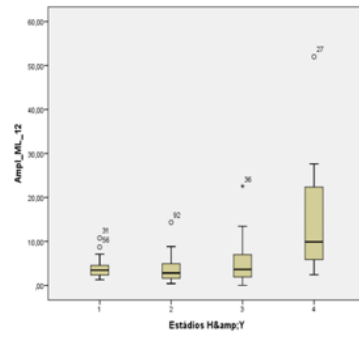
Ampl_AP_23



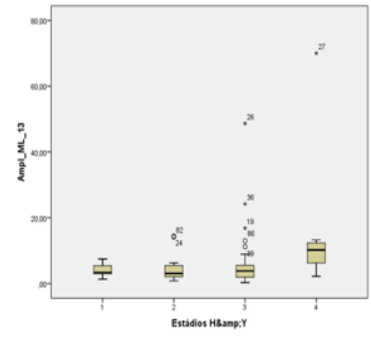
Ampl_ML_11



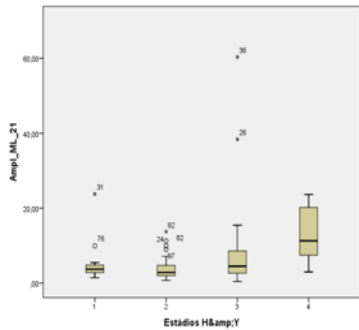
Ampl_ML_12



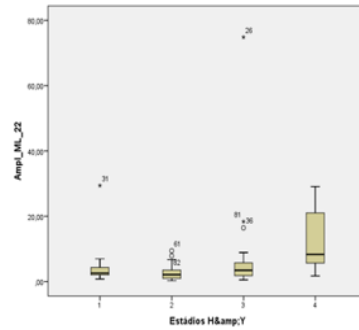
Ampl_ML_13



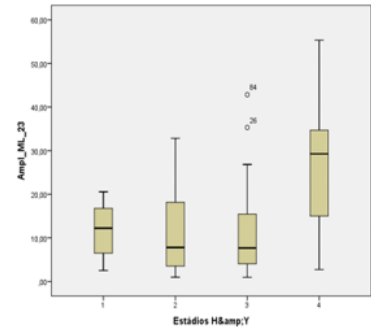
Ampl_ML_21



Ampl_ML_22

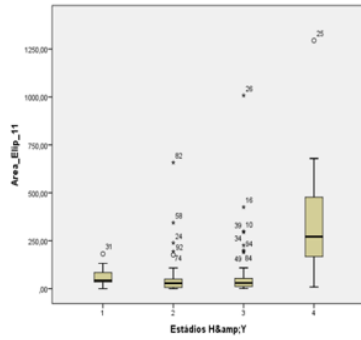


Ampl_ML_23

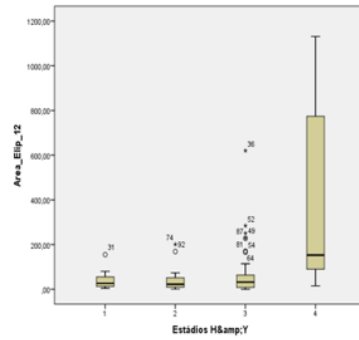


Boxplot para as features Área

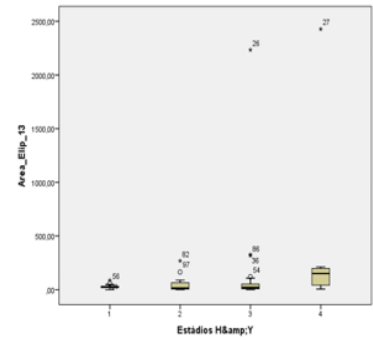
Area_Elip_11



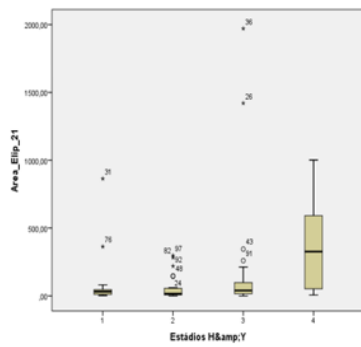
Area_Elip_12



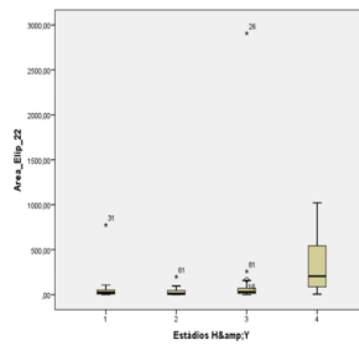
Area_Elip_13



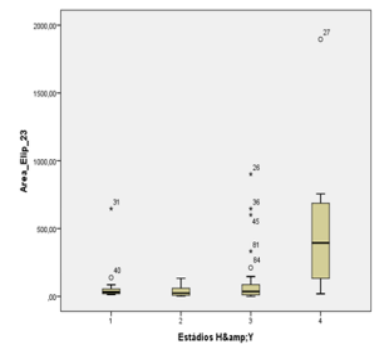
Area_Elip_21



Area_Elip_22

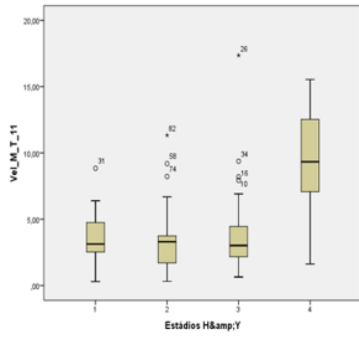


Area_Elip_23

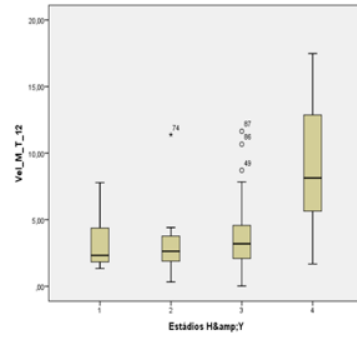


Boxplot para as features Velocidade

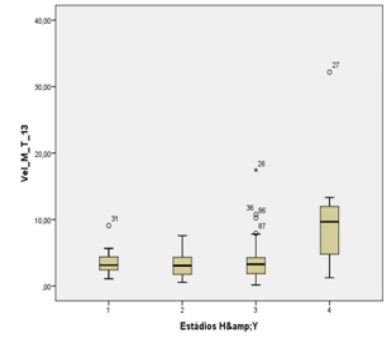
Vel_M_T_11



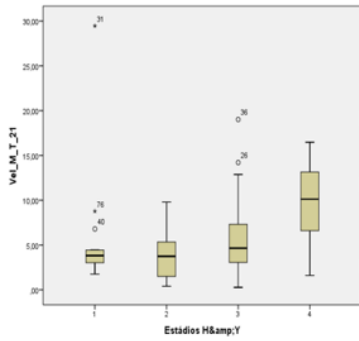
Vel_M_T_12



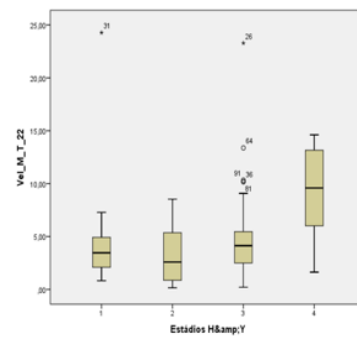
Vel_M_T_13



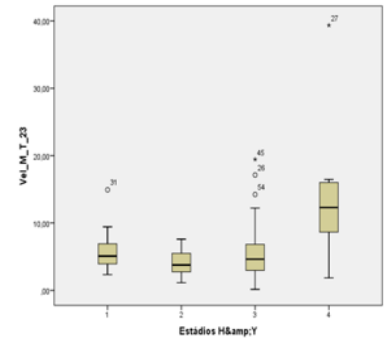
Vel_M_T_21



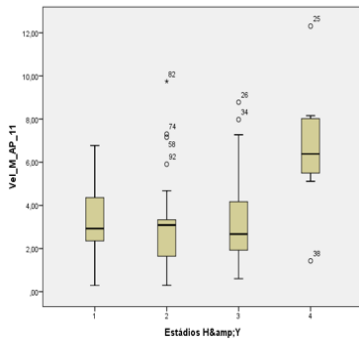
Vel_M_T_22



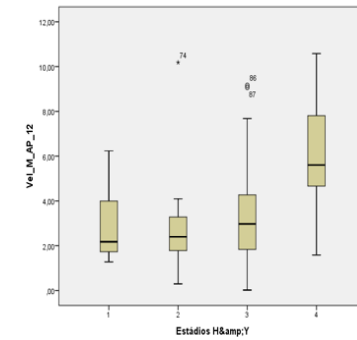
Vel_M_T_23



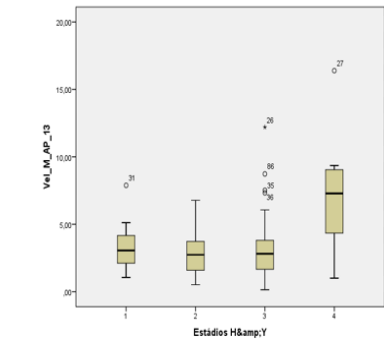
Vel_M_AP_11



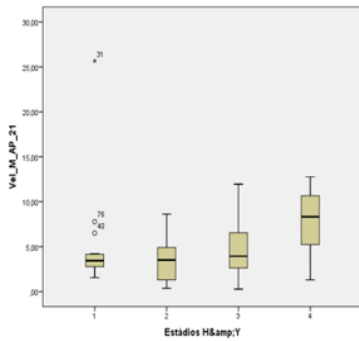
Vel_M_AP_12



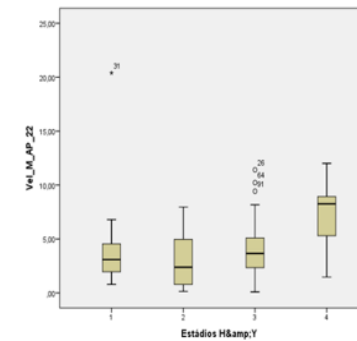
Vel_M_AP_13



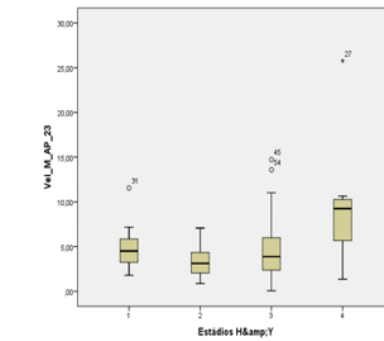
Vel_M_AP_21



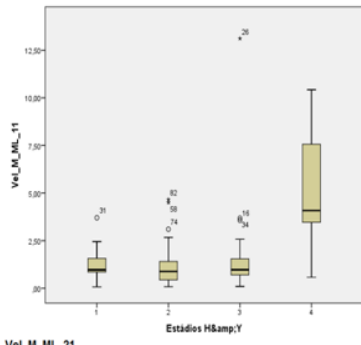
Vel_M_AP_22



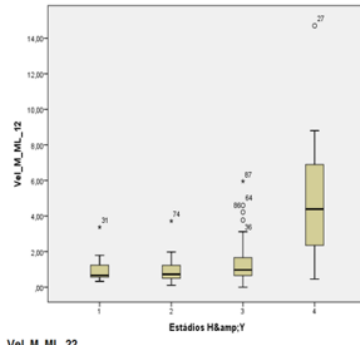
Vel_M_AP_23



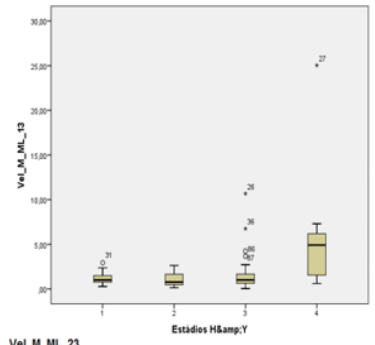
Vel_M_ML_11



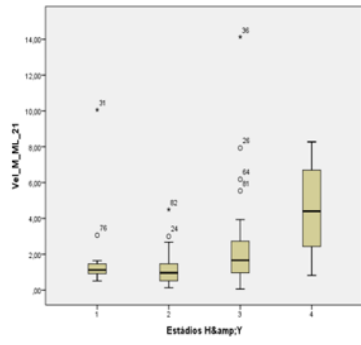
Vel_M_ML_12



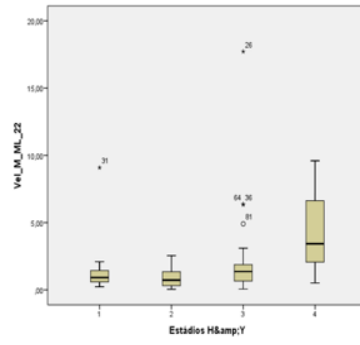
Vel_M_ML_13



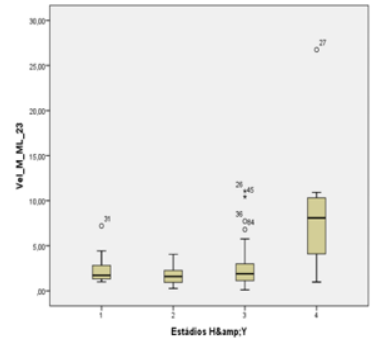
Vel_M_ML_21



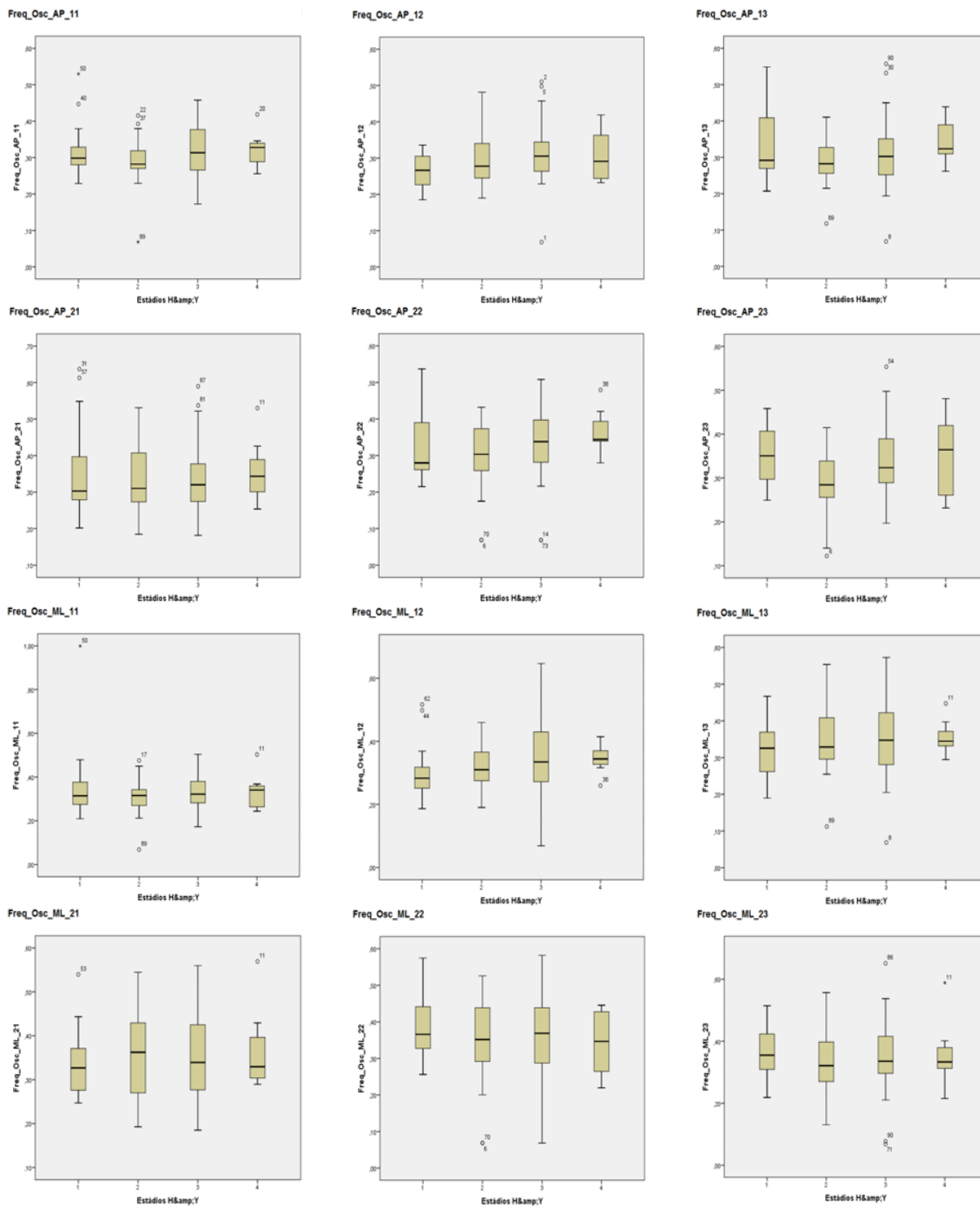
Vel_M_ML_22



Vel_M_ML_23



Boxplot para as features Frequência



Função geral para a validação cruzada do modelo SVM para a escolha dos parâmetros

```
Escolha_Dados = 1

if Escolha_Dados == 1
    Dados = xlsread('Dados_UAB.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)];

    X = Dados(:, [2:3 5]);
    Y = Dados(:,1);

    n = 161;

elseif Escolha_Dados == 2
    Dados = xlsread('UnifDados_UAB.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)];

    X = Dados(:, [2:3 5]);
    Y = Dados(:,1);

    n = 76;

elseif Escolha_Dados == 3
    Dados = xlsread('Equilibrio_UnifDados_UAB.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)];

    X = Dados(:,2:5);
    Y = Dados(:,1);

    n = 141;

elseif Escolha_Dados == 6
    Dados = xlsread('Dados_UABOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)];

    X = Dados(:,2:5);
    Y = Dados(:,1);

    n = 158;

elseif Escolha_Dados == 7
    Dados = xlsread('Dados_EquilibrioUABOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)];

    X = Dados(:,2:5);
    Y = Dados(:,1);

    n = 72;

elseif Escolha_Dados == 8
    Dados = xlsread('Dados_EquilibrioBinarioUABOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)];

    X = Dados(:,2:5);
```

```

Y = Dados(:,1);

n = 64;

elseif Escolha_Dados == 9
    Dados = xlsread('Postural_Dist.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:, [3 5]);
    Y = Dados(:,1);

    n = 98;

elseif Escolha_Dados == 10
    Dados = xlsread('Unif_Dist.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:, [3 5]);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 11
    Dados = xlsread('Equilibrio_BinarioDist.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:, [3 5]);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 12
    Dados = xlsread('Binario_Dist.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:, [2 5]);
    Y = Dados(:,1);

    n = 98;

elseif Escolha_Dados == 13
    Dados = xlsread('Postural_DistOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 93;

```

```

elseif Escolha_Dados == 14
    Dados = xlsread('Postural_EquilibrioDistOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 50;

elseif Escolha_Dados == 15
    Dados = xlsread('Postural_EquilibrioBinarioDistOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 50;

elseif Escolha_Dados == 16
    Dados = xlsread('Postural_BinarioDistOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 93;

elseif Escolha_Dados == 17
    Dados = xlsread('Postural_Ampl.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:, [2:4 7 11 13]);
    Y = Dados(:,1);

    n = 98;

elseif Escolha_Dados == 18
    Dados = xlsread('Unif_Ampl.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:, [2:4 7 11 13]);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 19
    Dados = xlsread('Equilibrio_BinarioAmpl.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)

```

```

        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:, [2:4 7 11 13]);
Y = Dados(:,1);

n = 52;

elseif Escolha_Dados == 20
    Dados = xlsread('Binario_Ampl.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
            Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:, [3 8:10 13]);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 21
    Dados = xlsread('Postural_AmplOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
            Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:13);
Y = Dados(:,1);

n = 91;

elseif Escolha_Dados == 22
    Dados = xlsread('Postural_EquilibrioAmpOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
            Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:13);
Y = Dados(:,1);

n = 46;

elseif Escolha_Dados == 23
    Dados = xlsread('Postural_EquilibrioBinarioAmpOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
            Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:13);
Y = Dados(:,1);

n = 46;

elseif Escolha_Dados == 24
    Dados = xlsread('Postural_Area.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)

```

```

        Dados(:,6) Dados(:,7)];

X = Dados(:, [2:3 5]);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 25
    Dados = xlsread('Unif_Area.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 26
    Dados = xlsread('Binario_Area.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 98;

elseif Escolha_Dados == 27
    Dados = xlsread('Equilibrio_BinarioArea.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7)];

    X = Dados(:, [2:3 5]);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 28
    Dados = xlsread('Postural_AreaOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

    n = 95;

elseif Escolha_Dados == 29
    Dados = xlsread('Postural_EquilibrioAreaOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
            Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);

```

```

Y = Dados(:,1);

n = 50;

elseif Escolha_Dados == 30
Dados = xlsread('Postural_EquilibrioBinarioAreaOutlier.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
         Dados(:,6) Dados(:,7)];

X = Dados(:,2:7);
Y = Dados(:,1);

n = 50;

elseif Escolha_Dados == 31
Dados = xlsread('Postural_BinarioAreaOutlier.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
         Dados(:,6) Dados(:,7)];

X = Dados(:,2:7);
Y = Dados(:,1);

n = 95;

elseif Escolha_Dados == 32
Dados = xlsread('Postural_Vel.xlsx','Dados');

Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
         Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
         Dados(:,11) Dados(:,12) Dados(:,13) Dados(:,14) Dados(:,15)
         Dados(:,16) Dados(:,17) Dados(:,18) Dados(:,19)];

X = Dados(:,2:19);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 33
Dados = xlsread('Unif_Vel.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
         Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
         Dados(:,11) Dados(:,12) Dados(:,13) Dados(:,14) Dados(:,15)
         Dados(:,16) Dados(:,17) Dados(:,18) Dados(:,19)];

X = Dados(:, [9 14:19]);
Y = Dados(:,1);

n = 52;

elseif Escolha_Dados == 34
Dados = xlsread('Equilibrio_BinarioVel.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
         Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
         Dados(:,11) Dados(:,12) Dados(:,13) Dados(:,14) Dados(:,15)
         Dados(:,16) Dados(:,17) Dados(:,18) Dados(:,19)];

```

```

X = Dados(:, [2:4 7 10 13:16 19]);
Y = Dados(:, 1);

n = 52;

elseif Escolha_Dados == 35
Dados = xlsread('Binario_Vel.xlsx', 'Dados');
Dados = [Dados(:, 1) Dados(:, 2) Dados(:, 3) Dados(:, 4) Dados(:, 5)
         Dados(:, 6) Dados(:, 7) Dados(:, 8) Dados(:, 9) Dados(:, 10)
         Dados(:, 11) Dados(:, 12) Dados(:, 13) Dados(:, 14) Dados(:, 15)
         Dados(:, 16) Dados(:, 17) Dados(:, 18) Dados(:, 19)];

X = Dados(:, 2:19);
Y = Dados(:, 1);

n = 98;

elseif Escolha_Dados == 36
Dados = xlsread('Postural_VelOutlier.xlsx', 'Dados');
Dados = [Dados(:, 1) Dados(:, 2) Dados(:, 3) Dados(:, 4) Dados(:, 5)
         Dados(:, 6) Dados(:, 7) Dados(:, 8) Dados(:, 9) Dados(:, 10)
         Dados(:, 11) Dados(:, 12) Dados(:, 13) Dados(:, 14) Dados(:, 15)
         Dados(:, 16) Dados(:, 17) Dados(:, 18) Dados(:, 19)];

X = Dados(:, [9 14:19]);
Y = Dados(:, 1);

n = 93;

elseif Escolha_Dados == 37
Dados = xlsread('Postural_EquilibrioVelOutlier.xlsx', 'Dados');
Dados = [Dados(:, 1) Dados(:, 2) Dados(:, 3) Dados(:, 4) Dados(:, 5)
         Dados(:, 6) Dados(:, 7) Dados(:, 8) Dados(:, 9) Dados(:, 10)
         Dados(:, 11) Dados(:, 12) Dados(:, 13) Dados(:, 14) Dados(:, 15)
         Dados(:, 16) Dados(:, 17) Dados(:, 18) Dados(:, 19)];

X = Dados(:, [9 14:19]);
Y = Dados(:, 1);

n = 47;

elseif Escolha_Dados == 38
Dados = xlsread('Postural_EquilibrioBinarioVelOutlier.xlsx', 'Dados');
Dados = [Dados(:, 1) Dados(:, 2) Dados(:, 3) Dados(:, 4) Dados(:, 5)
         Dados(:, 6) Dados(:, 7) Dados(:, 8) Dados(:, 9) Dados(:, 10)
         Dados(:, 11) Dados(:, 12) Dados(:, 13) Dados(:, 14) Dados(:, 15)
         Dados(:, 16) Dados(:, 17) Dados(:, 18) Dados(:, 19)];

X = Dados(:, [9 14:19]);
Y = Dados(:, 1);

n = 47;

```

```

elseif Escolha_Dados == 39
    Dados = xlsread('Postural_BinarioVelOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13) Dados(:,14) Dados(:,15)
             Dados(:,16) Dados(:,17) Dados(:,18) Dados(:,19)];

    X = Dados(:,[9 14:19]);
    Y = Dados(:,1);

    n = 93;

elseif Escolha_Dados == 40
    Dados = xlsread('Postural_Freq.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

    n = 98;

elseif Escolha_Dados == 41
    Dados = xlsread('Unif_Freq.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 42
    Dados = xlsread('Equilibrio_BinarioFreq.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

    n = 52;

elseif Escolha_Dados == 43
    Dados = xlsread('Binario_Freq.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

```

```

n = 98;

elseif Escolha_Dados == 44
    Dados = xlsread('Postural_FreqOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

n = 84;

elseif Escolha_Dados == 45
    Dados = xlsread('Postural_EquilibrioBinarioFreqOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

n = 45;

elseif Escolha_Dados == 46
    Dados = xlsread('Postural_BinarioFreqOutlier.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
             Dados(:,11) Dados(:,12) Dados(:,13)];

    X = Dados(:,2:13);
    Y = Dados(:,1);

n = 84;

elseif Escolha_Dados == 47
    Dados = xlsread('Discinesias_DistMembSup.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

    X = Dados(:,2:7);
    Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 48
    Dados = xlsread('Equilibrio_DiscinesiasDistMembSup.xlsx','Dados');
    Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
             Dados(:,6) Dados(:,7)];

```

```

X = Dados(:,2:7);
Y = Dados(:,1);

n = 52;

elseif Escolha_Dados == 49
Dados = xlsread('Discinesias_AmplMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:13);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 50
Dados = xlsread('Equilibrio_DiscinesiasAmplMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:13);
Y = Dados(:,1);

n = 52;

elseif Escolha_Dados == 51
Dados = xlsread('Discinesias_AreaMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7)];

X = Dados(:,2:7);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 52
Dados = xlsread('Equilibrio_DiscinesiasAreaMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7)];

X = Dados(:,2:7);
Y = Dados(:,1);

n = 52;

elseif Escolha_Dados == 53
Dados = xlsread('Discinesias_VelMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13) Dados(:,14) Dados(:,15)
        Dados(:,16) Dados(:,17) Dados(:,18) Dados(:,19)];

```

```

X = Dados(:,2:19);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 54
Dados = xlsread('Equilibrio_DiscinesiasVelMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13) Dados(:,14) Dados(:,15)
        Dados(:,16) Dados(:,17) Dados(:,18) Dados(:,19)];

X = Dados(:,2:19);
Y = Dados(:,1);

n = 52;

elseif Escolha_Dados == 55
Dados = xlsread('Discinesias_FreqMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:12);
Y = Dados(:,1);

n = 98;

elseif Escolha_Dados == 56
Dados = xlsread('Equilibrio_DiscinesiasFreqMembSup.xlsx','Dados');
Dados = [Dados(:,1) Dados(:,2) Dados(:,3) Dados(:,4) Dados(:,5)
        Dados(:,6) Dados(:,7) Dados(:,8) Dados(:,9) Dados(:,10)
        Dados(:,11) Dados(:,12) Dados(:,13)];

X = Dados(:,2:13);
Y = Dados(:,1);

n = 52;

end

k = 5;
grupos = amostra(n,k,Y);
lc = 4;
m = zeros(lc,lc,k);

c = [10.^(-4) 10.^(-3) 10.^(-2) 10.^(-1) 0.5 0.75 1 10 50 10.^2 10.^3 10.^4];
gamma= [10.^(-4) 10.^(-3) 10.^(-2) 10.^(-1) 0.5 0.75 1 10 50 10.^2 10.^3 10.^4];

Accuracy_mat = zeros(length(c),length(gamma));

for l = 1:length(c)
    for t = 1:length(gamma)

```

```

cmd = ['-t 2 -s 0 -c ', num2str(c(1)), ' -gamma ', num2str(gamma(t))];
soma = zeros(lc,lc);
m = zeros(lc,lc,k);

for ii = 1:k
    pos = posicao(grupos,n,ii);

    XTe = X(logical(pos),:);
    YTe = Y(logical(pos),:);

    XTr = X(logical(1-pos),:);
    YTr = Y(logical(1-pos),:);

    modelo = svmtrain(YTr,XTr, cmd);
    Yprevisto = svmpredict(YTe,XTe, modelo);

    for j = 1:length(YTe)

        m(Yprevisto(j),YTe(j),ii) = m(Yprevisto(j),YTe(j),ii)+1;

    end
end

soma = m(:,:,1);

for k = 2:length(m)
    % Soma das k matrizes para cada grupo de teste

    soma = soma + m(:,:,k);

end

diag_soma = sum(diag(soma));
sum_elementos_soma = sum(soma(:));
Accuracy_Global = diag_soma/sum_elementos_soma*100;
Accuracy_mat(1,t) = Accuracy_Global;
mat_corr = corr(X);

end
end
return

function v = posicao(grupos,n,k)

% Funcao para marcar 1's em varias posicoes do vetor de zeros

v = zeros(1,n);

grupo_k = grupos(k,:);

for j = 1:length(grupo_k)

    if isnan(grupo_k(j))
    else
        v(grupo_k(j)) = 1;
    end
end
end

```

```

    end
end

end

function grupos = amostra(n,k,Y)

% Funcao para uniformizar grupos

% k = numero de grupos
% n = numero de sujeitos em estudo

rank=[(1:n)',Y];
rank=sortrows(rank,2);
grupos=NaN(k,ceil(n/k));
grupos(:,1:floor(n/k))=reshape(rank(1:floor(n/k)*k,1),[k,floor(n/k)]);

finais=floor(n/k)*k+1;

grupos(1:(n-finais+1),end)=rank(finais:n,1);

end

```