



# Seleção de características baseada em margem flexível aplicada a problemas não linearmente separáveis

Alex da Silva Maroco

JUIZ DE FORA  
NOVEMBRO, 2019

# Seleção de características baseada em margem flexível aplicada a problemas não linearmente separáveis

ALEX DA SILVA MAROCO

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Bacharelado em Ciência da Computação

Orientador: Saulo Moraes Villela

Coorientador: Raul Fonseca Neto

JUIZ DE FORA  
NOVEMBRO, 2019

SELEÇÃO DE CARACTERÍSTICAS BASEADA EM MARGEM  
FLEXÍVEL APLICADA A PROBLEMAS NÃO LINEARMENTE  
SEPARÁVEIS

Alex da Silva Maroco

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS  
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-  
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE  
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

---

Saulo Moraes Villela  
D.Sc. em Engenharia de Sistemas e Computação

---

Raul Fonseca Neto  
D.Sc. em Engenharia de Sistemas e Computação

---

Luciana Brugiolo Gonçalves  
D.Sc. em Computação

---

Carlos Cristiano Hasenclever Borges  
D.Sc. em Engenharia Civil

JUIZ DE FORA  
26 DE NOVEMBRO, 2019

*Aos meus amigos e irmãos.*

*Aos pais, pelo apoio e sustento.*

## Resumo

A realização de tarefas de classificação binária sobre conjuntos de dados não linearmente separáveis são naturalmente complexas e são prejudicadas pela grande quantidade de características irrelevantes existentes em tais conjuntos de dados. Neste trabalho estuda-se uma forma de se reduzir esta complexidade ao se utilizar um classificador com margem flexível para identificar as amostras que tornam esta base não linearmente separável, e posterior remoção destas amostras, o que tornará o conjunto de dados linearmente separável, que também tornará a seleção de características mais eficiente. Foram realizados diferentes testes de forma a avaliar o processo de remoção de amostras e o de seleção de características, cujos resultados são comparados aos de outras abordagens para o mesmo problema.

**Palavras-chave:** Seleção de características, Classificação binária, Margem flexível, Busca Ordenada Admissível.

## Abstract

Binary classification tasks on non linearly separable datasets are naturally complex and are also impaired by irrelevant features in these datasets. In this work we study a way to reduce the complexity by using a soft margin classifier to identify the data samples which makes the dataset non linearly separable, so they can be removed, resulting in a linearly separable dataset, increasing the efficiency of feature selection algorithms. Several tests were performed to evaluate both sample removal and feature selection processes, which results were compared to existing approaches for this problem.

**Keywords:** Feature selection, Binary classification, Soft margin, Admissible Ordered Search.

## Agradecimentos

A todos os meus parentes, em especial meus pais e irmã, pelo amor, encorajamento e apoio.

Ao professores Saulo e Raul pela orientação, amizade e principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para o nosso enriquecimento pessoal e profissional.

# Sumário

<b>Lista de Figuras</b>	<b>6</b>
<b>Lista de Tabelas</b>	<b>7</b>
<b>Lista de Abreviações</b>	<b>8</b>
<b>1 Introdução</b>	<b>9</b>
<b>2 Classificadores de larga margem</b>	<b>12</b>
2.1 Classificação binária com norma $p$ . . . . .	12
2.2 Perceptron de margem fixa com norma $p$ – $FMP_p$ . . . . .	13
2.2.1 Margem flexível . . . . .	14
2.3 Algoritmo de margem incremental com norma $p$ – $IMA_p$ . . . . .	16
<b>3 Seleção de características</b>	<b>18</b>
3.1 <i>Admissible Ordered Search</i> – AOS . . . . .	18
<b>4 Abordagem proposta</b>	<b>21</b>
<b>5 Experimentos e resultados</b>	<b>24</b>
5.1 Dados . . . . .	24
5.2 Parâmetros . . . . .	24
5.3 Resultados . . . . .	25
5.3.1 Remoção de amostras . . . . .	25
5.3.2 Seleção de características . . . . .	28
<b>6 Considerações finais</b>	<b>42</b>
<b>Referências Bibliográficas</b>	<b>43</b>



## Lista de Figuras

2.1	Solução de um problema de classificação binária com norma $p$ . . . . .	13
4.1	Exemplo do processo de linearização. . . . .	22
4.2	Fluxograma da abordagem proposta. . . . .	23

## Lista de Tabelas

5.1	Informações dos conjuntos de dados. . . . .	24
5.2	Percentual de amostras removidas para cada parâmetro – $\text{IMA}_p$ . . . . .	27
5.3	Percentual de amostras removidas para cada flexibilização – SMO linear. . . . .	28
5.4	Comparação das margens $L_2$ . . . . .	29
5.5	Média e desvio padrão de 10 execuções do $10\text{-fold}$ na norma $L_2$ . . . . .	30
5.6	Resultados do $\text{IMA}_\infty$ (parte 1). . . . .	32
5.7	Resultados do $\text{IMA}_\infty$ (parte 2). . . . .	33
5.8	Média dos resultados – AOS com $\text{IMA}_\infty$ . . . . .	34
5.9	Resultados do SMO linear (parte 1). . . . .	35
5.10	Resultados do SMO linear (parte 2). . . . .	36
5.11	Média dos resultados – AOS com SMO linear. . . . .	37
5.12	Resultados do SMO com <i>kernel</i> gaussiano. . . . .	39
5.13	Média dos estados da execução do AOS. . . . .	40
5.14	Comparação nos conjuntos de dados. . . . .	41

## Lista de Abreviações

AM	Aprendizado de Máquina
AOS	<i>Admissible Ordered Search</i>
FMP	<i>Fixed Margin Perceptron</i>
IMA	<i>Incremental Margin Algorithm</i>
SMO	<i>Sequential Minimal Optimization</i>
SVM	<i>Support Vector Machine</i>

# 1 Introdução

O Aprendizado de Máquina é uma área de pesquisa que busca desenvolver métodos para extrair conhecimento a partir de conjuntos de dados. Geralmente estes métodos realizam tal extração através de classificadores, que rotulam as amostras do conjunto de dados em classes distintas. Com isto é possível, a partir de dados de treinamento, criar preditores que são capazes, idealmente, de dizer a qual classe, previamente definida, um novo dado se encaixa. Um dos paradigmas que definem a forma de aprendizado utilizada na criação de classificadores, abordado neste trabalho, é o aprendizado supervisionado, no qual o preditor utiliza um conjunto de dados previamente rotulado como base para o aprendizado. A qualidade de um preditor é avaliada pelo seu poder de generalização, sua capacidade de acertar o rótulo de um novo dado.

Em tarefas de classificação binária, onde deseja-se rotular as amostras em duas classes distintas, cada amostra é representada por um ponto num espaço  $d$ -dimensional, onde  $d$  é o número total de características, o conjunto de dados pode ser linearmente separável – quando existe uma função linear capaz de separar o conjunto em duas classes – ou não linearmente separável – quando não existe função linear capaz de realizar esta separação.

Para os dados linearmente separáveis, os métodos de classificação mais utilizados são os baseadas em classificadores de larga margem, como as máquinas de vetores suporte (*Support Vector Machines* – SVMs) (VAPNIK, 1995) e o algoritmo de margem incremental (*Incremental Margin Algorithm* – IMA) (LEITE; FONSECA NETO, 2008), capazes de encontrar um hiperplano, definido por uma função linear, que permite a separação do espaço de dados em duas classes. Tais métodos têm sua eficiência prejudicada quando os dados de treinamento apresentam um grande número de características irrelevantes, gerando baixo poder de generalização, induzindo diversos erros de classificação. E no caso de conjuntos não linearmente separáveis, onde não é possível encontrar o hiperplano, é necessário utilizar alguma técnica auxiliar.

Uma destas técnicas utiliza o conceito de flexibilização, onde o classificador tem

suas restrições relaxadas, permitindo que algumas amostras fiquem em uma classe à qual elas não pertencem, ou seja, do lado errado do hiperplano. Outra técnica baseia-se na projeção das características para um novo espaço, chamado espaço de características ou *kernel* (FACELI et al., 2011), para então realizar o treinamento. Mas esta técnica é prejudicada pela existência de muitas características irrelevantes, além de provocar um aumento na dificuldade do problema, pois esta projeção gera novas dimensões, tornando o problema mais complexo.

Como os conjuntos de dados comumente possuem muitas características e alta complexidade, o que torna o processo de classificação ineficiente, e dado que apenas um pequeno conjunto destas características são relevantes, possuindo alto poder de generalização, sendo assim úteis ao aprendizado (NG, 1998), é importante possuir um processo de seleção de características capaz de eficientemente encontrar um bom subconjunto de características que maximize o poder de generalização dos dados.

Para evitar os problemas relacionados ao espaço *kernel* no processo de seleção de características em conjuntos não linearmente separáveis, uma opção para se realizar a seleção de características é a realização de buscas de forma a avaliar os possíveis subconjuntos de características. Porém o número de possíveis combinações é de ordem exponencial, e portanto inviável para conjuntos com muitas características. Vapnik (1995) diz que a seleção de características em conjuntos não linearmente separáveis também é possível ao se utilizar um classificador que permita o relaxamento das restrições de classificação.

Neste trabalho é proposto o desenvolvimento de um processo que permita que algoritmos de seleção de características, em particular a busca ordenada admissível (*Admissible Ordered Search* – AOS) (VILLELA, 2011), trabalhem com dados não linearmente separáveis nas dimensões originais do problema. Porém, para garantir a admissibilidade do processo de seleção de características do AOS, é necessário possuir uma função de avaliação monótona decrescente, propriedade que só é garantida quando há separabilidade dos dados.

Para garantir esta separabilidade, foram utilizados classificadores de larga margem com margem flexível, permitindo que se realize o relaxamento das restrições de classificação, e depois remover as amostras que violem o hiperplano, caso a classificação fosse

---

realizada sem flexibilização, a fim de se obter um subconjunto de amostras linearmente separável. Com isto, é possível garantir a admissibilidade necessária para o AOS. Diferentes classificadores e parâmetros foram avaliados durante a criação dos subconjuntos, a fim de determinar qual combinação que gera a melhor solução. O classificador que obteve melhores resultados foi utilizado para criar os conjuntos que passaram pela seleção de características com o AOS. Os resultados deste processo serão comparados os obtidos pela utilização do *kernel* durante o processo de seleção de características do AOS.

Além desta introdução, este trabalho possui mais cinco capítulos, onde o segundo capítulo apresenta os classificadores de larga margem, o terceiro apresenta o processo de seleção de características, o quarto apresenta a abordagem proposta pelo trabalho, o quinto apresenta os conjuntos de dados trabalhados e os resultados obtidos, e o último apresenta as considerações finais e trabalhos futuros.

## 2 Classificadores de larga margem

Neste capítulo são apresentados os conceitos teóricos relacionados à classificação binária e classificadores de larga margem.

### 2.1 Classificação binária com norma $p$

Seja  $Z = \{z_i = (x_i, y_i), i \in \{1, \dots, m\}\}$  um conjunto de treinamento com  $d$  características, composto de  $m$  amostras  $x_i \in \mathbb{R}^d$  e classes  $y_i \in \{-1, +1\}$  e sejam definidos os subconjuntos  $Z^+ = \{(x_i, y_i) \in Z : y_i = +1\}$  e  $Z^- = \{(x_i, y_i) \in Z : y_i = -1\}$ . O problema de classificação binária consiste em encontrar um hiperplano, representado pelo seu vetor normal  $w \in \mathbb{R}^d$  e uma constante  $b \in \mathbb{R}$ , de forma a separar subconjuntos  $Z^+$  e  $Z^-$  nos dois semiespaços gerados por ele. Assim, define-se  $(w, b)$  tal que:

$$y_i(w \cdot x_i + b) \geq 0, \text{ para todo } (x_i, y_i) \in Z. \quad (2.1)$$

Mas este hiperplano pode não existir para alguns conjuntos  $Z$ . Supondo que exista, define-se também o conceito de margem, dada pela distância entre o hiperplano e a amostra mais próxima a ele. É dito que  $Z$  aceita uma margem  $\gamma \geq 0$  quando existe um hiperplano  $H := \{x \in \mathbb{R}^d : w \cdot x + b = 0\}$  tal que:

$$y_i(w \cdot x_i + b) \geq \gamma, \text{ para todo } (x_i, y_i) \in Z. \quad (2.2)$$

Nesse caso, define-se dois hiperplanos adicionais paralelos a  $H$ , dados por  $H^+ := \{x \in \mathbb{R}^d : w \cdot x + (b - \gamma) = 0\}$  e  $H^- := \{x \in \mathbb{R}^d : w \cdot x + (b + \gamma) = 0\}$ . A Figura 2.1 ilustra estes hiperplanos. A distância entre estes hiperplanos paralelos sob uma norma  $p$  é dada por:

$$\text{dist}_p(H^+, H^-) = \frac{-b + \gamma + b + \gamma}{\|w\|_q} = \frac{2\gamma}{\|w\|_q}, \quad (2.3)$$

onde  $p$  e  $q$  são normas conjugadas, tal que  $\frac{1}{p} + \frac{1}{q} = 1$ . Define-se  $\gamma_g := \text{dist}(H^+, H^-)/2$  a

margem geométrica entre  $H^+$  e  $H^-$ . Assim, é dito que  $Z$  aceita uma margem geométrica  $\gamma_g \geq 0$  quando existe um hiperplano com  $(w, b)$  tal que:

$$y_i(w \cdot x_i + b) \geq \gamma_g \|w\|_q, \text{ para todo } (x_i, y_i) \in Z. \quad (2.4)$$

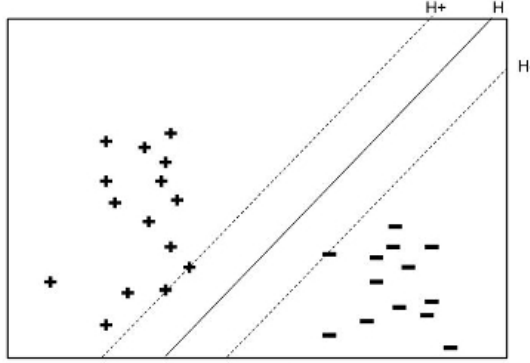


Figura 2.1: Solução de um problema de classificação binária com norma  $p$ .

## 2.2 Perceptron de margem fixa com norma $p$ – FMP $_p$

Dada uma  $p$ -margem geométrica fixa  $\gamma_f$  e um conjunto de treinamento  $Z$  que aceite  $\gamma_f$ , considere o problema de encontrar o hiperplano  $(w, b)$  tal que:

$$y_i(w \cdot x_i + b) \geq \gamma_f \|w\|_q, \text{ para todo } (x_i, y_i) \in Z. \quad (2.5)$$

Define-se também a função de erro  $J : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$  dada por:

$$J(w, b) := \sum_{(x_i, y_i) \in \mathcal{M}} \gamma_f \|w\|_q - y_i(w \cdot x_i + b), \quad (2.6)$$

onde  $\mathcal{M}$  é um subconjunto de  $Z$  que viola a Equação (2.5) na escolha de  $(w, b)$ , ou seja  $\mathcal{M} := \{(x_i, y_i) \in Z : y_i(w \cdot x_i + b) < \gamma_f \|w\|_q\}$ .

O perceptron de margem fixa com norma  $p$  (*Fixed  $p$ -Margin Perceptron* – FMP $_p$ ) é um algoritmo de minimização desta função de erro  $J$ . Seguindo a abordagem do gradiente local, o processo de minimização se inicia com valor  $(w^0, b^0)$ , normalmente  $(0, 0)$ . Em cada



iteração  $t$  do algoritmo, um par  $z_i = (x_i, y_i)$  é escolhido e verificado em  $(w^t, b^t)$ . Se este par  $z_i$  estiver errado, ou seja, se  $y_i(w^t \cdot x_i + b^t) < \gamma_f \|w^t\|_q$ , então um novo vetor normal  $w^{t+1}$  e constante  $b^{t+1}$  são calculados a partir do gradiente de  $J$  (VILLELA; LEITE; FONSECA NETO, 2016). Quando  $1 < q < \infty$ , tem-se a seguinte equação para o FMP $_p$  quando  $z_i$  estiver errado:

$$w^{t+1} \leftarrow w^t - \eta (\gamma_f \|w^t\|_q^{1-q} |w^t|^{q-1} \text{sinal}(w^t) - y_i x_i) \quad (2.7)$$

$$b^{t+1} \leftarrow b^t + \eta y_i, \quad (2.8)$$

onde  $\eta \in (0, 1]$  é a taxa de aprendizado,  $|w| := (|w_1|, \dots, |w_d|)'$ , e  $\text{sinal}(w) := (\text{sinal}(w_1), \dots, \text{sinal}(w_d))'$ . Define-se a norma euclidiana  $L_2$ , quando  $p = q = 2$  cuja formulação é apresentada por Leite e Fonseca Neto (2008):

$$w^{t+1} \leftarrow w^t (1 - (\eta \gamma_f) / \|w\|_2) + \eta \cdot x_i y_i. \quad (2.9)$$

Define-se também a norma  $L_\infty$ , onde  $p = \infty$ , correspondente a  $q = 1$ , a equação de atualização  $w^t$  é dada por:

$$w^{t+1} \leftarrow w^t - \eta (\gamma_f \text{sinal}(w^t) - y_i x_i). \quad (2.10)$$

### 2.2.1 Margem flexível

Em casos onde o conjunto de treinamento é não linearmente separável, é importante introduzir o conceito de flexibilização da margem ao FMP $_p$ . A estratégia para tal consiste em adaptar o método introduzido por Schölkopf e Smola (2002) para SVM, que considera a penalização quadrática das variáveis de folga associadas às restrições de classificação, cuja formulação para norma euclidiana é:

$$\begin{aligned} \min \quad & \frac{1}{2} w w^t + C \sum_i \varepsilon_i^2 \\ \text{s.a.} \quad & y_i (w \cdot x_i + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \text{ para todo } i \in \{1, \dots, m\}, \end{aligned} \quad (2.11)$$

onde  $\varepsilon$  é o vetor de variáveis de folga e  $C$  é um parâmetro de controle. De acordo com Schölkopf e Smola (2002), é possível reformular o problema de minimização definido na Equação (2.11) em um problema dual equivalente:

$$\begin{aligned} \max \quad & W(\alpha) = \sum_i \varepsilon_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_i \alpha_i^2 / 2C \\ \text{s.a} \quad & \sum_i \alpha_i y_i = 0, \alpha_i \geq 0 \text{ for all } i \in \{1, \dots, m\}. \end{aligned} \quad (2.12)$$

Fazendo  $\lambda = 1/2C$ , é possível resolver o problema da SVM com margem flexível através da soma deste valor constante à diagonal da matriz de produtos internos. Assim, reescreve-se a matriz da seguinte forma:

$$\begin{aligned} K(x_i, x_j) &= \langle x_i, x_j \rangle, \text{ para todo } j \neq i \\ K(x_i, x_j) &= \langle x_i, x_j \rangle + \lambda, \text{ caso contrário.} \end{aligned} \quad (2.13)$$

Considerando a representação expandida do vetor normal  $w = \sum_j \alpha_j y_j x_j$ ,  $\alpha_j \geq 0$ , pode-se realizar a mesma substituição no FMP $_p$ , para expressar as inequações em função do conjunto de multiplicadores  $\alpha$ . Assim, a Equação (2.5), em sua forma dual, torna-se:

$$y_i \left( \sum_{j=1}^m \alpha_j y_j K(x_i, x_j) + b \right) \geq \gamma_g \|w\|_q, \quad (2.14)$$

e para o caso flexível, introduz-se as variáveis de folga:

$$y_i \left( \sum_{j=1}^m \alpha_j y_j K(x_i, x_j) + b \right) + \alpha_i \lambda \geq \gamma_g \|w\|_q. \quad (2.15)$$

Assim, pode-se estimar o valor das variáveis de folga  $\alpha_i$ . Quanto maior o valor de  $\lambda$ , maior será a flexibilização das restrições. Utilizando esta ideia, acrescentam-se estas variáveis de folga à forma primal (Equação 2.5) do FMP $_p$ :

$$y_i (w \cdot x_i + b) \geq \gamma_f \|w\|_q - \alpha_i \lambda, \text{ para todo } (x_i, y_i) \in Z. \quad (2.16)$$

A variável  $\alpha$  requer correção a cada checagem. Para a norma euclidiana  $L_2$ ,

quando houver um erro associado a uma amostra  $x_i$ , a correção é dada por:

$$\alpha_i = \alpha_i + \eta \cdot 1, \quad (2.17)$$

sendo esta correção seguida de um escalonamento no vetor de multiplicadores  $\alpha$ :

$$\alpha^{t+1} \leftarrow \alpha^t (1 - (\eta\gamma_f) / \|w\|_2), \quad (2.18)$$

que, de acordo com Leite e Fonseca Neto (2008), equivale a correção feita pela Equação (2.9). Para uma norma arbitrária  $p$  é realizado o escalonamento do vetor  $\alpha$  baseado na respectiva norma conjugada  $q$ :

$$\alpha^{t+1} \leftarrow \alpha^t (1 - (\eta\gamma_f) / \|w\|_q). \quad (2.19)$$

## 2.3 Algoritmo de margem incremental com norma $p$ – IMA $_p$

Este classificador de larga margem, proposto por Villela, Leite e Fonseca Neto (2016), é uma extensão do IMA (LEITE; FONSECA NETO, 2008), utiliza um processo de maximização da margem iterativo a fim de evitar os altos custos computacionais que SVMs possuem em conjuntos de dados grandes. O problema de maximização sob uma  $p$ -margem é definido como o seguinte:

$$\begin{aligned} \max \quad & \gamma_g \\ \text{s.a} \quad & y_i(w \cdot x_i + b) \geq \gamma_g \|w\|_q. \end{aligned} \quad (2.20)$$

Para solucionar este problema, o algoritmo resolve sucessivos problemas de FMP $_p$ , com um aumento incremental da margem. Desta forma, seja  $\gamma_f^k$  a margem da iteração  $k$ , o algoritmo garante que  $\gamma_f^{k+1} > \gamma_f^k$  até uma iteração final  $K$  onde  $\gamma_f^K$  atinge um valor próximo ao máximo. Para calcular a margem da próxima iteração, são utilizadas duas regras. A primeira baseia-se no fato de que a solução ótima possui margens  $\gamma_+^*$  e  $\gamma_-^*$ , relativas aos hiperplanos adicionais, iguais. Portanto, quando as margens são diferentes,

calcula-se a nova margem através das seguintes equações:

$$\begin{aligned}\gamma_+^k &= \min y_i(w^k \cdot x_i + b^k) / \|w^k\|_q, \text{ para todo } x_i \in Z^+ \\ \gamma_-^k &= \min y_i(w^k \cdot x_i + b^k) / \|w^k\|_q, \text{ para todo } x_i \in Z^- \\ \gamma_f^{k+1} &= \frac{\gamma_+^k + \gamma_-^k}{2}.\end{aligned}\tag{2.21}$$

A segunda regra garante o incremento da margem quando a  $k$ -ésima solução do FMP $_p$  possui margens  $\gamma_+^k$  e  $\gamma_-^k$  iguais, pois não é possível afirmar que esta é a solução ótima. Para o cálculo da nova margem, utiliza-se a seguinte equação:

$$\gamma_f^{k+1} = (1 + \Delta)\gamma_f^k,\tag{2.22}$$

onde  $\Delta \in (0, 1)$  é um valor de incremento mínimo.

## 3 Seleção de características

A seleção de características tem como objetivo obter um subconjunto de características que terão maior relevância ao classificador. Realizar esta seleção manualmente normalmente gera melhores resultados, mas é inviável em grandes conjuntos de dados, pois existem muitas amostras e características, tornando difícil identificar visualmente as relações entre as características (FACELI et al., 2011). Portanto é necessário utilizar uma técnica automática, ou seja, um algoritmo, para realizar a seleção. O processo de seleção de características geralmente segue uma das seguintes abordagens: filtro, embutido ou envoltório.

Os métodos em filtro ordenam as características de acordo com algum critério, normalmente uma medida estatística. O ordenamento não utiliza informações de classificação, portanto independe do classificador. Por utilizar medidas estatísticas, a métrica observa individualmente cada característica, desconsiderando quaisquer relações entre elas. Nos métodos embutidos o próprio classificador escolhe o melhor subconjunto de características através de um mecanismo interno de avaliação. Já os métodos envoltórios utilizam um algoritmo de classificação como forma de avaliar a qualidade do subconjunto de características, normalmente acompanhados de uma estratégia heurística de busca para selecionar os subconjuntos mais promissores.

### 3.1 *Admissible Ordered Search* – AOS

Este método envoltório de seleção de características, proposto por Villela (2011), utiliza um classificador de larga margem para avaliar os subconjuntos de características, e consiste em realizar uma busca ordenada *best-first* pelo espaço de subconjuntos de características a fim de encontrar, para cada dimensão do problema, a melhor margem. O subconjunto de características que gerou esta margem é um conjunto mínimo de acordo com o critério de maximização da margem. Portanto, não existe outro subconjunto de mesmo tamanho que gere uma margem maior. Para garantir a admissibilidade do algoritmo, no entanto, é necessário que o algoritmo possua uma função de avaliação monótona (HART; NILSSON;

RAPHAEL, 1968), propriedade que só é garantida quando há separabilidade dos dados. Como o objetivo é maximizar a margem, esta função deve ser monótona decrescente. A admissibilidade do processo é mantida devido ao fato de que os valores da margem são sempre decrescentes quando a dimensão do problema é reduzida.

A estratégia de controle para o algoritmo *best-first* se dá pela inserção dos candidatos (subconjuntos) em uma fila de prioridade utilizando uma *heap*, ordenada pelo valor da margem. Para evitar a redundância dos candidatos, é utilizado uma tabela *hash*.

Como o valor a ser avaliado é a margem, seria necessário resolver, para todo candidato gerado, um problema de maximização da margem. Para contornar este problema, é feita uma estimativa da margem, de forma a manter a admissibilidade. Para isto, é feita a projeção da margem utilizando a mesma direção do vetor normal correspondente à remoção de uma característica do candidato pai. Para isto, define-se o vetor da margem geométrica com norma arbitrária  $p$  no espaço  $\mathbb{R}^d$  como:  $\gamma_g^d \cdot w / \|w\|_p$ . Seja  $\gamma_{p_j}^{d-1}$  a margem projetada no espaço  $\mathbb{R}^{d-1}$  de um candidato com  $d - 1$  características cuja  $j$ -ésima característica é removida. O valor é calculado através da seguinte equação:

$$\gamma_{p_j}^{d-1} = \frac{\gamma_g^d}{\|w\|_p} \left( \sum_{k \neq j} w_k^p \right)^{\frac{1}{p}}. \quad (3.1)$$

A cada iteração do algoritmo, quando um candidato é escolhido para avaliação e expansão, ou seja, gerar novos candidatos, existem duas possíveis situações. A primeira ocorre quando a margem do candidato escolhido for projetada. É necessário calcular o valor de margem real, e compará-la com a margem do próximo candidato na *heap*. Se ela ainda for maior, o candidato é removido e expandido. Se não, o seu valor de margem é atualizado e o candidato é reinserido na *heap*. No segundo caso, quando o candidato já possui uma margem real, o candidato é removido e expandido.

Para evitar a explosão combinatória decorrente da expansão dos candidatos, são utilizadas duas estratégias. A primeira controla o número de filhos gerados por um candidato durante a expansão, utilizando a estratégia de remoção das características que gerem as menores variações na função de avaliação.

A segunda estratégia realiza a poda da *heap* sempre que um candidato atinge

---

uma dimensão menor que a mínima atual, realizando a atualização do valor mínimo de margem e removendo os candidatos que estão abaixo do novo valor, além de remover os candidatos que possuem dimensão maior que a mínima de acordo com o parâmetro de profundidade de poda.

## 4 Abordagem proposta

Dado que verificar todas as possíveis combinações de características é inviável para conjuntos de dados com muitas características, e a projeção das características para o espaço *kernel* pode prejudicar o processo de seleção de características, propõe-se que os conjuntos de dados passem por um processo que os tornem linearmente separáveis, a fim de manter a admissibilidade do processo de seleção de características do AOS.

Para realizar a linearização dos conjuntos de dados, é necessário utilizar um classificador que permita o relaxamento das restrições de classificação. Para tal, foram utilizados os algoritmos  $\text{IMA}_p$  e otimização sequencial mínima (*Sequential Minimal Optimization – SMO*) (PLATT, 1998). Estes algoritmos foram executados utilizando valores distintos de flexibilização de forma a obter uma solução com margem flexível. Com esta solução, através do hiperplano  $(w, b)$  e da margem  $\gamma_g$ , torna-se possível identificar e remover as amostras do conjunto de treinamento cuja existência impedem que a base seja linearmente separável. Isto é possível baseando-se na Equação (2.1), que define a condição de existência da solução no caso da margem rígida. Assim, para encontrar quais amostras devem ser removidas, é feita a seguinte reformulação da equação:

$$y_i(w \cdot x_i + b) < 0 \quad (x_i, y_i) \in Z, \quad (4.1)$$

onde as amostras  $(x_i, y_i)$  que satisfazem esta equação estão do lado errado do hiperplano. A Figura 4.1 ilustra esta ideia. À esquerda, tem-se uma solução com margem flexível, onde as amostras em destaque satisfazem a Equação (4.1) e portanto devem ser removidas, tornando o conjunto resultante, à direita, linearmente separável.

Para avaliar a efetividade deste método, foram executados diversos testes para se definir os parâmetros ideais para cada um dos algoritmos. Para os critérios de avaliação, foram utilizadas a proporção de amostras removidas, a taxa percentual de erro pelo método de validação cruzada *k-fold cross validation* (KOHAVI, 1995) e a margem encontrada. Através destes critérios, selecionou-se o conjunto de parâmetros ideal, e os



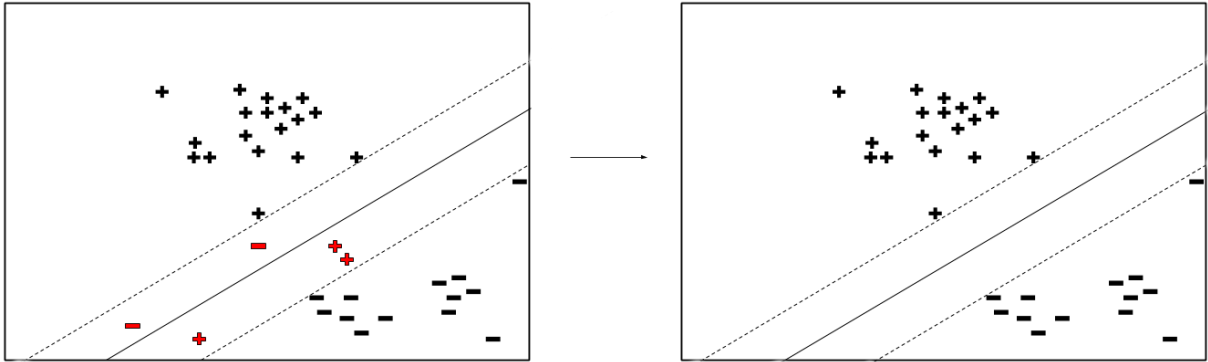


Figura 4.1: Exemplo do processo de linearização.

conjuntos de treinamentos gerados por ele foram utilizados para realizar a seleção de características.

No segundo passo, os conjuntos de dados foram separados pseudoaleatoriamente em subconjuntos de treino e teste, com  $2/3$  e  $1/3$  das amostras, respectivamente. O subconjunto de treino passa pelo processo de linearização, para o qual foram utilizados o  $\text{IMA}_\infty$  e o SMO linear. Em seguida, o problema passa por um algoritmo de programação linear e resolvido com o auxílio da ferramenta CPLEX<sup>1</sup>, como visto em Villela et al. (2019), de forma a destacar as componentes mais importantes do vetor normal do hiperplano  $w$ . Com a solução obtida, é possível identificar características com valor nulo, e portanto não contribuem no poder de generalização do conjunto de treinamento. Estas características são então removidas, a fim de simplificar o processo de seleção de características do AOS.

No terceiro passo é utilizado o algoritmo AOS para realizar seleção do melhor subconjunto de características. O classificador utilizado pelo AOS é o mesmo que realizou a linearização do subconjunto de treino. Por fim, avalia-se a solução encontrada pelo AOS utilizando o método de validação cruzada  $k$ -fold sobre o subconjunto de treino, seguido de uma validação final onde o classificador é treinado com o subconjunto de treino e sua solução é avaliada com o subconjunto de teste, previamente definidos. A Figura 4.2 apresenta um fluxograma da abordagem descrita nesta seção.

<sup>1</sup><https://www.ibm.com/analytics/cplex-cp-optimizer>

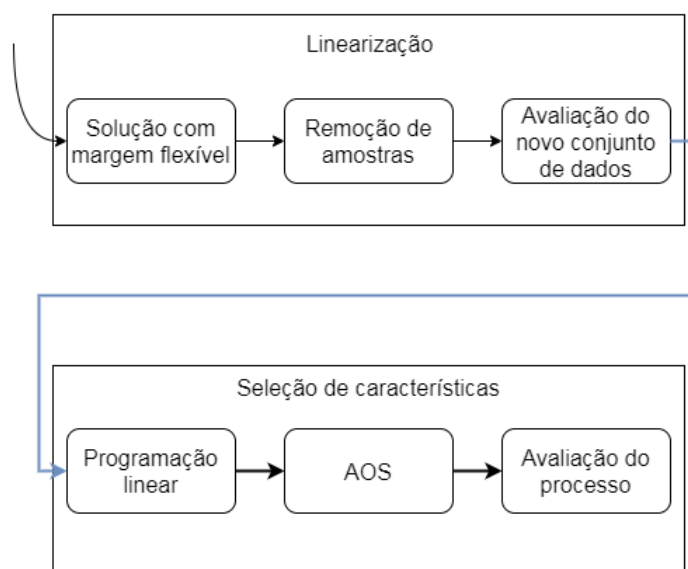


Figura 4.2: Fluxograma da abordagem proposta.

## 5 Experimentos e resultados

Neste capítulo são apresentados os conjuntos de dados e parâmetros utilizados nos experimentos, além dos resultados obtidos e comentários acerca destes resultados.

### 5.1 Dados

Para avaliar a abordagem proposta, foram utilizados nove conjuntos de dados considerados não linearmente separáveis, disponíveis no *UCI Machine Learning Repository* (BACHE; LICHMAN, 2013). A base *Sonar* é linearmente separável, porém classificadores SVM não conseguem encontrar uma solução linear para ela. Para a base *Wine*, que possui três classes, realizou-se a união das classes 1 e 3 para se trabalhar com uma classificação binária. Os detalhes sobre cada conjunto de dados são mostrados na Tabela 5.1. Os algoritmos usados para executar os testes foram implementados em C e os testes foram realizados em um Intel Core i5-3470 com 3,2 GHz e 8 GB de RAM.

Tabela 5.1: Informações dos conjuntos de dados.

Base	Atributos	Amostras		
		+1	-1	Total
Sonar	60	97	111	208
Ionosphere	34	225	126	351
WDBC	30	212	357	569
Bupa	6	145	200	345
Pima	8	268	500	768
Wine	13	107	71	178
Australian	14	307	383	690
German	24	300	700	1000
Heart	13	120	150	270

### 5.2 Parâmetros

Na etapa de linearização, para o  $\text{IMA}_p$ , foram utilizados os valores de norma  $p = 1$  ( $L_1$ ),  $q = 1$  ( $L_\infty$ ) e  $q = 2$  ( $L_2$ ). Foram escolhidas as menores potências da base 10 para a

flexibilização onde o treinamento é bem sucedido para cada base.

Para o SMO linear, foram selecionados os três menores potências na base 10 para a flexibilização – neste caso, o valor de flexibilização representa a rigidez da margem, sendo inversamente proporcional à flexibilização real, e o valor 10000 é considerado rígido – onde o treinamento é bem sucedido para cada base. Para a validação dos novos conjuntos de dados, foi utilizado o método de validação cruzada *k-fold*, realizando 10 execuções do *10-fold*, calculando o valor médio e o desvio padrão do erro para cada conjunto.

Já na etapa de seleção de características, os conjuntos de dados foram separados pseudoaleatoriamente, utilizando dez sementes distintas  $s \in \{0, 1, \dots, 9\}$  para todos os casos, em treino e teste com uma proporção de  $2/3$  e  $1/3$ , respectivamente, de forma a manter proporção original de amostras positivas e negativas. Para cada par de treino e teste gerado com uma das sementes  $s$ , executou-se o AOS sobre o conjunto de treino, utilizando os parâmetros de execução definidos por Villela et al. (2019), que são: fator de ramificação  $b = 5$ , profundidade de poda  $p = 2$  e profundidade do corte  $c = 5$ . Como objetivo foi definido o número de características (dimensão)  $n = 2$ . Como algoritmo de classificação acoplado ao AOS, foram utilizados o  $\text{IMA}_\infty$  e SMO linear, ambos com margem rígida. Para a validação dos resultados, foram realizadas dez execuções do *10-fold* sobre o conjunto de treino, e uma validação final onde o classificador é treinado com o conjunto de treino e sua solução é testada com conjunto de teste.

## 5.3 Resultados

Esta seção contém os resultados dos experimentos e é dividida em duas partes: primeiro apresentam-se os resultados referentes a parte da linearização dos conjuntos de dados, e em seguida apresentam-se os resultados da aplicação da seleção de características aos novos conjuntos.

### 5.3.1 Remoção de amostras

Nesta seção são apresentados três estudos acerca do processo de remoção de amostras, a fim de observar a eficácia dos algoritmos nesta abordagem. O primeiro estudo avalia

a quantidade percentual de amostras removidas por cada algoritmo. Os resultados são apresentados nas Tabelas 5.2 e 5.3, onde a colunas “ $\lambda$ ” e “ $C$ ” representam a flexibilização utilizada e as colunas “+1” e “-1” apresentam as porcentagens de amostras removidas das classes rotuladas como positivas e negativas, respectivamente. De forma geral, para ambos os algoritmos, quanto maior a flexibilização, maior o número de amostras removidas, pois o hiperplano não sofre grandes alterações. Os resultados encontrados utilizando o  $\text{IMA}_p$  (Tabela 5.2) para todas as normas são similares, e o algoritmo é capaz de encontrar, para a maioria dos casos, soluções com flexibilizações baixas. É notório que o SMO linear (Tabela 5.3) necessita de valores de flexibilização muito altos (para o SMO linear, quanto menor o valor de  $C$ , maior é a flexibilização) para sequer encontrar uma solução, resultando na remoção de muitas amostras, e na eventual remoção de uma classe inteira. Além disso, diversos subconjuntos gerados pelo SMO linear, demarcados por ‘\*’ na Tabela 5.3, apresentaram erros ao serem utilizados posteriormente.

Visto que o  $\text{IMA}_p$  obteve resultados melhores na remoção de amostras, realizou-se um estudo acerca da margem obtida após a remoção das amostras (Tabela 5.4), onde “ $\lambda$ ” define a flexibilização utilizada, e as colunas “ $\text{IMA}_1$ ”, “ $\text{IMA}_\infty$ ” e “ $\text{IMA}_2$ ” indicam o algoritmo utilizado na linearização. Para realizar uma comparação justa, todas as margens foram calculadas utilizando a norma  $L_2$ , em uma segunda execução do classificador, após a remoção das amostras, sem flexibilização. Nota-se que a margem encontrada pelo conjunto gerado pela remoção utilizando norma  $L_\infty$  é maior em 47,5% dos casos.

Por último, verificou-se o erro de treinamento associado à utilização dos conjuntos obtidos após a remoção das amostras (Tabela 5.5), onde “ $\lambda$ ” representa a flexibilização utilizada, e as colunas “ $\text{IMA}_1$ ”, “ $\text{IMA}_\infty$ ” e “ $\text{IMA}_2$ ” indicam os erros de validação cruzada sobre os conjuntos linearizados pelos respectivos algoritmos. Para obter uma comparação justa, todos os  $k$ -folds foram executados utilizando a norma  $L_2$ , e a remoção dos pontos ocorreu apenas nos conjuntos de treinamento gerados pelo  $k$ -fold. A validação é feita em uma segunda execução do classificador, após a remoção das amostras, sem flexibilização. As normas  $L_1$  e  $L_\infty$  obtiveram resultados similares, com 41% dos menores erros cada. De forma geral, observa-se que o erro é menor para as menores flexibilizações.

Após estes três estudos, conclui-se que os conjuntos gerados pelo  $\text{IMA}_\infty$  são, em

Tabela 5.2: Percentual de amostras removidas para cada parâmetro –  $IMA_p$ .

Base	$\lambda$	$IMA_1$		$IMA_\infty$		$IMA_2$	
		+1	-1	+1	-1	+1	-1
Sonar	0,0001	1,03	1,80	9,28	7,21	1,03	2,70
	0,001	9,28	4,50	13,40	17,12	6,19	6,31
	0,01	12,37	10,81	13,40	17,12	11,34	13,51
Ionosphere	0,00001	5,33	11,11	5,33	11,11	5,33	11,11
	0,0001	5,33	10,32	4,89	9,52	5,33	10,32
	0,001	4,00	8,73	5,33	11,90	6,22	10,32
	0,01	5,33	14,29	5,33	12,70	4,89	11,90
	0,1	4,89	11,90	4,00	14,29	6,67	11,90
WDBC	1	6,13	2,24	5,19	2,80	6,13	2,52
	10	6,13	3,64	8,49	3,92	6,60	3,64
	100	9,43	3,64	9,91	4,76	9,43	3,36
	1000	10,85	3,64	13,21	4,20	10,85	4,48
	10000	13,21	4,48	16,04	3,92	14,15	4,20
Bupa	0,01	46,90	25,50	46,90	25,50	46,90	25,50
	0,1	46,90	26,50	46,90	26,50	46,90	26,50
	1	46,90	32,00	46,90	32,00	46,90	32,00
	10	37,24	34,50	48,28	23,50	35,86	35,00
	100	69,66	6,50	69,66	6,50	69,66	6,50
Pima	1	32,84	20,20	38,06	14,60	34,33	18,00
	10	26,87	24,60	36,57	20,20	26,87	24,60
	100	46,27	19,20	51,12	15,60	46,27	19,20
	1000	67,16	12,20	78,73	8,00	69,78	11,40
Wine	1	5,61	2,82	4,67	2,82	3,74	4,23
	10	4,67	4,23	5,61	7,04	5,61	2,82
	100	6,54	4,23	13,08	15,49	5,61	14,08
Australian	10	20,85	12,01	18,24	12,01	17,92	12,01
	100	24,10	12,01	23,13	10,18	27,69	9,14
	1000	32,90	14,10	47,23	11,75	33,22	11,75
	10000	42,02	12,27	46,91	21,15	46,25	11,23
	100000	54,40	12,79	55,70	12,27	53,75	13,84
German	0,1	45,33	13,71	50,67	10,29	45,33	13,71
	1	48,67	10,29	56,33	10,86	57,67	7,86
	10	53,33	8,57	61,67	9,14	58,67	6,29
	100	51,00	12,14	73,33	7,71	63,33	6,86
	1000	94,00	0,29	100,00	0,00	91,00	0,29
Heart	0,1	18,33	15,33	20,83	11,33	20,83	11,33
	1	18,33	12,67	16,67	14,00	20,00	12,67
	10	16,67	16,00	22,50	10,00	15,00	18,00
	100	19,17	14,67	30,83	18,67	20,83	11,33
	1000	25,00	20,67	33,33	21,33	32,50	15,33

Tabela 5.3: Percentual de amostras removidas para cada flexibilização – SMO linear.

Base	$C$	+1	-1
Sonar	100	5,15	3,60
	10	15,46	9,91
	1	23,71	9,01
Ionosphere	10	2,67	11,11
	1	2,36	6,16
	0,1	3,45	13,50
WDBC	0,0001	11,32	2,80
	0,00001	15,09	2,80
	0,000001	18,40	1,68
Bupa	0,001	54,48	13,00
	0,0001	98,62	0,00
	0,00001*	100,00	0,00
Pima	0,0001	46,64	10,60
	0,00001*	55,60	7,40
	0,000001*	100,00	0,00
Wine	0,0001	18,69	26,76
	0,00001	19,63	29,58
	0,000001	21,50	28,17
Australian	0,00001	70,03	5,22
	0,000001	71,34	5,48
	0,0000001	72,96	4,70
German	0,001*	100,00	0,00
	0,0001	–	–
	0,00001	–	–
Heart	0,001*	33,33	12,67
	0,0001*	46,67	16,67
	0,00001	69,17	8,00

maioria, melhores que as demais normas, e portanto será o classificador utilizado durante o processo de seleção de características.

### 5.3.2 Seleção de características

Nesta seção apresenta-se os resultados do processo de seleção de características utilizando o algoritmo AOS. Inicialmente, o conjunto de amostras original é dividido, de forma pseudoaleatória com sementes  $s \in \{0, 1, \dots, 9\}$ , em treino e teste com proporção 2/3 para 1/3, respectivamente, mantendo a proporção de amostras positivas e negativas originais. O processo de remoção de amostras é executado apenas no subconjunto de treino, utilizando

Tabela 5.4: Comparação das margens  $L_2$ .

Base	$\lambda$	IMA <sub>1</sub>	IMA <sub>∞</sub>	IMA <sub>2</sub>
Sonar	0,0001	0,002666	0,003168	0,003168
	0,001	0,005825	0,036859	0,007758
	0,01	0,020842	0,036590	0,020817
Ionosphere	0,00001	0,047199	0,047315	0,047315
	0,0001	0,046035	0,028968	0,046035
	0,001	0,028242	0,045154	0,040747
	0,01	0,053539	0,064300	0,064505
	0,1	0,032810	0,062831	0,044311
WDBC	1	0,041094	0,082526	0,100658
	10	0,102200	0,305864	0,138838
	100	0,312630	0,479125	0,237067
	1000	0,484353	1,245431	0,641834
	10000	1,195426	2,303993	2,288650
Bupa	0,01	0,180399	0,180399	0,180399
	0,1	0,162614	0,162614	0,162614
	1	0,159541	0,159541	0,159541
	10	0,090100	0,372075	0,262975
	100	0,272163	0,272163	0,272163
Pima	1	0,038539	0,077619	0,056439
	10	0,090885	0,071095	0,090885
	100	0,068942	0,416894	0,068942
	1000	0,255041	0,449659	0,313873
Wine	1	0,306495	0,201961	0,255745
	10	0,267323	0,463979	0,362821
	100	0,428945	0,742119	0,499833
Australian	10	0,034071	0,036474	0,055163
	100	0,039781	0,061173	0,048402
	1000	0,286460	0,618778	0,286612
	10000	0,253448	0,873314	0,525091
	100000	2,972245	2,293152	2,584231
German	0,1	0,050400	0,033684	0,050400
	1	0,024233	0,040757	0,053931
	10	0,033439	0,124740	0,039934
	100	0,039681	0,133323	0,103621
	1000	0,452118	–	0,225582
Heart	0,1	0,093748	0,174721	0,184145
	1	0,187267	0,186738	0,191446
	10	0,235299	0,161451	0,179177
	100	0,305457	0,489569	0,194520
	1000	0,565285	0,818577	0,433618



Tabela 5.5: Média e desvio padrão de 10 execuções do 10-*fold* na norma  $L_2$ .

Base	$\lambda$	IMA <sub>1</sub>	IMA <sub><math>\infty</math></sub>	IMA <sub>2</sub>
Sonar	0,0001	25,20 (2,55)	25,89 (2,09)	25,65 (2,52)
	0,001	25,05 (2,26)	24,67 (1,75)	25,67 (3,33)
	0,01	26,02 (2,65)	23,19 (1,52)	26,30 (1,99)
Ionosphere	0,00001	12,16 (1,19)	12,50 (1,21)	12,56 (1,21)
	0,0001	13,65 (1,17)	13,39 (1,44)	13,91 (0,99)
	0,001	12,96 (1,30)	13,11 (0,79)	13,33 (1,42)
	0,01	12,96 (0,84)	12,56 (0,67)	13,02 (1,25)
	0,1	12,59 (0,65)	12,93 (1,13)	12,65 (0,85)
WDBC	1	4,84 (0,36)	4,66 (0,46)	4,98 (0,20)
	10	5,03 (0,47)	5,87 (0,27)	5,13 (0,28)
	100	6,52 (0,28)	6,94 (0,35)	6,47 (0,41)
	1000	6,96 (0,36)	7,75 (0,33)	7,21 (0,48)
	10000	7,89 (0,39)	7,81 (0,36)	7,98 (0,24)
Bupa	0,01	32,75 (4,46)	31,18 (5,24)	34,78 (3,78)
	0,1	33,71 (2,15)	32,17 (2,89)	33,38 (2,08)
	1	35,67 (0,87)	35,42 (1,33)	35,36 (1,36)
	10	33,53 (1,76)	35,12 (1,74)	33,56 (1,95)
	100	34,17 (1,90)	33,47 (1,40)	34,17 (1,44)
Pima	1	23,63 (1,64)	22,84 (1,79)	23,84 (1,49)
	10	24,20 (3,54)	24,32 (1,87)	24,94 (1,84)
	100	30,77 (0,67)	30,30 (0,54)	30,79 (1,16)
	1000	33,67 (0,87)	33,38 (0,61)	33,53 (0,75)
Wine	1	6,30 (0,83)	6,01 (0,42)	6,74 (0,85)
	10	6,74 (0,62)	6,80 (0,71)	6,12 (1,39)
	100	7,81 (1,05)	9,97 (1,16)	8,26 (1,05)
Australian	100	17,70 (1,54)	15,75 (2,42)	16,51 (1,93)
	1000	21,26 (0,58)	25,19 (1,54)	20,71 (0,62)
	10000	24,16 (0,94)	29,87 (0,79)	20,70 (0,64)
	100000	30,72 (0,65)	31,20 (0,61)	31,20 (0,82)
German	0,1	24,57 (0,51)	24,05 (2,40)	24,66 (0,55)
	1	24,60 (0,78)	24,93 (0,79)	24,60 (0,78)
	10	24,75 (0,32)	25,85 (0,62)	24,22 (0,58)
	100	25,35 (0,47)	27,70 (1,51)	25,45 (0,60)
	1000	27,70 (0,72)	–	28,29 (0,56)
Heart	0,1	16,15 (1,65)	16,41 (1,43)	17,26 (1,04)
	1	16,44 (1,42)	16,78 (1,40)	15,93 (1,31)
	10	17,07 (1,19)	16,89 (1,19)	16,52 (0,94)
	100	16,41 (1,18)	20,70 (1,02)	16,93 (1,10)
	1000	21,56 (1,35)	25,56 (0,94)	23,15 (0,94)

o IMA <sub>$\infty$</sub>  e o SMO linear com as flexibilizações que não apresentaram erros. O subconjunto de treino gerado pela remoção de amostras passa por um algoritmo de programação linear

e resolvido, como visto em Villela et al. (2019), cuja solução permite identificar características com dimensão nula, que serão removidas para auxiliar o AOS. Os parâmetros de execução do AOS são definidos na seção 5.2, e o classificador acoplado ao AOS é o mesmo utilizado no processo de remoção de pontos, porém utilizando margem rígida, de forma a garantir a admissibilidade do processo. Após a seleção de características, realizou-se dez execuções do *10-fold* utilizando o próprio algoritmo que foi acoplado ao AOS, com a norma  $L_2$  no caso o  $\text{IMA}_p$ . As amostras removidas tiveram sua dimensão igualada à obtida pelo AOS e foram reinseridas no conjunto de treino, de forma a realizar o *10-fold* com todas as amostras originais, com dimensão reduzida. Assim, o *fold* possui uma etapa onde seus conjuntos de treino passam pelo processo de linearização, e o teste é realizado numa segunda execução do classificador, sem flexibilização. Como critérios de avaliação são utilizados o erro no *10-fold* e a dimensão alcançada pelo AOS.

As Tabelas 5.6 e 5.7 apresentam os resultados do processo para cada uma das dez sementes utilizando o  $\text{IMA}_\infty$ . A coluna “ $n$ ” representa a dimensão alcançada na solução do AOS, a coluna “Margem” representa a margem obtida pela solução do AOS, “*10-fold*” representa o erro de validação cruzada sobre o conjunto de treino, e “Teste” representa o erro da validação final. Para a maioria das bases o algoritmo encerrou a execução antes de atingir o objetivo ( $n = 2$ ), indicando que o conjunto de treino utilizado retorna ao estado não linearmente separável, uma vez que o processo de geração de novos candidatos é condicionado à existência do hiperplano.

A Tabela 5.8 contém os valores médios e o desvio padrão dos resultados obtidos na seleção de característica nas dez sementes utilizando o  $\text{IMA}_\infty$ . A coluna “ $\lambda$ ” representa a flexibilização utilizada para linearizar o conjunto de treino, “Atrib.” representa a dimensão original da base, “Atrib. PL” representa a dimensão obtida pela aplicação da programação linear, “Atrib. AOS” representa a dimensão da solução obtida pelo AOS, “*10-fold*  $L_2$ ” representa o erro de validação cruzada sobre o conjunto de treino e “Teste  $L_2$ ” representa o erro de validação final. Para as bases Bupa e Pima, nota-se uma dificuldade em remover características utilizando esta técnica de remoção de amostras, principalmente em conjuntos onde foi utilizada uma flexibilização baixa. Para as demais, a utilização da programação linear para remover características nulas obteve resultados

Tabela 5.6: Resultados do IMA<sub>∞</sub> (parte 1).

Base	$\lambda_1$				$\lambda_2$				$\lambda_3$				
	<i>n</i>	Margem	10-fold	Teste	<i>n</i>	Margem	10-fold	Teste	<i>n</i>	Margem	10-fold	Teste	
Sonar	0,0001				0,001				0,01				
	14	0,000019	18,60 (2,72)	27,14	11	0,000967	18,87 (1,66)	28,57	9	0,002224	20,73 (1,98)	30,00	
	9	0,000971	18,91 (2,25)	32,86	8	0,000720	16,99 (1,40)	27,14	8	0,003788	18,70 (2,27)	34,29	
	11	0,000996	16,38 (1,67)	21,43	9	0,000636	20,80 (2,65)	32,86	7	0,003188	15,93 (1,91)	28,57	
	13	0,000646	16,40 (2,65)	34,29	13	0,000639	17,25 (1,66)	32,86	10	0,000089	21,93 (2,12)	25,71	
	12	0,000195	18,25 (2,20)	28,57	9	0,001644	15,69 (1,50)	25,71	8	0,002613	17,32 (1,86)	24,29	
	10	0,000656	18,25 (1,84)	30,00	7	0,001411	17,60 (1,25)	25,71	8	0,002088	20,51 (1,92)	28,57	
	14	0,000700	11,01 (1,76)	32,86	9	0,001520	12,77 (1,79)	40,00	7	0,005014	13,95 (1,43)	47,14	
	12	0,001030	22,98 (3,31)	17,14	10	0,000339	21,14 (2,10)	20,00	7	0,001517	25,24 (1,99)	17,14	
	13	0,000729	19,24 (2,62)	24,29	11	0,000664	17,75 (2,25)	28,57	7	0,001598	19,89 (2,05)	28,57	
	16	0,000370	21,32 (2,33)	22,86	9	0,000231	18,13 (2,36)	20,00	8	0,000994	21,97 (3,55)	25,71	
	Ionosphere	0,001				0,01				0,1			
		11	0,001999	9,61 (1,44)	13,68	11	0,002007	9,09 (0,82)	17,09	10	0,004171	8,64 (1,02)	11,97
		12	0,002015	9,91 (0,92)	12,82	12	0,002733	10,58 (0,76)	14,53	11	0,008089	8,48 (0,82)	13,68
		13	0,004662	11,82 (1,98)	15,38	11	0,002697	10,09 (1,54)	12,82	10	0,005107	10,73 (0,69)	13,68
		16	0,000932	10,89 (0,73)	18,80	11	0,004909	11,82 (1,39)	24,79	8	0,003106	10,37 (1,15)	16,24
13		0,003918	8,16 (0,90)	17,09	13	0,005155	8,55 (1,18)	18,80	11	0,002338	8,92 (1,10)	12,82	
13		0,001806	9,45 (1,22)	11,11	12	0,001538	9,83 (1,09)	17,09	11	0,003650	9,65 (1,12)	13,68	
12		0,003441	11,15 (0,86)	11,11	12	0,003284	11,38 (1,29)	16,24	11	0,003508	11,29 (0,98)	11,97	
13		0,004723	12,17 (1,62)	15,38	12	0,000831	11,75 (1,48)	9,40	11	0,002648	11,64 (1,11)	9,40	
13		0,001722	9,23 (0,72)	13,68	13	0,003759	7,62 (0,81)	14,53	11	0,002648	10,47 (0,90)	13,68	
11		0,003714	9,81 (0,86)	13,68	11	0,003265	10,09 (1,03)	14,53	8	0,001851	11,26 (1,27)	14,53	
WDBC	100				1000				10000				
	5	0,964877	9,90 (0,76)	6,32	3	1,295579	11,63 (1,55)	8,42	3	0,370529	9,55 (0,84)	5,79	
	4	3,803786	9,35 (1,48)	13,16	4	3,728755	11,43 (0,60)	10,53	2	0,422627	11,45 (1,20)	16,32	
	3	1,544474	7,62 (0,68)	10,00	2	4,752303	9,73 (0,69)	12,63	2	7,635848	10,74 (1,55)	15,26	
	5	1,657783	9,45 (0,39)	7,37	4	0,075616	8,08 (0,50)	7,37	3	4,221090	13,38 (0,69)	16,32	
	4	0,566759	7,99 (0,45)	8,42	3	2,211155	10,97 (1,47)	11,05	3	2,323487	8,36 (0,44)	8,42	
	4	3,110976	6,61 (3,62)	9,47	4	0,801287	7,13 (0,53)	9,47	2	6,656849	12,69 (0,82)	12,63	
	4	1,203467	8,61 (1,52)	14,74	3	9,315007	9,21 (1,21)	14,21	2	2,181483	10,90 (1,32)	12,11	
	6	1,061203	8,33 (0,86)	7,37	3	3,888627	8,60 (0,34)	8,42	2	2,162734	10,61 (1,28)	15,26	
	5	0,140996	8,05 (0,41)	5,26	3	1,145969	10,45 (1,39)	8,95	2	1,482595	12,24 (0,67)	15,26	
3	1,240811	4,15 (1,87)	10,53	3	4,244278	12,24 (0,87)	12,63	2	7,048889	11,19 (1,16)	14,21		
Eupa	1				10				100				
	6	0,177433	35,61 (2,57)	35,65	3	0,025447	37,13 (1,82)	38,26	5	0,259313	32,83 (2,26)	36,52	
	6	0,163897	34,87 (1,91)	29,57	5	0,189076	36,87 (2,17)	33,04	4	1,338736	34,61 (1,88)	29,57	
	5	0,233698	37,48 (2,88)	32,17	4	0,437435	34,35 (3,05)	39,13	5	0,215538	32,83 (2,18)	35,65	
	5	0,055694	36,39 (2,29)	29,57	4	0,015060	35,48 (3,90)	34,78	3	0,134495	38,35 (2,73)	40,87	
	5	0,161060	37,00 (2,43)	34,78	4	0,129507	37,26 (1,73)	37,39	5	0,271459	33,65 (2,03)	36,52	
	6	0,215131	31,22 (2,98)	35,65	5	0,303770	33,09 (2,21)	37,39	5	0,110200	31,57 (2,16)	36,52	
	6	0,144613	33,39 (2,38)	42,61	5	0,228060	33,22 (2,25)	40,87	5	0,236721	32,22 (1,27)	38,26	
	5	0,221338	36,09 (2,39)	30,43	6	0,144406	32,61 (2,12)	33,91	5	0,368961	32,04 (2,01)	33,04	
	5	0,075227	33,87 (1,48)	33,91	5	0,178886	32,61 (2,04)	30,43	5	0,214738	30,78 (1,00)	31,30	
6	0,201143	37,30 (2,07)	40,00	5	0,101775	36,26 (1,97)	31,30	5	0,262522	35,39 (1,94)	37,39		
Pima	10				100				1000				
	6	0,120577	23,38 (2,05)	25,78	5	0,675237	30,21 (1,72)	35,55	5	0,451632	32,56 (1,17)	36,33	
	5	0,073441	25,35 (1,81)	27,73	5	0,926482	31,58 (1,67)	33,59	5	0,972786	33,91 (0,88)	32,81	
	7	0,045445	22,80 (3,55)	30,08	3	1,520595	31,94 (1,72)	32,81	5	1,731031	35,08 (0,89)	34,38	
	8	0,144942	24,19 (3,37)	23,44	5	0,203507	32,72 (1,95)	32,42	3	1,897754	35,12 (1,68)	34,38	
	8	0,083403	22,67 (4,22)	28,91	4	0,327918	31,30 (0,94)	33,20	4	1,262036	32,33 (0,73)	31,64	
	7	0,032144	25,16 (1,90)	26,95	5	0,022708	30,87 (1,58)	32,42	4	0,309922	31,42 (0,80)	33,20	
	6	0,003903	19,61 (2,80)	30,47	4	0,353621	29,44 (1,07)	34,38	3	1,609907	32,37 (0,86)	34,77	
	6	0,148002	21,19 (2,27)	32,03	5	0,616912	27,60 (2,39)	34,77	6	0,719175	31,70 (0,91)	33,98	
	6	0,044072	23,39 (1,40)	27,34	5	0,385514	31,90 (1,49)	33,59	4	0,207730	31,75 (1,22)	35,94	
7	0,024525	20,76 (5,14)	28,91	4	0,063644	31,51 (1,38)	35,16	5	1,711500	33,62 (0,74)	34,77		

satisfatórios, que foram melhorados com a utilização do AOS. Os melhores resultados são visíveis para as bases WDBC, Wine e Australian, onde o número de dimensões é igual ou próximo ao objetivo.

Tabela 5.7: Resultados do IMA<sub>∞</sub> (parte 2).

Base	$\lambda_1$			$\lambda_2$			$\lambda_3$					
	$n$	Margem	10-fold	Teste	$n$	Margem	10-fold	Teste	$n$	Margem	10-fold	Teste
	1			10			100					
Wine	3	0,095755	3,27 (1,01)	10,17	3	0,195143	5,22 (0,83)	6,78	4	0,171816	8,33 (1,66)	13,56
	3	0,085014	5,48 (0,93)	5,08	6	0,295601	7,90 (1,17)	10,17	5	0,052415	10,43 (1,37)	18,64
	3	0,009987	4,18 (0,67)	5,08	3	0,010244	5,77 (1,13)	6,78	5	0,150742	9,73 (1,24)	8,47
	4	0,112345	7,13 (2,35)	3,39	4	0,113979	7,80 (2,96)	0,00	3	0,005097	15,86 (1,86)	13,56
	3	0,235247	4,52 (0,83)	5,08	3	0,246181	9,26 (1,18)	6,78	3	0,524589	13,27 (1,84)	18,64
	3	0,099585	4,27 (1,50)	1,69	6	0,307151	9,66 (1,82)	3,39	6	0,279933	10,76 (1,58)	5,08
	4	0,015563	5,68 (0,53)	6,78	5	0,335719	7,06 (0,98)	3,39	4	0,144047	11,17 (1,48)	15,25
	3	0,053281	4,61 (0,81)	1,69	3	0,006879	5,55 (0,97)	6,78	3	0,031586	14,64 (2,31)	20,34
	3	0,082370	3,95 (1,02)	6,78	3	0,037541	6,71 (1,44)	10,17	3	0,037141	12,31 (2,25)	15,25
	3	0,147248	5,70 (0,87)	1,69	3	0,167087	11,64 (2,14)	5,08	5	0,681749	12,67 (2,09)	8,47
Australian	1000			10000			100000					
	7	0,290252	21,98 (1,18)	24,35	5	1,137020	30,07 (1,15)	26,96	3	2,049067	31,71 (1,62)	33,91
	9	0,295076	22,15 (0,78)	19,13	4	1,497518	29,87 (0,70)	26,96	2	5,946953	32,46 (0,89)	32,17
	7	0,268315	22,48 (0,65)	24,35	4	2,233211	28,39 (1,17)	32,61	2	15,905512	31,22 (1,32)	34,78
	6	0,395414	21,83 (3,03)	23,91	8	0,701854	27,20 (0,82)	28,70	3	3,362901	31,11 (1,25)	31,30
	5	0,212613	23,00 (1,66)	26,52	5	0,781367	28,78 (0,72)	32,17	2	8,921569	30,52 (1,50)	34,35
	8	0,099300	23,89 (1,44)	24,35	5	1,163951	29,09 (1,04)	26,09	2	13,800786	32,15 (1,47)	35,22
	3	1,038333	26,54 (1,28)	24,35	6	0,071908	28,02 (1,12)	32,17	3	2,399481	30,48 (0,25)	32,17
	4	0,132992	27,43 (0,90)	27,83	4	1,469122	32,52 (1,42)	28,26	3	5,966270	31,93 (0,76)	30,43
	9	0,372758	24,67 (1,51)	23,04	5	0,755971	30,57 (0,82)	26,96	3	2,041609	32,22 (1,95)	32,61
10	0,297151	22,50 (2,18)	21,30	3	1,161441	34,35 (1,13)	32,61	2	8,014706	33,33 (0,46)	30,43	
German	10			100			1000					
	9	0,003051	27,33 (0,76)	27,93	3	0,078013	16,72 (6,53)	28,83				
	9	0,036803	25,31 (0,59)	25,83	3	0,218446	26,97 (2,17)	29,13				
	9	0,002670	25,65 (1,46)	27,03	8	0,074584	26,49 (0,47)	27,33				
	6	0,052211	26,00 (1,06)	28,23	4	0,057049	28,43 (0,50)	28,53				
	12	0,004101	24,87 (0,79)	25,53	9	0,071830	26,61 (0,65)	27,03				
	10	0,004367	26,76 (0,65)	26,43	—	—	—	—				
	11	0,003394	24,35 (0,62)	27,93	8	0,020870	27,15 (0,78)	27,63				
	12	0,031918	25,29 (0,61)	24,02	5	0,085405	11,90 (5,73)	30,03				
	14	0,027489	24,89 (0,97)	24,62	8	0,010815	25,90 (0,80)	27,33				
14	0,012030	24,88 (0,66)	25,53	5	0,171392	27,23 (1,12)	30,63					
Heart	10			100			1000					
	9	0,078105	15,92 (1,39)	14,44	4	0,005181	20,81 (1,10)	18,89	5	0,422916	26,83 (1,79)	27,78
	8	0,079417	15,83 (1,29)	15,56	5	0,252721	18,44 (1,54)	15,56	5	0,715643	26,83 (1,28)	24,44
	5	0,034154	14,44 (1,41)	23,33	4	0,215652	18,50 (1,48)	27,78	5	0,329957	24,06 (0,98)	31,11
	5	0,016829	14,50 (0,76)	20,00	3	0,057652	19,83 (0,91)	24,44	5	0,359855	27,39 (1,72)	27,78
	7	0,043909	17,56 (1,68)	18,89	6	0,217640	20,56 (1,70)	16,67	5	0,181775	25,78 (1,87)	23,33
	5	0,002721	18,06 (1,80)	14,44	4	0,136570	19,06 (0,53)	17,78	5	0,373172	26,22 (1,25)	23,33
	4	0,056658	21,33 (1,18)	15,56	4	0,150647	24,00 (0,97)	15,56	4	0,370463	28,94 (1,35)	23,33
	6	0,023910	14,39 (1,09)	18,89	4	0,155439	25,44 (2,16)	20,00	4	1,242278	28,83 (1,56)	24,44
	7	0,027225	18,83 (0,61)	13,33	5	0,178357	20,67 (1,01)	24,44	3	0,552920	25,00 (0,98)	33,33
7	0,050356	15,33 (1,46)	20,00	4	0,228446	22,22 (1,14)	28,89	4	0,172923	23,61 (1,55)	34,44	

As Tabelas 5.9 e 5.10 apresentam os resultados obtidos para cada semente ao se utilizar os conjuntos de treino linearizados pelo SMO linear utilizando as flexibilizações que obtiveram solução (Tabela 5.3). A coluna “ $n$ ” representa a dimensão alcançada na solução do AOS, “ $C$ ” representa a flexibilização, “Margem” representa a margem obtida pela solução do AOS, “10-fold” representa o erro de validação cruzada sobre o conjunto de treino, e “Teste” representa o erro da validação final. Desta vez, um número ainda menor de flexibilizações obteve solução (lacunas representam falha ao obter solução). As flexibilizações que obtiveram menos de seis de dez soluções, referentes a cada semente

Tabela 5.8: Média dos resultados – AOS com  $\text{IMA}_\infty$ .

Base	$\lambda$	Atrib.	Atrib. PL	Atrib. AOS	10-fold $L_2$	Teste $L_2$
Sonar	0,0001		32,90 (3,84)	12,40 (2,07)	18,13 (3,21)	27,14 (5,64)
	0,001	60	29,60 (1,51)	9,60 (1,71)	17,7 (2,41)	28,14 (6,06)
	0,01		26,30 (2,06)	7,90 (0,99)	19,62 (3,27)	29,00 (7,77)
Ionosphere	0,001		28,4 (1,65)	12,7 (1,42)	10,22 (1,25)	14,27 (2,45)
	0,01	34	28,70 (2,16)	11,80 (0,79)	10,08 (1,38)	15,98 (4,05)
	0,1		25,60 (2,32)	10,20 (1,23)	10,15 (1,16)	13,17 (1,81)
WDBC	100		4,90 (0,88)	4,30 (0,95)	8,01 (1,66)	9,26 (2,98)
	1000	30	4,10 (1,10)	3,20 (0,63)	9,95 (1,68)	10,37 (2,23)
	10000		3,20 (1,03)	2,30 (0,48)	11,11 (1,47)	13,16 (3,54)
Bupa	1		6,00 (0,00)	5,50 (0,53)	35,32 (2,00)	34,43 (4,34)
	10	6	5,90 (0,32)	4,60 (0,84)	34,89 (1,93)	35,65 (3,48)
	100		5,80 (0,42)	4,70 (0,67)	33,43 (2,21)	35,56 (3,37)
Pima	10		7,90 (0,32)	6,80 (1,03)	22,85 (1,87)	28,16 (2,48)
	100	8	6,80 (0,42)	4,50 (0,71)	30,91 (1,49)	33,79 (1,13)
	1000		6,30 (1,06)	4,40 (0,97)	32,99 (1,37)	34,22 (1,41)
Wine	1		7,00 (0,67)	3,20 (0,42)	4,88 (1,12)	4,74 (2,75)
	10	13	6,20 (0,63)	3,90 (1,29)	7,66 (2,05)	5,93 (3,12)
	100		5,80 (0,92)	4,10 (1,10)	11,92 (2,29)	13,73 (5,02)
Australian	1000		7,90 (1,66)	6,80 (2,30)	23,65 (1,98)	23,91 (2,44)
	10000	14	5,70 (1,16)	4,90 (1,37)	29,89 (2,16)	29,35 (2,72)
	100000		2,80 (1,03)	2,50 (0,53)	31,71 (0,90)	32,74 (1,75)
German	10		17,70 (2,50)	10,60 (2,50)	25,53 (0,93)	26,31 (1,45)
	100	24	11,78 (2,82)	5,89 (2,37)	24,16 (5,75)	28,50 (1,28)
	1000		–	–	–	–
Heart	10		11,20 (1,14)	6,30 (1,57)	16,62 (2,29)	17,44 (3,23)
	100	13	7,20 (1,14)	4,30 (0,82)	20,95 (2,32)	21,00 (5,01)
	1000		6,30 (1,06)	4,50 (0,71)	26,35 (1,80)	27,33 (4,29)

utilizada, foram excluídas de posteriores análises.

Observando o resumo apresentado na Tabela 5.11, onde coluna “C” representa a flexibilização utilizada para linearizar o conjunto de treino, “Atrib.” representa a dimensão original da base, “Atrib. PL” representa a dimensão obtida pela aplicação da programação linear, “Atrib. AOS” representa a dimensão da solução obtida pelo AOS, “10-fold” representa o erro de validação cruzada sobre o conjunto de treino e “Teste” representa o erro de validação final, para as bases Sonar e Ionosphere foi possível reduzir consideravelmente a dimensão, mas apenas as bases WDBC e Australian atingiram uma dimensão próxima à objetivo, onde foi possível remover muitas características nulas

Tabela 5.9: Resultados do SMO linear (parte 1).

Base	$C_1$				$C_2$				$C_3$			
	$n$	Margem	10-fold	Teste	$n$	Margem	10-fold	Teste	$n$	Margin	10-fold	Teste
Sonar	100				10				1			
	20	0,008859	14,20 (3,54)	28,57	25	0,042323	18,36 (1,25)	22,86	8	0,013486	22,79 (2,63)	27,14
	20	0,013894	7,68 (4,40)	27,14	11	0,011241	17,68 (2,90)	32,86	8	0,012201	24,97 (2,00)	37,14
	22	0,00889	11,99 (4,95)	25,71	11	0,005484	17,18 (3,66)	24,29	6	0,01395	23,27 (2,18)	27,14
	21	0,010296	9,48 (3,39)	37,14	9	0,006604	17,81 (2,07)	21,43	7	0,012897	24,85 (2,40)	32,86
					9	0,011915	13,98 (2,18)	38,57	2	0,096995	24,26 (1,77)	30,00
					11	0,005894	17,34 (1,67)	27,14	2	0,079016	24,99 (1,83)	25,71
					8	0,012183	10,97 (1,67)	40,00	7	0,015445	16,07 (1,57)	40,00
					11	0,010898	23,13 (2,09)	20,00	2	0,063984	26,29 (0,73)	22,86
					10	0,00861	16,11 (2,20)	22,86	7	0,012076	18,97 (2,09)	20,00
					9	0,008298	17,33 (3,83)	24,29	6	0,00864	21,28 (2,27)	25,71
	Ionosphere	10				1				0,1		
23		0,065181	7,42 (2,25)	9,40	7				7	0,020251	13,25 (0,91)	16,24
					13	0,037533	7,86 (0,96)	14,53	8	0,034675	13,87 (1,16)	16,24
					10	0,019152	9,68 (2,00)	14,53	8	0,035090	14,87 (0,60)	16,24
					10	0,029478	10,18 (1,51)	11,11	6	0,044370	10,48 (0,81)	16,24
					12	0,024694	8,66 (1,52)	17,09	9	0,016673	12,08 (1,23)	16,24
					7				7	0,031800	11,22 (1,18)	14,53
					28	0,100203	13,32 (1,80)	12,82	7	0,049305	12,00 (0,88)	11,97
					10	0,016491	11,00 (1,49)	10,26	6	0,021558	14,70 (0,84)	9,40
					13	0,028662	9,36 (1,52)	13,68	8	0,015831	12,36 (0,77)	15,38
					11	0,020111	12,50 (0,86)	13,68				
	WDBC				100				1000			
4		3,801895	6,52 (0,93)	4,21	2	6,851316	8,52 (1,11)	6,32	2	8,746906	10,00 (0,42)	6,84
4		2,402466	4,70 (1,81)	6,84	2	7,619445	7,87 (0,54)	7,89	2	5,542485	8,84 (0,36)	10,53
5		3,351838	4,85 (0,48)	8,95	2	6,503676	6,62 (0,54)	11,05	2	11,768819	6,86 (0,18)	10,53
5		4,145381	6,65 (0,39)	6,32	3	7,282391	8,00 (0,62)	7,37	2	8,327752	9,26 (0,31)	7,89
2		2,858699	6,15 (0,71)	7,89	3	12,447887	7,18 (0,21)	7,37	2	26,216633	8,44 (0,22)	10,00
					3	5,854589	6,68 (0,71)	9,47	2	9,211801	8,23 (0,40)	9,47
2		8,857971	3,91 (1,14)	11,05	2	6,384947	6,63 (0,43)	10,00	2	4,549124	8,87 (0,22)	8,95
4		3,293405	4,81 (1,43)	7,89	3	3,968431	7,28 (1,38)	8,42	2	21,611133	8,13 (0,17)	9,47
2		22,25027	6,51 (1,87)	4,21	3	5,854662	8,23 (1,17)	5,26	2	10,121585	9,45 (0,35)	8,42
				2	2,758874	7,13 (1,32)	7,89	2	13,929361	9,87 (0,31)	6,84	
Eupa	0,001				0,0001				-			
	5	0,454647	21,26 (5,14)	27,83								
	6	0,274216	20,30 (5,62)	32,17								
	6	0,897372	24,57 (4,84)	33,04								
	6	0,842117	16,74 (4,10)	34,78								
	6	0,576952	20,43 (2,64)	30,43								
	6	0,448215	18,96 (4,03)	32,17								
	6	0,741195	26,91 (4,81)	34,78								
Pima	0,0001				-				-			
	3	1,667144	21,81 (1,81)	24,61								
	6	0,514011	21,80 (1,82)	23,05								

através da programação linear. Assim, conclui-se que o SMO linear não é efetivo para realizar este processo nestas bases.

A Tabela 5.12 apresenta os resultados da utilização do SMO utilizando o *kernel*

Tabela 5.10: Resultados do SMO linear (parte 2).

Base	$C_1$				$C_2$				$C_3$			
	$n$	Margem	10-fold	Teste	$n$	Margem	10-fold	Teste	$n$	Margin	10-fold	Teste
	0,0001				0,00001				0,000001			
Wine					2	8,250307	24,03 (1,27)	22,03				
					2	7,632169	21,45 (0,58)	23,73				
									2	6,380686	19,40 (1,93)	30,51
					2	16,186261	23,34 (1,36)	18,64				
					2	8,276473	23,36 (1,50)	20,34				
Australian	0,00001				0,000001				0,0000001			
	5	7,749152	23,30 (5,00)	46,52					2	49,407920	32,46 (1,87)	37,39
	2	12,638756	7,89 (6,59)	33,91					2	61,018902	34,50 (1,57)	33,91
									2	28,632925	33,87 (1,53)	36,09
					2	50,213518	32,74 (4,61)	36,09	2	65,44925	34,57 (1,01)	35,65
									3	8,261757	33,57 (1,78)	36,09
										2	72,542503	34,28 (0,15)
									2	69,856484	34,20 (1,10)	36,52
German	-				-				-			
Heart	-				-				0,00001			
									2	12,539936	76,56 (4,36)	44,44
									4	3,371468	35,72 (2,01)	42,22
									2	17,156106	90,61 (6,33)	43,33
									4	2,606266	61,11 (5,86)	40,00

Gaussiano  $\sigma$  (FACELI et al., 2011). A coluna “ $n$ ” representa a dimensão alcançada na solução do AOS, “Margem” representa a margem obtida pela solução do AOS, “10-fold” representa o erro de validação cruzada sobre o conjunto de treino, e “Teste” representa o erro da validação final. Para as bases Ionosphere, WDBC, Bupa, Pima e Wine os resultados foram retirados da publicação de Villela et al. (2019), que utiliza a largura de *kernel*  $\sigma$  que obteve melhores resultados, e realiza a seleção de características utilizando o AOS com os mesmos parâmetros deste trabalho. Devido a limitações de tempo, este

Tabela 5.11: Média dos resultados – AOS com SMO linear.

Base	$C$	Atrib.	Atrib. PL	Atrib. AOS	10-fold	Teste
Sonar	100		–	–	–	–
	10	60	27,70 (2,79)	11,40 (4,90)	16,99 (3,11)	27,43 (7,18)
	1		24,90 (2,33)	5,50 (2,51)	22,77 (3,17)	28,86 (6,24)
Ionosphere	10		–	–	–	–
	1	34	26,75 (2,31)	13,38 (6,05)	10,32 (1,87)	13,46 (2,13)
	0,1		23,44 (2,30)	7,33 (1,00)	12,76 (1,52)	14,72 (2,45)
WDBC	0,0001		4,13 (1,13)	3,50 (1,31)	5,51 (1,06)	7,17 (2,32)
	0,00001	30	3,40 (0,70)	2,50 (0,53)	7,41 (0,70)	8,1 (1,72)
	0,000001		2,50 (0,71)	2,00 (0,00)	8,80 (0,94)	8,89 (1,37)
Bupa	0,001	6	5,86 (0,38)	5,86 (0,38)	21,31 (3,42)	32,17 (2,46)
	0,0001		–	–	–	–
Pima	0,0001	8	–	–	–	–
Wine	0,0001		–	–	–	–
	0,00001	13	–	–	–	–
	0,000001		–	–	–	–
Australian	0,00001		–	–	–	–
	0,000001	14	–	–	–	–
	0,0000001		2,14 (0,38)	2,14 (0,38)	33,92 (0,73)	36,02 (1,08)
German	–	24	–	–	–	–
Heart	0.00001	13	–	–	–	–

processo não foi executado para as bases Sonar, Australian e German. Para a base Heart foram testados três larguras de *kernel*.

Para analisar o comportamento do AOS, foram verificadas algumas métricas: “Ins.” representa o número de candidatos inseridos na *heap*, “Reins.” representa o número de candidatos retirados da *heap* que posteriormente retornaram a ela, “Exp.” representa o número de candidatos avaliados e expandidos, “Max.” representa o tamanho máximo atingido pela *heap*, “Ñ.Trein.” representa o número de candidatos que não foram avaliados, “Poda” representa o número de candidatos que foram podados, “PL” representa a dimensão atingida com a utilização da programação linear, “ $n$ ” representa a dimensão atingida pelo AOS, “Margem” representa o valor da margem da solução, e “Segs.” representa o tempo em que a solução foi encontrada. Observando a Tabela 5.13 que para algumas bases, o AOS realiza poucas iterações, resultando na geração de poucos candidatos. Isto se deve ao fato do conjunto de dados utilizado voltar a ser não linearmente



separável após a eliminação de algumas características.

Por fim, a Tabela 5.14 apresenta uma comparação dos resultados gerais de cada algoritmo, onde “ $n$ ” representa a dimensão média atingida, “10-*fold*” representa o erro médio de validação cruzada sobre o conjunto de treino, e “Teste” representa a média do erro de validação final. Para o  $\text{IMA}_p$  e SMO linear, foram utilizados os resultados obtidos pela menor flexibilização testada. Para as bases Ionosphere, WDBC, Bupa, Pima e Wine os resultados do *kernel* gaussiano foram retirados da publicação de Villela et al. (2019), que realiza a seleção de características utilizando o AOS com os mesmos parâmetros deste trabalho. Para as bases WDBC, Bupa, Pima, Wine e Heart os resultados obtidos utilizando a abordagem proposta são promissores, pois mesmo que a dimensão seja mais alta que as obtidas com a utilização do *kernel*, o erro de teste é menor. É possível levantar duas hipóteses para a causa disto: a utilização do *kernel* descaracterizou as relações entre as características, resultando em uma solução ruim, ou estes conjuntos de fato possuem muitas características relevantes ao aprendizado. Para a Ionosphere, a abordagem proposta obteve resultados piores em relação ao *kernel*, visto que há uma grande diferença na dimensão encontrada, mesmo que o erro seja similar.

Tabela 5.12: Resultados do SMO com *kernel* gaussiano.

Base	$\sigma_1$				$\sigma_2$				$\sigma_3$			
	<i>n</i>	Margem	10-fold	Teste	<i>n</i>	Margem	10-fold	Teste	<i>n</i>	Margin	10-fold	Teste
1												
Ionosphere	4	0,015449	8,75 (1,10)	12,82								
	4	0,011025	13,92 (1,20)	5,98								
	4	0,013044	10,80 (0,98)	9,40								
	4	0,011863	6,85 (2,02)	8,55								
	4	0,014916	8,45 (1,58)	6,84								
	4	0,015808	9,79 (2,00)	8,55								
	4	0,016387	8,67 (0,42)	10,26								
	4	0,010953	10,55 (1,30)	10,26								
	4	0,009643	10,89 (0,89)	9,40								
4	0,008056	13,00 (1,28)	5,98									
0,01												
WDBC	2	0,056291	32,00 (2,13)	7,89								
	2	0,056255	32,53 (2,03)	7,37								
	2	0,059294	11,88 (2,63)	13,16								
	2	0,059848	22,52 (2,52)	10,53								
	2	0,057023	14,14 (3,26)	9,47								
	2	0,059923	10,93 (1,43)	11,05								
	2	0,059894	17,59 (2,89)	13,68								
	2	0,060103	14,90 (1,56)	13,16								
	2	0,056163	25,99 (2,90)	7,89								
2	0,059751	11,19 (2,73)	9,47									
0,01												
Bupa	3	0,007142	37,13 (2,02)	41,74								
	3	0,004627	37,83 (2,30)	49,57								
	4	0,030070	37,96 (0,58)	39,13								
	3	0,006906	36,48 (1,34)	46,09								
	3	0,005240	43,65 (2,24)	40,00								
	3	0,012001	39,78 (2,07)	39,13								
	3	0,007938	42,39 (2,96)	46,09								
	4	0,022128	44,83 (1,58)	37,39								
	3	0,006744	46,35 (2,30)	46,09								
	3	0,005524	44,39 (1,57)	38,26								
	1											
Pima	2	0,067648	35,99 (0,67)	36,72								
	2	0,067354	36,34 (0,69)	35,94								
	2	0,069616	35,68 (0,64)	33,98								
	2	0,068366	34,61 (0,77)	33,98								
	2	0,060415	39,55 (0,75)	31,25								
	2	0,064711	35,75 (1,14)	32,81								
	2	0,068143	36,52 (0,62)	36,72								
	2	0,068214	34,98 (0,55)	35,55								
	2	0,065315	35,35 (0,97)	35,16								
2	0,064862	35,04 (0,95)	33,59									
1												
Wine	2	0,059726	22,53 (2,23)	18,64								
	2	0,059663	27,16 (1,93)	13,56								
	2	0,071163	23,63 (2,04)	27,12								
	2	0,069243	20,56 (3,34)	25,42								
	2	0,073070	16,39 (1,49)	20,34								
	2	0,068094	21,98 (1,22)	23,73								
	2	0,069433	14,17 (0,65)	18,64								
	2	0,068340	13,69 (1,43)	15,25								
	2	0,055393	20,14 (2,36)	18,64								
	2	0,081591	16,22 (1,97)	23,73								
0,1												
Heart	2	0,062785	41,33 (1,09)	38,89	2	0,076089	40,56 (0,87)	46,67	2	0,076254	41,83 (0,53)	44,44
	2	0,061807	38,94 (1,32)	46,67	2	0,075146	48,44 (0,78)	46,67	2	0,088221	42,06 (0,83)	37,78
	2	0,062330	40,72 (1,05)	50,00	2	0,075455	44,50 (1,18)	48,89	2	0,075868	44,06 (0,53)	45,56
	2	0,062911	43,44 (1,48)	48,89	2	0,082281	34,56 (1,71)	43,33	2	0,098135	43,44 (1,41)	45,56
	2	0,066977	41,83 (1,05)	45,56	2	0,083426	40,94 (1,41)	41,11	2	0,075868	43,83 (0,18)	43,33
	2	0,065123	43,78 (1,33)	41,11	2	0,075259	46,28 (0,83)	45,56	2	0,076315	45,78 (0,39)	44,44
	2	0,059725	46,50 (1,11)	46,67	2	0,077647	40,67 (1,30)	33,33	2	0,085568	42,56 (1,29)	36,67
	2	0,063483	40,00 (1,05)	45,56	2	0,076199	44,67 (0,95)	40,00	2	0,075431	44,94 (0,18)	43,33
	2	0,061262	43,61 (1,09)	42,22	2	0,078744	41,78 (1,41)	37,78	2	0,075000	44,44 (0,00)	43,33
	2	0,063314	42,50 (1,51)	42,22	2	0,085053	34,00 (1,87)	43,33	2	0,075699	44,11 (0,47)	45,56

Tabela 5.13: Média dos estados da execução do AOS.

Base	Clas.	Parâmetro	Ins.	Reins.	Exp.	Max.	Ñ.Trei.	Poda	PL	$n$	Margem	Segs.	
Sonar	SMO	$C$	100										
			10	472	301	325	56	109	19	28	11	0,012345	38,27
			1	891	645	658	132	135	78	25	6	0,032869	34,62
	IMA $_{\infty}$	$\lambda$	0,0001	739	533	552	110	104	31	33	12	0,000631	377,20
			0,001	583	407	469	80	75	12	30	10	0,000877	104,94
			0,01	505	343	395	79	58	13	26	8	0,002311	22,72
Ionosphere	SMO	$C$	10										
			1	216	134	148	26	57	7	27	13	0,034541	30,57
			0,1	511	376	386	75	90	19	23	7	0,029950	36,03
	IMA $_{\infty}$	$\lambda$	0,001	272	211	221	44	23	5	28	13	0,002893	58,75
			0,01	557	435	449	83	63	15	29	12	0,003018	82,24
			0,1	227	156	184	37	24	1	26	10	0,003712	9,63
WDBC	SMO	$C$	0,0001	1	0	0	1	0	0	4	4	6,370241	0,22
			0,00001	1	0	1	1	0	0	3	3	6,552622	0,12
			0,000001	1	0	0	1	0	0	3	2	12,00256	0,08
	IMA $_{\infty}$	$\lambda$	100	1	0	1	1	1	0	5	4	1,529513	3,25
			1000	1	0	1	1	0	0	4	3	3,145858	2,76
			10000	2	0	1	1	1	0	3	2	3,450613	2,07
Bupa	SMO	$C$	0,001	0	0	0	0	0	0	6	6	0,604959	0,05
			0,0001										
	IMA $_{\infty}$	$\lambda$	1	1	0	1	1	0	0	6	6	0,164923	0,98
			10	3	0	2	2	1	0	6	5	0,175342	1,54
			100	2	0	1	2	1	0	6	5	0,341268	1,00
Pima	SMO	$C$	0,0001										
	IMA $_{\infty}$	$\lambda$	10	2	0	1	1	0	0	8	7	0,072045	4,40
			100	4	0	3	3	1	0	7	5	0,509614	1,63
			1000	5	1	3	2	1	0	6	4	1,087347	0,92
Wine	SMO	$C$	0,0001										
			0,00001										
			0,000001										
	IMA $_{\infty}$	$\lambda$	1	4	0	4	1	0	0	7	3	0,093640	8,32
			10	3	0	2	1	0	0	6	4	0,171553	5,69
			100	2	0	2	1	0	0	6	4	0,207912	4,10
Australian	SMO	$C$	0,00001										
			0,000001										
			0,0000001	0	0	0	0	0	0	2	2	50,738534	0,03
	IMA $_{\infty}$	$\lambda$	1000	2	0	1	1	1	0	8	7	0,340220	2,49
			10000	1	0	1	1	0	0	6	5	1,097336	2,20
			100000	1	0	0	1	0	0	3	3	6,840885	0,86
German	SMO	$C$											
	IMA $_{\infty}$	$\lambda$	10	33	13	24	8	8	0	18	11	0,017803	51,93
			100	19	5	13	5	5	0	12	6	0,087600	12,29
Heart	SMO	$C$	0,00001										
	IMA $_{\infty}$	$\lambda$	10	12	3	8	5	3	1	11	6	0,041328	16,97
			100	7	1	5	3	2	0	7	4	0,159831	3,62
			1000	3	0	2	2	1	0	6	5	0,472190	1,23

Tabela 5.14: Comparação nos conjuntos de dados.

Base	Classificador	$n$	10-fold	Teste
Sonar	SMO gaussiano	–	–	–
	SMO linear	11,40 (4,90)	16,99 (3,11)	27,43 (7,18)
	IMA <sub>∞</sub>	12,04 (2,07)	18,13 (3,21)	27,14 (5,64)
Ionosphere	SMO gaussiano	4,00 (0,00)	10,17 (2,15)	8,80 (2,13)
	SMO linear	13,38 (6,05)	10,32 (1,87)	13,46 (2,13)
	IMA <sub>∞</sub>	12,70 (1,42)	10,22 (1,25)	14,27 (2,45)
WDBC	SMO gaussiano	2,00 (0,00)	19,37 (8,37)	10,37 (2,36)
	SMO linear	3,50 (1,31)	5,51 (1,06)	7,17 (2,32)
	IMA <sub>∞</sub>	4,30 (0,95)	8,01 (1,66)	9,26 (2,98)
Bupa	SMO gaussiano	3,20 (0,40)	41,08 (3,65)	42,35 (4,24)
	SMO linear	5,86 (0,38)	21,31 (3,42)	32,17 (2,46)
	IMA <sub>∞</sub>	5,50 (0,53)	35,32 (2,00)	34,43 (4,34)
Pima	SMO gaussiano	2,00 (0,00)	35,98 (1,39)	34,57 (1,77)
	SMO linear	–	–	–
	IMA <sub>∞</sub>	6,80 (1,03)	22,85 (1,87)	28,16 (2,48)
Wine	SMO gaussiano	2,00 (0,00)	19,65 (4,41)	20,51 (4,41)
	SMO linear	–	–	–
	IMA <sub>∞</sub>	3,20 (0,42)	4,88 (1,12)	4,74 (2,44)
Australian	SMO gaussiano	–	–	–
	SMO linear	2,14 (0,38)	33,92 (0,73)	36,02 (1,08)
	IMA <sub>∞</sub>	6,80 (2,30)	23,65 (1,98)	23,91 (2,44)
German	SMO gaussiano	–	–	–
	SMO linear	–	–	–
	IMA <sub>∞</sub>	10,60 (2,50)	25,53 (0,93)	26,31 (1,45)
Heart	SMO gaussiano	2,00 (0,00)	42,27 (2,19)	44,78 (3,56)
	SMO linear	–	–	–
	IMA <sub>∞</sub>	6,30 (1,57)	16,62 (2,29)	17,44 (3,23)

## 6 Considerações finais

Este trabalho apresenta uma abordagem alternativa ao *kernel* que se mostrou promissora para se trabalhar com conjuntos de dados não linearmente separáveis. Através dos diferentes estudos, observou-se a efetividade do processo de linearização dos conjuntos de dados, cujos resultados são satisfatórios. Algumas bases conseguem atingir o estado linear com valores baixos de flexibilização, enquanto outras requerem valores muito altos, resultando na remoção de muitas amostras. O algoritmo  $\text{IMA}_p$ , por realizar um processo incremental da margem, consegue encontrar soluções mais facilmente, tornando-se ideal para a abordagem deste trabalho.

Nos estudos da aplicação da abordagem proposta à seleção de características utilizando o AOS, é possível notar que poucos conjuntos atingiram o número de características definido como objetivo, pois o problema, após a remoção de algumas características, retorna ao estado não linearmente separável. Porém não é possível relacionar o baixo número de características à qualidade da solução, visto que várias soluções apresentaram erros de teste menores mesmo possuindo mais características.

Para trabalhos futuros, é possível igualar a dimensão obtida pelo *kernel* gaussiano à obtida neste trabalho, de forma a comparar os resultados na mesma dimensão, para reforçar a efetividade desta abordagem. Além disso, podem ser estudadas outras formas de se reduzir ainda mais o número de características, como a reaplicação da abordagem deste trabalho ou até mesmo métodos exaustivos de busca, visto que o número de características atingido neste trabalho é baixo, permitindo que se avalie todas as possíveis combinações.

## Referências Bibliográficas

BACHE, K.; LICHMAN, M. *UCI Machine Learning Repository*. 2013. Disponível em: <http://archive.ics.uci.edu/ml>.

FACELI, K.; LORENA, A. C.; GAMA, J.; DE CARVALHO, A. C. P. L. *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. 1. ed. Rio de Janeiro: LTC, 2011.

HART, P.; NILSSON, N.; RAPHAEL, B. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, v. 4, n. 2, p. 100–107, 1968.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., 1995. v. 2, p. 1137–1143.

LEITE, S. C.; FONSECA NETO, R. Incremental margin algorithm for large margin classifiers. *Neurocomputing*, Elsevier, v. 71, n. 7-9, p. 1550–1560, 2008.

NG, A. Y. On feature selection: Learning with exponentially many irrelevant features as training examples. In: *Proceedings of the Fifteenth International Conference on Machine Learning*. Madison, Wisconsin, USA: Morgan Kaufmann Publishers Inc., 1998. p. 404–412.

PLATT, J. C. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. [S.l.], 1998. 21 p.

SCHÖLKOPF, B.; SMOLA, A. J. *Learning with Kernels*. [S.l.]: MIT Press, 2002.

VAPNIK, V. N. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

VILLELA, S. M. *Seleção de características com busca ordenada e classificadores de larga margem*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2011.

VILLELA, S. M.; LEITE, S. C.; FONSECA NETO, R. Incremental p-margin algorithm for classification with arbitrary norm. *Pattern Recognition*, v. 55, p. 216–272, 2016.

VILLELA, S. M.; LEITE, S. C.; XAVIER, A. E.; FONSECA NETO, R. An ordered search with a large margin classifier for feature selection. Submitted. 2019.