

Relatório de Implementação do LSSVM

Angélica Alves Viana

Julho 2019

1 Introdução

Este relatório visa descrever os passos de implementação do classificador LSSVM e apresentar os resultados obtidos a partir dos testes das bases da Iris, Coluna Vertebral e Artificial I (Todas com duas classes). Além disso, serão apresentados os resultados obtidos com os algoritmos implementados anteriormente, para fins de comparação.

2 Classificador LSSVM

A versão least-squares do classificador SVM é obtida pela formulação do problema de minimização dado por:

$$\begin{aligned} \bullet \min J(w, e) &= \frac{1}{2}w^T w + \gamma \frac{1}{2} \sum_{i=1}^n e_i^2 \\ \text{s.a. } d_i[w^T x_i + b] &= 1 - e_i, i = 1, \dots, n. \end{aligned}$$

Em que J é a função-custo a ser minimizada e e_i são as variáveis de folga.

A solução do LS-SVM será obtida depois de construída a função lagrangeana, que é apresentada abaixo

$$L(w, b, e, \alpha) = \frac{1}{2}w^T w + \gamma \frac{1}{2} \sum_{i=1}^n e_i^2 - \sum_{i=1}^n \alpha_i (d_i(x_i^T w + b) - 1 + e_i)$$

Um fator importante a se destacar é que os valores dos elementos pertencentes aos conjuntos $\{\alpha_i\}_{i=1}^n$, que representam os multiplicadores de Lagrange, são quase sempre não-nulos. Tendo em vista essa formulação, as condições de optimalidade são dadas por:

- $\bullet \frac{\partial L}{\partial w} = 0 \mapsto w = \sum_{i=1}^n \alpha_i d_i x_i$
- $\bullet \frac{\partial L}{\partial b} = 0 \mapsto \sum_{i=1}^n \alpha_i d_i = 0$
- $\bullet \frac{\partial L}{\partial e_i} = 0 \mapsto \alpha_i = \gamma e_i, i = 1, \dots, N.$
- $\bullet \frac{\partial L}{\partial \alpha_i} = 0 \mapsto d_i(x_i^T w + b) - 1 + e_i = 0.$

A partir das condições descritas acima, é possível formular um sistema de equações lineares ($Ax = b$), em que as variáveis são os parâmetros α e b que devem ser encontrados. Esse sistema pode ser representado em forma matricial conforme pode ser visto abaixo.

$$\begin{bmatrix} 0 & \mathbf{d}^T \\ \mathbf{d} & \Omega + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}$$

Onde $\Omega_{i,j} = y_i y_j x_i^T x_j$, $d = [d_1 d_2 \dots d_n]^T$, $\alpha = [\alpha_1 \alpha_2 \dots \alpha_n]^T$ e $\mathbf{1} = [1 1 \dots 1]^T$, de modo que o tamanho do vetor $\mathbf{1}$ é igual a n . Pode-se redefinir $\Omega_{i,j}$ para $\Omega_{i,j} = y_i y_j K(x_i, x_j)$. A saída do classificador pode ser calculada como

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b \right)$$

O γ que está envolvido no sistema linear consiste em um parâmetro de regularização que se relaciona ao tamanho da margem do classificador. Esse parâmetro na implementação foi encontrado a partir de uma busca em grade com valores de γ variando no intervalo $[1, 10]$.

Os testes a seguir são referentes a utilização do LSSVM sobre as bases das Iris, Coluna e Artificial com kernels linear, quadrático e rbf, que são definidos pelas fórmulas abaixo:

- **Kernel Linear:** $K(x, x_i) = x_i^T x$
- **Kernel Quadrático:** $K(x, x_i) = (1 + \frac{x_i^T x}{c})^2$
- **Kernel RBF:** $K(x, x_i) = \exp(\frac{-\|x - x_i\|^2}{\sigma^2})$

3 Testes

Nesta seção serão apresentados os resultados obtidos para as bases da Iris, Coluna Vertebral e Artificial I utilizando todos os classificadores implementados até agora. Os parâmetros avaliados são referentes às médias de acurácia (A.M.) e o desvio padrão dessas acurácias (D.P.) de 20 realizações de treinamento e teste, sendo 80% dos dados destinados ao treinamento e 20% a teste.

- **Base de Dados da Íris - 2C:** Contém 150 padrões, 4 atributos e 2 classes.

Table S1: Tabela contendo as métricas de avaliação para os quinze algoritmos.

Classificador	Treinamento		Teste	
	A.M.	D.P.	A.M.	D.P.
LSSVM - RBF	100,00	0,00	99,17	1,44
LSSVM - Quadrático	100,00	0,00	100,00	0,00
LSSVM - Linear	100,00	0,00	99,33	1,33
Bayesiano GMM	98,29	0,61	96,83	2,46
Bayesiano Parzen	98,41	0,91	95,33	3,23
Bayesiano Rejeicao	92,79	0,19	90,83	8,00
Bayesiano Puro	98,25	0,69	97,5	2,07
Disc. Quadrático	98,21	0,71	97,33	3,26
Disc. Linear 1	86,54	1,56	83,67	7,22
Disc. Linear 2	92,67	1,66	91,16	6,43
Disc. Linear 3	86,46	1,64	85,33	5,31
Disc. Linear 4	93,46	1,06	92,67	3,59
Disc. Linear 5	86,75	1,23	85,50	5,60
KNN	97,33	0,77	97,33	3,05
DMC	92,59	3,97	92,59	3,97

Fonte: Autoria Própria.

As realizações cujos valores de acurácia mais se aproximaram das acurácias médias de testes para os classificadores LSSVM-Linear, LSSVM-Quadrático e LSSVM-RBF respectivamente na base da Iris, é representada a partir de sua matriz de confusão abaixo:

	-1	1
-1	12	0
1	0	18

(a) LSSVM - Linear

	-1	1
-1	9	0
1	0	21

(b) LSSVM - Quadrático

	-1	1
-1	10	0
1	0	20

(c) LSSVM - RBF

Abaixo vê-se a superfície de decisão do classificador LSSVM com os kernels linear, quadrático e rbf, respectivamente, para os atributos 1 e 2 da base da Iris.

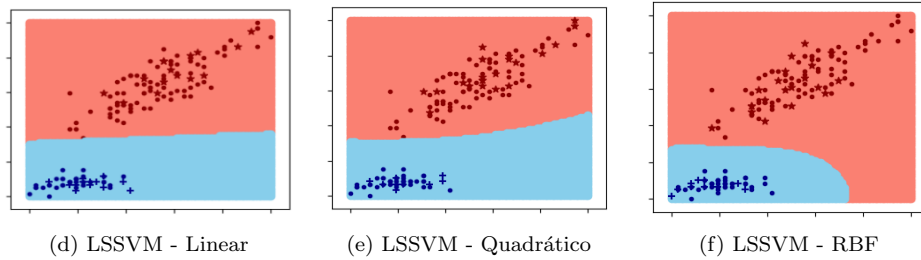


Figure 1: Superfície de decisão - Iris

- Base da Coluna vertebral - 2C: 310 padrões, 6 atributos e 2 classes.

Table S2: Tabela contendo as métricas de avaliação para os quinze algoritmos.

Classificador	Treinamento		Teste	
	A.M.	D.P.	A.M.	D.P.
LSSVM - RBF	94,94	2,26	81,80	4,18
LSSVM - Quadrático	84,43	1,70	87,46	3,92
LSSVM - Linear	82,50	1,50	84,09	1,33
Bayesiano GMM	84,41	1,22	81,72	4,31
Bayesiano Parzen	83,02	1,12	77,95	5,85
Bayesiano Rejeicao	100,00	0,00	100,00	0,00
Bayesiano Puro	86,88	1,12	83,52	4,03
Disc. Quadrático	87,56	1,04	82,95	3,57
Disc. Linear 1	83,06	1,41	80,90	5,68
Disc. Linear 2	69,70	1,11	70,49	5,61
Disc. Linear 3	79,86	2,18	78,03	6,27
Disc. Linear 4	75,02	1,47	72,05	5,06
Disc. Linear 5	79,41	1,80	76,56	5,05
KNN	78,90	1,10	78,90	4,37
DMC	74,18	4,91	74,18	4,91

Fonte: Autoria Própria.

As realizações cujos valores de acurácia mais se aproximaram das acurácias médias de testes para os classificadores LSSVM-Linear, LSSVM-Quadrático e LSSVM-RBF respectivamente na base da Coluna, é representada a partir de sua matriz de confusão abaixo:

	-1	1		-1	1		-1	1
-1	9	8	-1	14	6	-1	10	0
1	2	42	1	2	39	1	0	20

(a) LSSVM - Linear (b) LSSVM - Quadrático (c) LSSVM - RBF

Abaixo vê-se a superfície de decisão do classificador LSSVM com os kernels linear, quadrático e rbf, respectivamente, para os atributos 1 e 5 da base da Coluna.

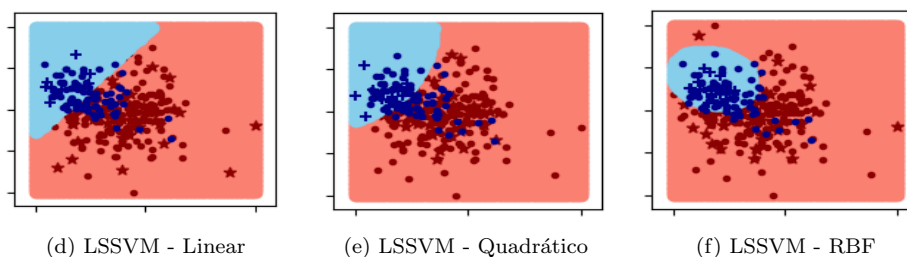


Figure 2: Superfície de decisão - Iris

- Base Artificial I: Essa base foi criada a partir da geração de dados aleatoriamente ao redor dos pontos (0,25;0,75), (0,5; 0,5) e (0,75; 0,75). Totalizando 40 padrões (30 da classe 0 e 10 da classe 1).

Table S3: Tabela contendo as métricas de avaliação para os quinze algoritmos.

Classificador	Treinamento		Teste	
	A.M.	D.P.	A.M.	D.P.
LSSVM - RBF	100,00	0,00	99,33	1,33
LSSVM - Quadrático	100,00	0,0	100,00	0,0
LSSVM - Linear	100,00	0,0	100,00	0,0
Bayesiano GMM	100,00	0,0	100,00	0,0
Bayesiano Parzen	100,00	0,0	100,00	0,0
Bayesiano Rejeição	100,00	0,0	100,00	0,0
Bayesiano Puro	100,00	0,0	100,00	0,0
Disc. Quadrático	100,00	0,0	100,00	0,0
Disc. Linear 1	100,00	0,0	100,00	0,0
Disc. Linear 2	100,00	0,0	100,00	0,0
Disc. Linear 3	100,00	0,0	100,00	0,0
Disc. Linear 4	100,00	0,0	100,00	0,0
Disc. Linear 5	100,00	0,0	100,00	0,0

Fonte: Autoria Própria.

As realizações cujos valores de acurácia mais se aproximaram das acurácias médias de testes para os classificadores LSSVM-Linear, LSSVM-Quadrático e LSSVM-RBF respectivamente na base Artificial, é representada a partir de sua matriz de confusão abaixo:

	-1	1
-1	5	0
1	0	3

	-1	1
-1	7	0
1	0	1

	-1	1
-1	6	0
1	0	2

(a) LSSVM - Linear (b) LSSVM - Quadrático (c) LSSVM - RBF

Abaixo vê-se a superfície de decisão do classificador LSSVM com os kernels linear, quadrático e rbf, respectivamente, para os atributos 1 e 5 da base da Coluna.

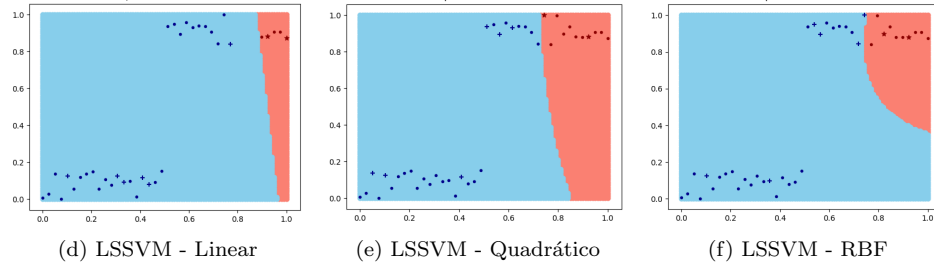


Figure 3: Superfície de decisão - Iris

4 Conclusão

Foram obtidos resultados ótimos para todas as bases testadas e para os três tipos de kernel LSSVM. De um modo geral, o classificador LSSVM com kernel quadrático obteve uma melhor performance que os demais, permitindo que houvesse uma variação menor entre as acurácias obtidas no treinamento e no teste. Os demais classificadores obtiveram resultados semelhantes aos do LSSVM, sendo possível observar, no entanto, que quanto maior a complexidade do algoritmo, maior a sua capacidade de resolver o problema com sucesso, de modo, que é possível ver que um classificador simples como o DMC para a base da Coluna não apresenta resultados tão bons quanto os demais, sendo necessário, portanto um classificador mais robusto.