

Exercício Reconhecimento de Padrões Classificação com SVM

Lucas Ribeiro da Silva - 2022055564

Universidade Federal de Minas Gerais

Belo Horizonte - Minas Gerais - Brasil

lucasrsilvak@ufmg.br

1 Introdução

Nesse relatório consta a implementação de um classificador SVM, a separação do banco de dados em treinamento e teste e a subsequente validação cruzada nas partições criadas. Depois, foi efetuado o cálculo da média e do desvio padrão das amostras.

2 Método

Foram separadas as classes em treinamento e teste utilizando o k-folds com k setado em 10 e repetido o algoritmo por 10 vezes, então foi calculada a média das acurácias parciais e totais do programa. O hiperparâmetro C foi configurado para 237 enquanto o hiperparâmetro σ foi configurado para 0,1.

As sementes foram configuradas como aleatórias.

3 Resultados

Os resultados ficaram tais como se apresentam abaixo:

Tabela 1: Acurácia média e desvio padrão para os folds

Repetição	Acurácia Média	Desvio Padrão
1	0.7004	0.0824
2	0.6970	0.0497
3	0.7208	0.1037
4	0.7190	0.1038
5	0.6957	0.0545
6	0.7013	0.1208
7	0.7149	0.0806
8	0.7203	0.0695
9	0.7004	0.0732
10	0.7286	0.0881

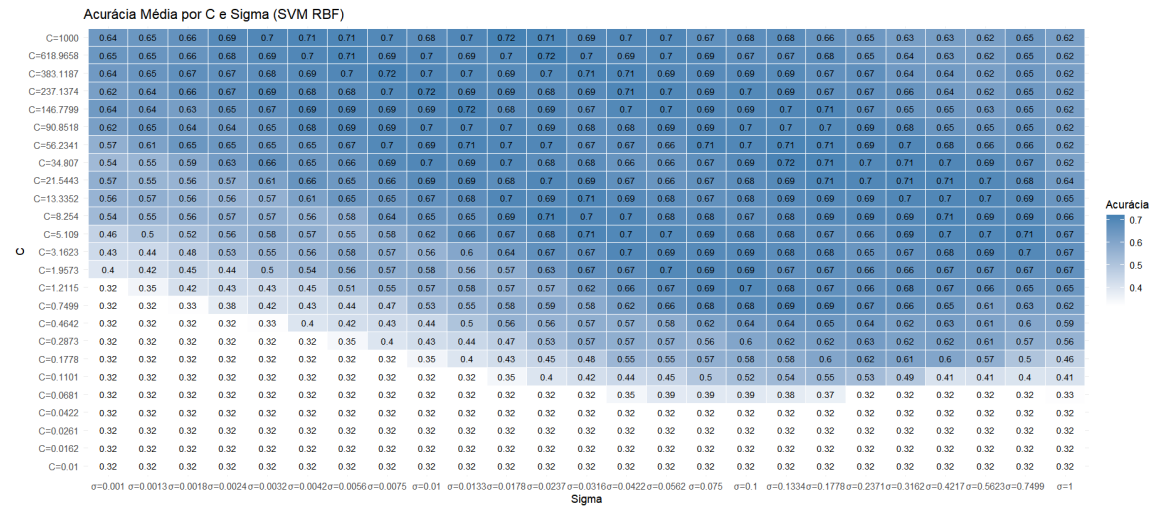
Que por sua vez, tem como acurácia média geral e desvio padrão médio geral os seguintes valores;

Acurácia média geral: **0.7098**

Desvio padrão médio geral: **0.0826**

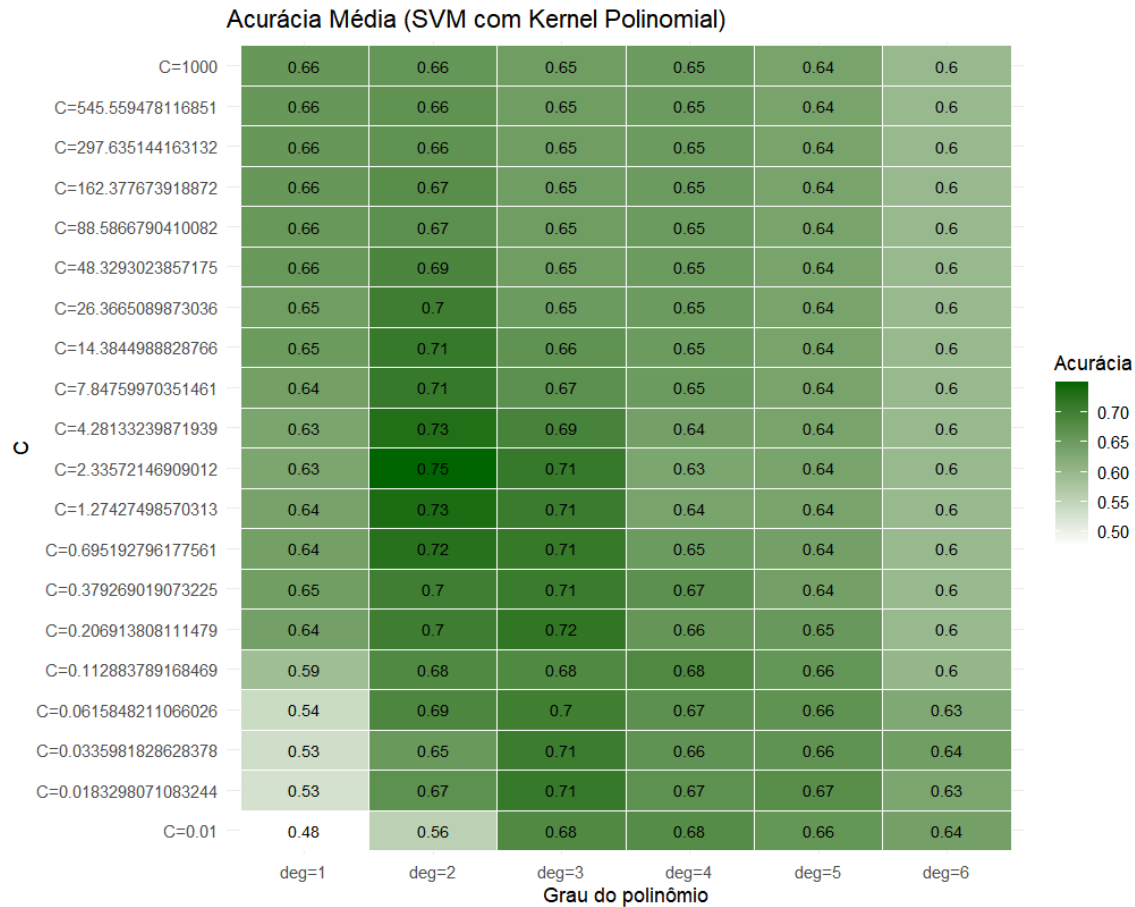
4 Escolha dos Parâmetros C e Sigma

Para testar os parâmetros, foram configurados 25 valores em escala logarítmica para encontrar o **C** e **σ** ideais.



5 Escolha nos Modelos Lineares e Polinomiais

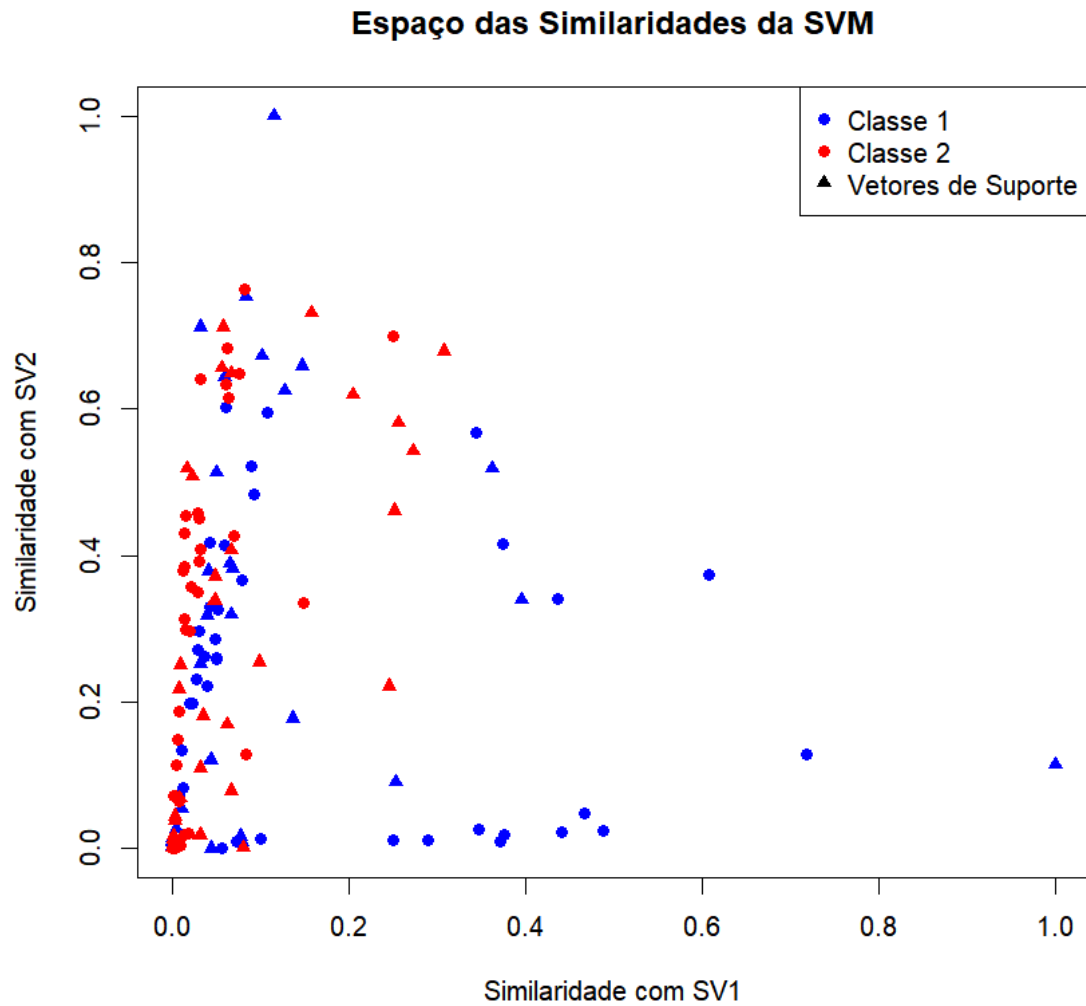
Em modelos polinomiais (o linear é o caso grau = 1), a tabela registrou os valores abaixo:



Tendo sido o melhor valor linear aquele com $C = 1000$ e para os modelos polinomiais, o de grau 2 com $C = 2.335$

6 Espaço de Verossimilhanças

A projeção das classes separadas pela majoritária (Classe 1) e as demais (Classe 2) tal como se encontram no espaço das verossimilhanças e destacadas com os vetores de suporte é a abaixo:



Pelo gráfico, fica perceptível a não-linearidade do problema e também a concentração dos Vetores de Suporte nas áreas de maior densidade do problema, sendo esta a região mais crítica por localizar-se na fronteira.

7 Conclusões

O problema de classificação apresentado foi tratado a partir do classificador SVM e obteve uma acurácia razoável para a solução do problema, isso pode ser identificado pelo desbalanceamento de elementos entre as classes no primeiro teste e pode ser observado pela dificuldade de separação entre as classes no espaço de similaridades da SVM, onde a classe majoritária não se separa com facilidade das demais.