

Exercício Reconhecimento de Padrões KNN

Lucas Ribeiro da Silva - 2022055564

Universidade Federal de Minas Gerais

Belo Horizonte - Minas Gerais - Brasil

lucasrsilvak@ufmg.br

1 Introdução

Nesse relatório, implementaremos o método de agrupamento KNN e variaremos tanto as amostras de treinamento quanto o parâmetro K para observar como o algoritmo se comporta dado seu funcionamento.

2 Variação do Desvio Padrão

Para as próximas imagens, variaremos os desvios padrões (0.3, 0.5, 0.7, 0.9 e 1.1) utilizando $K = 3$.

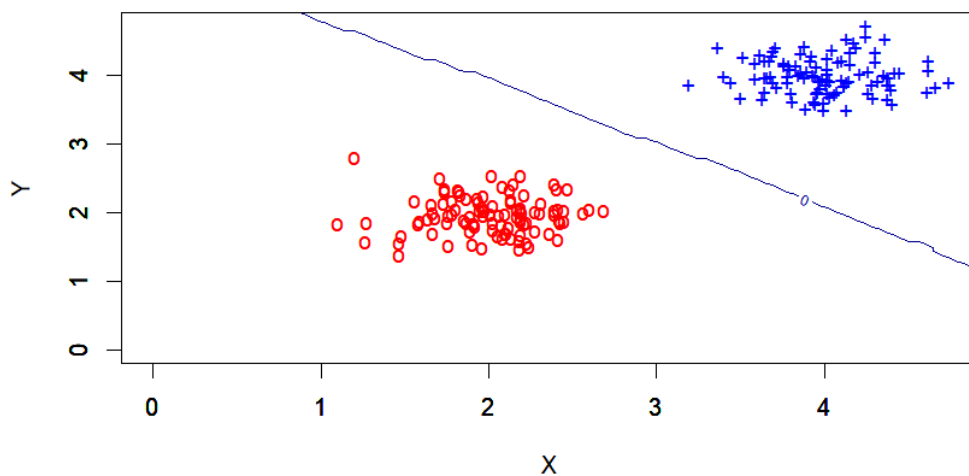


Figura 1: $sd = 0.3$

Para desvio padrão = 0.3, podemos perceber que os dados ficam muito próximos e a superfície de separação divide muito bem os dados.

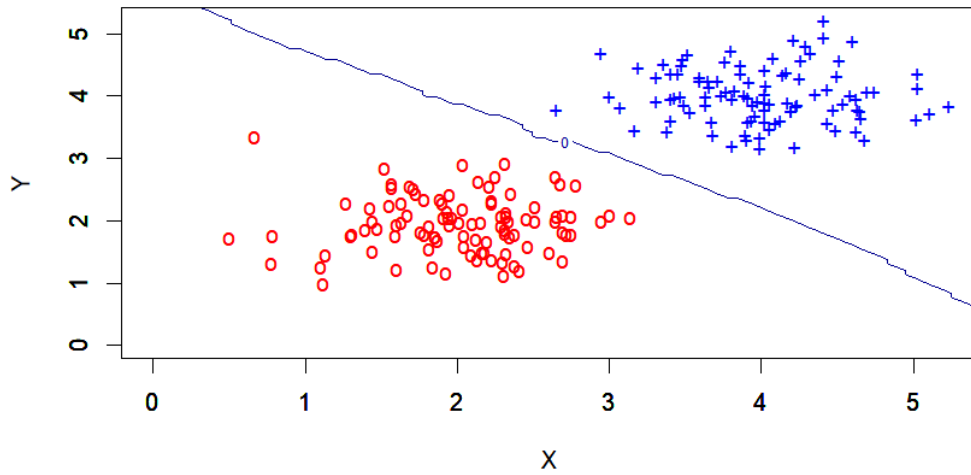


Figura 2: $sd = 0.5$

Para desvio padrão = 0.5, podemos perceber que os dados ainda estão próximos e que a superfície de separação ainda divide muito bem os dados.

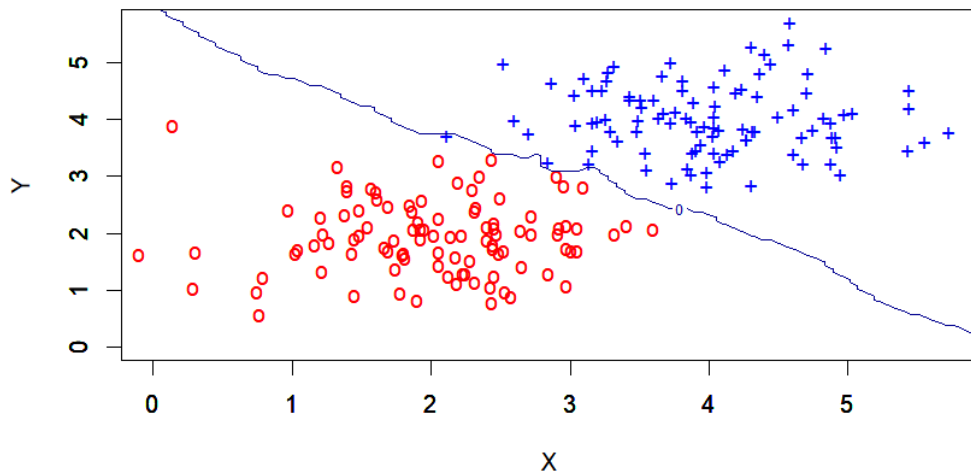


Figura 3: $sd = 0.9$

Para desvio padrão = 0.7, podemos perceber que os dados começam a se amontoar

nas regiões de interseção e que a superfície de separação apresenta diversas ranhuras para se ajustar aos dados de amostragem.

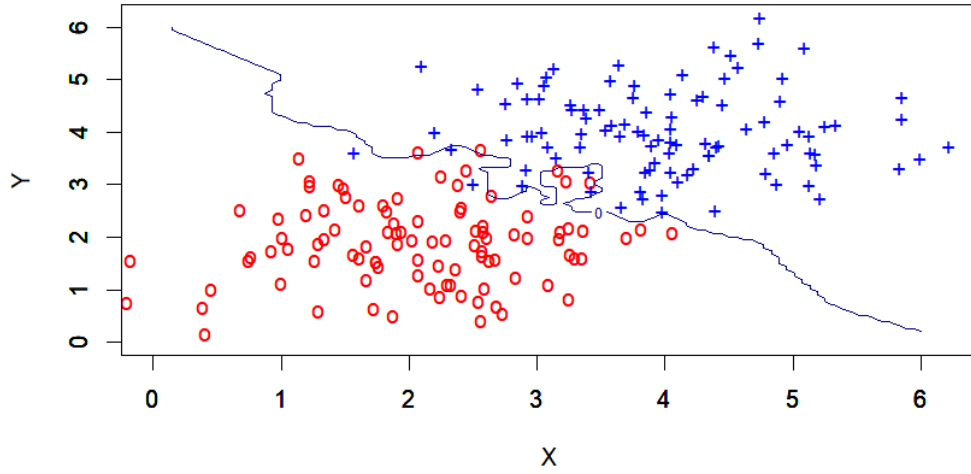


Figura 4: $sd = 0.9$

Para desvio padrão = 0.9, podemos perceber que os dados começam a se amontoar bastante nas regiões de interseção e que a superfície de separação já apresenta sinais de *overfitting* para $K = 3$.

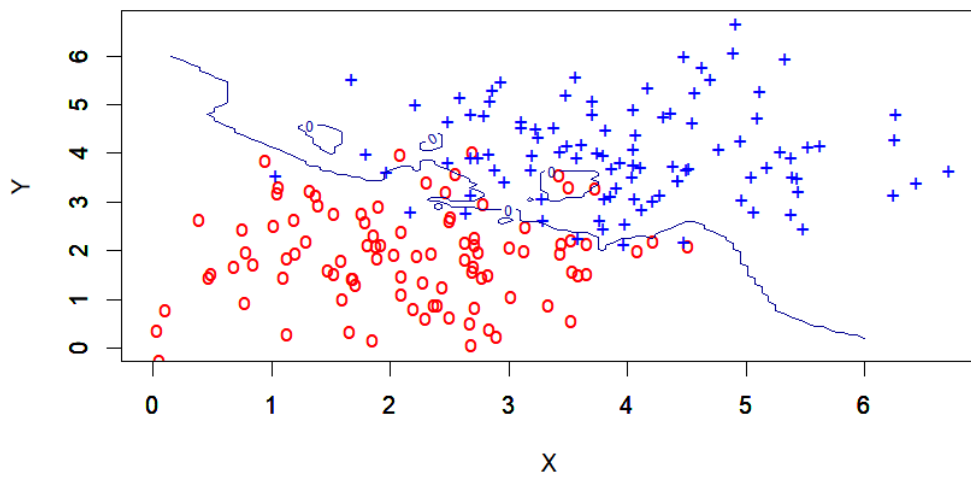


Figura 5: $sd = 1.1$

Para desvio padrão = 1.1, podemos perceber que os dados começam a se misturar muito nas regiões de interseção e que a superfície de separação já apresenta muitos sinais de *overfitting* para $K = 3$, inclusive com a formação de bolsas.

3 Variação do K

Para as próximas imagens, variaremos o K utilizando diferentes esquemas de desvios padrões e plotaremos 10 valores para K, (3, 13, 23, 33, 43, 53, 63, 73, 83, 93 e 103)

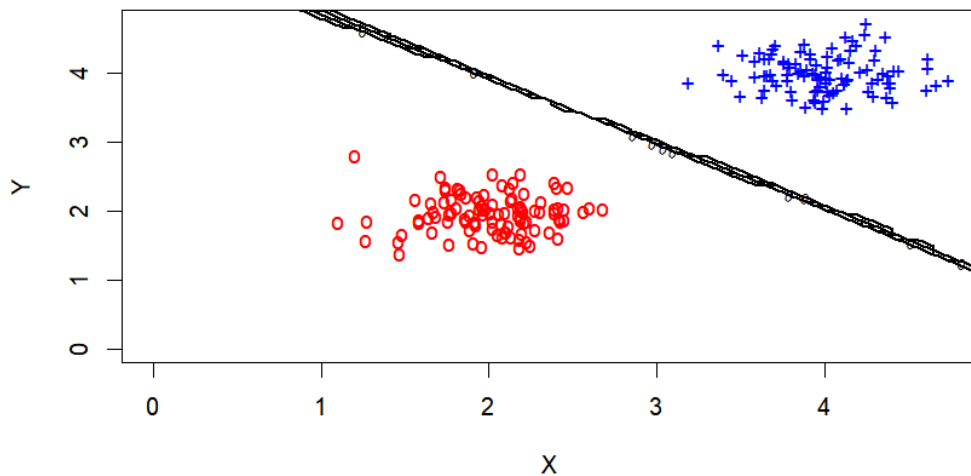


Figura 6: 10 K em 0.3

Para o exemplo acima, podemos perceber que quando o desvio padrão é muito baixo e há pouca interseção entre os dados, o valor de K não influencia no resultado.

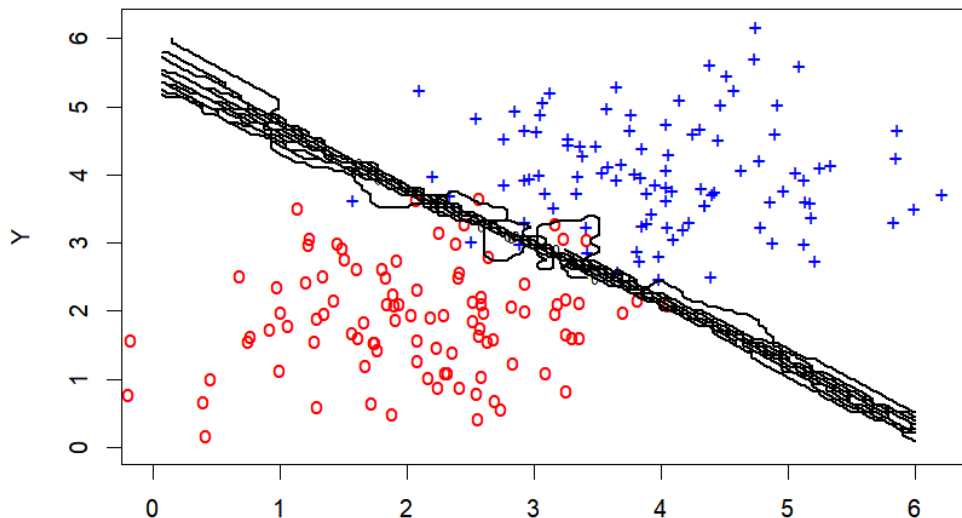


Figura 7: 10 K em 0.9

Para o exemplo acima, podemos perceber que há uma linha que apresenta *overfitting*, sendo a mesma linha demonstrada no gráfico de variação do desvio padrão e é possível perceber que as linhas subsequentes de K mais alto tendem a ser mais suaves, evitando a especialização nos dados de treinamento.

4 Conclusões

O KNN é um algoritmo que nos permite identificar um elemento e classificá-lo em uma das duas classes. Como o K é um parâmetro ajustável, deve ser configurado sabiamente pelo desenvolvedor. Pelos testes, é notável dizer que um K em valores baixos tende a especializar o modelo nos dados de treinamento, levando a um *overfitting*, enquanto valores altos de K tendem a suavizar o modelo e criar uma solução bem generalizada, especialmente quando a interseção entre os dois grupos é muito densa.