

Classificador bayesiano.

Arthur Felipe Reis Souza
Electrical Engineering Department,
Federal University of Minas Gerais,
Belo Horizonte, Brazil
arthurfreisouza@gmail.com

Antônio de Pádua Braga and Frederico Gualberto Ferreira Coelho
Electrical Engineering Department,
Federal University of Minas Gerais,
Belo Horizonte, Brazil
apbraga@cpdee.ufmg.br, fredgfc@ufmg.br

November 3, 2024

Abstract

1 Introdução

Este relatório tem por objetivo mostrar o processo de classificação utilizando a regra de Bayes como base.

2 Geração dos dados

A geração dos dados consiste em, para o primeiro conjunto, uma função normal com 240 pontos, de media 2 e desvio padrão 0,8. Enquanto para a segunda base de dados contendo apenas 120 pontos, e uma média em 4 e desvio padrão de 0,4.

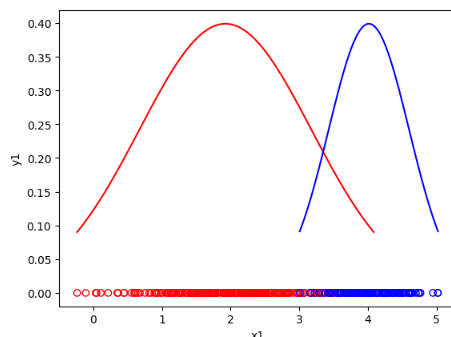


Figure 1: Geração de dados de acordo com as distribuições normais.

3 Aplicação do algoritmo

O classificador bayesiano leva em consideração a independência entre os atributos, bem como a probabilidade a priori de ocorrência da classe ao realizar a classificação. Ele se baseia no teorema de Bayes, que é descrito pela seguinte equação:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

O classificador, inicialmente, calcula a probabilidade de ocorrência de cada classe e, em seguida, multiplica esse resultado pela probabilidade da amostra pertencer a essa classe.

4 Resultados

Após aplicar o classificador bayesiano na base de dados simples, obtemos a seguinte matriz de confusão :

O classificador obteve um bom desempenho pois as amostras seguem duas distribuições normais, com médias em 2 e em 4, e também pois as colunas são independentes entre si, levando o classificador a uma clara separação. A imagem abaixo mostra os dados separados em treino e em teste.

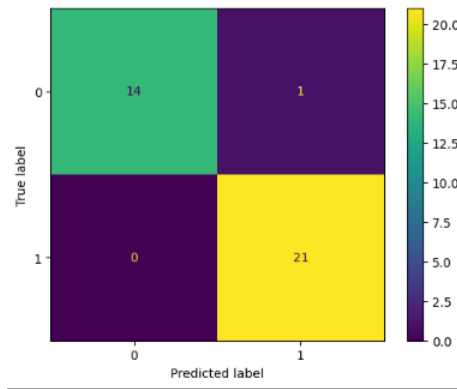


Figure 2: Matriz de confusão classificador de bayes.

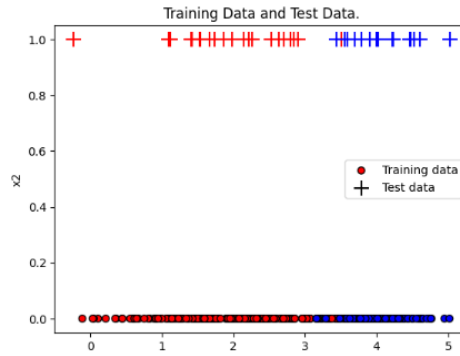


Figure 3: Pontos divididos em treino e teste.

5 Conclusão

Portanto, com esse relatório foi possível introduzir conceitos relativos ao classificador bayesiano, cuja classificação se baseia no teorema de bayes e também na independência das colunas. É um classificador muito importante, pois é consideravelmente mais rápido do que os demais e também tem um bom desempenho para dados que seguem distribuições normais.