

Introdução à Ciência de Dados

Questionário 3 - Teoria da Decisão Bayesiana

Prof. Francisco Rodrigues
Universidade de São Paulo

1 - Considerando a base de dados vertebralcolumn-2C.csv, encontre o melhor classificador dentre os métodos classificador Bayesiano, classificador Bayesiano não-paramétrico e Naive Bayes.

2 - Considerando a base de dados "winequality-red", encontre o valor do hiperparâmetro h que oferece a maior acurácia no Bayesiano não-paramétrico.

3 - Faça a classificação dos dados gerados artificialmente com o código abaixo. Compare os resultados para os métodos Naive Bayes, Classificador Bayesiano paramétrico e o classificador Bayesiano não-paramétrico (escolha um valor para h).

```
from sklearn import datasets

import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))

n_samples = 1000

data = datasets.make_moons(n_samples=n_samples, noise=.5)

X = data[0]

y = data[1]

plt.scatter(X[:,0], X[:,1], c=y, cmap='viridis', s=50, alpha=0.7)

plt.show(True)
```