

Bibliotecas

```
In [ ]: # Instalação manual de bibliotecas
!pip install numpy matplotlib pandas seaborn
# Bibliotecas

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import math
```

Requirement already satisfied: numpy in ./venv/lib/python3.10/site-packages (1.26.2)
Requirement already satisfied: matplotlib in ./venv/lib/python3.10/site-packages (3.8.2)
Requirement already satisfied: pandas in ./venv/lib/python3.10/site-packages (2.1.4)
Requirement already satisfied: seaborn in ./venv/lib/python3.10/site-packages (0.13.0)
Requirement already satisfied: contourpy>=1.0.1 in ./venv/lib/python3.10/site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in ./venv/lib/python3.10/site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in ./venv/lib/python3.10/site-packages (from matplotlib) (4.46.0)
Requirement already satisfied: kiwisolver>=1.3.1 in ./venv/lib/python3.10/site-packages (from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in ./venv/lib/python3.10/site-packages (from matplotlib) (23.2)
Requirement already satisfied: pillow>=8 in ./venv/lib/python3.10/site-packages (from matplotlib) (10.1.0)
Requirement already satisfied: pyparsing>=2.3.1 in ./venv/lib/python3.10/site-packages (from matplotlib) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in ./venv/lib/python3.10/site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in ./venv/lib/python3.10/site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in ./venv/lib/python3.10/site-packages (from pandas) (2023.3)
Requirement already satisfied: six>=1.5 in ./venv/lib/python3.10/site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)

Questões teóricas

1) Explique o porquê do classificador Naive Bayes ser robusto a um atributo com valor constante ao longo de todas as amostras de um conjunto de dados.

O classificador Naive Bayes parte do pressuposto de que os valores dos atributos de

cada amostra são independentes entre si, ou seja, que não há correlação entre um atributo x_1 e outro atributo qualquer da mesma amostra, x_2 .

A partir dessa premissa, ele realiza a classificação da seguinte forma:

- Calcula-se a probabilidade de, dado os atributos de uma amostra, aquela amostra pertencer a uma classe determinada
- Repete-se o processo, para cada uma das classes possíveis
- A classe com maior probabilidade, dado o conjunto de atributos de uma amostra, é escolhida para a classificação.

A probabilidade de uma amostra pertencer à cada classe é calculada a partir da frequência e cada classificação dado os atributos das amostras do grupo de treinamento.

Um atributo cujo valor é o mesmo para todas as amostras de treinamento não afeta de forma tão significativa o método, pois a probabilidade se torna a mesma para todas as classes. O método não irá tender a nenhuma classe em particular por consequência desse atributo.

2) Explique com suas palavras como funciona o algoritmo DBSCAN para agrupar as amostras.

O Algoritmo DBSCAN requer dois hiperparâmetros: Eps e $MinPts$.

Eps representa o raio de agrupamento, e $MinPts$ representa um número mínimo de pontos.

Munido desses dois valores, o algoritmo irá analisar ponto por ponto. Para um determinado ponto P_i , o algoritmo irá verificar quantos outros pontos existem dentro do raio Eps de distância desse ponto. Se o número de pontos for maior ou igual a $MinPts$, o ponto P_i é denominado como um ponto de núcleo. O processo é repetido até que todos os pontos tenham sido avaliados.

Quando dois pontos de núcleo estão dentro do raio de agrupamento um do outro, eles formam um grupo, ou cluster. Pontos dentro do raio de agrupamento de pontos de núcleo mas que não são eles mesmos pontos de núcleo são denominados pontos de borda. Pontos fora do raio de borda de qualquer ponto de núcleo são outliers, e não pertencem a agrupamentos.

Partindo dessas definições, o algoritmo constrói grupos (clusters), e cada grupamento define uma classe distinta.