

Questionário: Regressão logística e Knn

1 - Considere os dados obtidos a partir do código a seguir. Encontre o melhor valor de k e a métrica de distância no algoritmo knn. A classe a ser considerada está na primeira (fruit_label) ou segunda coluna (fruit_name). A que não for usada, precisa ser removida. Notem que os atributos nominais precisam ser convertidos em numéricos usando OneHotEncoder.

Dicas:
 Use vm = ['euclidean', 'manhattan', 'chebyshev', 'minkowski']
 Não precisa usar validação cruzada, use diretamente o treinamento e teste uma vez.

```
import random
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
random.seed(42) # define the seed (important to reproduce the results)

import pandas as pd

# URL do CSV "raw" no GitHub
url = (
    "https://raw.githubusercontent.com/rocksaint/fruit-data-with-colours/"
    "master/fruit_data_with_colours.csv"
)

data = pd.read_csv(url)

nrow, ncol = data.shape
print("Matriz de atributos: Número de linhas:", nrow, " colunas: ", ncol)
attributes = list(data.columns)
data.head(10)
```

2 - Faça a classificação usando regressão logística.