

Laboratório 3

Agrupamento (Clustering) k-means

Fernando Marcos e Paulo Vieira

Maio de 2023

Laboratório 3

Atividade 1 - No desenvolvimento deste trabalho use o dataset *digits* (relacionado com a leitura e reconhecimento dos dígitos de 0 a 9 escritos à mão) através de:

```
from sklearn.datasets import load_digits
data, labels = load_digits(return_X_y=True)
```

Atividade 2 - Adapte, para este dataset, o código Python usado no problema de Clustering de bolhas aleatórias (algoritmo K-means), apresentado na aula e descrito na Figura 3.1,

Atividade 3 - Trabalhando com o conceito de redução de dimensão do vetor input, usando a ferramenta PCA, descrita no parágrafo 2.2.3, construa um gráfico de dispersão que represente o agrupamento K-means dos dados em questão bem como os centroides dos respectivos clusters.

Atividade 4 - (Objetivo Final) Faça a estratificação deste dataset em dados de treino e de teste, importando a livreria `train_test_split` e usando o código Python,

```
from sklearn.model_selection import train_test_split
X, y = load_digits(return_X_y=True)
X_train, X_test, y_train, y_test =
    train_test_split(X, y, stratify=y, test_size=0.33, random_state=42)
```

e crie os respectivos códigos,

- Treino.py
- Teste.py

replicando a atividade 2 para estas duas situações e usando a ferramenta `pickle` para a criação/serialização de um ficheiro `Preditor`, bem como das matrizes de dados `X_train` e `X_test`.