

O objetivo deste exercício é realizar a atividade de agrupamento (ou clusterização) em um conjunto de dados artificial.

Utilizando os dados provenientes do dataset “Breast Cancer Wisconsin (original)”, disponível em UCI machine Learning (<https://archive.ics.uci.edu/ml/index.php>)

Para este dataset, realize as seguintes atividades e responda as perguntas:

- 1) Verifique se existem dados faltantes. Caso existam, elimine as linhas.
- 2) Apresente os dados do arquivo em um gráfico de três dimensões (escolha as melhores dimensões). Os pontos parecem estar de alguma forma agrupados? Visualmente, se você tivesse que agrupá-los, quantos grupos sugeriria?
- 3) Apresente os histogramas de cada uma das dimensões, e verifique se existem outliers.
- 4) Faça o agrupamento utilizando o algoritmo k-means, com $k = 2$.
- 5) Analise o resultado em relação às classes esperadas.
- 6) Utilize o método do cotovelo (elbow method). O valor de k sugerido pelo método está de acordo com o esperado?
- 7) Quais as suas conclusões sobre a utilização de clusterização para este problema?

Entrega:

Data de entrega: 01/06, durante o horário da aula.

O trabalho poderá ser realizado em duplas.

A dupla deverá entregar um relatório com o código fonte, e as perguntas respondidas, via Teams.