

UNIDADE 4

Aprendizado de Máquina

Agrupamento

Exercícios

Disciplina: Tópicos Especiais III (DEC7553)

Prof. Alexandre L. Gonçalves

E-mail: a.l.goncalves@ufsc.br

Exercícios

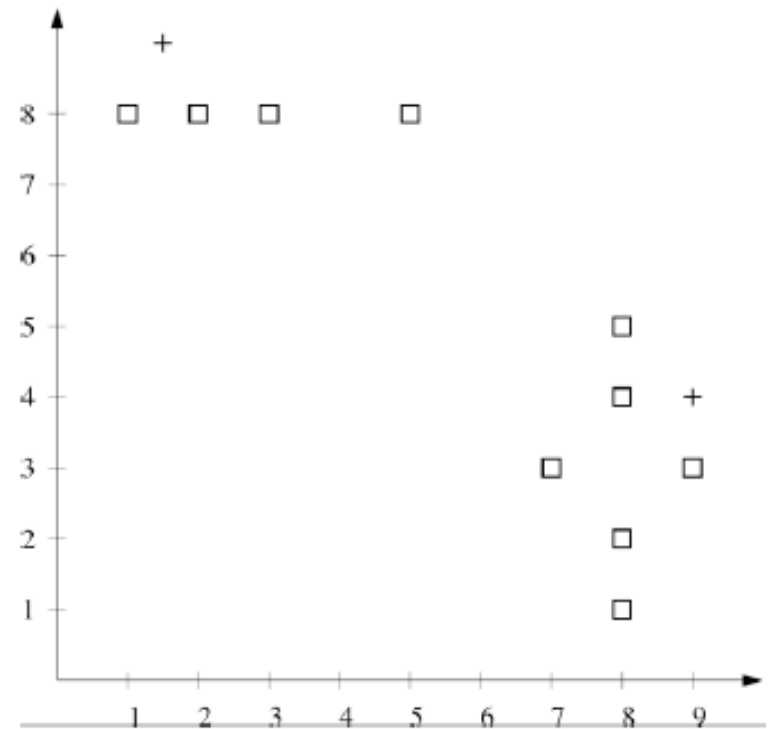
1. Considerando o conjunto de dados a seguir com 10 instâncias e, utilizando o algoritmo *k-means*, desenvolva os cálculos manualmente utilizando a Distância Euclidiana:

$$\bar{x}_i = \left\{ \begin{pmatrix} 1 \\ 8 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 8 \end{pmatrix}, \begin{pmatrix} 5 \\ 8 \end{pmatrix}, \begin{pmatrix} 8 \\ 5 \end{pmatrix}, \begin{pmatrix} 8 \\ 4 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \end{pmatrix}, \begin{pmatrix} 9 \\ 3 \end{pmatrix}, \begin{pmatrix} 8 \\ 2 \end{pmatrix}, \begin{pmatrix} 8 \\ 1 \end{pmatrix} \right\}$$

E, sendo $k=2$ e os centróides iniciais:

$$c_1 = \begin{pmatrix} 1 \\ 9 \end{pmatrix}, c_2 = \begin{pmatrix} 9 \\ 4 \end{pmatrix}$$

Quais são os valores dos centróides depois que o algoritmo *k-means* convergir usando a função de Distância Euclidiana?



Exercícios

2. Sugira um pequeno conjunto de dados (algumas linhas, entre 10 e 20, e algumas colunas) e, a partir disso, apresente os cálculos com base no algoritmo *k-means* para determinar os agrupamentos com um $k=3$. Ao final apresente os valores dos centróides.
3. Utilizando o Weka aplique o algoritmo *K-means* sobre um conjunto de dados obtidos na web que possua pelo menos 3 classes. Utilize diferentes valores de K para verificar o comportamento do algoritmo em relação a curva de erro. Ao final apresente um gráfico da curva de erros.
4. Utilizando um conjunto de dados obtido na *web* (exercício 3) e com base no exemplo em Python realize uma implementação para a tarefa de agrupamento. Desenvolva também uma estratégia de determinação do valor de K , assim como apresente a acurácia. Note que a determinação da acurácia somente é possível devido a natureza do conjunto de dados, ou seja, um conjunto de dados para classificação.