**Exercícios: algoritmo k - médias**

1. Dados os exemplos indicados na tabela, caracterizados por 2 atributos, utilize o algoritmo k médias para determinar uma partição adequada dos dados em 2 grupos.



Realize 3 execuções do algoritmo, partindo dos diferentes centros iniciais a seguir indicados:

* 1. C1= (0, -2), C2 = (0, 2)
  2. C1 = (-5, 2), C2 = (5, -2)
  3. C1 = (0, -2), C2 = (5, 2)

1. Pretende-se agrupar os seguintes 8 pontos (onde (x,y) representa a localização espacial) em 2 grupos, usando o algoritmo k-means:



A medida de distância a utilizar é a distância euclideana.

* 1. Usando como centros iniciais os pontos A2 e A1, aplique o algoritmo e indique os grupos obtidos e a avaliação dessa partição pelo critério intra-grupos da soma do quadrado dos erros e pelo critério inter-grupos da distância entre centros.
  2. Seria possível obter uma melhor partição? No caso afirmativo, apresente-a.

1. Se pretende agrupar oito pontos: x1 = (2, 10), x2 = (2, 5), x3 = (8, 4), x4 = (5, 8), x5 = 7, 5), x6 = (6, 4), x7 = (1, 2), x8 = (4, 9). Suponhamos que x1, x4 e x7 constituem as sementes iniciais dos *clusters* para o algoritmo k – médias (k = 3). Utilizando o algoritmo k – médias com a distância Manhattan calcule os três *clusters* para cada época até que o algoritmo convirja.
2. Use o algoritmo k – médias e a distância euclidiana para agrupar os oito exemplos seguintes em três *clusters*: x1 = (2, 10), x2 = (2, 5), x3 = (8, 4), x4 = (5, 8), x5 = (7, 5), x6 = (6, 4), x7 = (1, 2) e x8 = (4, 9). A matriz de distâncias, baseada na distância euclidiana, é dada a continuação.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **X1** | **X2** | **X3** | **X4** | **X5** | **X6** | **X7** | **X8** |
| **X1** | 0 |  |  |  |  |  |  |  |
| **X2** |  | 0 |  |  |  |  |  |  |
| **X3** |  |  | 0 |  |  |  |  |  |
| **X4** |  |  |  | 0 |  |  |  |  |
| **X5** |  |  |  |  | 0 |  |  |  |
| **X6** |  |  |  |  |  | 0 |  |  |
| **X7** |  |  |  |  |  |  | 0 |  |
| **X8** |  |  |  |  |  |  |  | 0 |

Suponhamos que as sementes iniciais (centros de cada *cluster*) são x1, x4 e x7. Execute o algoritmo k – médias por uma época. No fim dessa época mostre:

* 1. Os novos *clusters* (ou seja, as instâncias que pertencem a cada cluster).
  2. Os centroides dos novos *clusters*
  3. Ilustre num gráfico de 10 x 10 as oito instâncias e mostre os *clusters* no final da primeira época e os correspondentes centroides.
  4. Quantas iterações adicionais seriam necessárias para que converja o algoritmo? Mostre o resultado para cada época.

1. Considere o seguinte conjunto de exemplos não etiquetados:

Utilize o algoritmo k – médias para encontrar as três classes existentes. Calcule e mostre as posições dos centroides durante todas as iterações do algoritmo.

