

Exercício teórico Aprendizado não supervisionado

November 7, 2023

Vitor Albuquerque de Paula Aula 10 - Exercício teórico Aprendizado não supervisionado

1 Execute o passo a passo do algoritmo *complete linkage* nos dados a seguir:

	1	2	3	4	5
1	0				
2	2	0			
3	3	6	5	0	
4	4	10	9	4	0
5	5	9	8	5	3

1.1 Passo 1

Encontramos o menor par de distância, que é entre os pontos 1 e 2 (distância = 2). Agrupamos 1 e 2 juntos.

Agora, atualizamos a tabela de distâncias considerando o novo cluster $\{1,2\}$:

	$\{1,2\}$	3	4	5
$\{1,2\}$	0	6	10	9
3	6	0	4	5
4	10	4	0	3
5	9	5	3	0

1.2 Passo 2

O próximo menor par de distância é entre 4 e 5 (distância = 3). Agrupamos 4 e 5.

Atualizamos a tabela de distâncias com o novo cluster $\{4,5\}$:

	$\{1,2\}$	3	$\{4,5\}$
$\{1,2\}$	0	6	10
3	6	0	5
$\{4,5\}$	10	5	0

1.3 Passo 3

Agora, o menor par de distância é entre o cluster $\{4,5\}$ e o ponto 3 (distância = 5). Agrupamos $\{4,5\}$ e 3.

A tabela de distâncias atualizada com o cluster $\{3,4,5\}$ é:

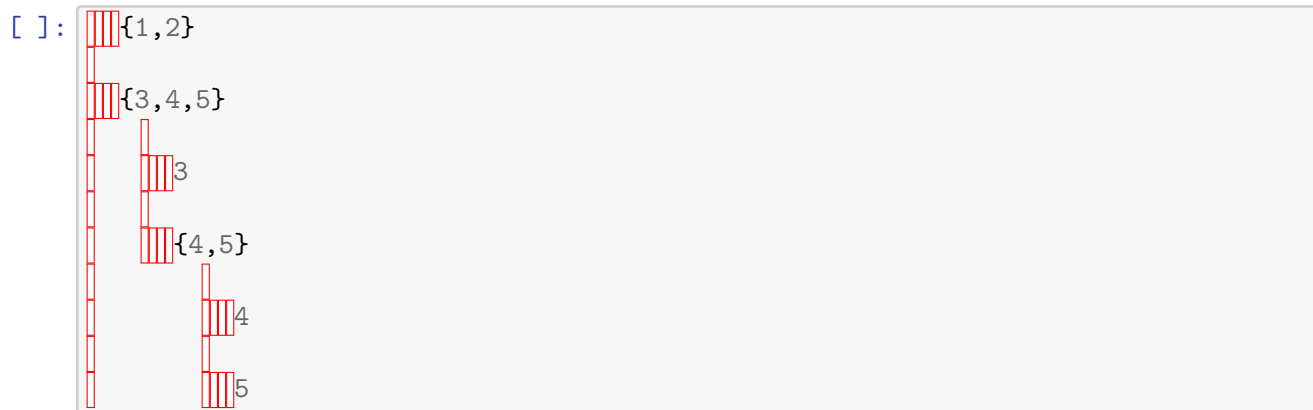
	$\{1,2\}$	$\{3,4,5\}$
$\{1,2\}$	0	10
$\{3,4,5\}$	10	0

1.4 Passo Final

Por fim, restam apenas dois clusters: $\{1,2\}$ e $\{3,4,5\}$. Eles são agrupados juntos para formar um único cluster $\{1,2,3,4,5\}$.

A tabela final não é mais necessária, pois todos os pontos foram agrupados em um único cluster.

Este processo resulta em uma árvore de agrupamento (dendrograma) que mostra as etapas de agrupamento dos pontos de dados.



2 Execute o passo a passo do algoritmo *k-means* com $k=3$ nos dados abaixo a partir dos protótipos $[6, 6]$, $[4, 6]$ e $[5, 10]$:

	A1	A2
1	1	2
2	2	1
3	1	1
4	2	2
5	8	9
6	9	8
7	9	9
8	8	8
9	1	15
10	2	15

	A1	A2
11	1	14
12	2	14

Vamos aplicar o algoritmo *k-means* com $k=3$ nos dados fornecidos usando os protótipos iniciais [6, 6], [4, 6] e [5, 10].

Etapas: 1. Atribuir pontos aos clusters mais próximos dos protótipos - Protótipo P1 = [6,6]: pontos 5, 6, 7, 8 - Protótipo P2 = [4,6]: pontos 1, 2, 3, 4 - Protótipo P3 = [5,10]: pontos 9, 10, 11, 12

2. Atualizar os protótipos calculando a média dos pontos pertencentes a cada cluster
 - P1 (Cluster 1): $(8+9+9+8)/4 = 8.5$, $(9+8+9+8)/4 = 8.5$
 - P2 (Cluster 2): $(1+2+1+2)/4 = 1.5$, $(2+1+1+2)/4 = 1.5$
 - P3 (Cluster 3): $(1+2+1+2)/4 = 1.5$, $(15+15+14+14)/4 = 14.5$
3. Verificar se há alguma mudança na atribuição dos pontos aos clusters dos novos protótipos. Nesse caso, todas as atribuições permaneceram as mesmas, então podemos considerar que o algoritmo convergiu.

Resultado final:

- Cluster 1: pontos 5, 6, 7, 8; Protótipo [8.5, 8.5]
- Cluster 2: pontos 1, 2, 3, 4; Protótipo [1.5, 1.5]
- Cluster 3: pontos 9, 10, 11, 12; Protótipo [1.5, 14.5]

3 Considere a seguinte lista de compras de 6 clientes. Aplique o algoritmo *Apriori* considerando suporte ≥ 2 e confiança $> 60\%$

- 1 biscoito, cerveja, pão, salaminho
- 2 cerveja, couve, linguiça, pão, queijo
- 3 café, brócolis, couve, pão
- 4 brócolis, café, cerveja, couve, pão, salaminho
- 5 brócolis, café, couve, pão, refrigerante
- 6 couve, linguiça

Para aplicar o algoritmo Apriori, primeiramente devemos encontrar os conjuntos de itens frequentes, considerando suporte ≥ 2 . Em seguida, avaliaremos as regras de associação com confiança maior que 60%.

Etapas 1: Encontrar os conjuntos de itens frequentes

1. Contar o suporte dos itens individuais:
 - biscoito: 1
 - cerveja: 3
 - pão: 5
 - salaminho: 2
 - couve: 5
 - linguiça: 2

- queijo: 1
 - café: 3
 - brócolis: 3
 - refrigerante: 1
2. Remover itens com suporte menor que 2:
- queijo
 - refrigerante
 - biscoito
3. Contar o suporte dos pares de itens:
- cerveja, pão: 2
 - cerveja, salaminho: 2
 - cerveja, couve: 2
 - pão, salaminho: 2
 - pão, couve: 4
 - pão, linguiça: 2
 - pão, café: 3
 - pão, brócolis: 3
 - salaminho, couve: 1
 - couve, linguiça: 2
 - couve, café: 2
 - couve, brócolis: 3
 - café, brócolis: 2
4. Remover pares de itens com suporte menor que 2:
- salaminho, couve

Etapas 2: Estados de regras de associação

1. Gerar regras para os pares frequentes e calcular a confiança:
- cerveja \rightarrow pão: $2/3 = 66,7\%$
 - pão \rightarrow cerveja: $2/5 = 40\%$ (removida por ter confiança menor que 60%)
 - cerveja \rightarrow salaminho: $2/3 = 66,7\%$
 - salaminho \rightarrow cerveja: $2/2 = 100\%$
 - cerveja \rightarrow couve: $2/3 = 66,7\%$
 - couve \rightarrow cerveja: $2/5 = 40\%$ (removida por ter confiança menor que 60%)
 - pão \rightarrow salaminho: $2/5 = 40\%$ (removida por ter confiança menor que 60%)
 - salaminho \rightarrow pão: $2/2 = 100\%$
 - pão \rightarrow couve: $4/5 = 80\%$
 - couve \rightarrow pão: $4/5 = 80\%$
 - pão \rightarrow linguiça: $2/5 = 40\%$ (removida por ter confiança menor que 60%)
 - linguiça \rightarrow pão: $2/2 = 100\%$
 - pão \rightarrow café: $3/5 = 60\%$
 - café \rightarrow pão: $3/3 = 100\%$
 - pão \rightarrow brócolis: $3/5 = 60\%$
 - brócolis \rightarrow pão: $3/3 = 100\%$
 - couve \rightarrow linguiça: $2/5 = 40\%$ (removida por ter confiança menor que 60%)

- linguiça -> couve: $2/2 = 100\%$
- couve -> café: $2/5 = 40\%$ (removida por ter confiança menor que 60%)
- café -> couve: $2/3 = 66,7\%$
- couve -> brócolis: $3/5 = 60\%$
- brócolis -> couve: $3/3 = 100\%$
- café -> brócolis: $2/3 = 66,7\%$
- brócolis -> café: $2/3 = 66,7\%$

Resultado das regras de associação:

- cerveja -> pão (66,7%)
- cerveja -> salaminho (66,7%)
- salaminho -> cerveja (100%)
- salaminho -> pão (100%)
- pão -> couve (80%)
- couve -> pão (80%)
- linguiça -> pão (100%)
- pão -> café (60%)
- café -> pão (100%)
- pão -> brócolis (60%)
- brócolis -> pão (100%)
- linguiça -> couve (100%)
- café -> couve (66,7%)
- couve -> brócolis (60%)
- brócolis -> couve (100%)
- café -> brócolis (66,7%)
- brócolis -> café (66,7%)