

Lista de Exercícios 1 – Tópicos Especiais em Banco de Dados (SCC-0542)

Não é necessário entregar a resolução destes exercícios.

1. Descreva a diferença entre agrupamento e classificação.

2. Considerando o conjunto X abaixo, faça:

x_1	2	2	3
x_2	7	9	8
x_3	7	2	1
x_4	3	9	9
x_5	6	2	2
x_6	6	9	5
x_7	8	1	3
x_8	4	9	6
x_9	6	1	3
x_{10}	2	8	6

a) Simule a execução do algoritmo k-médias com os centróides iniciais $m_1=[1,0,0]$ e $m_2=[0,0,1]$ até convergir (os centróides serem iguais em duas iterações consecutivas). Quais as coordenadas finais dos centróides e o valor de SEQ da partição final?

b) Simule a execução do algoritmo hierárquico Single-Linkage e apresente o dendrograma obtido.

c) Simule a execução do algoritmo hierárquico Complete-Linkage e apresente o dendrograma obtido.

3. Cite uma vantagem em utilizar o algoritmo hierárquico Single-linkage ao invés do algoritmo Complete-linkage.

4. Considere uma aplicação que você sabe que o número de grupos é 10. A base de dados é formada por informações referentes a registros de 1.000.000 de transações financeiras descritas por 100 atributos. Qual algoritmo, entre os vistos em aula, você acha mais apropriado para esta base de dados? Justifique sua resposta.

5. No algoritmo *Bisecting K-means* é necessário um critério para definir a ordem em que os grupos devem ser divididos. Sugira um critério e justifique.

6. Simule a execução de uma iteração do algoritmo *Spherical K-means* nos dados abaixo e apresente os centróides obtidos. Considere como centróides iniciais: $m_1=[0,1,0,1,0,1,0,1,0,1]$ e $m_2=[1,0,1,0,1,0,1,0,1,0]$. Aplique a normalização apropriada.

x_1	0	1	1	1	0	0	1	0	0	1
x_2	1	0	1	0	1	0	1	0	0	1
x_3	1	0	1	0	1	0	1	1	0	0
x_4	0	1	1	0	1	0	0	1	1	0
x_5	1	0	1	1	0	1	1	0	0	0
x_6	0	1	0	1	1	0	0	1	0	1
x_7	1	0	1	0	0	0	1	1	0	1
x_8	0	0	0	1	0	1	0	1	1	1
x_9	0	0	0	1	1	0	1	1	1	0
x_{10}	0	1	0	0	0	1	0	1	1	1