



Universidade Federal de Santa Maria – Centro de Tecnologia

Peso: 4

Disciplina: Mineração de Dados (ELC1098)

Professor: Dr. Joaquim Assunção

6.7

Atenção: Leia atentamente as questões e escreva de forma clara e legível.

Nome: Denise Costa Pereira

Curso: Ciência da Computação Data: 26/11/19

- 1) Ao longo do semestre, estudamos 4 tipos diferentes de algoritmos para mineração de dados; associação, regressão, classificação e agrupamento. Quais destes não precisam ser treinados com exemplos rotulados? (1) Cite dois algoritmos e diga a qual categoria cada um pertence (1).

Associação = Associação, K-means = Agrupamento.

7.7

- 2) Um algoritmo de agrupamento consegue identificar, automaticamente, grupos de dados com base em determinados atributos. O processo de formar grupos pode ser ambíguo? Por quê? (2).

Sim ou Não?

Por durante o processo podem se formar n grupos.

Como um grupo tem certos atributos, desta forma, podemos, além dos atributos a serem agrupados, passar também a quantidade de clusters esperados, como é o caso do algoritmo K-means.

7.7

- 3) Associe os tipos de grupos com as suas descrições. (2)

- a. Bem separados
- b. Baseados em centro
- c. Contínuos
- d. Baseados em densidade

4.7

a ( ) Conjunto de pontos de modo que qualquer ponto do grupo seja mais próximo de todos os outros pontos do grupo do que de qualquer ponto que não esteja no grupo.

b ( ) Grupo em que para cada ponto do grupo existe outro ponto cuja distância é menor do que a distância deste ponto para todos os outros pontos do outro grupo. De um ponto A para um ponto B passando por todos os pontos, existe claramente uma distância menor que as demais.

d ( ) Dados grupos P e Q, e pontos de referência  $q_r$  e  $p_r$   $\forall Q, P \mid dist(Q, q_r) < dist(P, q_r)$

b ...Tendem a separar pontos próximos a um centroide devido ao vazio de objetos entre este ponto e este centroide.

- 4) Considere a tabela abaixo como dados de entrada para o algoritmo APRIORI. Qual o suporte (1,5) e a Confiança (1,5) da regra " $var1 \rightarrow classe$ "?

Var1	Var2	Var3	Classe
1	1	0	1
0	0	0	0
1	1	0	1
1	0	0	0
1	1	1	1

SUP

4/5

3/5

1/5

3/5

A confiança da Regra é de  $3/4 = 75\%$ , onde 3 é o número de ocorrências da regra e 4 o número total da variável a esquerda. O suporte é de  $3/5 = 60\%$ .

Regra / total de transações.

- 5) Em um grupo de *data science*, dois times usam o algoritmo *AdaBoost.M1* (algoritmo de *Boosting*) para classificar dados e tentar validar suas hipóteses. Usando os mesmos parâmetros numéricos, os times chegam a resultados diferentes. O time A obtém melhor acurácia mas seu algoritmo demora bem mais para rodar que o algoritmo do time B. Tendo hardware, SO e infraestrutura iguais, em poucas palavras, o que pode ter acontecido para que tais resultados sejam discrepantes? (1)