

EXAM Yuteng WANG

1.

Question 1 : La différence entre k-means et k-medoids s'entend comme la différence entre la moyenne et la médiane d'un échantillon de données : la première peut prendre n'importe quelle valeur dans l'espace continu, tandis que la seconde ne peut prendre que les échantillons de l'intervalle de l'échantillon de données. La raison en est que k-means est trop exigeant pour les échantillons de données, exigeant que tous les échantillons de données soient dans un espace européen, ce qui peut conduire à une énorme erreur s'il y a beaucoup de bruit dans les données, donc les résultats ne sont pas les mêmes !

Question 2 : Les k moyennes initiales des k centres sont d'abord choisies au hasard, puis les k clusters finales déterminées par itération sont fortement corrélées avec les k centres initiales, de sorte que cet algorithme en calcule deux qui donneront des résultats différents.

2.

- A) NON
- B) OUI
- C) OUI

3.

1. $\text{Coverage}(X \rightarrow Y) < \text{Coverage}(X' \rightarrow Y)$
 $\text{LIFT}(X \rightarrow Y) < \text{LIFT}(X' \rightarrow Y)$

2. La couverture est le support du côté gauche de la règle, c'est-à-dire $\text{supp}(X)$. La couverture est calculée rapidement à partir des mesures de qualité des règles (soutien et confiance) stockées dans le créneau de qualité.

4.

A) $\text{Min_support} = 2$

L'algorithme Apriori

C1: item1: 4 item2: 4 item3 : 3 item4 : 3 item5: 4

L1: item1: 4 item2: 4 item3 : 3 item4 : 3 item5: 4

C2 12: 2 13: 1 14: 2 15: 3 23: 2 24: 2 25: 3 34: 1 35: 2 45: 1

L2 12: 2 14: 2 15: 3 23: 2 24: 2 25: 3 35: 2

C3 124: 1 125: 2 145: 1 234: 1 235: 2 245: 1

L3 125: 2 235: 2

L4 = \emptyset

Donc 1 2 3 4 5 12 14 15 23 24 25 35 125 235

B) Min_confiance = 0.8

1->2 confiance: $\text{support}(1 \text{ and } 2) / \text{support}(1) = 1/2$

1->3 confiance: $\text{support}(1 \text{ and } 3) / \text{support}(1) = 1/4$

1->4 confiance: $\text{support}(1 \text{ and } 4) / \text{support}(1) = 2/4$

1->5 confiance: $\text{support}(1 \text{ and } 5) / \text{support}(1) = 3/4 = 0.75$

2->1 confiance: $\text{support}(2 \text{ and } 1) / \text{support}(2) = 1/2$

3->1 confiance: $\text{support}(3 \text{ and } 1) / \text{support}(3) = 1/3$

4->1 confiance: $\text{support}(4 \text{ and } 1) / \text{support}(4) = 2/3 = 0.666$

5->1 confiance: $\text{support}(5 \text{ and } 1) / \text{support}(5) = 3/4 = 0.75$

12->5 confiance: $\text{support}(1 \text{ and } 2 \text{ and } 5) / \text{support}(12) = 2/2 = 1$

5->12 confiance: $\text{support}(1 \text{ and } 2 \text{ and } 5) / \text{support}(5) = 1/2$

1->25 confiance: $\text{support}(1 \text{ and } 2 \text{ and } 5) / \text{support}(1) = 1/2$

25->1 confiance: $\text{support}(1 \text{ and } 2 \text{ and } 5) / \text{support}(25) = 2/3 = 0.666$