Université Mohamed Boudiaf M'sila
Faculté MI, Dépt Informatique

M2 SIGL
M'sila, Le 24/02/2021
Durée : 60 Minutes

Examen Final

Data Mining et Recherche d'Information

Exercice 1 (5 points)

Répondre brièvement à ces questions ***(1 Pt/question)***
Question 1.1 : Donner le principe général de l'algorithme K-means.
Question 1.2 : Quelle est la complexité de K-means ?
Question 1.3 : Donner le principe général de l'algorithme K-medoids.
Question 1.4 : Quel est l'avantage principal de K-medoids par rapport au K-means?
Question 1.5 : Quel est l'inconvénient principal de K-medoids ?

Exercice 2 (7 points)

Soit l'ensemble de données suivant :

X1 = 0 X2 = 2 X3 = 6 X4 = 11

Question 2.1 : Appliquer un CAH sur ces données en utilisant le lien minimum (single link). ***(3 Pts)***
Question 2.2 : Représenter le dendrogramme. ***(2 Pts)***
Question 2.3 : Discuter les différents clusters obtenus selon la distance d'agrégation. ***(2 Pts)***

Exercice 3 (08 points)

A l'approche du mois de Ramadhan, l'épicerie de la rue décide de lancer une grande opération de promotion. Son patron veut appliquer les techniques de data mining sur ses données. Il vous demande d'utiliser les règles d'associations pour trouver des règles intéressantes pour ses futures promotions. Un extrait des données de ventes est présenté dans le tableau suivant.

Achats	Produit 1	Produit 2	Produit 3	Produit 4	Produit 5
A1	X			X	X
A2	X	X			X
A3					X
A4			X	X	X
A5	X	X	X	X	X
A6	X				X
A7	X			X	X
A8		X	X		

Question 3.1 : Extraire les règles d'association avec un support de 0.5 ***(6 Pts)***
Question 3.2 : Que pouvez-vous conseiller comme promotion au patron ? ***(2 Pts)***

BONNE REUSSITE

Dr. Tahar Mehenni

Correction de l'Examen Final

Data Mining et Recherche d'Information

Exercice 1 (5 points)

Question 1.1 : Donner le principe général de l'algorithme K-means.

Réponse : 1. Choisir k objets formant ainsi k clusters
2. (Ré)affecter chaque objet O au cluster Ci de centre Mi tel que dist(O,Mi) est minimal
3. Recalculer Mi de chaque cluster (le barycentre)
4. Aller à l'étape 2 si on vient de faire une affectation

(1 Pt)

Question 1.2 : Quelle est la complexité de K-means ?

Réponse : O(tkn), où n est # objets, k est # clusters, et t est # itérations ***(1 Pt)***

Question 1.3 : Donner le principe général de l'algorithme K-medoids.

Réponse : - On choisit initialement k medoids aléatoirement.
- Tant que c'est possible, on remplace un medoid par un autre point, de telle façon qu'on obtienne la plus grande amélioration possible du score (réduction de la distance globale).
- Quand aucune amélioration n'est plus possible, on arrête.

(1 Pt)

Question 1.5 : Quel est l'avantage principal de K-medoids par rapport au K-means?

Réponse : Les clusters sont mieux représentés, ce qui donne un clustering de qualité. ***(1 Pt)***

Question 1.6 : Quel est l'inconvénient principal de K-medoids ?

Réponse : Le calcul est long et complexe. ***(1 Pt)***

Exercice 2 (7 points)

X1 = 0 X2 = 2 X3 = 6 X4 = 11

Question 2.1 : Appliquer un CAH sur ces données en utilisant le lien minimum (single link). ***(3 Pts)***

	x_1	x_2	x_3	x_4
x_1	0			
x_2	**2**	0		
x_3	6	4	0	
x_4	11	9	5	0

grouper (x_1, x_2) **(1 pt)**

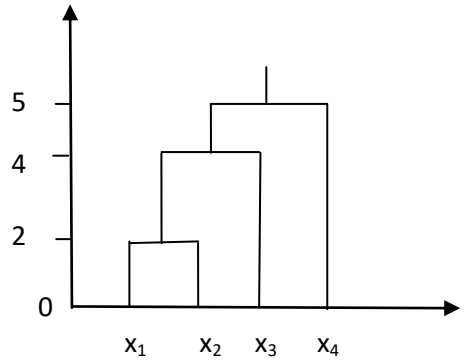
	x_1x_2	x_3	x_4
x_1x_2	0		
X_3	**4**	0	
X_4	9	5	0

grouper(x_3,x_1x_2) **(1 pt)**

	$x_1x_2x_3$	X_4
$x_1x_2x_3$	0	
X_4	**5**	0

grouper (x_4, x_1x_2 x_3) **(1 pt)**

Question 2.2 : Représenter le dendrogramme. ***(2 Pt)***



Question 2.3 : Discuter les différents clusters obtenus selon la distance d'agrégation. ***(2 Pts)***

CAH a donné les clusters suivants selon la distance d'agrégation d :

2<d<4 : trois clusters : (x1, x2), (x3) et (x4) ***(1 Pt)***

4<d<5 ; deux clusters : (x1, x2, x3) et (x4) ***(1 Pt)***

Exercice 3 (8 points)

Achats	Produit 1	Produit 2	Produit 3	Produit 4	Produit 5
A1	X			X	X
A2	X	X			X
A3					X
A4			X	X	X
A5	X	X	X	X	X
A6	X				X
A7	X			X	X
A8		X	X		

Question 3.1 : Extraire les règles d'associations avec un support de 0.5 ***(6 Pts)***

1- On commence par trouver tous les 1-itemsets (itemsets de taille 1) et leurs supports. **(4pts)**

{P1}– 5/8, support = 0.625 {P2} – 3/8, support = 0.375 {P3} – 3/8, support = 0.375

{P4} – 4/8, support = 0.5 {P5} – 7/8, support = 0.875

On retient les itemsets ayant un support 0.5 (P1, P4, P5)

{P1}– 5/8, support = 0.625 {P4} – 4/8, support = 0.5 {P5} – 7/8, support = 0.875 **(1.5 pt)**

Sur la base des 1-itemsets frequents, on génère les 2-itemsets.

{P1, P4} – 3/8, support = 0.375 {P1,P5} – 4/8, support = 0.5 {P4, P5} – 4/8, support = 0.5 **(1.5 pt)**

On retient les itemsets ayant un support 0.5 (P1,P5) et (P4,P5)

Sur la base des 2-itemsets frequents, on génère les 3-itemsets.

{P1, P4, P5} – 3/8, support = 0.375. **Non accepté**. **(0.5 pt)**

Ceci termine le processus de génération des itemsets fréquents. {P1, P4, P5, (P1,P5), (P4,P5)} **(0.5 pt)**

2- Génération des règles à partir des itemsets fréquents : **(2 pts)**

P1 $\rightarrow$ P5 confiance= 4/5=0.8 , P5 $\rightarrow$ P1 confiance= 4/7=0.57

P4 $\rightarrow$ P5 confiance= 4/4=1 , P5 $\rightarrow$ P4 confiance= 4/7=0.57

Question 3.2 : Que pouvez-vous conseiller comme promotion au patron ? ***(2 Pts)***

La règle P4 $\rightarrow$ P5 a une confiance égale à 1. Le patron peut lancer une promotion concernant les produits P4 et P5.

Le patron peut aussi lancer une promotion sur les produits P1 et P5 (règle P1 $\rightarrow$ P5 confiance=0.8)

Rédigé par T. Mehenni