

Correction de l'Examen Final

Data Mining et Recherche d'Information

Exercice 1 (6 points)

$E = \{3, 7, 10, 12, 15, 17\}$

Etape 1 : (2 pts)

$c1 = 10$ et $c2 = 12$, $C1 = \{10\}$, $C2 = \{12\}$ (0.5 pt)

$d(3, c1) = |3 - 10| = 7$, $d(3, c2) = |3 - 12| = 9$, $3 \rightarrow C1$ (0.25 pt)

$d(7, c1) = |7 - 10| = 3$, $d(7, c2) = |7 - 12| = 5$, $7 \rightarrow C1$ (0.25 pt)

$d(15, c1) = |15 - 10| = 5$, $d(15, c2) = |15 - 12| = 3$, $15 \rightarrow C2$ (0.25 pt)

$d(17, c1) = |17 - 10| = 7$, $d(17, c2) = |17 - 12| = 5$, $17 \rightarrow C2$ (0.25 pt)

$C1 = \{3, 7, 10\}$, $C2 = \{12, 15, 17\}$ (0.5 pt)

Etape 2 : (4 pts)

$c1 = (3 + 7 + 10) / 3 = 6.66$ et $c2 = (12 + 15 + 17) / 3 = 14.66$ (0.5 pt)

$d(3, c1) = |3 - 6.66| = 3.66$, $d(3, c2) = |3 - 14.66| = 11.66$, $3 \rightarrow C1$ (0.25 pt)

$d(7, c1) = |7 - 6.66| = 0.34$, $d(7, c2) = |7 - 14.66| = 7.66$, $7 \rightarrow C1$ (0.25 pt)

$d(10, c1) = |10 - 6.66| = 3.34$, $d(10, c2) = |10 - 14.66| = 4.66$, $10 \rightarrow C1$ (0.25 pt)

$d(12, c1) = |12 - 6.66| = 5.34$, $d(12, c2) = |12 - 14.66| = 2.66$, $12 \rightarrow C2$ (0.25 pt)

$d(15, c1) = |15 - 6.66| = 8.34$, $d(15, c2) = |15 - 14.66| = 0.34$, $15 \rightarrow C2$ (0.25 pt)

$d(17, c1) = |17 - 6.66| = 10.34$, $d(17, c2) = |17 - 14.66| = 2.34$, $17 \rightarrow C2$ (0.25 pt)

$C1 = \{3, 7, 10\}$, $C2 = \{12, 15, 17\}$ (0.5 pt)

Aucun déplacement (stabilité) (0.5 pt), d'où $C1 = \{3, 7, 10\}$, $C2 = \{12, 15, 17\}$ (1 pt)

Exercice 2 (8 points)

	Produit A	Produit B	Produit C	Produit D	Produit E	Produit F	Produit G	Produit H
Achat 1	X	X	X	X				
Achat 2	X	X	X		X	X		
Achat 3	X	X	X			X	X	
Achat 4		X		X				X
Achat 5					X		X	X

1- On commence par trouver tous les 1-itemsets (itemsets de taille 1) et leurs supports. (2pts)

$\{A\} - 3/5$, support 0.6 $\{B\} - 4/5$, support 0.8 $\{C\} - 3/5$, support 0.6 $\{D\} - 2/5$, support 0.4

$\{E\} - 2/5$, support 0.4 $\{F\} - 2/5$, support 0.4 $\{G\} - 2/5$, support 0.4 $\{H\} - 2/5$, support 0.4

On retient les itemsets ayant un support 0.6

$\{A\} - 3/5$, support 0.6 $\{B\} - 4/5$, support 0.8 $\{C\} - 3/5$, support 0.6 (0.75 pt)

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

$\{A, B\} - 3/5$, support 0.6 $\{B, C\} - 3/5$, support 0.6 $\{A, C\} - 3/5$, support 0.6 (0.75 pt)

Tous ces itemsets ont un support 0.6, on les utilise pour générer les 3-itemsets. Il existe un seul itemset.

$\{A, B, C\} - 3/5$, support 0.6 (0.25 pt)

Ceci termine les processus de génération des itemsets fréquents.

$\{A\} - 3/5$, support 0.6 $\{B\} - 4/5$, support 0.8 $\{C\} - 3/5$, support 0.6

$\{A, B\} - 3/5$, support 0.6 $\{B, C\} - 3/5$, support 0.6 $\{A, C\} - 3/5$, support 0.6

$\{A, B, C\} - 3/5$, support 0.6

2- Puisque la règle demandée comporte 3 items, on génère les règles d'association possibles à partir de l'itemset qui contient 3 items, c-à-d l'itemset {A, B, C} et on calcule pour chaque règle générée sa confiance. **(0.5 pt)**

$A \rightarrow \{B, C\}$ - confiance $3/3 = 1.0$ $C \rightarrow \{A, B\}$ - confiance $3/3 = 1.0$ $B \rightarrow \{A, C\}$ - confiance $3/4 = 0.75$ **(1.5 pts)**

Toutes ces règles satisfont la confiance minimale de 0.7. **(0.25 pt)**

3- Les règles d'association certaines sont:

$A \rightarrow \{B, C\}$ - confiance $3/3 = 1.0$ et $C \rightarrow \{A, B\}$ - confiance $3/3 = 1.0$. **(2 pts)**

Exercice 3 : (06 points)

Single Link (Lien minimum)

	x_1	x_2	x_3	x_4	x_5
x_1	0				
x_2	0.4	0			
x_3	0.9	0.3	0		
x_4	0.6	0.8	0.3	0	
x_5	0.5	0.7	0.2	<u>0.1</u>	0

grouper (x_4, x_5) **(1 pt)**

	x_1	x_2	x_3	x_4x_5
x_1	0			
x_2	0.4	0		
x_3	0.9	0.3	0	
x_4x_5	0.5	0.7	<u>0.2</u>	0

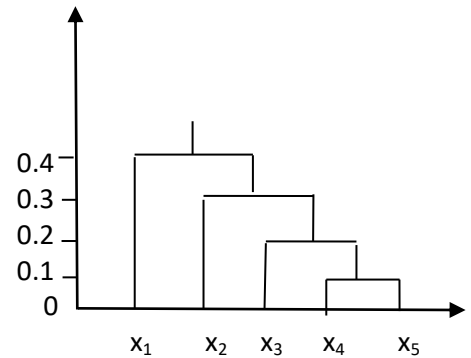
grouper(x_3, x_4x_5) **(1 pt)**

	x_1	x_2	$x_3x_4x_5$
x_1	0		
x_2	0.4	0	
$x_3x_4x_5$	0.5	<u>0.3</u>	0

grouper ($x_2, x_3x_4x_5$) **(1 pt)**

	x_1	$x_2x_3x_4x_5$
x_1	0	
$x_2x_3x_4x_5$	<u>0.4</u>	0

(1 pt)



Dendrogramme (2 Pts)