

## EXAMEN S1

### Exercice 1

Soit la matrice des transactions suivantes

	A	B	C	D	E	F
T1	1	1	1	1	1	1
T2	1	1	1	1	1	0
T3	1	0	0	0	1	0
T4	0	1	0	1	0	1
T5	1	1	1	0	1	1

- 1) Donner les étapes d'extraction des itemsets fréquents fermées pour Minsup=0.4?
- 2) Donner les règles d'association relatives aux 2-itemsets fréquents fermés relevés?
- 3) Donner les étapes d'extraction des itemsets fréquents maximaux pour Minsup=0.33?

$$0.4 = \frac{6}{10} = \frac{3}{5}$$

### Exercice 2

Soient les données suivantes :

A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), ~~A7=(1,1)~~

- 1) Faire une classification de ces données en utilisant l'algorithme k-means pour k=3. On considère les classes de départ C1=(A1,A2,A3), C2=(A4,A5), C3=(A6,~~A7~~). On utilisera city-blocs pour mesurer la distance entre les données.

### Exercice 3

- 1) Mentionner deux avantages et deux inconvénients de l'algorithme k-means?
- 2) Dans quel cas l'extraction des itemsets maximaux est intéressante?
- 3) Quelles sont les conditions à satisfaire pour considérer qu'une règle d'association est valide?
- 4) Définir la classification non supervisée?

Exercice N°1 (8,6) Min sup: 60% (3/5)

	sup	
A	4/5	AE
B	4/5	B
C	3/5	ABCE
D	3/5	BD
E	4/5	AE
F	3/5	BF

frequent

①

AB	3/5	ABCE - freq
AD	2/5	ABCDE - infreq
AF	2/5	ABCEF - freq
BE	3/5	ABCEF - freq
CF	2/5	ABCDE
DE	2/5	ABCDE
DF	2/5	BDF - infreq
EF	2/5	ABCEF
CD	2/5	ABCDE
<del>AC</del>	<del>2/5</del>	<del>ABCE</del>
<del>AD</del>	<del>2/5</del>	<del>ABCDE</del>
<del>AE</del>	<del>4/5</del>	<del>ABCE</del>
<del>AF</del>	<del>2/5</del>	<del>ABCEF</del>
<del>BD</del>	<del>2/5</del>	<del>ABDE</del>
<del>BE</del>	<del>3/5</del>	<del>ABCEF</del>
<del>BF</del>	<del>2/5</del>	<del>ABEF</del>
<del>CE</del>	<del>2/5</del>	<del>ABCE</del>
<del>DE</del>	<del>2/5</del>	<del>ABCDE</del>
<del>EF</del>	<del>2/5</del>	<del>ABCEF</del>

②/5

③

2) règles d'association pour les 2-ensembles fermés fréquents


① { AB → CE      confiance = 100%    sup = 3/5  
 { AB → CEF      confiance = 100%    sup = 3/5

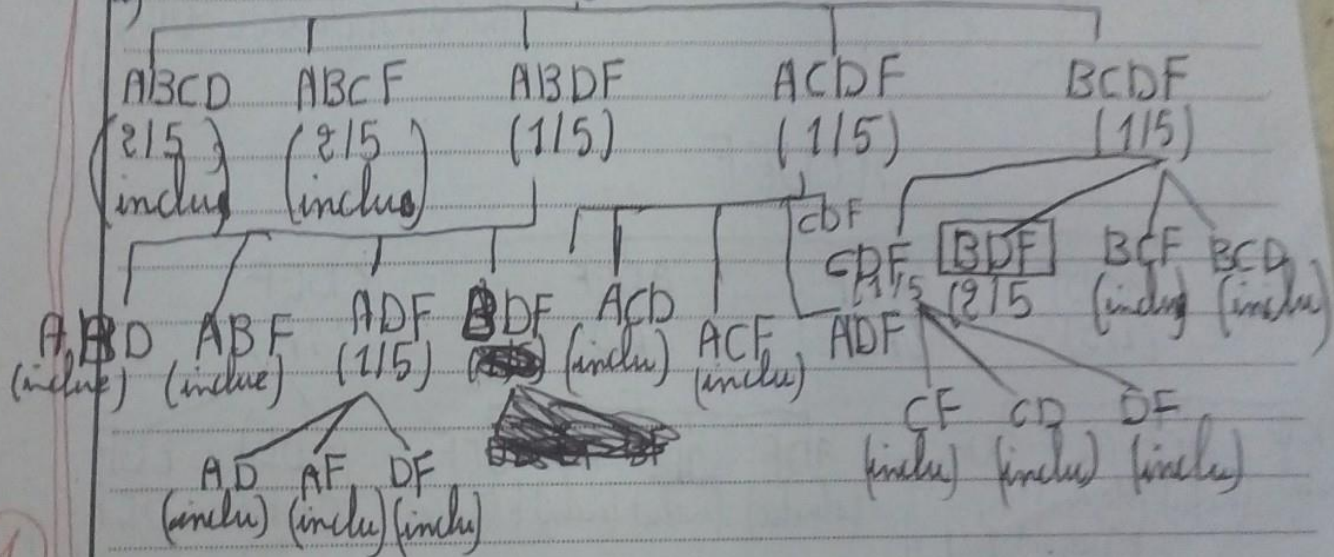


## Extraction des maximaux.

ABCEDEF (sup = 1/5) infre

$\boxed{ABCDE}$   $\boxed{ABCEF}$   $ABCDF$   $ABDEF$   $ACDEF$   $BCDEF$   
 $\text{sup} = 2/5$   $\text{sup} = 2/5$   $(1/5)$   $(\text{sup} = 1/5)$   $(\text{sup} = 1/5)$   $(\text{sup} = 1/5)$

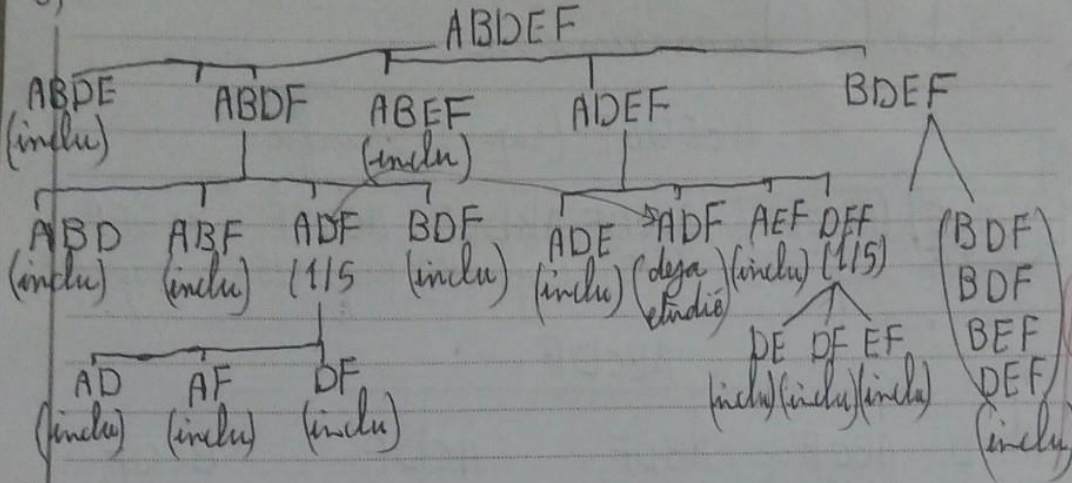
a)  ABCDEF



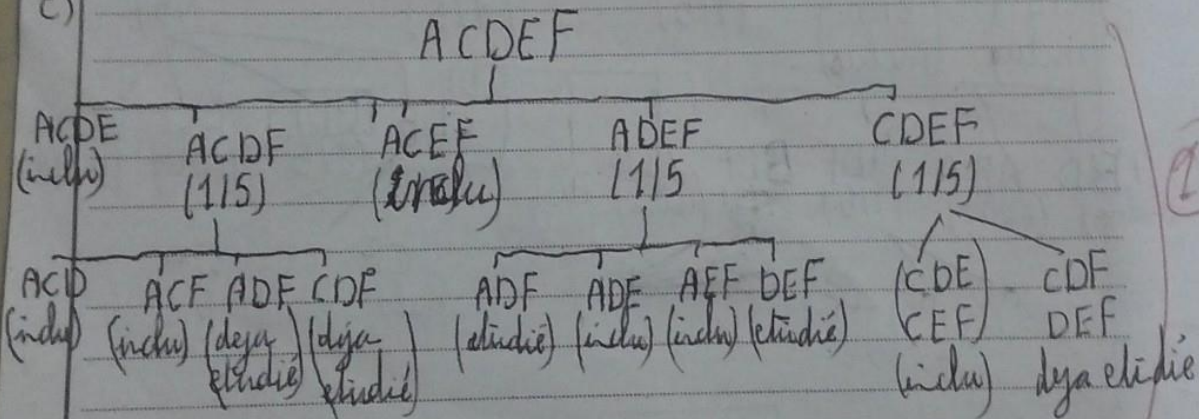
Maximaux :  $\begin{array}{|l} ABCDE \\ ABCEF \\ BDEF \end{array}$



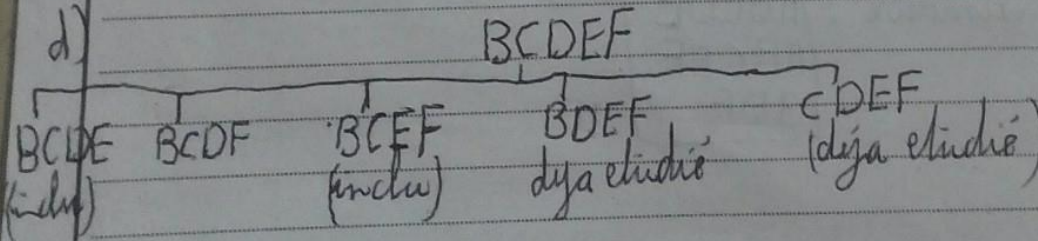
b)



c)



d)



Les stemsets maximaux

ABCDE  
ABCEF  
BDF

Ministère de l'enseignement  
Supérieur et de la Recherche  
Scientifique

Date :

Nom et Prénom :

Matricule :

Filière :

Note 20

Correcteur :

Surveillants :

05





Date : \_\_\_\_\_  
Nom et Prénom : \_\_\_\_\_  
Matricule : \_\_\_\_\_  
Filière : \_\_\_\_\_ Section : \_\_\_\_\_ Groupe : \_\_\_\_\_

Note 20

Correcteur : \_\_\_\_\_

Module : \_\_\_\_\_

Exercice N°2 : (7pts)  
calcul des centres des 3 sous

$$M1 = ((2+2+8)/3; (10+5+4)/3) = (4; 6.33)$$

$$M2 = ((5+7)/2; (8+5)/2) = (6; 6.5)$$

$$M3 = (6, 4)$$

$$d(A1, M1) = |2-4| + |10-6.33| = 2+3.67 = 5.67$$

$$d(A2, M1) = |2-4| + |5-6.33| + 1.33 = 3.33$$

$$d(A3, M1) = |8-4| + |4-6.33| = 4+2.33 = 6.33$$

$$d(A4, M1) = |5-4| + |8-6.33| = 1+1.67 = 2.67$$

$$d(A5, M1) = |7-4| + |5-6.33| = 3+1.33 = 4.33$$

$$d(A6, M1) = |6-4| + |4-6.33| = 2+2.33 = 4.33$$

$$d(A1, M2) = |2-6| + |10-6.5| = 4+3.5 = 7.5$$

$$d(A2, M2) = |2-6| + |5-6.5| = 4+1.5 = 5.5$$

$$d(A3, M2) = |8-6| + |4-6.5| = 2+2.5 = 4.5$$

ملاحظة : كل استعمال خارج عن الامتحانات يعرض صاحبه إلى تدابير تأديبية.



$$d(A2, M2) = |2 - 5| + |5 - 8| = 3 + 3 = 6$$

$$d(A3, M2) = |8 - 5| + |14 - 8| = 7$$

$$d(A4, M2) = |5 - 5| + |8 - 8| = 0$$

$$d(A5, M2) = |7 - 5| + |5 - 8| = 5$$

$$d(A6, M2) = |6 - 5| + |14 - 8| = 5$$

$$d(A1, M3) = |2 - 7| + |10 - 4.33| = 5 + 5.67 = 10.67$$

$$d(A2, M3) = |2 - 7| + |5 - 4.33| = 5 + 0.67 = 5.67$$

$$d(A3, M3) = |8 - 7| + |14 - 4.33| = 1 + 0.33 = 1.33$$

$$d(A4, M3) = |5 - 7| + |8 - 4.33| = 2 + 3.67 = 5.67$$

$$d(A5, M3) = |7 - 7| + |5 - 4.33| = 0 + 0.67 = 0.67$$

$$d(A6, M3) = |6 - 7| + |14 - 4.33| = 1.33$$

0,5 Les clusters sont:  $C1 = \{A1, A2\}$ ;  $C2 = \{A4\}$ ;  $C3 = \{A3, A5, A6\}$   
 arrêt de l'algorithme.

Exercice N°3:

1) Avantages de K-means

- complexité de l'algorithme très basse.
- simple à mettre en oeuvre

Inconvénients de K-means

- Valeur K introduite par l'utilisateur
- clustering dur. un objet appartient à un et un seul cluster à la fois.



- (1) 2) L'extraction des itemsets maximale est intéressante dans un contexte fortement corrélié et la taille des itemsets fréquents sont très long
- 3) Les conditions à satisfaire pour considérer qu'une règle d'association est valide:
- Les itemsets formant la règle sont fréquents ( $\text{sup}(x) \geq \text{min\_sup}$ )
  - La confiance de la règle est  $\geq \text{min\_confiance}$
  - La intersection entre l'antécédant et le conséquent de la règle est vide ( $X \rightarrow Y$  avec  $X \cap Y = \emptyset$ )
- 4) La classification non supervisée

(10)

# Exercise N°3

$$\pi_1 = (4, 6.33)$$

$$\pi_2 = (6, 6.5)$$

$$\pi_3 = (5, 4) \quad (2.5)$$

$$d(A_1, \pi_1) = 5.67$$

$$d(A_1, \pi_2) = 7.5$$

$$d(A_1, \pi_3) = 10$$

$$d(A_2, \pi_1) = 3.33$$

$$d(A_2, \pi_2) = 5.5$$

$$d(A_2, \pi_3) = 5$$

$$d(A_3, \pi_1) = 6.33$$

$$d(A_3, \pi_2) = 4.5$$

$$d(A_3, \pi_3) = 2 \quad (2.5)$$

$$d(A_4, \pi_1) = 2.67$$

$$d(A_4, \pi_2) = 2.5$$

$$d(A_4, \pi_3) = 5$$

$$d(A_5, \pi_1) = 4.33$$

$$d(A_5, \pi_2) = 2.5$$

$$d(A_5, \pi_3) = 2$$

$$d(A_6, \pi_1) = 4.33$$

$$d(A_6, \pi_2) = 2.5$$

$$d(A_6, \pi_3) = 0$$

$$c_1(A_1, A_2); c_2(A_4)$$

$$c_3 = (A_3, A_5, A_6) \quad (2.5)$$



$$M1 = (2, 7.5) \quad M2 = (5, 8) \quad M3 = (7, 4.33) \quad (0.5)$$

$$d(A1, M1) = 2.5$$

$$d(A1, M2) = 5$$

$$d(A1, M3) = 10.67$$

$$d(A2, M1) = 2.5$$

$$d(A2, M2) = 6$$

$$d(A2, M3) = 5.67$$

$$d(A3, M1) = 9.5$$

$$d(A3, M2) = 7$$

$$d(A3, M3) = 1.33$$

$$d(A4, M1) = 3.5$$

$$d(A4, M2) = 0$$

$$d(A4, M3) = 5.67$$

$$d(A5, M1) = 7.5$$

$$d(A5, M2) = 5$$

$$d(A5, M3) = 0.67$$

$$d(A6, M1) = 7.5$$

$$d(A6, M2) = 5$$

$$d(A6, M3) = 1.33$$

les clusters sont:

$$C1 = (A1, A2), C2 = (A4) \quad C3 = (A3, A5, A6) \quad (0.5)$$