

**Découverte de  
Connaissances  
Examen du 13/11/2017**

*Durée : 60 minutes.*

*Documents autorisés. Les questions faisant apparaître le symbole ♣ peuvent présenter zéro, une ou plusieurs bonnes réponses. Les autres ont une unique bonne réponse.  
Des points négatifs pourront être affectés à de très mauvaises réponses.*

Pour le barème :

- motifs : environ 55%
- clustering : environ 45%

Prix	10	20	30	40	50	60
<i>Id</i>	A	B	C	D	E	F
<i>t</i> <sub>1</sub>	X	X	X	X		X
<i>t</i> <sub>2</sub>	X	X	X		X	
<i>t</i> <sub>3</sub>	X	X		X	X	
<i>t</i> <sub>4</sub>	X		X	X	X	
<i>t</i> <sub>5</sub>	X		X	X	X	X

Table 1: Une base de données de transactions

**Question 1 [ind]** Soit la base de transactions de la table ci-dessus. Quel est le nombre maximal d'ensembles fréquents que l'on peut extraire dans cette base ?

- $2^6$      
  16     
   $2^7 - 1$      
   $\frac{7 \times 6}{2}$      
   $6^2$

**Question 2 [passes]** Quel est le nombre maximal de passages sur les données (Tab. ??) réalisé par APriori ?

- 6     
  7     
   $2^6 - 1$      
  15     
  5

**Question 3 [apriori-freq]** Combien de motifs vérifiant la contrainte ( $support(X) \geq 3$ ) sont extraits avec Apriori ? (on ne compte pas l'ensemble vide)

- 15     
  21     
  18     
  17     
  12

**Question 4 [apriori-candidat]** Combien de motifs candidats sont générés avec Apriori lors de l'extraction des motifs  $X$  vérifiant la contrainte ( $support(X) \geq 3$ ) ? (on ne compte pas l'ensemble vide)

- 20     
  22     
   $2^6 - 1$      
  23     
  17

**Question 5 [apriori-closed]** Combien de motifs de support supérieur ou égal à 3 sont fermés?

- 8     
  12     
  16     
  14

**Question 6 [apriori-libre]** Soit  $I$  l'ensemble des items. On dit qu'un itemset  $X \subseteq I$  (noté  $free(X)$ ) est libre si  $\nexists Y \subset X$  tel que  $support(X) = support(Y)$ . On a également la propriété suivante:

$$\forall Y \subset X, free(X) \Rightarrow free(Y).$$

Exploiter cette propriété afin d'extraire les motifs libres dont le support est supérieur ou égal à 2 sans effectuer un post-traitement des motifs fréquents. Quel est le nombre de candidats considérés ?

- 20       17       22       25       21

**Question 7 [libre-motif] ♣** Parmi les motifs suivants, indiquer ce qui sont libres (avec support supérieur ou égal à 2).

- {}       A       CE       DF       CF       C

**Question 8 [apriori-librebis]** Combien de motifs libres sont découverts?

- 12       14       6       8       15

**Question 9 [init]** Soit l'ensemble de points  $\mathcal{P} = \{A : (1, 5), B : (7, 3), C : (6, 4), D : (3, 6), E : (2, 7), F : (5, 7), G : (7, 6), H : (5, 1), I : (9, 1), J : (2, 9), K : (9, 5)\}$ . On souhaite partitionner l'ensemble de ces points en deux groupes en utilisant la méthode des k-moyennes ( $k$ -means) avec  $E$  et  $F$  comme centroïdes de départ. Nous considérerons la distance euclidienne comme distance. Combien de points ont été affectés à la partition de centre  $F$  ( $F$  compris) à la fin de la première étape ?

- 7       6       2       3       4       5

**Question 10 [final] ♣** Au final, quels sont les points qui sont dans la même partition que B ?

- C       K       H       G A       I E       K J  
 G       I       F D       H B       J F

**Question 11 [nb-iteration]** Au bout de combien d'étapes, l'algorithme retourne le résultat ?

- 3       1       2       6       4       5

**Question 12 [ouverte]** Quelles sont les partitions obtenues et les centres respectifs à la fin de l'exécution de l'algorithme?

**Question 13 [hierar-clust]** On souhaite faire un clustering hiérarchique jusqu'à obtenir 2 classes. Comme distance entre clusters, nous considérons la distance de Manhattan couplée à la complete link method. Combien d'itérations a-t-on pour obtenir ces 2 classes ?

- 9       7       6       8       4       5

Feuille de réponses :

Nom et prénom : .....
--------------------------

*Les réponses aux questions sont à donner exclusivement sur cette feuille : les réponses données sur les feuilles précédentes ne seront pas prises en compte.*

QUESTION 1 :  B  C  D  E

QUESTION 2 :  B  C  D  E

QUESTION 3 :  B  C  D  E

QUESTION 4 :  B  C  D  E

QUESTION 5 :  B  C  D

QUESTION 6 :  B  C  D  E

QUESTION 7 :  B  D  E

QUESTION 8 :  B  C  D  E

QUESTION 9 :  B  C  D  E  F

QUESTION 10 :      F  G  H  I  J  K

QUESTION 11 :  B  C  D  E  F

QUESTION 12 :  F  P

..... ..... ..... ..... .....
---

QUESTION 13 :  B  C  D  E  F