

Université de Mascara

Faculté des Sciences Exactes

Département d'Informatique



DATAMINING

Master 1 : SITW

1h00

Exercice 1 (10 points) :

Soit l'ensemble D des entiers suivants :

$D = \{ 2, 5, 8, 10, 11, 18, 20 \}$

On veut répartir les données de D en trois (3) clusters, en utilisant l'algorithme Kmeans. La distance d entre deux nombres a et b est calculée ainsi : $d(a, b) = |a - b|$ (la valeur absolue de a moins b)

1/ Appliquez Kmeans en choisissant comme centres initiaux des 3 clusters respectivement : 8, 10 et 11. Montrez toutes les étapes de calcul.

2/ Donnez le résultat final et précisez le nombre d'itérations qui ont été nécessaires.

3/ Peut-on avoir un nombre d'itérations inférieur pour ce problème ? Discutez.

Exercice 02 (10 points)

Soit la base de transaction suivante :

TID	Items
1	I1, I3, I4
2	I2, I3, I5
3	I1, I2, I3, I5
4	I2, I5

1 : Appliquer l'algorithme Apriori pour extraire les sous-ensembles d'items fréquents avec $minSupp = 1$, $minSupp = 2$, et $minSupp = 3$