

Examen semestriel

Modules "Fouille et extraction de données" & "Datamining"

Durée : 01H30

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 :

$$d(a, b) = |a - b| \quad (\text{la valeur absolue de a moins b})$$

Travail à faire :

- 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 2 (10 points) :

Le tableau suivant contient des données sur les résultats obtenus par des étudiants de Tronc Commun (première année à l'Université). Chaque étudiant est décrit par 3 attributs : Est-il doublant ou non, la série du Baccalauréat obtenu et la mention. Les étudiants sont répartis en deux classes : Admis et Non Admis.

On veut construire un arbre de décision à partir des données du tableau, pour rendre compte des éléments qui influent sur les résultats des étudiants en Tronc Commun. Les lignes de 1 à 12 sont utilisées comme données d'apprentissage. Les lignes restantes (de 13 à 16) sont utilisées comme données de tests.

	Doublant	Série	Mention	Classe
1	Non	Maths	ABien	Admis
2	Non	Techniques	ABien	Admis
3	Oui	Sciences	ABien	Non Admis
4	Oui	Sciences	Bien	Admis
5	Non	Maths	Bien	Admis
6	Non	Techniques	Bien	Admis
7	Oui	Sciences	Passable	Non Admis
8	Oui	Maths	Passable	Non Admis
9	Oui	Techniques	Passable	Non Admis
10	Oui	Maths	TBien	Admis
11	Oui	Techniques	TBien	Admis
12	Non	Sciences	TBien	Admis
13	Oui	Maths	Bien	Admis
14	Non	Sciences	ABien	Non Admis
15	Non	Maths	TBien	Admis

16	Non	Maths	Passable	Non Admis
----	-----	-------	----------	-----------

Travail à faire :

1/ Utiliser les données des lignes de 1 à 12 pour construire l'arbre en utilisant l'algorithme ID3. Montrez toutes les étapes de calcul. Dessinez l'arbre final.

2/ Quels sont les résultats de test de l'arbre obtenu sur les données des lignes de 13 à 16 ?.

Dr Mourad LOUKAM