

Série de Td N°4

Data Mining – Classification Automatique – 2022/2023 – R.M : Ali LALOUCI

Exercice 1

- 1) A quoi consiste le classement ?
- 2) A quoi consiste le clustering ?
- 3) Donner trois méthodes courantes du classement.
- 4) Quelle est la différence entre la méthode Kppv et celle du centroïde ?
- 5) Donner les plus importantes approches du clustering.
- 6) Donnez les étapes de l'algorithme de K-means ?

Exercice 2

Soit les points A(1,1), B(2,1), C(1,2), D(4,3), E(3,4), F(4,4) :

1. Donnez la répartition géométrique de ces points. Quels sont les clusters qu'on peut identifier visuellement.
2. En prenant les centroides initiaux A et B, appliquer l'algorithme de K-means pour regrouper les points en deux clusters (utiliser la distance du Manhattan).

Exercice 3

On a l'ensemble des points suivants :

$A1=(3,10)$, $A2=(1,2)$, $A3=(9,5)$, $A4=(4,8)$, $A5=(8,6)$, $A6=(7,5)$, $A7=(2,3)$, $A8=(1,9)$

1. Par l'utilisation de la distance euclidienne, appliquer l'algorithme de K-means pour une seule itération avec les centres $A8$, $A6$, et $A4$.
2. Combien d'itérations sont-elles nécessaires pour arriver à une stabilisation ?
3. Donner le résultat final du Clustering K-means avec les centres finaux.
4. Faites une représentation graphique montrant les points d'étude, les clusters et les centres des clusters.