TD3
Apprentissage Non-Supervisé

Exercice 1

1. Dites sur quel principe se base un algorithme de classification automatique non-supervisée.

2. Quel est l'inconvénient de l'algorithme K-means ?

3. Quelle est la différence majeure entre l'algorithme K-means et le regroupement hiérarchique ascendant ?

4. Dans un tableau, comparer les classifieurs K-means et groupement hiérarchique en donnant leurs <u>avantages</u> et leurs <u>inconvénients</u> ?

5. Dans un tableau, donner 3 différences majeures entre les classifieurs K-nn et K-means.

Exercice 2

1. Soient 8 objets de la nature suivants, décrits chacun par 2 paramètres : $A_1 = (2, 10)$; $A_2 = (2, 5)$; $A_3 = (8, 4)$; $A_4 = (5, 8)$; $A_5 = (7, 5)$; $A_6 = (6, 4)$; $A_7 = (1, 2)$ et $A_8 = (4, 9)$.
2. Calculer la matrice des distances Euclidienne entre les différents objets mutuels.
3. Utilisez l'algorithme k-means et la distance Euclidienne pour regrouper en 3 groupes les 8 exemples. Au début les centres de chaque groupe sont $A1$, $A4$ et $A7$. Exécuter l'algorithme k-means pour un nombre suffisant d'itération.

A la fin de chaque période montrez :

- Les nouveaux groupes d'exemples.
- Les nouveaux centres.
- Dessinez les 8 points (exemples) dans un espace 10x10 puis montrer les groupes et leurs nouveaux centroîdes.
- Mettez à jour les distances des objets par rapport aux nouveaux centroîdes.
- Dites combien d'itérations sont nécessaires pour converger ?

Exercice 2 (suite)

4. Appliquez l'algorithme du regroupement hiérarchique pour réaliser la classification des 8 éléments et tracez à chaque étape le dendrogramme correspondant, selon les critères suivants :

   a) à *saut minimal*

   b) à *diamètre maximal*

   c) le critère de Ward (inertie)

   d) le Lien moyen (consiste à calculer la moyenne des distances entre les individus des groupes)

   Choisissez convenablement le critère d'arrêt, soit, par exemple, un seuil de 6 ou nombre d'itérations égal à 3, pour obtenir le meilleur clustering possible. Comparez.

Exercice 3

- Un algorithme de *regroupement hiérarchique* comprend essentiellement 4 procédures :
  - ✓ Initialisation.
  - ✓ Déterminer la distance minimale.
  - ✓ Comparaison des distances et regroupement des éléments ou Arrêt (avec seuil).
  - ✓ Mise à jour de la matrice des distances.
- En prenant comme éléments à classifier les *N* points $a_i = (x_i, y_i)$ d'un espace 2D et en utilisant la distance Euclidienne entre les points. Écrivez un algorithme qui permet de réaliser un *regroupement hiérarchique* avec *saut minimal*.

Exercice 4

- Soient 6 objets de la nature représentés par 2 valeurs (x_i, y_i) chacun : $A_1 = (2, 10)$; $A_2 = (2, 5)$; $A_3 = (8,\ 4)$; $A_4 = (5, 8)$; $A_5 = (7, 5)$ et $A_6 = (1, 2)$.

  1. Calculer la matrice des distances Manhattan entre les différents objets mutuels.

  2. Utilisez l'algorithme *k-means* et la distance Manhattan pour regrouper en 3 groupes les 6 exemples. Au début les centres de chaque groupe sont A_1, A_4 et A_6.

     - A chaque itération :

     – Préciser les nouveaux groupes et leurs nouveaux centres.

     – Dessinez dans un espace 10x10, les 6 points (exemples) puis montrer les groupes et leurs nouveaux centroîdes.

     – Mettez à jour les distances des objets par rapport aux nouveaux centroîdes.

     – Dites s'il y a convergence ou pas ?

Exercice 4 (suite)

3. Appliquez l'algorithme de *groupement hiérarchique* à critère *saut minimal*, selon les deux cas ci-dessous, pour déterminer le nombre de classes du problème, grouper les 6 éléments et reconnaitre leurs classes. Tracez à chaque étape le dendrogramme correspondant.

    – Nombre d'itérations = 3.

    – Seuil d'agrégation = 9,5.

4. Comparer les deux méthodes : *k-means* et *regroupement hiérarchique*. Que peut-on dire ?

Exercice 5

1. Décrivez de deux manières (probabiliste et géométrique) l'algorithme de clustering *Mean-Shift*.

2. Donner un exemple simple pour expliquer son fonctionnement.

3. Programmer l'algorithme *Mean-shift* sur un ensemble de *n* points (formant l'ensemble d'apprentissage) du plan 2D euclidien orthonormé (*0, i, j*).