DATA MINING

Test : SEGMENTATION

Exercice 1 (10 pts)

On désire appliquer la méthode CAH sur les données suivantes :

	X_1	X_2	X_3
I_1	10	5	14
I_2	8	16	12
I_3	14	5	12
I_4	8	10	10

1. Citer l'avantage de la classification hiérarchique ascendante par rapport à la méthode K-means
2. Déterminer, pour chaque phase de l'algorithme, la mise à jour des individus et la matrice des distances. NB : Utiliser la distance de Manhattan **d(I, J) = |X1(I) – X1(J)| + |X2(I) – X2(J)|**
3. Tracer le dendrogramme de la classification hiérarchique ascendante en graduant l'axe vertical
4. On suppose maintenant qu'on va travailler avec une autre métrique de distance en utilisant la formule **suivante d(I, J) = MAX (|X1(I) – X1(J)| , |X2(I) – X2(J)| , |X3(I) – X3(J)|)**
   Refaire le même travail : matrices de distance et dendrogramme obtenu.
5. Interpréter l'obtention des deux dendrogrammes.

Exercice 2 (10 pts)

Soit le tableau1 de six individus caractérisés par 3 variables. On souhaite construire deux groupes homogènes à partir de ces individus via la méthode K-means.

1. Décrire les étapes de l'algorithme de la méthode K-means.
2. Citer deux inconvénients de la méthode K-means.

On propose de commencer la construction à partir des deux groupes du tableau2.

3. Continuer la construction des groupes en utilisant la distance euclidienne pour mesurer la similarité entre individus.

Individus	V1	V2	V3
I1	4.5	2	1
I2	3.5	4	2.5
I3	5	7	3.5
I4	3	4	2
I5	1.5	2	1
I6	1	1	0.5

	Individus
Groupe 1	I3
Groupe 2	I6