

Série d'exercices: Classification automatique. Mme Hamdad

Exercice1: Soit le tableau de distance défini sur $I=\{a,b,c,d,e,f,g\}$ suivant:

	a	b	c	d	e	f	g
a	0						
b	1	0					
c	3	2	0				
d	6	5	3	0			
e	7	6	4	1	0		
f	11	10	8	5	4	0	
g	16	15	13	10	9	5	0

Soit l'indice d'agrégation défini par:

$$\Delta(A, B) = \sup\{d(i, j), i \in A, j \in B\}$$

- 1- En partant de la partition discrète et en utilisant l'indice Δ , construire les partitions de I : P_1, P_2, \dots
- 2- Construire l'arbre hiérarchique correspondant en précisant les nœuds et les indices d'agrégation.

Exercice2: Soit le tableau de données défini sur $I=\{W_1, W_2, W_3, W_4, W_5\}$ suivant:

	W_1	W_2	W_3	W_4	W_5
X^1	1	2	3	-2	-3
X^2	2	1	2	-2	-2

Soit l'indice d'agrégation défini par:

$$\Delta(A, B) = I(A) + I(B) + \frac{p_A p_B}{p_A + p_B} d^2(g_A, g_B).$$

- Construire une hiérarchie indicée à partir de cette indice.

. $I(A)$ est l'inertie de la partie A . p_A, p_B sont respectivement les poids des classes A et B .

Exercice3 Soient les 10 points à classer, du plan dont les coordonnées sont les colonnes du tableau:

X^1	1	2	3	4	5	6	7	8	9	10
X^2	2	5	2	4	7	7	3	4	1	

- 1- Donner les partitions stabilisées correspondant à:

- aux noyaux (1,2,9)
- aux noyaux (1,4,10)

- 2- Calculer l'inertie intra classes pour les deux partitions obtenues. Comparer.

Exercice4: Soit E un nuage de R^2 , on agrège les individus qui font le moins varier l'inertie intra classes. Dans ce cas on cherche un indice de dissimilarité égale à la perte d'inertie interclasse

résultant du regroupement, il est donné par

$$\Delta(A, B) = \frac{p_A p_B}{p_A + p_B} d^2(g_A, g_B).$$

Soit le tableau de données suivant:

	W_1	W_2	W_3	W_4	W_5
X^1	2	1	1	2	1
X^2	1	2	1	2	4

1- Construire le tableau de distance euclidienne

2- En partant de la partition discrète et en utilisant l'indice Δ , construire la hiérarchie H et construire l'arbre hiérarchique correspondant.

Exercice5: Voici un tableau de données sur 5 individus avec 2 variables quantitatives :

	X^1	X^2
W1	0	0
W2	3	3
W3	9	0
W4	3	6
W5	9	8

1- Donner le tableau de distance euclidienne.

2- en Utilisant comme indice d'agrégation le saut minimal, effectuer la classification ascendante hiérarchique et représenter le dendrogramme.

3- En déduire une classification à deux classes.

4- En utilisant comme centres initiaux les centres de gravités des deux classes obtenus. donner par la méthode des k means une partition en 2 classes, comparer à celle obtenue dans la question(3).

Exercice6: On peint les drapeaux de 4 pays avec trois pots de peinture, bleu, blanche et rouge. Dans le tableau T, on met 1 lorsque la couleur est utilisée pour colorier le drapeau du pays 0 sinon.

	Bleu	Blanc	Rouge
Luxembourg	1	1	1
Danemark	0	1	1
Suisse	0	1	1
Grèce	1	1	0

1- Donner le tableau de distance selon la distance suivante

$$d(p_j, p_k) = \sum_i (p_{ij} - p_{ik})^2$$

2- Classer ces pays selon la classification hiérarchique, en utilisant l'indice d'agrégation suivant:

$$\Delta(A, B) = \frac{\sum_{x \in A, y \in B} d(x, y)}{\text{card}A \times \text{card}B}.$$

Exercice7: Soit I un ensemble de 5 objets décrits par deux variables quantitatives X et Y: a(6,4), b(5,2), c(1,7), d(0,8), e(3,10).

Une mesure de dissimilarité est défini sur ces objets par

$$d(i, j) = \max(|x_i - x_j|, |y_i - y_j|), \text{ pour } i, j \in I.$$

Le tableau de distance euclidienne est donné par:

A9									0	17.02
----	--	--	--	--	--	--	--	--	---	-------

En utilisant l'indice du lien minimal suivant:

$$\Delta(h, h') = \min_{x \in h, y \in h'} d(x, y)$$

- 1- Effectuer la classification ascendante hiérarchique et représenter le dendrogramme.
- 2- En déduire une classification en trois classes.

Exercice11: Ci dessous un tableau de données de 6 individus selon 2 variables quantitatives:

	W1	W2	W3	W4	W5	W6
X ¹	-2	-2	-2	2	2	1
X ²	3	1	-1	-1	1	0

- 1- Donner le tableau de distance euclidienne d².
- 2- En utilisant comme centres initiaux les objets W1 et W6, donner la partition en deux classes, en utilisant k-means.
- 3- En utilisant l'indice d'agrégation suivant:

$$\Delta(h, h') = \frac{\sum_x \sum_y d^2(x, y)}{|h||h'|}$$

Donner une classification hiérarchique avec dendrogramme.

- 4- En déduire la partition en 2 classes et comparer à la partition obtenue avec k-means.

Exercice12:

On considère le tableau ci-après représentant 10 fleurs Iris caractérisées par deux variables quantitatives; longueur des pétales et largeur des pétales.

Individus	1	2	3	4	5	6	7	8	9	10
Longueur	1.4	1.5	1.7	1.6	1.1	7	6.4	6.9	5.5	6.5
Largeur	0.2	0.3	0.4	0.2	0.1	3.2	3.2	3.1	2.3	2.8

- 1- Donner une représentation graphique de ces 10 fleurs.
- 2- En utilisant comme centres initiaux les objets Iris1 et Iris6, donner la partition en deux classes, en utilisant k-means. Reprenez en considérant comme centres initiaux Iris1 et Iris4.
- 3- Calculer le pourcentage d'inertie interclasse relativement à l'inertie totale dans les deux cas. Que constatez vous?

On donne l'inertie totale égale à: 82,72 et g=(3.96,1.58)^t.