## Série d'exercices: Classification automatique. Mme Hamdad

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

	а	b	С	d	е	f	g
а	0						
b	1	0					
С	3	2	0				
d	6	5	3	0			
е	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  $\Delta$ , construire les partitions de l:  $P_1, P_2...$
- 2- Construire l'arbre hiérarchique correspondant en précisant les nœuds et les indices d'agrégation.

**Exercice2**: Soit le tableau de données définit sur I={ W<sub>1</sub>, W<sub>2</sub>, W<sub>3</sub>, W<sub>4</sub>, W<sub>5</sub>} suivant:

	$W_1$	W <sub>2</sub>	$W_3$	$W_4$	$W_5$
X <sup>1</sup>	1	2	3	-2	-3
X <sup>2</sup>	2	1	2	-2	-2

Soit l'indice d'agrégation définit 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 <sup>1</sup>	1	2	3	4	5	6	7	8	9	10
X <sup>2</sup>	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<sup>2</sup>, 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 <sub>2</sub>	$W_3$	W <sub>4</sub>	$W_5$
X <sup>1</sup>	2	1	1	2	1
X <sup>2</sup>	1	2	1	2	4

- 1- Construire le tableau de distance euclidienne
- 2- En partant de la partition discrète et en utilisant l'indice  $\Delta$ , 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 la tableau de distance selon la distance suivante

$$d(\mathbf{p}_{j},\mathbf{p}_{k}) = \sum_{i} (\mathbf{p}_{ij} - \mathbf{p}_{ik})^{2}$$

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

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

**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 définit sur ces objets par

$$d(i,j) = max(|x_i - x_i|, |y - y_i|)$$
, pour i,  $j \in I$ .

Calculer la matrice de dissimilarité associée à I.

- 2- Effectuer une classification ascendante hiérarchique sur I, en utilisant comme stratégie d'agrégation le saut minimum.
- 3- Représenter le dendogramme. Donner la partition en 3 classes.

**Exercice8 :** Soient E un ensemble des 8 objets suivants et le tableau de distance euclidienne associé :  $A1=(2,10), A_2=(2,5), A_3=(8,4), A_4=(5,8), A_5=(7,5), A_6=(6,4), A_7=(1,2), A_8=(4,9).$ 

$D^2$	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	25	36	13	50	52	65	5
A2		0	37	18	25	17	10	20
A3			0	25	2	2	53	41
A4				0	13	17	52	2
A5					0	2	45	25
A6						0	29	29
A7							0	58
A8								0

1- En utilisant l'indice d'agrégation suivant:

$$\Delta(h,h')=d^2(g_h,g_{h'}),$$

tels que  $g_h$ ,  $g_{h'}$  sont les centres de gravité des parties h et h' de E respectivement, effectuer la classification ascendante hiérarchique et représenter le dendrogramme.

2- En déduire une classification en trois classes et calculer son inertie inter classe.

## Exercice9:

Ci dessous un tableau de données de 5 individus selon 2 variables quantitatives:

	W1	W2	W3	W4	W5
$X^1$	0	3	9	3	9
X <sup>2</sup>	0	3	0	6	8

- 1- Donner le tableau de distance euclidienne.
- 2- En utilisant comme centres initiaux les objets W3 et W1, donner la partition en deux classes, en utilisant k-means.
- 3- Refaire la même chose avec W1 et W5. Obtient-on la même partition?

## Exercice10:

On voudrait classifier 10 voitures selon leurs vitesses et la distance nécessaire pour stopper:

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Speed	4	4	7	7	8	9	10	10	10	11
Dist	2	10	4	22	16	10	18	26	34	17

Le tableau de distance euclidienne est donné par:

D	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	0	8	3.6	20.22	14.56	9.43	17.08	24.73	32.55	16.55
A2		0	6.70	12.36	7.21	5	10	17.08	24.73	9.89
A3			0	18	12.04	6.32	14.31	22.20	30.14	13.60
A4				0	6.08	12.16	5	5	12.36	6.40
A5					0	6.08	2.82	10.19	18.11	3.16
A6						0	8.06	16.03	24.02	7.28
A7							0	8	16	1.41
A8								0	8	9.06

A9					0	17.02

En utilisant l' indice du lien minimal suivant:

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

- 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 <sup>1</sup>	-2	-2	-2	2	2	1
X <sup>2</sup>	3	1	-1	-1	1	0

- 1- Donner le tableau de distance euclidienne d<sup>2</sup>.
- 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:

$$\Lambda(h,h') = \frac{\sum_{x} \sum_{y} d^{2}(x,y)}{|h||h'|}$$

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  $g=(3.96,1.58)^{t}$ .