O.de Cambry **MA 412**

ESIEE **09/10**

Traitement statistique des données

TD n^0 2

Dans la suite on considérera les points suivants de $\mathcal{R}^2$:
$M_1(0;0)$; $M_2(1;1)$; $M_3(1;3)$; $M_4(2;3)$; $M_5(3;0)$; $M_6(5;1)$; $M_7(5;4)$

- A -

On utilise la distance euclidienne. Les distances entre les points sont données par le tableau suivant :

	M_1	M_2	M_3	M_4	M_5	M_6
M_2	1.41					
M_3	3.16	2.00				
M_4	3.61	2.24	1.00			
M_5	3.00	2.24	3.61	3.16		
M_6	5.10	4.00	4.47	3.61	2.24	
M_7	6.40	5.00	4.12	3.16	4.47	3.00

1. (a) *Déterminer un partitionnement en 2 classes par l'algorithme des centres mobiles en prenant comme centres initiaux :* $G^0 = \{M_1, M_2\}$.
   (b) *Déterminer la hiérarchie en utilisant la distance entre les centres de gravité et représenter le dendrogramme.*
2. *Déterminer la hiérachie du saut minimum de 2 manières différentes* :
   (a) *Par construction ascendante hiérarchique*
   (b) *Par l'algorithme de Prim*

- B -

Soient M et M' les points de $\mathcal{R}^p$ de coordonnées respectives $x = (x_1, \ldots, x_p)$ et $x' = (x'_1, \ldots, x'_p)$. On note $d(M, M')$ une distance entre ces 2 points.

1. Soient A et B, 2 groupements de points disjoints, de $\mathcal{R}^p$ et n_A et n_B leurs effectifs respectifs. On définit la distance entre ces 2 groupements par :
   $D(A, B) = \dfrac{1}{n_A\, n_B} \sum_{M \in A} \sum_{M' \in B} d(M, M')$
   (a) *Montrer que si C est un autre groupement de points disjoint de A et B d'effectif n_C, on a* : (**1**) $\qquad D(A \cup B, C) = \dfrac{n_A\, D(A, C) + n_B\, D(B, C)}{n_A + n_B}$

(b) *En déduire que D est réductible.Quelle conséquence peut-on en tirer pour la construction d'une hiérachie ?*

2. On veut construire une hiérarchie sur un ensemble $\mathcal{C} = (M_1, \ldots, M_n)$ de points de $\mathcal{R}^p$ par construction ascendante hiérarchique (C.A.H) en utilisant la distance D.

(a) On utilise à chaque étape de la C.A.H, l'égalité (**1**) pour calculer itérativement les distances entre classes nécessaires à l'algorithme.
Combien d'appels à l'égalité (**1**) *seront-ils nécessaires ?*

(b) On se place dans $\mathcal{R}^2$.
Soient $M(x, y)$ et $M'(x', y')$ 2 points de $\mathcal{R}^2$.
On utilise la distance : $d(M, M') = |x - x'| + |y - y'|$
Les distances entre les points $M_1, \ldots, M_7$ sont données par le tableau suivant :

M_2	2					
M_3	4	2				
M_4	5	3	1			
M_5	3	3	5	4		
M_6	6	4	6	5	3	
M_7	9	7	5	4	6	3
	M_1	M_2	M_3	M_4	M_5	M_6

Construire la hiérarchie en utilisant l'algorithme des plus proche voisins réciproques (P.P.V.R) et représenter le dendrogramme.