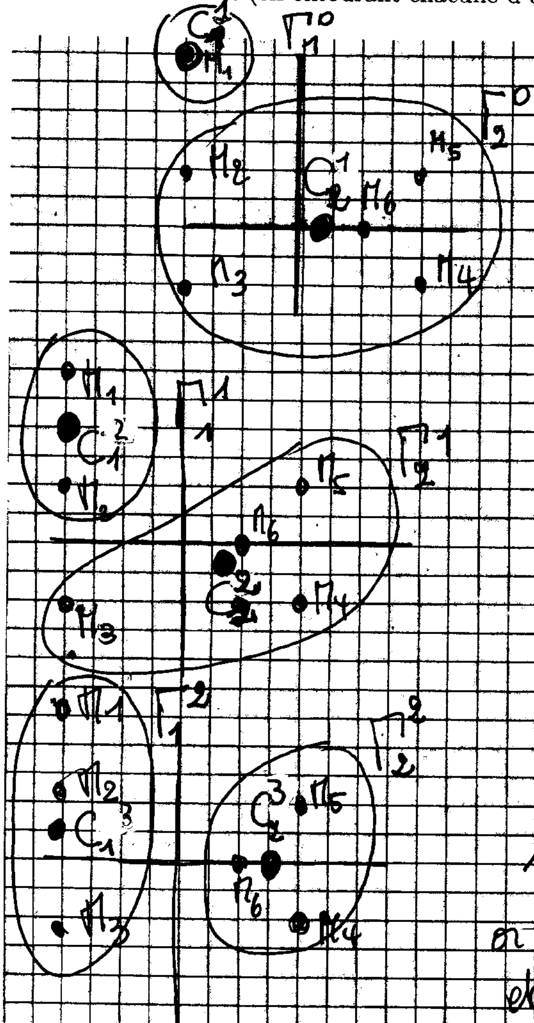
NOM : Corrigé
PRENOM :

Date :
Groupe :

Mathématiques pour la Biologie (semestre 2) : Feuille-réponses du TD 8
Classification par la méthode des centres mobiles

Exercice 1 : On considère les 6 points $M_1 = (-2,3)$, $M_2 = (-2,1)$, $M_3 = (-2,-1)$, $M_4 = (2,-1)$, $M_5 = (2,1)$ et $M_6 = (1,0)$. En supposant que les deux premiers points M_1 et M_2 sont les centres initiaux, décrire par une succession de dessins, les étapes de l'algorithme des centres mobiles en représentant les centres et les classes (en entourant chacune d'un rond) à chaque itération.



$C_1^0 = M_1$ $\quad C_2^0 = M_2$ centres initiaux

$\Gamma_1^0 = \{M_1\}$ $\quad \Gamma_2^0 = \{M_2, M_3, M_4, M_5, M_6\}$

$C_1^1 = \begin{pmatrix}-2\\3\end{pmatrix}$ $\quad C_2^1 = \begin{pmatrix}\frac{-2-2+2+2+1}{5}\\ \frac{1-1-1+1+0}{5}\end{pmatrix} = \begin{pmatrix}1/5\\0\end{pmatrix}$

On regroupe les points autour des nouveaux centres pour obtenir les nouvelles classes

$\Gamma_1^1 = \{M_1, M_2\}$ $\quad \Gamma_2^1 = \{M_3, M_4, M_5\}$

On calcule les nouveaux centres

$C_1^2 = \begin{pmatrix}-2\\2\end{pmatrix}$ $\quad C_2^2 = \begin{pmatrix}\frac{-2+2+2+1}{4}\\ \frac{-1-1+1+0}{4}\end{pmatrix} = \begin{pmatrix}3/4\\-1/4\end{pmatrix}$

Pour déterminer les nouvelles classes il faut savoir si $d(M_3, C_1^2)$ est plus grand ou plus petit que $d(M_3, C_2^2)$.

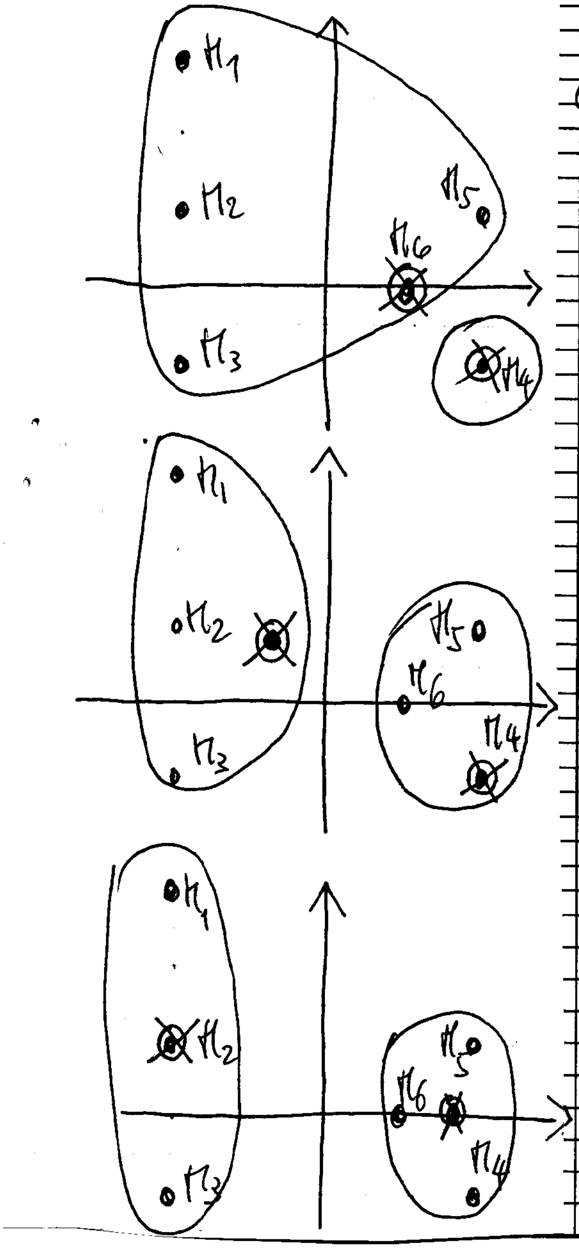
or $d(M_3, C_1^2) = \sqrt{(-2+2)^2 + (-1-2)^2} = \sqrt{9} = 3$

et $d(M_3, C_2^2) = \sqrt{(-2-\frac{3}{4})^2 + (-1+\frac{1}{4})^2} = \sqrt{8{,}125} < 3$

On calcule les nouveaux centres $C_1^3 = (-1, \frac{1}{2})$ et $C_2^3 = (\frac{3}{2}, 0)$ et on constate que les nouvelles classes sont les mêmes que les précédentes.

Donc l'algorithme n'évoluant plus, on le stoppe.

Exercice 2 : Recommencer sur cette page en choisissant différemment les centres initiaux. Obtient-on la même classification ?



On choisit les centres initiaux

$C_1^0 = M_6 = \begin{pmatrix}1\\0\end{pmatrix}$ $\quad C_2^0 = M_4 = \begin{pmatrix}2\\-1\end{pmatrix}$

On aglomère autour de ces centres

$\Gamma_1 = \{M_1, M_2, M_3, M_5, M_6\}$ $\quad \Gamma_2 = \{M_4\}$

On calcule les nouveaux centres

$C_1^1 = \begin{pmatrix}-3/5\\4/5\end{pmatrix}$ $\quad C_2^1 = \begin{pmatrix}2\\-1\end{pmatrix}$

On aglomère autour de ces centres

(On vérifie que M_6 est plus proche de C_2^1 que de C_1^1)

Si on recommence, la partition ne change plus ; on stoppe

Exercice 3 : Classifier les points du nuage précédent par une classification hiérarchique ascendante et représenter le dendrogramme (à noter que lorsqu'on doit regrouper les deux points les plus proches et qu'il existe deux couples de points satisfaisant cette condition, on convient de choisir les deux points dont les numéros sont les plus petits).

On calcule la matrice des écarts de Ward selon la formule

$$d(M_i, M_j) = \frac{p_i p_j}{p_i + p_j} d_2^2(M_i, M_j)$$

	M_1	M_2	M_3	M_4	M_5	M_6
M_1	0	2	8	16	10	9
M_2		0	2	16	8	9
M_3			0	8	10	5
M_4				0	2	1
M_5					0	1
M_6						0

plus petits écarts (on choisit le 1er)

A la 1ère étape, on agglomère M_4 et M_6 en 1 classe ayant pour centre de gravité $M_7 = \begin{pmatrix} 3/2 \\ -1/2 \end{pmatrix}$

A la 2e étape, on agglomère M_5 et M_7 en une classe dont le centre de gravité est $M_8 = \begin{pmatrix} 5/3 \\ 0 \end{pmatrix}$

	M_1	M_2	M_3	M_7	M_5
M_1	0	2	8	16,33	10
M_2		0	2	9,66	8
M_3			0	8,33	10
M_7				0	1,66
M_5					0

plus petit écart

A la 3e étape, on agglomère M_1 et M_2 en une classe de centre de gravité $M_9 = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$ puis M_3 avec M_8 en... $M_{10} = \begin{pmatrix} 3/4 \\ -1/4 \end{pmatrix}$ puis tous les points.

	M_1	M_2	M_3	M_8
M_1	0	2	8	16,8
M_2		0	2	10,8
M_3			0	10,8
M_8				0

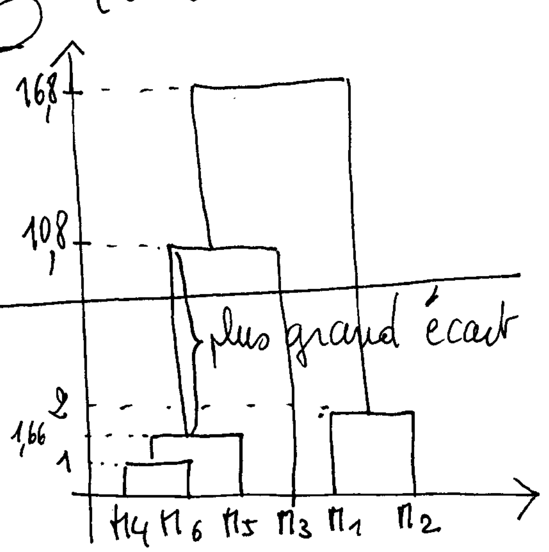
	M_9	M_3	M_8
M_9	0	16,66	20,98
M_3		0	10,8
M_8			0

plus petits écarts (on choisit le 1er)

plus petit écart

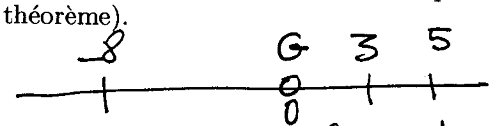
	M_9	M_{10}
M_9	0	16,8
M_{10}		0

la coupure du dendrogramme au "plus grand écart" conduit à 3 classes $\{M_4, M_6, M_5\}$, $\{M_3\}$ et $\{M_1, M_2\}$.



Exercice 4 (pour amateur de mathématiques...) : En choisissant un nuage de trois points alignés sur l'axe des x regroupés en deux classes, calculer l'inertie totale, l'inertie intraclasse et l'inertie interclasse. Vérifier le théorème de Huygens dans cet exemple.

En considérant cette fois trois points du plan non nécessairement alignés, montrer le théorème de Huygens (on pourra utiliser le fait que leurs projections sur les deux axes de coordonnées vérifient le théorème).



On considère les 3 points de coordonnées -8, 3 et 5 donc le centre de gravité est $G = 0$.

L'inertie totale est égale à $\frac{(-8)^2+(3)^2+(5)^2}{3} = \frac{98}{3}$

L'inertie du nuage $\Gamma_1 = \{-8\}$ et celle du nuage $\Gamma_2 = \{3, 5\}$ valent respectivement

$J(\Gamma_1) = 0$ et $J(\Gamma_2) = \frac{1^2+1^2}{3} = \frac{2}{3}$ donc $J_{intra} = \frac{2}{3}$

L'inertie inter pour la partition $\Gamma = \Gamma_1 + \Gamma_2$ vaut

$J_{inter} = \frac{64}{3} + (16)\frac{2}{3} = \frac{96}{3}$

On a donc bien, dans cet exemple, $\frac{98}{3} = \frac{96}{3} + \frac{2}{3}$ donc le théorème de Huygens est vérifié.

Supposons le théorème satisfait lorsque les points considérés sont alignés sur une droite. Pour en déduire le théorème lorsque les points sont non nécessairement alignés, dans le plan, on raisonne sur chacune des 2 coordonnées. Pour les projections des points sur chacun des axes de coordonnées le théorème s'applique.

Or l'inertie d'un nuage se décompose en "inertie sur l'axe des x plus inertie sur l'axe des y" en vertu du théorème de Pythagore.