

Correction CAH

Préparé par Sarra Zouari

Exercice 2: Classification Ascendante Hiérarchique CAH

1. Donner le **principe** algorithmique de la méthode de segmentation CAH. [2pts]

- chercher pour chaque phase les deux individus les plus proches => les fusionner en un seul élément représentatif
- calculer de nouveau la matrice des distances
- jusqu'à obtenir une seule classe

2. On désire appliquer la méthode CAH sur les données suivantes : 5 individus caractérisés par deux variables X1 et X2. De terminer, pour chaque phase de l'algorithme, la **mise à jour des individus** et la **matrice des distances**. [2pts]

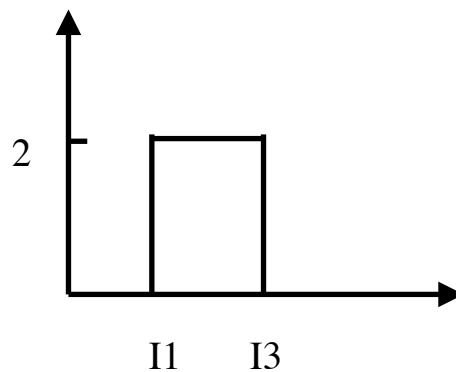
NB : Utiliser la distance de Manhattan $d(I, J) = |X1(I) - X1(J)| + |X2(I) - X2(J)|$

Mise à jour des Individus	Matrice des distances																																																						
Phase 1 : <table><tr><td></td><td>X₁</td><td>X₂</td></tr><tr><td>I₁</td><td>12</td><td>5</td></tr><tr><td>I₂</td><td>8</td><td>16</td></tr><tr><td>I₃</td><td>14</td><td>5</td></tr><tr><td>I₄</td><td>8</td><td>10</td></tr><tr><td>I₅</td><td>2</td><td>20</td></tr></table>		X ₁	X ₂	I ₁	12	5	I ₂	8	16	I ₃	14	5	I ₄	8	10	I ₅	2	20	<table><tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>1</td><td></td><td>15</td><td>2</td><td>9</td><td>25</td></tr><tr><td>2</td><td></td><td></td><td>17</td><td>6</td><td>10</td></tr><tr><td>3</td><td></td><td></td><td></td><td>11</td><td>27</td></tr><tr><td>4</td><td></td><td></td><td></td><td></td><td>16</td></tr><tr><td>5</td><td></td><td></td><td></td><td></td><td></td></tr></table>		1	2	3	4	5	1		15	2	9	25	2			17	6	10	3				11	27	4					16	5					
	X ₁	X ₂																																																					
I ₁	12	5																																																					
I ₂	8	16																																																					
I ₃	14	5																																																					
I ₄	8	10																																																					
I ₅	2	20																																																					
	1	2	3	4	5																																																		
1		15	2	9	25																																																		
2			17	6	10																																																		
3				11	27																																																		
4					16																																																		
5																																																							
Phase 2 : <table><tr><td></td><td>X₁</td><td>X₂</td></tr><tr><td>13</td><td>13</td><td>5</td></tr><tr><td>2</td><td>8</td><td>16</td></tr><tr><td>4</td><td>8</td><td>10</td></tr><tr><td>5</td><td>2</td><td>20</td></tr></table>		X ₁	X ₂	13	13	5	2	8	16	4	8	10	5	2	20	<table><tr><td></td><td>13</td><td>2</td><td>4</td><td>5</td></tr><tr><td>13</td><td></td><td>16</td><td>10</td><td>27</td></tr><tr><td>2</td><td></td><td></td><td>6</td><td>10</td></tr><tr><td>4</td><td></td><td></td><td></td><td>16</td></tr><tr><td>5</td><td></td><td></td><td></td><td></td></tr></table>		13	2	4	5	13		16	10	27	2			6	10	4				16	5																		
	X ₁	X ₂																																																					
13	13	5																																																					
2	8	16																																																					
4	8	10																																																					
5	2	20																																																					
	13	2	4	5																																																			
13		16	10	27																																																			
2			6	10																																																			
4				16																																																			
5																																																							
Phase 3 : <table><tr><td></td><td>X₁</td><td>X₂</td></tr><tr><td>13</td><td>13</td><td>5</td></tr><tr><td>24</td><td>8</td><td>13</td></tr><tr><td>5</td><td>2</td><td>20</td></tr></table>		X ₁	X ₂	13	13	5	24	8	13	5	2	20	<table><tr><td></td><td>13</td><td>24</td><td>5</td></tr><tr><td>13</td><td></td><td>13</td><td>26</td></tr><tr><td>24</td><td></td><td></td><td>13</td></tr><tr><td>5</td><td></td><td></td><td></td></tr></table>		13	24	5	13		13	26	24			13	5																													
	X ₁	X ₂																																																					
13	13	5																																																					
24	8	13																																																					
5	2	20																																																					
	13	24	5																																																				
13		13	26																																																				
24			13																																																				
5																																																							
Phase 4 :																																																							

Phase 1

Matrice des distances :

- Calculer les distances entre tous les points en utilisant la distance de Manhattan. (Exp : $d(I1, I2) = |12 - 8| + |5 - 16| = 4 + 11 = 15$)
- Choisir la valeur la plus petite (2) (la distance entre l'individu 3 et 1).
- Agrégation de 3 et 1 : on dessine un palier dans le dendogramme avec une hauteur de 2.



Phase 2

Mise à jour des individus :

- On calcule la moyenne de I1 et I3 : $(12 + 14) / 2, (5 + 5) / 2 = (13, 5)$

Matrice des distances :

- On calcule les autres distances entre tous les points en utilisant la distance de Manhattan.
- On choisit la valeur la plus petite (6) (la distance entre l'individu 2 et 4).
- Agrégation de 2 et 4 : on dessine un palier dans le dendogramme avec une hauteur de 6.

....

....

....