

TP: Non-supervisée

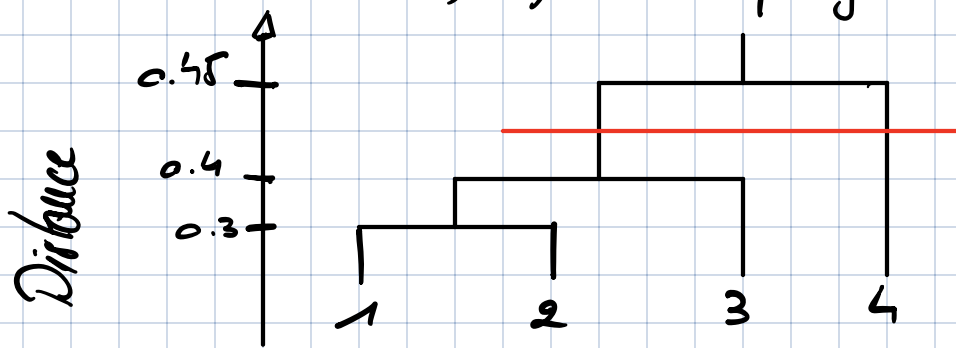
Exercice 1:

Matrice de distance.

	1	2	3	4
1	0	0.3	0.4	0.7
2	0.3	0	0.5	0.8
3	0.4	0.5	0	0.45
4	0.7	0.8	0.45	0

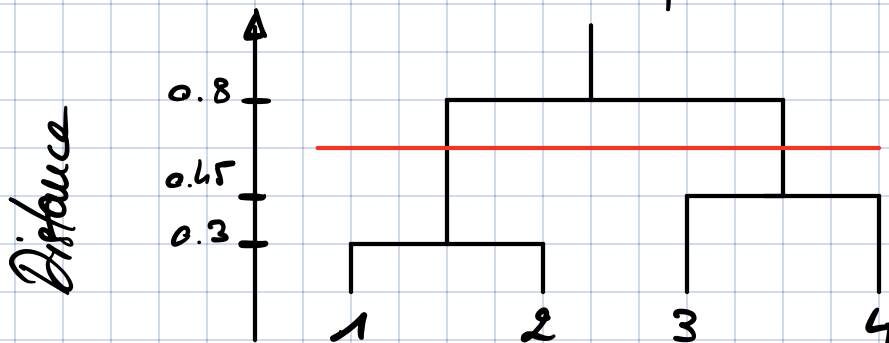
1. Méthode du plus proche voisin :

$$d(A, B) = \min \{ d_{ij} : i \in A, j \in B \}.$$



2. Méthode du voisin le plus distant :

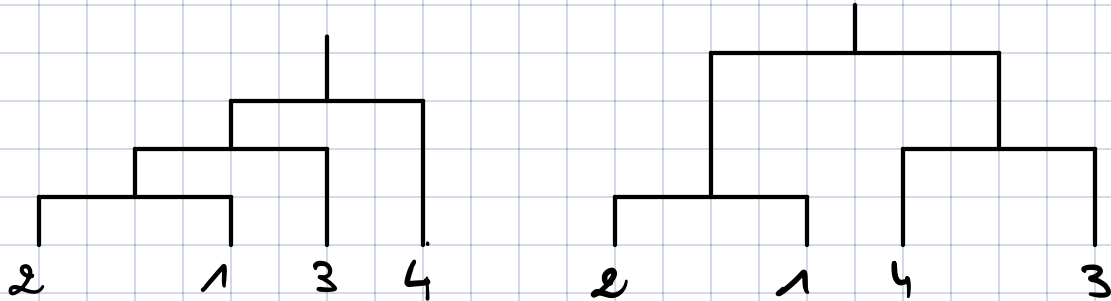
$$d(A, B) = \max \{ d_{ij} : i \in A, j \in B \}$$



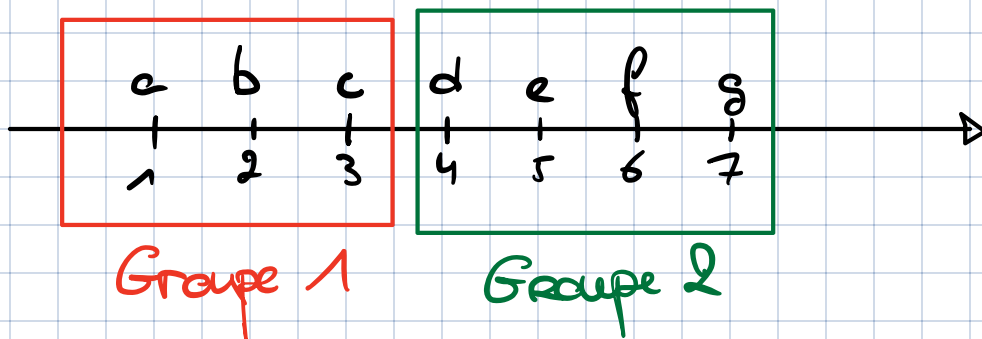
3. Si on coupe le premier arbre de sorte à avoir 2 groupes:
 $\{1, 2, 3\}$, $\{4\}$

Si on coupe le deuxième arbre de sorte à avoir 2 groupes:
 $\{1, 2\}$, $\{3, 4\}$.

4.



Exercice 2:



$$1. S(\{c\}) = \frac{b_c - a_c}{\max(a_c, b_c)}$$

avec a_c = distance moyenne entre $\{c\}$ et $\{a\}$ et $\{b\}$.

$$d(a, c) = 2, d(b, c) = 1 \rightarrow a_c = \frac{1}{2} (2+1) = \frac{3}{2}$$

b_c = distance moyenne entre $\{c\}$ et $\{d\}$, $\{e\}$, $\{f\}$ et $\{g\}$.

$$d(d, c) = 1, d(e, c) = 2, d(f, c) = 3 \text{ et } d(g, c) = 4.$$

$$\rightarrow b_c = \frac{1}{4} (1+2+3+4) = \frac{5}{2}$$

$$\text{donc } S(\{c\}) = \frac{\frac{5}{2} - \frac{3}{2}}{\frac{5}{2}} = \frac{2}{5} \rightarrow \text{ds bien classé.}$$

$$2. \text{Pseudo-}R^2 = \frac{I_{\text{inter}}}{I_{\text{tot}}}, \quad G = \{d\}.$$

$$\begin{aligned} I_{\text{tot}} &= \frac{1}{n} \sum_{i=1}^n d(X_i, G) \\ &= \frac{1}{7} (3 + 2 + 1 + 0 + 1 + 2 + 3) \\ &= \frac{12}{7} \end{aligned}$$

$$\begin{aligned} I_{\text{inter}} &= \frac{1}{n} \sum_{k=1}^K m_k d(G_k, G) \quad \text{avec } G_1 = \{b\} \\ &= \frac{1}{7} \left(3 \times 2 + 4 \times \frac{3}{2} \right) \quad G_2 = \{5.5\} \\ &= \frac{12}{7} \end{aligned}$$

$\text{Pseudo-}R^2 = 1 \rightarrow$ le clustering est bon.

Exercice 3 :

Obs	1	2	3	4	5
$x_{i,1}$	-1	-0.5	0	0.5	1
$x_{i,2}$	-1	0	0.5	-0.5	1

1. On assigne arbitrairement les obs. 1, 2 et 5 au groupe 1 et les observations 3 et 5 au groupe 2.

$$\text{Donc } C(1) = C(2) = C(5) = 1$$

$$C(3) = C(4) = 2$$

$$2. \mu_1 = \frac{1}{3} \left\{ \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -0.5 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} -1/6 \\ 0 \end{pmatrix}$$

$$\mu_2 = \frac{1}{2} \left\{ \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix} \right\} = \begin{pmatrix} 1/4 \\ 0 \end{pmatrix}$$

3.

i	1	2	3	4	5
$d(i, \mu_1)$	<u>1.69</u>	<u>0.11</u>	0.28	<u>0.69</u>	2.36
$d(i, \mu_2)$	2.56	0.56	<u>0.31</u>	0.31	<u>1.87</u>

$$\begin{aligned}
 W(C) &= \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{j: C(j)=k} d(x_i, x_j) \\
 &= d(x_1, x_2) + d(x_1, x_4) + d(x_2, x_4) \\
 &\quad + d(x_3, x_4)
 \end{aligned}$$

4. On attribue \bar{l} à i au groupe le plus proche, donc si $d(i, \mu_1) < d(i, \mu_2)$, on attribue au groupe 1, et au groupe 2, sinon.

Groupe 1 devient: $\{1, 2, 3\}$

Groupe 2 devient: $\{4, 5\}$.

$$5. \mu_1 = \frac{1}{3} \left\{ \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -0.5 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.5 \end{pmatrix} \right\} = \begin{pmatrix} -1/2 \\ -1/6 \end{pmatrix}.$$

$$\mu_2 = \frac{1}{2} \left\{ \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix}.$$

6.

i	1	2	3	4	5
$d(i, \mu_1)$	<u>0.94</u>	<u>0.03</u>	<u>0.69</u>	1.11	3.61
$d(i, \mu_2)$	4.63	1.63	0.63	<u>0.63</u>	<u>0.63</u>

$$\begin{aligned}
 W(C) &= \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{j: C(j)=k} d(x_i, x_j) \\
 &= d(x_1, x_2) + d(x_1, x_3) + d(x_2, x_3) + d(x_4, x_5)
 \end{aligned}$$

Les groupes deviennent :

Groupe 1 \rightarrow $\{1, 2\}$

Groupe 2 \rightarrow $\{3, 4, 5\}$.

Si on refait les calculs, les obs. ne changent plus de groupe après cet étape.