



## **TP N°3 (CAH)**

### **Exercice :**

On considère la matrice de données X dans  $R^2$  définie par:

$\Omega$	$X_1$	$X_2$
$\omega_1$	2	2
$\omega_2$	7.5	4
$\omega_3$	3	3
$\omega_4$	0.5	5
$\omega_5$	6	4

On va regrouper les individus avec l'algorithme **CAH** et la méthode du voisin le plus éloigné munie de la distance euclidienne.

- D'abord, on met les données dans une matrice et on trace le nuage de points :

```
x = c(2, 7.5, 3, 0.5, 6, 2, 4, 3, 5, 4)
m = matrix(x, ncol = 2, nrow = 5)
plot(m)
```

- On calcule les distances euclidiennes :

```
dist(m)
```

- On met en œuvre l'algorithme CAH avec la méthode du voisin le plus éloigné (complete linkage)

```
hc = hclust(dist(m), "complete")
ag = agnes(dist(x), method = "average")
```

- On affiche les regroupements :

```
hc$merge
```

- On affiche les écarts de regroupements :

```
hc$height
```

Cela renvoie : 1:414214 1:500000 3:354102 7:071068, rejoignant ainsi la conclusion de l'exercice, à savoir :

- les éléments  $\{\omega_1\}$  et  $\{\omega_3\}$  ont été regroupés avec un écart de 1:41,
- les éléments  $\{\omega_2\}$  et  $\{\omega_5\}$  ont été regroupés avec un écart de 1:50,
- les éléments  $A = \{\omega_1; \omega_3\}$  et  $\{\omega_4\}$  ont été regroupés avec un écart de 3:35,
- les éléments  $C = \{\omega_4; A\}$  et  $B = \{\omega_2; \omega_5\}$  ont été regroupés avec un écart de 7:07.

- 
- On trace le dendrogramme :

**plot(hc, hang = -1)**

- On peut demander à quel groupe chaque individu appartient suivant la hauteur des sauts avec la commande `cutree`. Avec 2 groupes, on a :

**cutree(hc, k = 2)**

- On peut alors afficher clairement les groupes :

**rect.hclust(hc, 2)**