

## TP6 : Clustering hiérarchique ascendant

L'algorithme de segmentation hiérarchique ascendante est disponible au travers de la fonction *hclust()* de R. Elle s'applique à une matrice de distance. On peut facilement obtenir cette matrice pour un data frame à l'aide de la fonction *dist()* qui calcule la distance euclidienne entre chaque paire de données du data frame.

```
dist(irisN [,1:4])
```

Pour effectuer une classification hiérarchique ascendante sur les 4 premières colonnes, on appelle les deux fonctions et on mémorise le résultat :

```
iris.hclust <- hclust (dist (irisN [,1:4]))
```

La fonction *hclust()* a construit un dendrogramme. On peut le visualiser avec la commande :

```
plot (iris.hclust)
```

Pour obtenir 3 groupes à partir de la segmentation hiérarchique précédente, on utilise la commande :

```
cutree(iris.hclust, 3)
```

Si on souhaite découper en 3 groupes. Cette commande fournit en sortie un vecteur de  $N$  entiers : l'élément  $i$  indique le numéro de groupe de la donnée  $i$ . Pour visualiser le clustering obtenu, on peut utiliser la fonction *rect.hclust()*

```
plot (iris.hclust)
rect.hclust(iris.hclust,5)
rect.hclust(iris.hclust, 5, border = c ("blue", "green",
"red", "pink", "black"))
rect.hclust(iris.hclust, 5, which = c (2,5))
```

Testez la commande suivante et expliquez ce que contient l'objet *iris.hclust.5*

```
iris.hclust.5 <- rect.hclust (iris.hclust, 5)
```

On peut aussi afficher le dendrogramme et indiquer à la souris où (les segments verticaux) on veut couper. On utilise pour cela la fonction *identify* :

```
plot (iris.hclust)
identify (iris.hclust)
```

Vous pouvez récupérer la composition de chacun des groupes en affectant le résultat de la fonction `identify` à un objet :

```
plot (iris.hclust)
iris.hclust.id <- identify (iris.hclust)
iris.hclust.id
```