

Fouille de Données T.P. N° 2

Etude des algorithmes de clustering

15 janvier 2023

Etude pratique avec R

Le but de cette section est d'appliquer le classifieur bayésien sur le dataset IRIS :

1. Exploration de l'ensemble des données IRIS :

Caractéristiques :

```
dim(iris)
names(iris)
str(iris)
attributes(iris)
```

Contenu par ligne ou par colonne :

```
iris[1:5,]
head(iris)
tail(iris)
iris$Sepal.Length[1:10]
```

Statistiques :

```
summary(iris)
var(iris$Sepal.Length)
hist(iris$Sepal.Length)
```

```
quantile(iris$Sepal.Length)
quantile(iris$Sepal.Length, c(.1, .3, .65))
table(iris$Species)
pie(table(iris$Species))
```

Plot :

```
plot(density(iris$Sepal.Length))
barplot(table(iris$Species))
boxplot(Sepal.Length~Species, data=iris)
with(iris, plot(Sepal.Length, Sepal.Width,
col=Species, pch=as.numeric(Species)))

library(scatterplot3d)
scatterplot3d(iris$Petal.Width, iris$Sepal.Length, iris$Sepal.Width)

library(MASS)
parcoord(iris[1:4], col=iris$Species)
```

Exercice 1 : étude du clustering

1. Appliquer l'algorithme K-means avec k=7

```
iris2 <- iris
iris2$Species <- NULL
(kmeans.result <- kmeans(iris2, 7))
```

2. Relever les attributs du modèle K-Means et leurs valeurs. Que représente ces attributs.
3. Trouver la matrice de correspondance classe cluster.
4. représenter graphique les résultats :

```
plot(iris2[c("Sepal.Length", "Sepal.Width")], col = kmeans.result$cluster)
# plot cluster centers
points(kmeans.result$centers[,c("Sepal.Length", "Sepal.Width")],
col = 1:3, pch = 8, cex=2)
```

5. Appliquer la méthode classification hiérarchique ascendant sur 40 exemples :

```
idx <- sample(1:dim(iris)[1], 40)
irisSample <- iris[idx,]
irisSample$Species <- NULL
hc <- hclust(dist(irisSample), method="ave")
```

6. Représenter graphiquement es résultats avec une coupure permettant d'obtenir 3 clusters

```
plot(hc, labels=iris$Species[idx])  
# cut tree into 3 clusters  
rect.hclust(hc, k=3)  
groups <- cutree(hc, k=3)
```

Exercice 2 Refaire le même travail sur le dataset *PimaIndiansDiabetes* de la librairie *mlbench*