TP-Classification Ascendante Hiérarchique

2023-2024

Contents

L'objectif de ce TP est d'illustrer les notions abordées en classification hiérarchique. Les librairies R nécessaires pour ce TP :

```
library(mclust)
library(clusterSim)
library(factoextra)
library(FactoMineR)
library(ggplot2)
library(reshape2)
library(circlize)
library(viridis)
```

Clustering des données de vin par CAH

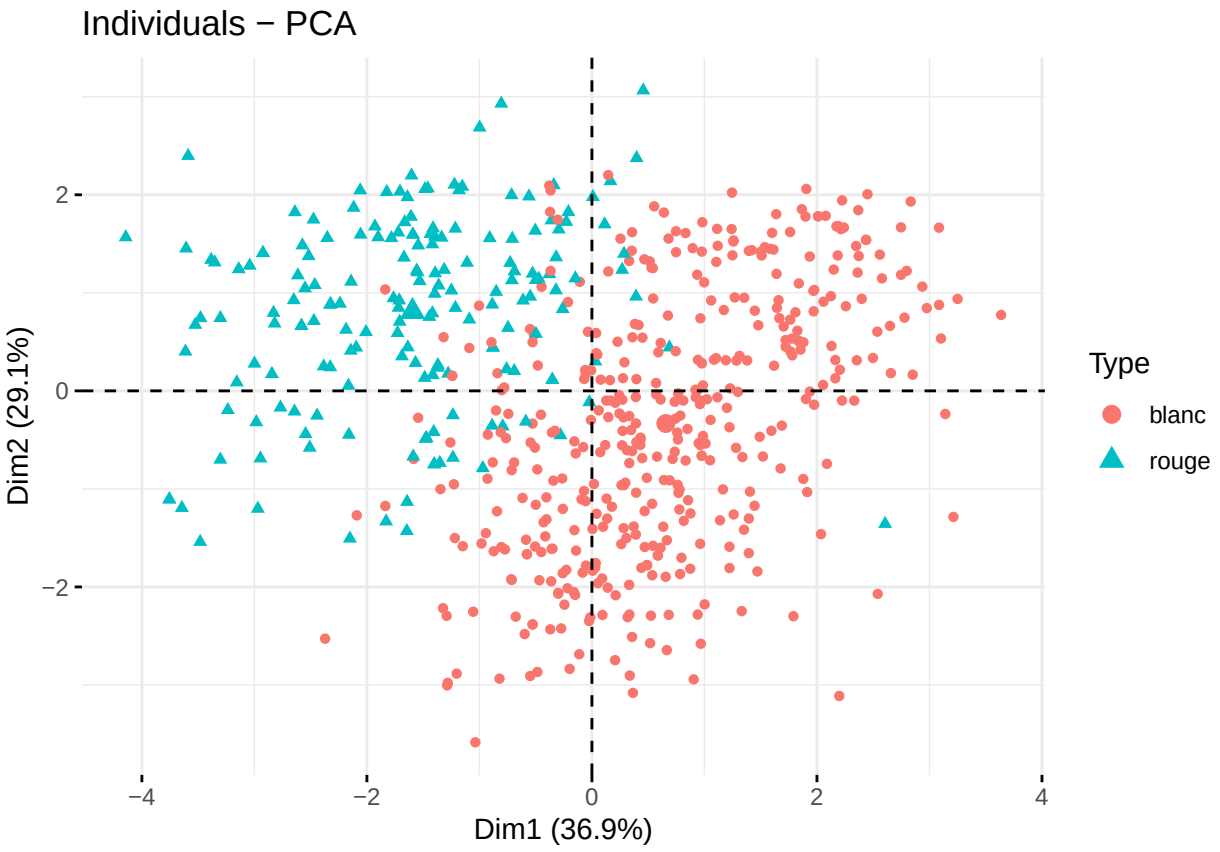
On reprend dans ce TP les données `wine` disponibles sur la page moodle du cours. On charge ici les données.

```
wine<-read.table("wine.txt",header=T)
wine$Qualite = as.factor(wine$Qualite)
wine$Type = factor(wine$Type, labels = c("blanc", "rouge"))
wineinit<-wine
wine[,-c(1,2)]<-scale(wine[,-c(1,2)],center=T,scale=T)
head(wine)
```

```
##      Qualite  Type      AcidVol    AcidCitr     SO2lbr      SO2tot     Densite
## 1352  medium rouge  1.638714588 -1.92626362 -1.2083376 -1.15967786 -0.46497450
## 5493  medium blanc -0.068544417 -1.35617574 -0.7004747 -0.85707581 -0.33499781
## 5153  medium blanc -0.800226847 -0.59605856  0.5409681 -0.02047014  1.32391517
## 5308  medium blanc -0.007570881  0.92417581  1.7824108  1.27893867  1.08790487
## 3866  medium blanc  0.419243870  0.03737243 -0.5311870  0.99413674  0.03783006
## 694   medium rouge  0.785085086  0.03737243 -0.4747578  0.19313131  1.27260858
##           Alcool
## 1352  1.14546909
## 5493 -1.12092616
## 5153 -1.29526426
## 5308 -1.29526426
## 3866  0.09944051
## 694  -0.94658806
```

On fait une ACP pour la visualisation des résultats dans la suite

```
resacp<-PCA(wine,quali.sup=c(1,2), scale.unit = TRUE,graph=FALSE)
fviz_pca_ind(resacp,geom=c("point"),habillage=2)
```



Question : A l'aide de la fonction `hclust`, faites une classification hiérarchique des données de vins avec les mesures d'agrégation `single`, `complete` et `average` respectivement. Comparez visuellement les dendrogrammes associés. Commentez.

```
# A COMPLETER
d<-dist(...,method="euclidean")
hclustsingle<-hclust(...)
hclustcomplete<-hclust(...)
hclustaverage<-hclust(...)

# Dendrogramme
plot(hclustsingle,hang=-1,labels=FALSE)
...

fviz_dend(hclustsingle,show_labels=FALSE)
...
```

Question : Déduisez du dendrogramme avec la mesure d'agrégation `complete` une classification en 3 classes. Vous pouvez utiliser la fonction `cutree()`. A l'aide d'une table de contingence et de l'`adjustedRandIndex` comparez-la avec les variables *Qualité* et *Type*. Commentez.

```
# A COMPLETER
ClassK3<-cutree(...,k=...)
...
...
```

Question : Dans cette question et pour les suivantes, on se focalise sur la mesure d'agrégation de Ward. Ajustez une classification hiérarchique avec la mesure de Ward (`ward.D2`). Que représentent les hauteurs du

dendrogramme dans ce cas ?

```
# A COMPLETER
hward<-hclust(...)
fviz_dend(...)
```

Question : Déterminez le nombre de classes à retenir avec l'indice de Calinski-Harabasz. Vous pouvez vous aider de la fonction `index.G1()` de la librairie `clusterSim`. Tracez la classification obtenue sur le dendrogramme et sur le premier plan factoriel de l'ACP.

```
# A completer
CH<-NULL
Kmax<-20
for (k in 2:Kmax){
  clusters=...
  CH<-c(CH,index.G1(...))
}
daux<-data.frame(NbClust=2:Kmax,CH=CH)
ggplot(daux,aes(x=NbClust,y=CH))+geom_line()+geom_point()

ClustCH<-cutree(...)
fviz_dend(...,show_labels=FALSE,k=...)
fviz_pca_ind(...,habillage=...)
```

Question : Déterminez le nombre de classes à retenir avec le critère Silhouette. Vous pouvez vous aider de la fonction `index.S()` de la librairie `clusterSim`. Comparez avec la classification de la question précédente.

```
# A COMPLETER
```

```
daux<-data.frame(NbClust=2:Kmax,Silhouette=...)
ggplot(daux,aes(x=NbClust,y=Silhouette))+geom_line()+geom_point()
```

Question : Comparez la classification obtenue avec la méthode des Kmeans dans le TP précédent et celle obtenue à la question précédente.

```
# A COMPLETER
reskmeans<-kmeans(wine[,-c(1,2)],4)
table(...,...)
adjustedRandIndex(...,...)

#library(circlize)
#library(viridis)
clust1F<-paste("ClKm-",reskmeans$cluster,sep="")
clust2F<-paste("ClCAH-",cutree(hward,4),sep="")

chordDiagram(table(...))
```