Machine Learning - 1100-ML0ENG (Ćwiczenia informatyczne Z-23/24)

Hierarchical clustering

Hierarchical methods allow the clustering algorithm to be observed, as they represent the clustering structure in the form of a tree (dendrogram).

```
df <- swiss
```

```
m.dist <- dist(df)
library(cluster)
library(factoextra)
#complete
```

```
tree.swiss<-hclust(m.dist, method="complete")
fviz_dend(tree.swiss, cex = 0.5 , main = "Swiss dataset tree - complete")
fviz_dend(tree.swiss, k=3, cex = 0.5 , main = "Swiss dataset tree - complete")
```

```
#fviz_silhouette(tree.swiss) error
gr.h.eclust<- eclust(swiss, "hclust", k = 3, stand = TRUE,hc_method="complete")
```

```
fviz_silhouette(gr.h.eclust)
```

```
sapply(2:10, function(x){gr.h.eclust<-eclust(swiss, "hclust", k = x, stand = TRUE,
hc_method="complete")
print(x)
print(gr.h.eclust$silinfo$avg.width)})
```

```

```

```
#ward.D2
```

```
tree.swiss2<-hclust(m.dist, method="ward.D2")
fviz_dend(tree.swiss2, cex = 0.5 , main = "Swiss dataset tree - ward")
fviz_dend(tree.swiss2, k=3, cex = 0.5 , main = "Swiss dataset tree - ward")
fviz_dend(tree.swiss2, k=4, cex = 0.5 , main = "Swiss dataset tree - ward")
```

```
sapply(2:10, function(x){gr.h.eclust<-eclust(swiss, "hclust", k = x, stand = TRUE,
hc_method="ward.D2")
print(x)
print(gr.h.eclust$silinfo$avg.width)})
```

```
gr.h.eclust2<- eclust(swiss, "hclust", k = 3, stand = TRUE, hc_method="ward.D2")
fviz_silhouette(gr.h.eclust2)
```

```
#single
```

```
tree.swiss3<-hclust(m.dist, method="single")
fviz_dend(tree.swiss3, cex = 0.5 , main = "Swiss dataset tree - single")
sapply(2:10, function(x){gr.h.eclust<-eclust(swiss, "hclust", k = x, stand = TRUE,
hc_method="single")
print(x)
print(gr.h.eclust$silinfo$avg.width)})
gr.h.eclust3<- eclust(swiss, "hclust", k = 4, stand = TRUE, hc_method="single")
fviz_silhouette(gr.h.eclust3)
```

```
#average
```

```
tree.swiss4<-hclust(m.dist, method="average")
fviz_dend(tree.swiss4, cex = 0.5 ,k=3, main = "Swiss dataset tree - average")
sapply(2:10, function(x){gr.h.eclust<-eclust(swiss, "hclust", k = x, stand = TRUE,
hc_method="average")
print(x)
print(gr.h.eclust$silinfo$avg.width)})
gr.h.eclust4<- eclust(swiss, "hclust", k = 5, stand = TRUE, hc_method="average")
fviz_silhouette(gr.h.eclust4)
```

Clustering

3 clusters are selected

```
clust.swiss<-cutree(tree.swiss2,3)
```

```
#description of the clusters
swiss4<-cbind(swiss,clust.swiss)
aggregate(swiss4, by=list(cluster = clust.swiss), mean)


library(dplyr)
swiss4%>%
  filter(clust.swiss==1)%>%
  summary()
```

```
fviz_cluster(list(data = df, cluster = clust.swiss),
 ellipse.type = "convex",
 repel = TRUE,
 main = "Clusters in Swiss4",
 show.clust.cent = FALSE, ggtheme = theme_minimal())
```

Visualizing Dendrograms

```
fviz_dend(tree.swiss2, cex = 0.5, horiz = TRUE, k=4, main = "Swiss tree - average")
```

```
fviz_dend(tree.swiss2, cex = 0.5, k = 3,
 type = "circular")
require("igraph")
fviz_dend(tree.swiss2, k = 3, k_colors = "jco",
 type = "phylogenic", repel = TRUE)
```

```
require("igraph")
fviz_dend(tree.swiss4, k = 3,
 k_colors = "jco",
 type = "phylogenic", repel = TRUE,
 phylo_layout = "layout.gem")
```

Heatmap

A **heatmap** is another way to visualize hierarchical clustering.
Heatmaps allow us to simultaneously visualize clusters of samples and features:

- First hierarchical clustering is done of both the rows and the columns of the data matrix.
- The columns/rows of the data matrix are re-ordered according to the hierarchical clustering result, putting similar observations close to each other. The blocks of high and low values are adjacent in the data matrix.
- Finally, a color scheme is applied for the visualization and the data matrix is displayed.

```
install.packeges("pheatmap")
library(pheatmap)
pheatmap(as.data.frame(swiss), clustering_method="average", cutree_rows = 3)
```

Last modified: czwartek, 30 listopada 2023, 11:15

Accessibility settings

Przetwarzanie danych osobowych

Platformą administruje Komisja ds. Doskonalenia Dydaktyki wraz z Centrum Informatyki Uniwersytetu Łódzkiego Więcej

Informacje na temat logowania

Na platformie jest wykorzystywana metoda logowania za pośrednictwem Centralnego Systemu Logowania.

Studentów i pracowników Uniwersytetu Łódzkiego obowiązuje nazwa użytkownika i hasło wykorzystywane podczas logowania się do systemu USOSweb.

Deklaracja dostępności