

D3: Local Session 3 - Solutions for Part 1

Description

This script contains the solutions for the local session 3.

Section 1: Exercises

1. You are able to apply the techniques of PCA and clustering (k-means, hierarchical, fuzzy c??means) to examine the structure of high-dimensional data

Exercise 1

```
#load data
spectra <- read.csv("spectra.csv", sep=";", header=T, stringsAsFactors = T)
fruits=c('Apple', 'Apricot', 'Banana', 'Carrot', 'Peach')
spectra1=spectra[spectra$bag=='none' & spectra$class %in% fruits,]

spec_dist_man=dist(spectra1[,-c(1,2)],method="manhattan")
mds = cmdscale (spec_dist_man , k=2)
plot(mds,col=spectra1$class)
legend('topleft', legend = unique(spectra1$class), text.col = as.numeric(unique(spectra1$class)))

#kmeans
k=5
cl=kmeans(spectra1[,-c(1,2)],k)
actual_class=as.matrix(spectra1$class)
predicted_cluster=cl$cluster
confusion_matrix=table(actual_class,predicted_cluster)
confusion_matrix

# hierarchical clustering
d= dist(spectra1[,-c(1,2)])
hc= hclust(d,method ="ward.D")
plot (hc , labels=spectra1$class,hang = -1)
actual_class=as.matrix(spectra1$class)
predicted_cluster=cutree(hc,k=5) #cut at 5 clusters
#you may again use a confusion matrix to assess the result
confusion_matrix=table(actual_class,predicted_cluster)
confusion_matrix
```

Exercise 2

```

library(e1071)
library(rgl)
cl_result = cmeans(iris[,1:2], 3, 100, m=2, method="cmeans")
plot(iris[,1], iris[,2], col=cl_result$cluster)
points(cl_result$centers[,c(1,2)], col=1:3, pch=8, cex=2)
points(iris[135,1:2], cex=2)
cl_result$membership[135,]
cl_result$membership[136,]

iris[135,5]
x=iris[,1]; y=iris[,2]; z=iris[,3];
plot3d(x,y,z,col=as.numeric(iris$Species))
points3d(x[135],y[135],z[135],col="blue",size=10)

```

Note that in the 2D projection the membership is wrong. This is an example where relevant information is lost in a projection.

Exercise 3

```

# run Program a first
k=7
cl_isis=cmeans(isis,k, iter.max=100, m=2, method="cmeans")
cl_isis
#comparison
isislist=isisall$class #only labels
actual_classes=as.matrix(as.numeric(isislist)) #actual labels as matrix
predicted_classes = cl_isis$cluster #predicted cluster labels
confusion_matrix = table(actual_classes, predicted_classes)
confusion_matrix

```
