Presenting: Group 3, Group 4, Date: 12.05.2021

Exercise 1:

Consider the following data generating process in which n observations belong to one of two classes. There are two covariates, drawn from the multivariate normal distribution $\mathbf{X} = (X_1, X_2)^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with class specific means. The class means are $\boldsymbol{\mu}_1 = (-3,\ 3)$ for class 1, and $\boldsymbol{\mu}_2 = (5,\ 5)$ for class 2 and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Initially, you may set $\boldsymbol{\Sigma} = \begin{pmatrix} 16 & -2 \\ -2 & 9 \end{pmatrix}$ and $n_1 = 300$ and $n_2 = 500$.

The goal of this exercise is to compare the performance of linear discriminant analysis and quadratic discriminant analysis when classifying observations.

a) Generate the covariates from a multivariate normal distribution using the $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}$ as described above and an indicator variable indicating class dependence for n observations and combine these in a data frame.

b) Calculate the linear discriminant analysis and quadratic discriminant analysis, estimating all relevant quantities.

c) Calculate the mean training error for both methods and compare.

Exercise 2 (Simulation Study)**:**

a) Evaluate the difference between the two methods through calculating classification training error in a simulation study for 100 different samples.

b) Consider the theoretical properties of lda and qda that we discussed in the lecture: Which properties of the initial simulation set up could we manipulate in order to increase the difference between the classification error of lda and qda? Test your intuition by performing a suitable simulation study.

c) Design (and perform!) a simulation study where you illustrate the problem with reducing the total error vs. sensitivity and specificity by manipulating the relevant quantities in the initial simulation set-up.

You do not have to program the `lda` and `qda` functions yourself (although you may, of course). Some helpful packages, libraries and commands:

```
library(mvtnorm) ###Random draws from a multivariate normal distribution

library(MASS) ###required to fit lda and qda commands below

lda(x, ...) ###performs lda, see help-file

qda(x,...) ##performs qda, see help-file

#### Both commands require the data to be saved as a data frame.
#### For converting matrices and vectors into
#### data frames, see the command below and the help file
data.frame(..., row.names = NULL, check.rows = FALSE,
           check.names = TRUE, fix.empty.names = TRUE,
           stringsAsFactors = default.stringsAsFactors())
```