

Exercise 1:

Consider the following data generating process in which n observations belong to one of two classes. There are two covariates, drawn from normal distribution $x_1 \sim \mathcal{N}(\cdot)$ and $x_2 \sim \mathcal{N}(\cdot)$ with class specific means. The class means are $\mu_1 = (-3 \ 3)$ for class 1, and $\mu_2 = (5 \ 5)$ for class 2 and $\Sigma_1 = \Sigma_2$. Initially, you may set $\Sigma = \begin{pmatrix} 16 & -2 \\ -2 & 9 \end{pmatrix}$ and $n_1 = 300$ and $n_2 = 500$.

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

- Generate the covariates from a multivariate normal distribution using the μ_k and Σ as described above and an indicator variable indicating class dependence for n observations and combine these in a data frame.
- Calculate the linear discriminant analysis and the logistic regression model, estimating all relevant quantities.
- Calculate the mean training and test error for both methods and compare.
- What other performance dimension could you analyze?

Exercise 2 (Simulation Study):

- Evaluate the difference between the two methods through calculating classification training and test error in a simulation study for 100 different samples. What other statistic could we consider?
- Consider the differences in the *objective* of both lda and the logit model. Test your intuition by performing a suitable simulation study and manipulating the dgp above to increase or decrease the advantage of the lda.
- Design 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 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())
```