Exercise 1

For the exercises in this section, we again use the `Default` data set.

You can reuse your R script from the last session and just append the following code to it. Particularly, we reuse our division in training- and test data:

```
##############   LDA - Linear Discriminant Analyses   ##########################

    # Fitting the model to the training data

      (lda.fit <- lda(default~balance,data=training.data))
      # Interpretation:
      # "Group means" ... class mean estimates
      # "Coefficients of linear discriminants" ... slope k of the discriminant function d(s)=kx+d
      plot(lda.fit) # histograms of the linear discriminants
      # "Discriminant" ...  k*x
      # Discriminants are used to build the decision rule for classification
      # (because the intercept does not depend on x):
          # kx small -> No
          # kx big   -> Yes
```

- Stop the video and try it for yourself.
- Try to interpret the group means and the slope of the discriminant function: What does it tell you about our data?
- Try to understand the outputs of the plots and interpret them.

Exercise 1

```
# Predicting test data

  lda.pred <- predict(lda.fit, test.data)
  # Interpretation:
  # class ... predicted class label (Yes of No)
      head(lda.pred$class)
  # posterior ... posterior probability to belong to a class
      head(lda.pred$posterior)
  # x ... linear discriminants
      head(lda.pred$x)

  # Plotting the predicted classes
  lda.class <- lda.pred$class
  lda.class.df <- data.frame(balance=test.data$balance,lda.class=lda.class) # make it a data frame for plotting
  (p1 <- ggplot() + geom_point(data = lda.class.df, aes(x=balance, y=lda.class, col=test.data$default), size=5))

  # Calculating the validation error rate (percentage of incorrectly classified samples) as an estimate of the test error rate
  mean(lda.class != test.data$default)

  # Confusion matrix
  table(test.data$default,lda.class)
```

- Inspect the result of the predict function.
- Interpret the plot. (A look into the test error and the confusion matrix will help you understand it.)
- Try to understand the code that is used to caluclate the validation (i.e., test) error rate?