

Assignment 9

Jenni Simon
09-116-005

Exercise 1 I performed logistic regression using all the predictors (no model-selection performed). The resulting model achieved a training accuracy of 84.8%. The associated confusion-matrix is shown in Table 1.

Tab. 1: Confusion matrix for Logistic Regression on the training data

	Predicted: Abnormal	Predicted: Normal
Actual: Abnormal	186	23
Actual: Normal	24	77

Exercise 2 I performed linear discriminant analysis using all the predictors (no model-selection performed). The resulting model achieved a training accuracy of 85.8%. The associated confusion-matrix is shown in Table 1.

Tab. 2: Confusion matrix for Linear Discriminant Analysis on the training data

	Predicted: Abnormal	Predicted: Normal
Actual: Abnormal	193	27
Actual: Normal	17	73

Exercise 3 To compare the models I performed 10-fold cross validation with identical folds and used the resulting per fold error-rates to perform a paired t-test. The results of the t-test are shown in Listing 1. No significant difference in the mean error-rate could be detected, suggesting that the models are comparable in performance. The mean error-rate of the logistic regression model was 15.5% and the mean error-rate achieved with LDA was 16.1%. I would therefor favour the logistic regression model due to the lower mean error-rate in cross-validation.

Listing 1: Summary of the logistic regression model.

```
Paired t-test

data:  glm.errors and lda.errors
t = -0.32733, df = 9, p-value = 0.7509
alternative hypothesis: difference in means is not equal to 0
95 percent confidence interval:
 -0.05103874  0.03813551
sample estimates:
mean of the differences
 -0.006451613
```