## Practical 3

Statistical Learning, 2011

Niels Richard Hansen September 21, 2011

## Logistic regression

Consider again the iris data and restrict attention to the two last species for this exercise, which you can find as iris[51:150, ].

1. Use glm in R to fit a logistic regression model to this data set – modeling species as a binary variable conditionally on the four other measurements. Predict the species of the new measurement

Sepal Length Sepal Width Petal Length Petal Width 5.9 2.7 4.6 1.3

Can you do any simplifications of the model?

- 2. Compute the LDA classifier using only the two species. Compute and compare the conditional probabilities of the species on the data set when we use logistic regression as above with the conditional probabilities found by LDA.
- 3. Recall the parameter transformation that gives the  $\beta$  parameters in the conditional distribution from the parameters in the LDA setup. Compute the plug-in estimate of the  $\beta$ 's from the LDA-method and compare with  $\hat{\beta}$  from logistic regression.

**Hint:** The group means can be extracted from the lda-object in R, or they can be computed from scratch. Subtracting these from the observations you can compute the common empirical covariance matrix and compute the value of the parameter transformation.