

# Practical 1

Statistical Learning, BGC - 2011

Niels Richard Hansen  
September 18, 2011

## Linear discriminant analysis

Load the `iris` data from the `datasets` package. Consult the R help files for information on the use of the function `lda` from the `MASS` library.

1. Use `lda` to compute the linear discriminant for the full dataset. Use both the formula method and the data matrix method (the “x”-method). Predict the species for a new observation with

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

2. Make a plot of the 2D-projection of the 4D dataset based on the linear discriminant. Use different symbols for the three different species and color code the points plotted according to whether they are correctly classified by the linear discriminator or not. Does the 2D-plot provide all the information needed for the linear discrimination?
3. Assume that we know that *virginica* is the most commonly occurring iris in a given part of the country with 75% of all the flowers being *virginica*, 20% being *setosa* and only 5% being *versicolor*. With these prior probabilities compute a new linear discriminant and make a corresponding 2D-plot as above. Has anything changed?
4. Is the assumption of multivariate normality reasonable? If not, what can we do?