

# M5MS10 Machine Learning

## Computer Lab 2

Dr Ben Calderhead

### 1 Laboratory Exercise

In this computer lab you will use R to perform supervised learning with a discriminative classification approach that uses a linear model for the discriminant function, as we saw in lecture 3. We note that when we use a 1st order polynomial to model the discriminant function, the resulting approach is known as *logistic regression*. You will investigate the effect that increasing model complexity has on the classification rate (defined as the percentage of correct predictions the algorithm makes using the testing data).

#### 1.1 Classification Errors

You can download the code and dataset for this laboratory from the Imperial Blackboard. The dataset consists of a 2 dimensional feature vector and a 1 dimensional class label vector.

- Compute the classification rate on the test set and associated test set predictive likelihood for a range of polynomial orders from 1 to 20. What happens to the log-likelihood of the training set and of the testing set? Which is a better measurement of performance: log-likelihood or classification rate? What are the differences, does one give us any different information from the other?
- Implement a leave-one-out cross validation approach using the average log-likelihood for determining the appropriate order of polynomial to use for making predictions.