

# M5MS10 Machine Learning 2017

## Assessed Coursework 1

**Deadline for submission: by 4pm Tuesday 14th February 2017.**

**Email a typed report (preferably in PDF format) of no more than 6 pages containing well-commented computer code for carrying out the tasks detailed below to: [b.calderhead@imperial.ac.uk](mailto:b.calderhead@imperial.ac.uk)**

### Assessed Coursework 1

In this coursework, you will investigate the use of dimensionality reduction via principal component analysis (PCA) within a classification algorithm for face recognition. A data set is provided containing training and test images of faces, each of which is 112 pixels tall and 92 pixels wide. By computing a PCA basis we can construct a lower dimensional representation of these images, which we can then use within a nearest neighbour classifier.

Remember to adequately comment your code! Marks will be deducted for unclear code.

Carry out the following tasks:

- Calculate and plot the average face of the training set, then write a function to find a PCA basis of size  $M$ , where the inputs will be  $M$  and  $X$ , the matrix containing the training set. Clearly describe all aspects of your function, then use it to plot the first 5 eigenfaces of the training set.
- Choose a single face and project it into a PCA basis for dimension  $M = 5, 10, 50$ , then plot the results.
- Plot a graph of the mean squared error of each lower dimensional approximation of this chosen face, with the dimensionality plotted along the x-axis. Is there a clear point at which we can choose a good approximation? Discuss how we should choose the appropriate dimensionality of the approximation.
- Write a function implementing a K-nearest neighbour classifier and investigate its use on the face recognition dataset. Make some recommendations regarding how to best set up this algorithm for this particular application.