(http://www.ox.ac.uk/)

FRANK WOOD

)X.AC.UK/~FWOOD/TEACHING/AIMS_CDT_ML/HOMEWORK/INDEX.HTML)

C.UK/~FWOOD/TEACHING/AIMS_CDT_ML/PEOPLE/INDEX.HTML)

JK/~FWOOD/TEACHING/AIMS_CDT_ML/INDEX.HTML)

ASSIGNMENT 2: EXPECTATION-MAXIMIZATION

In this homework you will need to implement the expectation and maximization steps of the EM algorithm for Bayesian linear regression and a classical Gaussian mixture model.

Background reading: Bishop chapters 3, 9

GETTING STARTED

- Download the skeleton code for the assignment (hw_2.tar.gz)
- Unzip the downloaded material in an appropriate folder, something like
  `~/Documents/AIMS_CDT_ML/HW2/`
- Open MATLAB and navigate to the folder containing the downloaded material

PART 1: BAYESIAN LINEAR REGRESSION

Implement the the functions `e_step_linear_regression` and `m_step_linear_regression` to implement the EM algorithm for Bayesian linear regression.
The model being fit here is as follows

$$\begin{aligned} t_n &\sim \text{Normal}(\mathbf{w}^T \phi(\mathbf{x}_n), 1/\beta) && \forall\, n \in \{1, \ldots, N\} \\ w_m &\sim \text{Normal}(0, 1/\alpha) && \forall\, m \in \{1, \ldots, \text{length}(\mathbf{x}_n)\}. \end{aligned}$$

The parameters of interest are α and β, and we wish to fit their values using maximum likelihood. This requires the EM algorithm because you will integrate over the values of the weights w.

You will need to consult the book (Bishop, sections 9.3 and 3.3) and the programs provided to understand the function signatures. Data is provided in the main file which you should use while developing your code. However, make sure the functions will run on any dimensional design matrix. We recommend you test your functions by testing them on synthetic data.

PART 2: GAUSSIAN MIXTURE MODEL

Implement the functions `e_step_gaussian_mixture`, `m_step_gaussian_mixture`, and `log_likelihood_gaussian_mixture` to implement the EM algorithm for a Gaussian mixture model. The model we are fitting here is defined Bishop section 9.2, and can be described as

$$p(z_n = k) = \pi_k$$
$$\mathbf{x}_n \Big| z_n = k \sim \text{Normal}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \right)$$

The data you are using is the Fisher iris data. Each row of data consists of four measurements made regarding an iris flower. You will need to cluster the measurements using EM on a Gaussian mixture model to esimate the parameters π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ for $k = 1, \ldots, K$. You should make sure that the program runs no matter what choice of K you make. Also, make sure the functions will run on any dimensional real vector valued data with any chosen number of components. We recommend you test your functions by testing them on synthetic data.