# 6CCE3MLE Machine Learning for Engineers(21~22 SEM2 000001)

## Prepared by (Prof Osvaldo Simeone, module leader)

*You will be required to undertake this alternative assessment in place of your deferred coursework component or as reassessment for your failed coursework attempt. Please note, this also includes in class tests which are scheduled within term outside of the examination window.*

*This alternative assessment for 6CCE3MLE (2021/22 academic year) is designed to replace the outstanding components below.*

*Coursework 1 - Least square*
*Coursework 2 – Logistic Regression and PCA*

**Alternative Assessment: Binary classification using Least Squares and Logistic Regression**

Data set: USPS (full data set containing all digits, images as 256-length vectors, available on keats)

1. [10 points] Pre-processing
    a. Focus only on the digits 3 and 8.
    b. Split the data set (inputs x as 256 dimensional and target t (0 for digit three and 1 for digit 8) into training and validation. Training will have the first 80% images for each label, and the validation data set will have the last 20% of each.
    c. Add a bias pixel to form the feature u(x)=[1,x], expanding the input into a vector of size 257.
2. [20 points] Least Squares
    a. For each 1<=M<=257, crop the first M features of the vector u(x). Solve the least squares theta_M for the cropped feature vector.
    b. Produce a graph: training and validation quadratic errors of the predictor (as two curves), as validated on the entire test sets, versus number of components M.
3. [30 points] PCA
    a. Calculate the PCA dictionary full 256x256 matrix acquired over the training data set (without the bias pixel). You can assume the data has no bias and needs no correction.
    b. Compress the training and validation data set independently, using the dictionary matrix for all (1<=M<=256) principled components.
    c. Solve a Least square problem to obtain the M-length model parameters theta, using as input the M-length representation vectors z from the training set, and their true labels.
    d. Produce a graph: training and validation quadratic errors of the PCA predictor (as two curves), versus number of components M.
    e. Discuss, using a printout code `disp('…')`, why this graph is different from the one in Section 2.
4. [40 point] Logistic Regression with one layer
    a. Train using 25 iterations of stochastic gradient descent a logistic regression (with one layer) with input size of 4, being the PCA representation (M=3) and a bias input (similar to Section 1c). Use the following parameters: initial model parameter vector `theta= 1/sqrt(M+1)*randn(M+1,1);` learning rate=0.1; and minibatch size of 10 (`randperm(n,k)` can be helpful).
    b. Produce a graph of logistic loss and detection-error loss vs iteration index.

**Submission requirements:**

*The submitted work should have a single m-file without any other files attached. The m-file should self include all code, it must run without raising any errors and must not use any toolboxes (verify using license('inuse')). You may use auxiliary functions, as you see fit, at the end of the file or as anonymous function (using the operator @) to support modularity. When the grader runs the m-file, all three figures must be shown, all with title, axis labels and legend. No printouts are allowed (use ';' at the end of assignments), apart from the requested discussion in Section 3.c . Submit the file into the proper keats item.*

Coursework is due **4pm, Thursday 4th August 2022 (BST)**. As clarified in the email sent to students in June (Engineering Assessments Period 3 – August information, 16/06/22), no extensions will be permitted. For those with approved PAA, please follow the separate email sent. Submissions will not be accepted after the deadline has passed, please ensure that you leave plenty of time to check over and upload your work. If you are unable to submit by this deadline, email engineering-ug@kcl.ac.uk and apply for mitigation as soon as possible.

Further information on the submission area and where to submit can be found via – Course: Engineering Coursework Submissions - August 2022 (kcl.ac.uk)

Support for carrying out this assessment is available via Course padlet website https://kings.padlet.org/osvaldosimeone/lljjvgxcoaa1ijy1