**Homework Set 5**

**Problem 1 (LASSO and Ridge regression)**

A dataset is provided in the file “05HW1\_diabetes.mat”. The file contains training data variables x\_train and y\_train and test data variables x\_test and y\_test. Download the LASSO regression package Glmnet for MATLAB at <http://web.stanford.edu/~hastie/glmnet_matlab/>.

A simple instruction for using the Glmnet is given below. Suppose we want to solve the following LASSO linear regression problem:

Type the following command in MATLAB:

>> fit = glmnet(x\_train, y\_train);

where x\_train is the training data attributes and y\_train is the training data label. The returned value fit is a structure containing the LASSO solver parameters. The model coefficients of the LASSO regression varies with the regularization parameter . To obtain the model coefficients for a given (my\_lambda), type:

>> beta = glmnetCoef( fit, my\_lambda);

To make a prediction on test data x\_test using the developed LASSO model and a given (my\_lambda), type:

>> pred = glmnetPredict( fit, x\_test, my\_lambda, 'link');

You are asked to develop LASSO models on the diabetes dataset using  = {0.01, 0.1, 1, 10, 100, 1000}.

1. Plot the weights at different λ. Discuss how the weights are changed by varying λ.
2. Plot test error at different λ. Discuss how the test error is changed by varying λ.
3. Repeat the same procedure by using the ridge regression (built-in function provided by MATLAB) with the same set of λ. Comparing the solution obtained by the ridge regression to that by LASSO.`

**Problem 2 (LASSO regression)**

This problem explores the prediction of wine quality using physicochemical factors. Two files “05HW2\_wine\_training.txt” and “05HW2\_wine\_test.txt” are the dataset of the Red Wine Quality Rating available at UCI (<https://archive.ics.uci.edu/ml/datasets/Wine+Quality>). Each row in the files are the 11 different factors (first 11 columns) and the wine quality (the 12th column) of a wine. Read “05HW2\_note.txt” for more details.

Develop LASSO models on “05HW2\_wine\_training.txt” using  = {0.0001, 0.0005, 0.0025, 0.0125, 0.0625, 0.3125, 1.5625, 7.815, 39.0625, 195.3125}. Finish the following tasks:

1. Plot the (square) training and test errors versus different .
2. Briefly explain the fitting behavior (i.e., over-fitting and under-fitting) of the models with different .
3. Which would you choose to train your final model? Why?