

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:

$$\boldsymbol{\beta} = \min_{\boldsymbol{\beta}} (\|\mathbf{y} - \mathbf{x}\boldsymbol{\beta}\|^2 + \lambda \|\boldsymbol{\beta}\|).$$

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 $\boldsymbol{\beta}$ of the LASSO regression varies with the regularization parameter λ . To obtain the model coefficients $\boldsymbol{\beta}$ for a given λ (`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 λ (`my_lambda`), type:

```
>> pred = glmnetPredict( fit, x_test, my_lambda, 'link');
```

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

- Plot the weights $\boldsymbol{\beta}$ at different λ . Discuss how the weights are changed by varying λ .
- Plot test error at different λ . Discuss how the test error is changed by varying λ .
- Repeat the same procedure by using the ridge regression (built-in function provided by MATLAB) with the same set of λ . Comparing the solution $\boldsymbol{\beta}$ 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 $\lambda = \{0.0001, 0.0005, 0.0025, 0.0125, 0.0625, 0.3125, 1.5625, 7.815, 39.0625, 195.3125\}$. Finish the following tasks:

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