Homework 1 (Question 2 and 3 Results)

Question 2:

2a) Below are two kinds of plots for each of the six datasets:

* The first plot is the training set Mean Squared Error for a given lambda ranging from 0 to 150 (in my code I labeled lambda as “delta”).
* The second plot is a combined graph of both the training set MSE and the test MSE for each of the original six datasets. Once again, the training set MSE and the test MSE were calculated for each lambda value ranging from 0 to 150.

A graph of a training set

Description automatically generated 100x10 **Train** MSE 100x10 **Train and Test** MSE

A graph of a training and testing test

Description automatically generatedA graph of a training set

Description automatically generated 100x100 **Train** MSE 100x100 **Train and Test** MSE

A graph with a line

Description automatically generatedA graph of a training set

Description automatically generated 1000x100 **Train** MSE 1000x100 **Train and Test** MSE

2a) For each dataset, the best λ value that gives the least test set MSE is (Note: In my project λ = Delta):

A table with numbers and a few different sets of data

Description automatically generated with medium confidence

2b) Below are the training and test set plots for 50(1000)-100, 100(1000)-100, and 150(1000)-100 datasets with λ ranging from 1 to 150 (these are the three additional datasets that were created from the 1000-100 train dataset.

A graph with numbers and a line

Description automatically generated with medium confidenceA graph with numbers and lines

Description automatically generated 50(1000)x100 **Train** MSE 50(1000)x100 **Train** **and Test** MSE

A graph with numbers and lines

Description automatically generatedA graph with a line

Description automatically generated 100(1000)x100 **Train** MSE 100(1000)x100 **Train** **and Test** MSE

A graph of a training and test

Description automatically generatedA graph with a line

Description automatically generated 150(1000)x100 **Train** MSE 150(1000)x100 **Train** **and Test** MSE

2c) λ = 0 gives abnormally large MSEs for the three training datasets in question 2b because the training models have no regularization (penalty) being applied at λ = 0. The linear regression models are solely focused on minimizing the mean squared error, but there is no restriction being placed on the kind of weights the models can use to fit to the y-values in the data. Thus, the models are at the greatest risk of overfitting to the training data. They are extremely biased at λ = 0. This is especially the case when considering how these specific models are fitting their regression line to 50 or 100 or 150 datapoints, but then are being directly tested on 1000 new data points.

Question 3:

3a) This is the best choice of λ and the corresponding test set MSE for each of the six datasets.

A table with numbers and letters

Description automatically generated

3b) The λ values were higher for the six datasets after performing Cross Validation, compared with the λ values in question 2a. Note how the mean of the lambda values after performing Cross Validation is 28.5 and the median is 27, whereas in question 2a the mean of the lambda values is 17.83 and the median is 20.5. Despite these differences in result for the λ values, the test Mean Squared Errors from Cross Validation did not very much differ from the test Mean Squared Errors in question 2a. The test MSE results were very similar, which shows a major advantage of using Cross Validation. CV was able to get us low and comparable MSE values without ever having to touch the Testing Sets during the process of finding the optimal λ values.

3c) A noticeable drawback of Cross Validation is that the algorithm to train the data takes more computation time and energy than the training algorithm without Cross Validation. To obtain the λ and corresponding MSE results for Cross Validation, we have to compute (and then re-compute in a loop) the training algorithm on k folds, where k is the number of disjoint folds we are splitting our data into. For this reason, Cross Validation requires many times more computation compared with not using Cross Validation (exactly k times more computation). This could be computationally costly when trying to make an evaluation for the optimal λ value to use.