**Report**

**CSL 407 Machine Learning Assignment 1**

**1.Question 1**

1. Removing least significant features

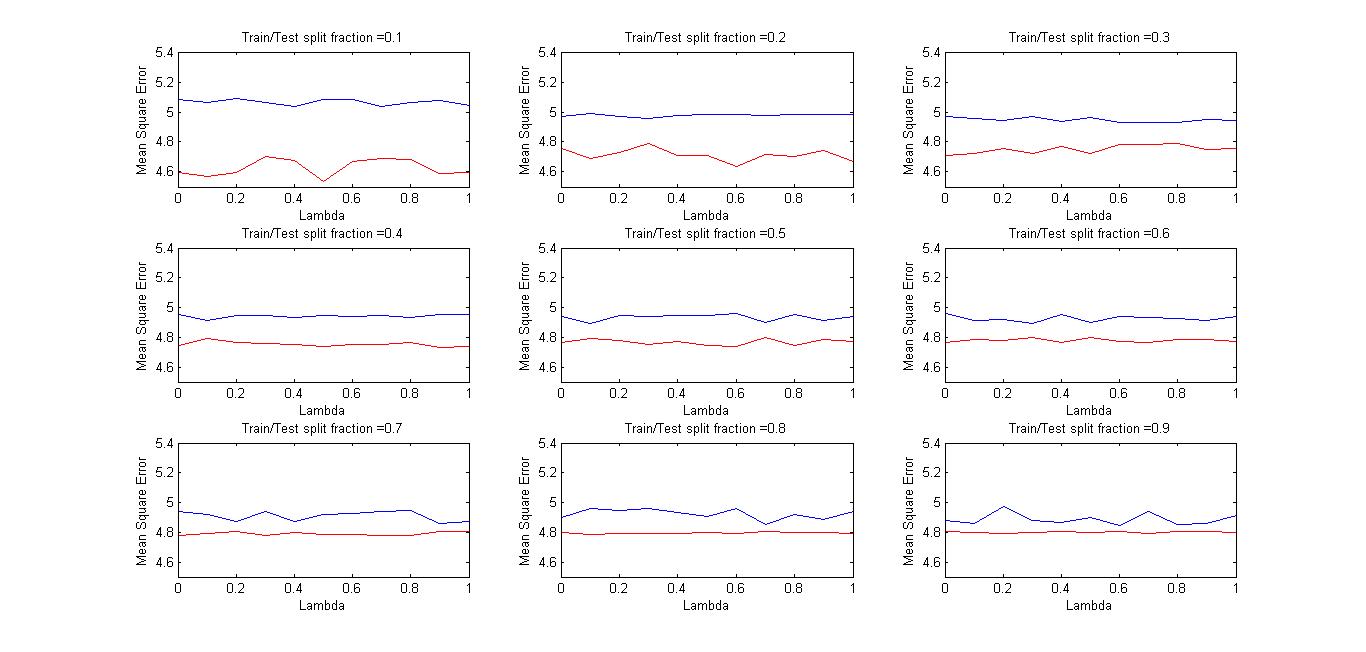
We try to find the weights learned when the training test split fraction is 0.8 and the values of the lambda is 0.2.

We get the weights as [ 9.9315 0.1142 -0.2677 0.1485 -0.0586 1.1871 0.3746 4.1069 -4.3840 -0.8967 1.1970] so we can see that 9th feature with coefficient -4.3840 is the most significant feature. Also 2nd ,5th and 4th features are the least significant attributes .The present training and test error is 4.7772 and 4.9421 respectively .When we remove these three features and then retrain using the remaining features then we get the training and testing error as 4.8245 and 4.8746 respectively .

**Observation :** Removing the least significant features does not causes not much change in the training and test mean square error .

1. **Plotting graph of mean square error vs lambda for each fraction**

Now the Average mean square error for both test and training set is found for each fraction from 0.1 to 0.9 and the graph is plotted between the values of lambda and the mean square error .

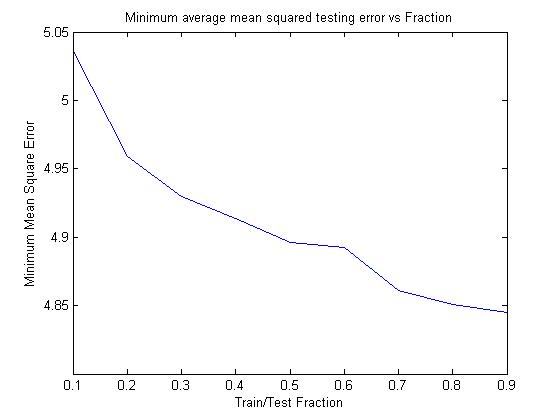


**Observation :** It is observed that the training set mean square error s always lower than test set mean square error in general though the difference is varying .

**Plot :** The given graph is stored as figure1.jpg

3.**Plotting graph minimum error vs fraction**

Now we try to find the minimum average mean square error for each fraction and plot it on the graph .

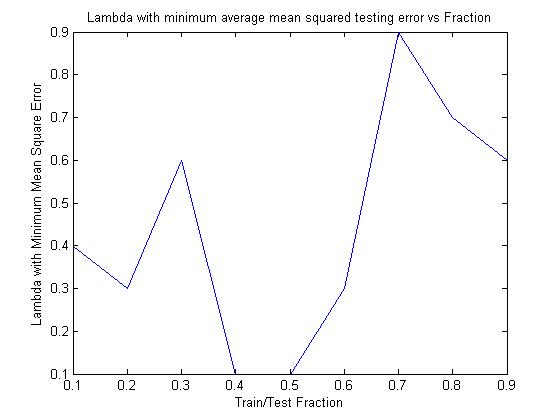


**Observation** : From the graph we can say that the mean square error decreases in general with increasing number of training samples .

**Plot :** The given graph is stored as figure2.jpg

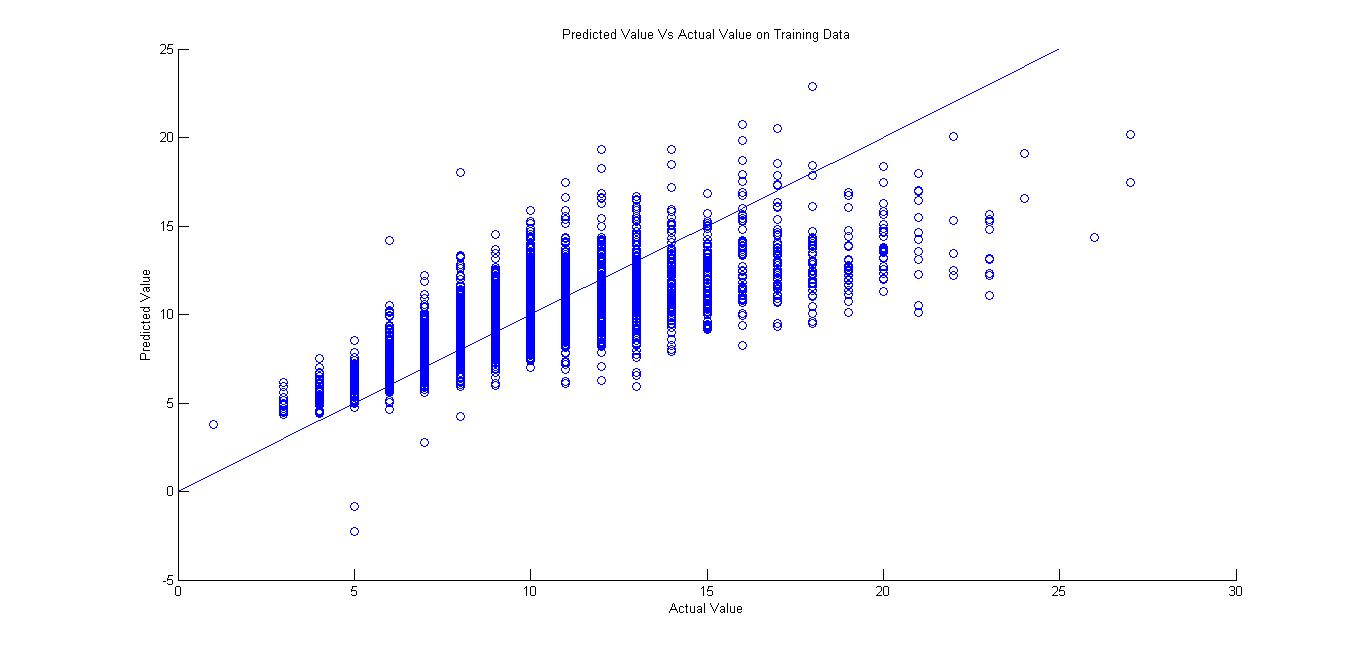
4.Plotting minimum error Lambda against the fraction

Now we find the lambda for which we got the minimum average mean square error versus the fraction of the training test split .



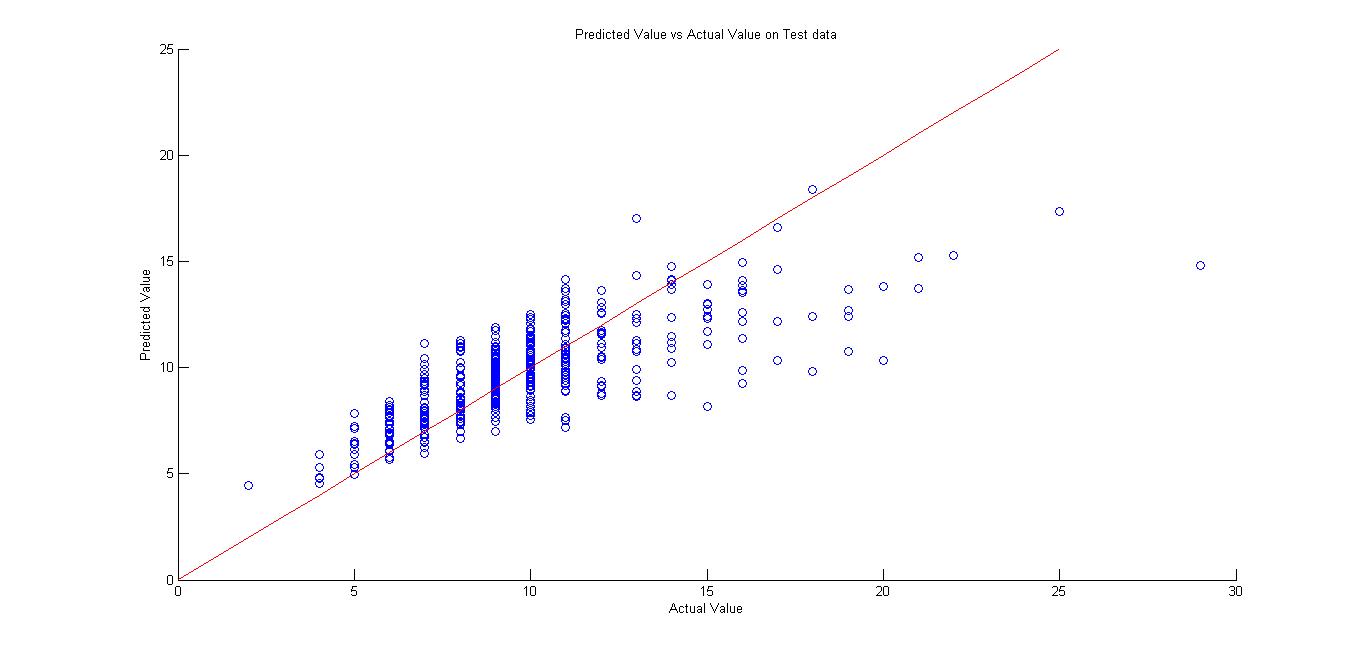
**Plot :** The given graph is stored as figure3.jpg

**5.Plotting the predicted vs actual value for training data**

Now we try to visualize the value predicted as compared to the actual value .This is the visualization for the training data 

**Plot :** The given graph is stored as figure4.jpg

6. **Plotting the predicted vs actual value for test data**

Now we try to visualize the value predicted as compared to the actual value .This is the visualization for the test data 

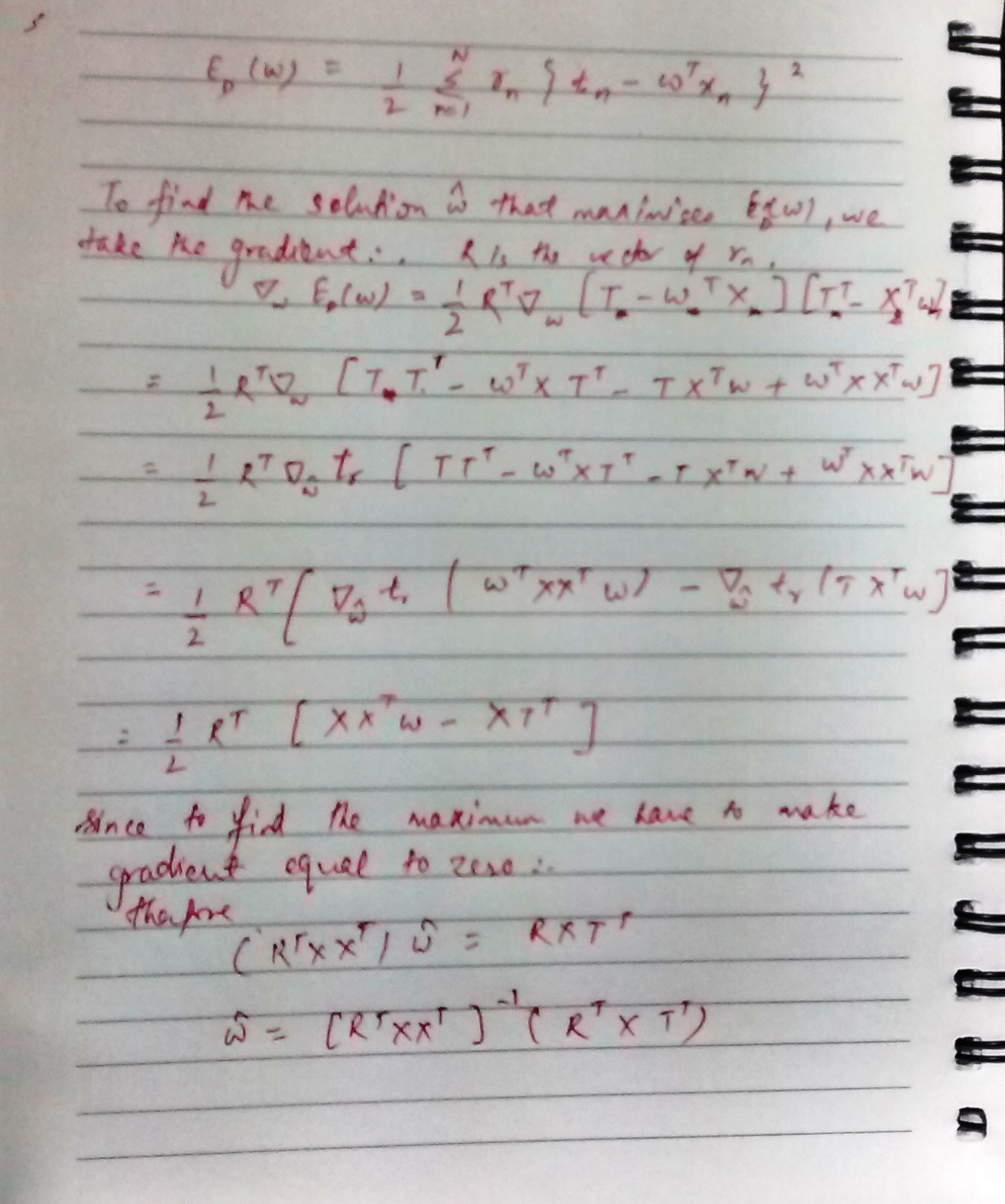
**Observation** : Most of the points are close to the x=y line but there are some point which are a bit far and contribute more to the total error .

**Plot :** The given graph is stored as figure1.jpg

**Question 2:**

1. We would expect the training residual sum of squares (RSS) for both linear and quartic to be **same** . The reason is since the actual relation is linear therefore the linear model would be able to fit perfectly and since the degree of actual relation is less than the degree of quartic model therefore the quartic model would also be able to fit perfectly therefore the residual sum of squares for both would be same and zero .
2. We would expect the test residual sum of squares of quartic to **be more or equal to that of linear** . The reason is the actual relation is linear so linear model would predict perfectly on test data and hence zero error . But in the case of quartic model learned in (a)can be linear(coefficient of higher degree zero ) or higher degree less than four . If the model was linear than both model have same error but if model learnt was not linear than the eoor for quartic model would be greater since the actual relation is linear .
3. We would expect the training residual sum of squares for quartic **model to be less than linear model** . The reason is since the the quartic model is of higher degree it can fit more closely to the points than a linear model . Hence lower error.
4. **There is not enough information to tell** . We don’t know the actual relation how far is the relation is from the linear . If the relation is almost close to linear(lower degree features with high coefficient and the higher degree feature having low coefficient ) then the linear model would have lower error then quartic .In this case the quartic model may fit well on the training data but perform worse on test data . Otherwise quartic model would have lower error .

**Question 3:**

Solution**:**