CSL-407 (Machine Learning)
Homework – 1

Submitted By:
Rajeev Kumar
(2011CS1031)

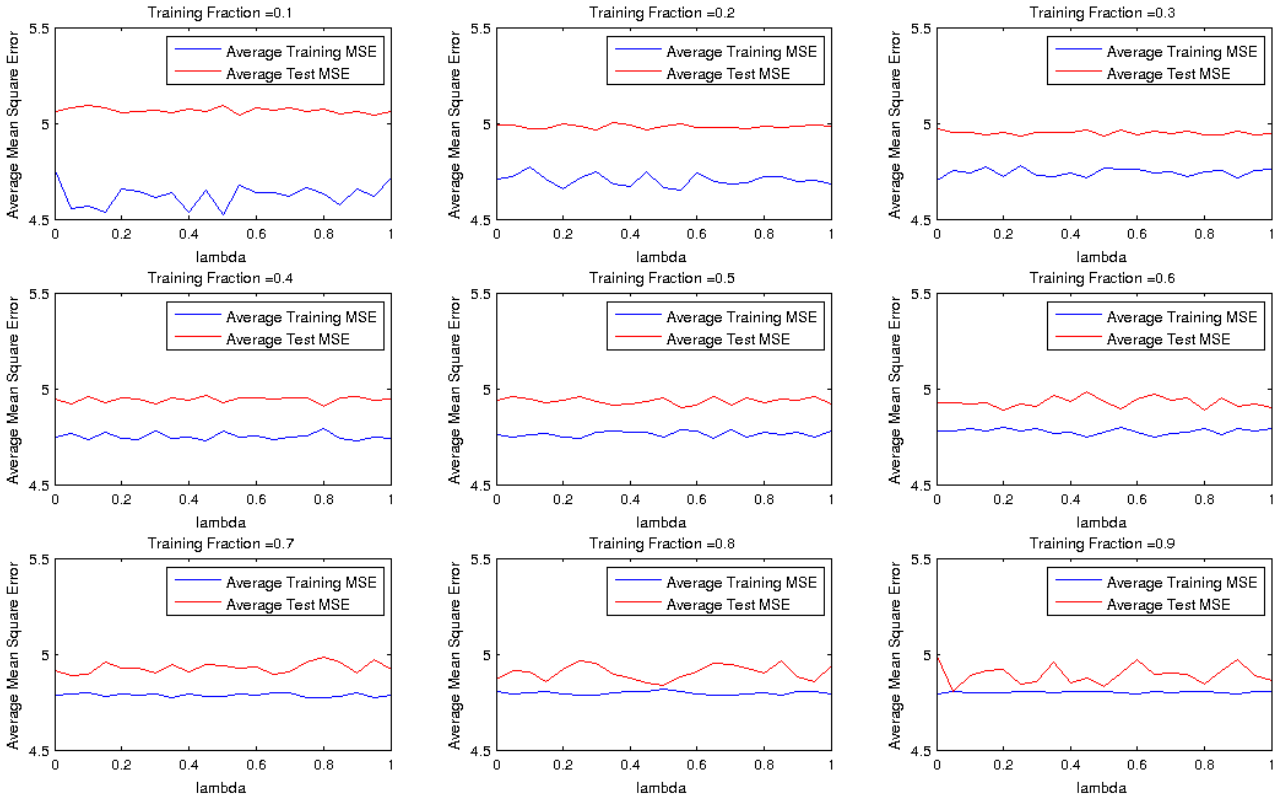
Question 1

Answer 1.6

For lambda = 0.4 and Training fraction = 0.8

the weight vector comes out to be {9.9351 0.1217 -0.2536 0.1278 -0.0798 0.8997 0.8100 4.4604 -4.4701 -1.1309 1.0887 }. Here the highest weight coresponds to W0 (9.9351) but since X0 is always 1. Considering this, the most significant attributes would be W7 and W8 (weights corresponding to whole weight and shucked weight attributes). The least significant weights would be W4, W1 and W3 (weights corresponding to length, female and male attributes). Removing the least significant attributes does not create much change in the average MSE.

Fig 1) Plot between Average MSE and lambda for different Training Fraction.



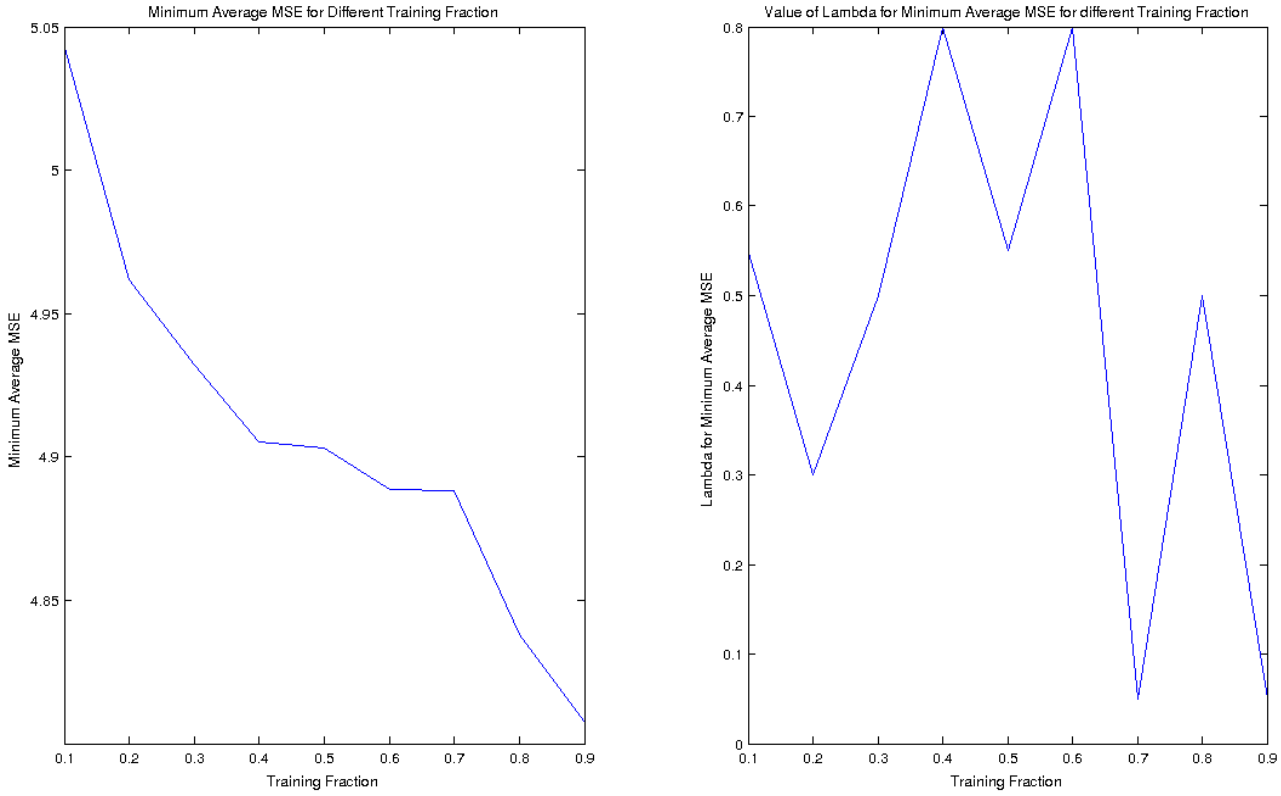
A good learning model should have lower average MSE for both training and test data set.

Observations:

1) The average MSE of test set is mostly greater than the average MSE of training data set.
2) As the training fraction increases, the average test MSE decrease. This is because as the

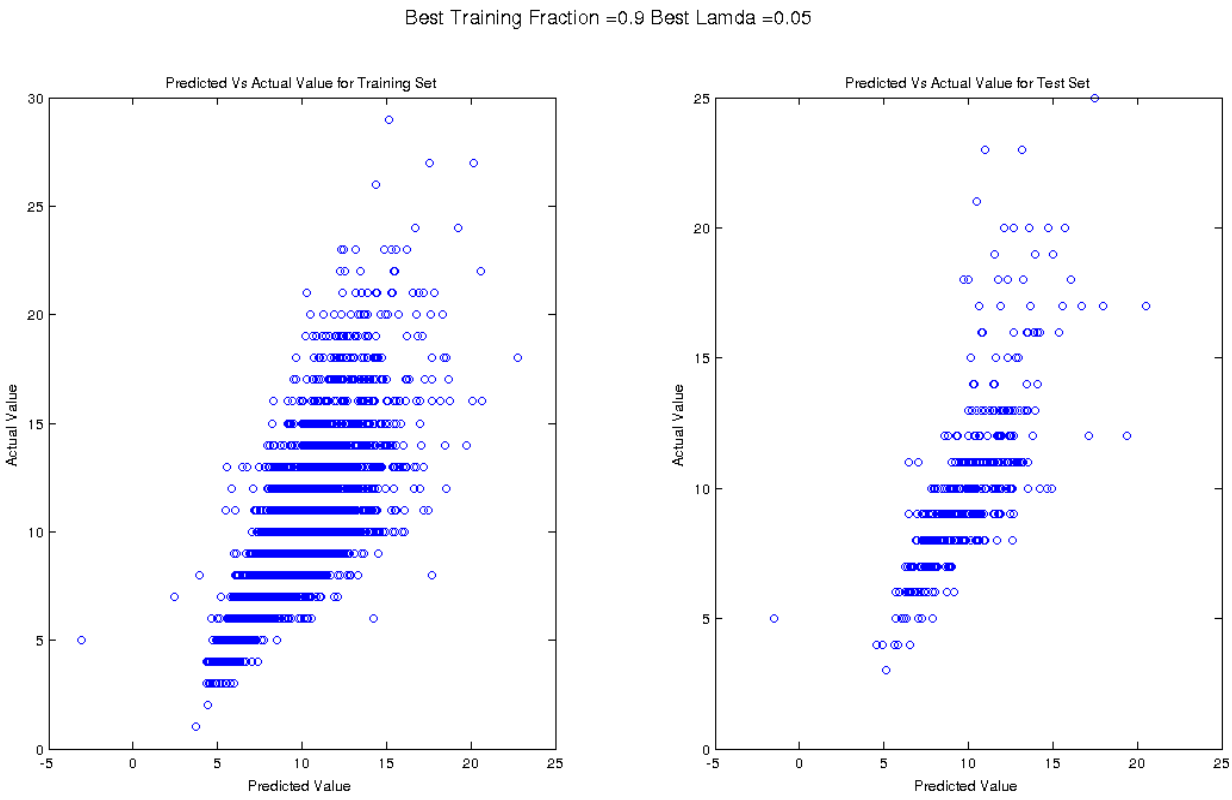
number of training instances increase, the model is trained better for predicting the new instances. But after a certain training fraction, the average MSE of test set does not decrease much due to the attainment of saturation in the learning model.
3) When the value of lambda is between 0 and 1, it does not have much effect on the value of average MSE but for higher values of lambda (above 50 or so), the average MSE increase with increase in lambda.

Fig 2) : Minimum Average MSE (for test set) vs Training Fraction (subplot 1) and Lambda for minimum Average MSE vs Training Fraction.



Observations:
1) Minimum Average MSE for test set usually decrease with increase in the training fraction.
2) Lamda for Minimum MSE is usually different for different runs of the code.

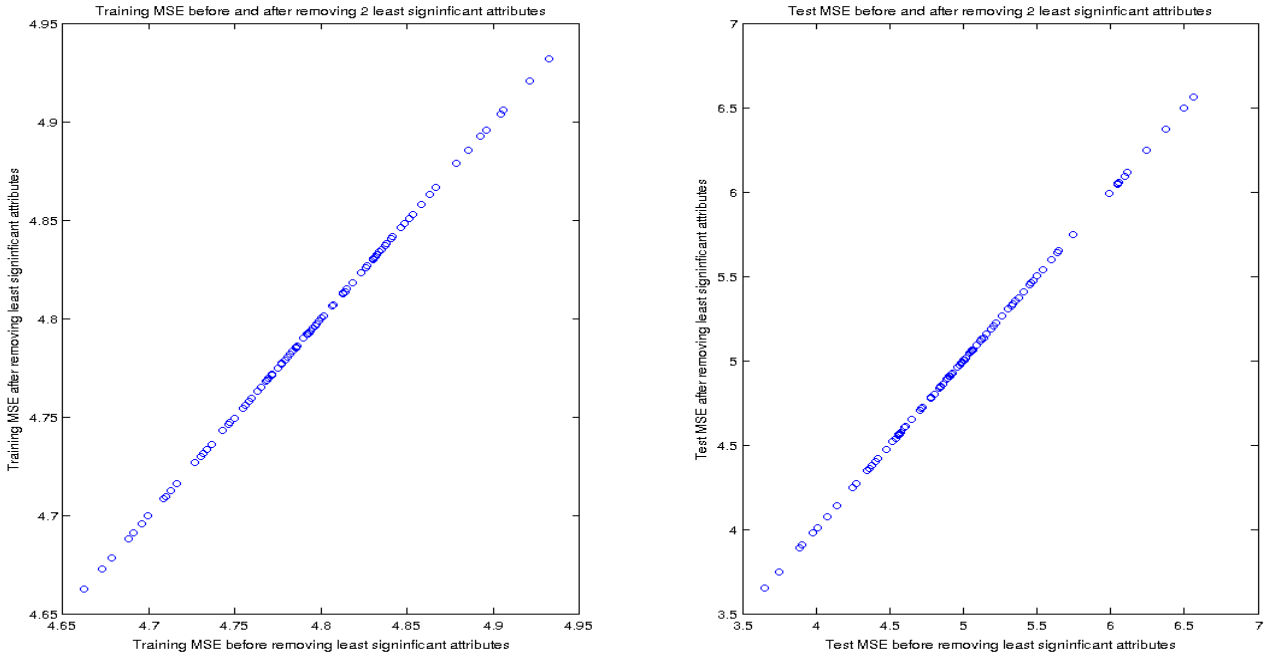
Fig3): The value of predicted Target Attribute and Actual value of target attribute for Training set and Test set.



Observations:

1) The closer the points are to the 45 degree line, the better model of prediction we have.
2) The predicted and Acutal value for Test set is fairly near the 45 degree line.

Fig 4: Effect of removing least significant attributes on the average MSE.



The above figure plots the average MSE value before and after removing two least significant attributes for some 100 random partitions of dataset into training set and test set. As mostly the points are near the 45 degree line. So, we can say the removing the least significant attributes does not have much effect on the predicted value of target attribute.

Question 2)

a) Since we have more variables in quartic regression, so a quartic equation should be able to fit the training data better than a linear equation. So quartic equation is expected to have lower RSS.

b) On test data, a linear regression model would have lower RSS due to overfitting problem in quartic equation.

c) A quartic regeression model should have lower RSS value for the training data.

d) Since we don't know how far the actual relation between X and T is from linear, so we can't say which regression model would have lower RSS value. Given information is not enough.