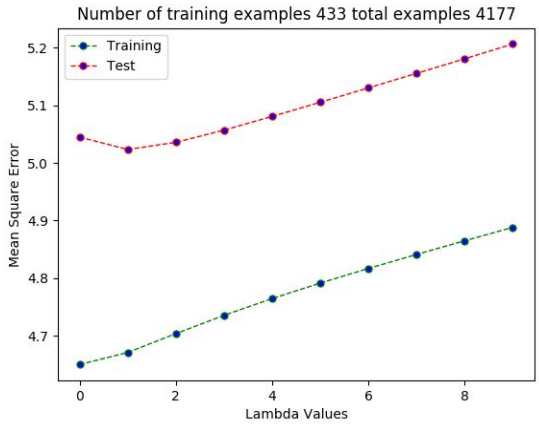
Linear Regression Model and Analysis

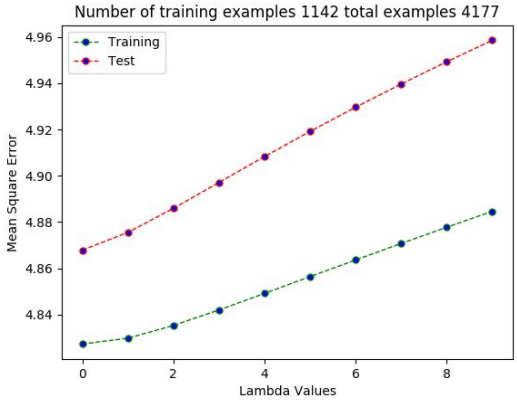
The analysis is based on convergence of cost function by regularized analytical normal equations. The program is implemented in linear.py inside code folder.

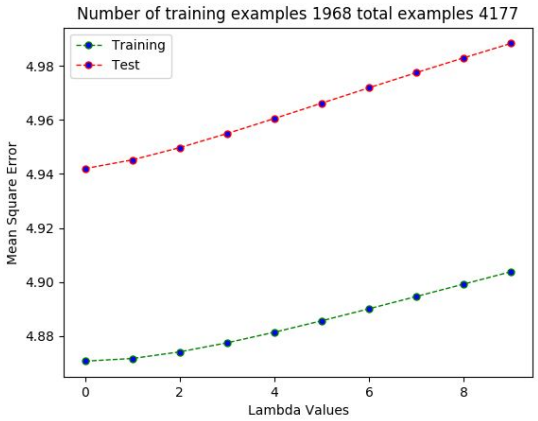
Approach and graphs plotted

First I created 100 fractions of training of training set and each training set is trained for lambda values of 0 to 9. All the generated graphs are stored in graphs folder.

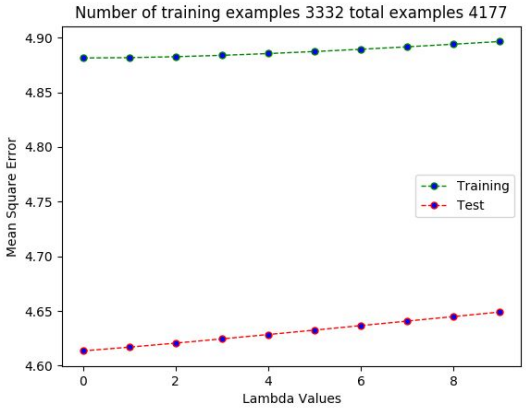
Graph of some particular mean square error vs lambda values are:-

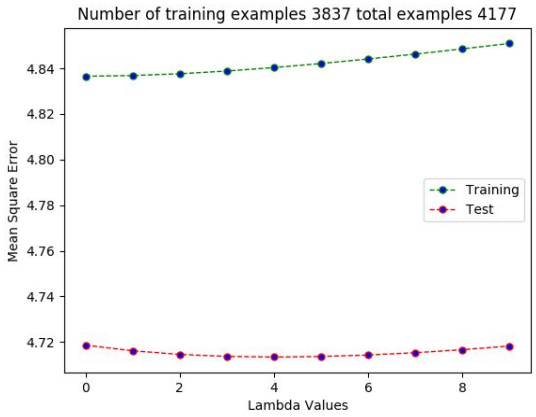

Number of training examples 433 total examples 4177
Training
Test
Mean Square Error
Lambda Values


Number of training examples 1142 total examples 4177
Training
Test
Mean Square Error
Lambda Values


Number of training examples 1968 total examples 4177
Training
Test
Mean Square Error
Lambda Values


Number of training examples 2702 total examples 4177
Training
Test
Mean Square Error
Lambda Values

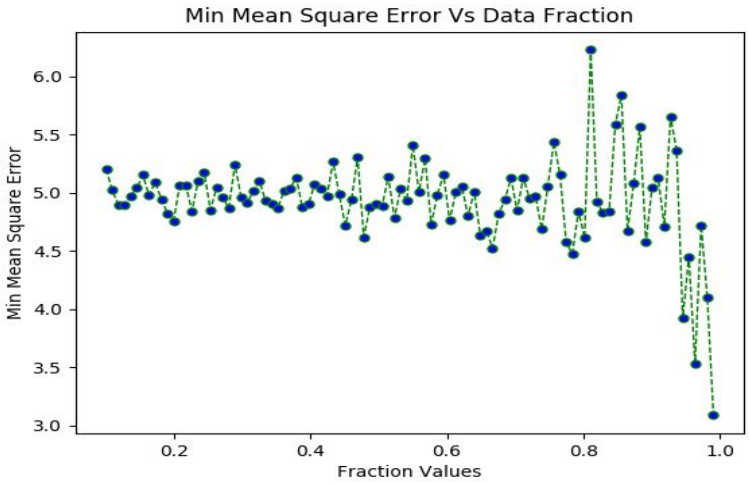

Number of training examples 3332 total examples 4177
Training
Test
Mean Square Error
Lambda Values

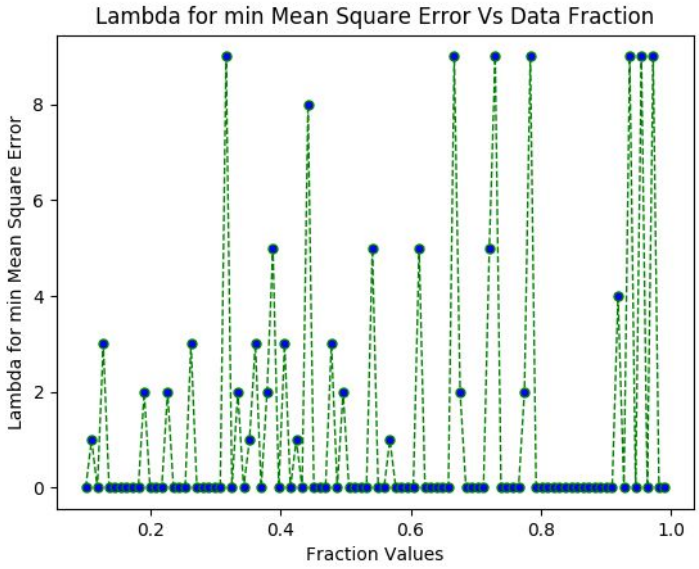

Number of training examples 3837 total examples 4177
Training
Test
Mean Square Error
Lambda Values

Observations

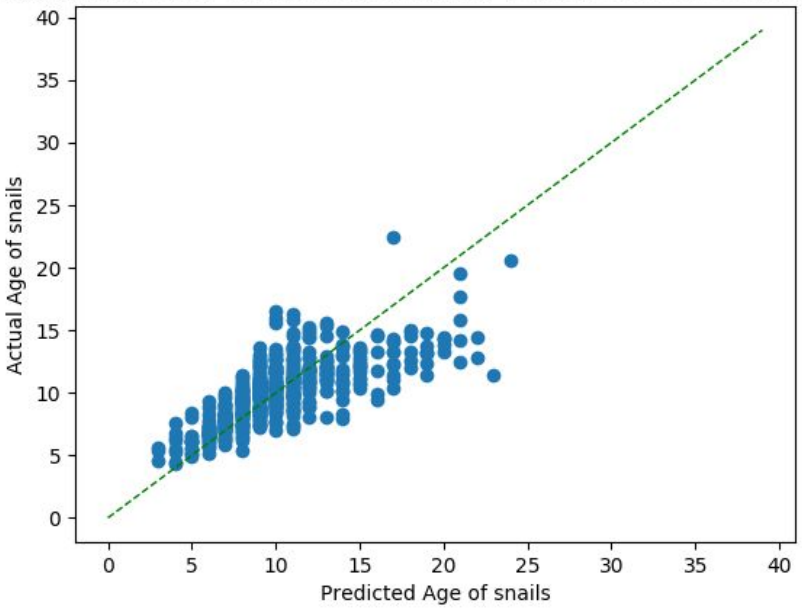
1. As lambda increases error increases, since the bias increases.
2. When fraction increases the error becomes almost constant, since the variance increases with number of training examples.
3. When fraction values are low error changes drastically because lambda tries to regularize the high variance.

Graph for minimum squared error and Lambda VS fraction values and graph for predicted VS actual value

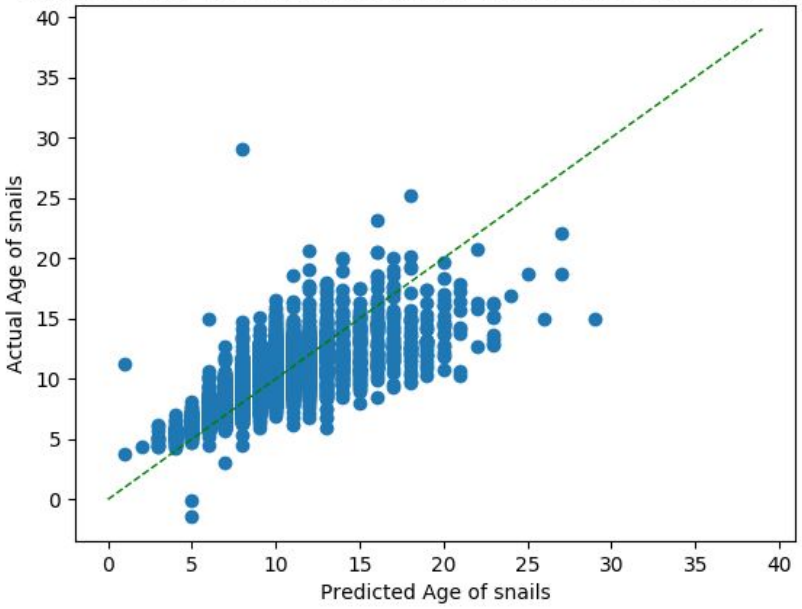




Min Training Error fraction: 0.91 lambda: 4 Error: 3.7108228841546675



Min Test Error fraction: 0.91 lambda: 4 Error: 3.7108228841546675



Observations

1. Minimum squared error decreases as we train our model with larger training dataset, since variance decreases with more training examples.
2. The graph is decreasing because bias is kept constant while variance decreases which increases the accuracy of the model.
3. For a good model the predicted and actual values are should lie around the 45 degree line which can be seen in our graph.