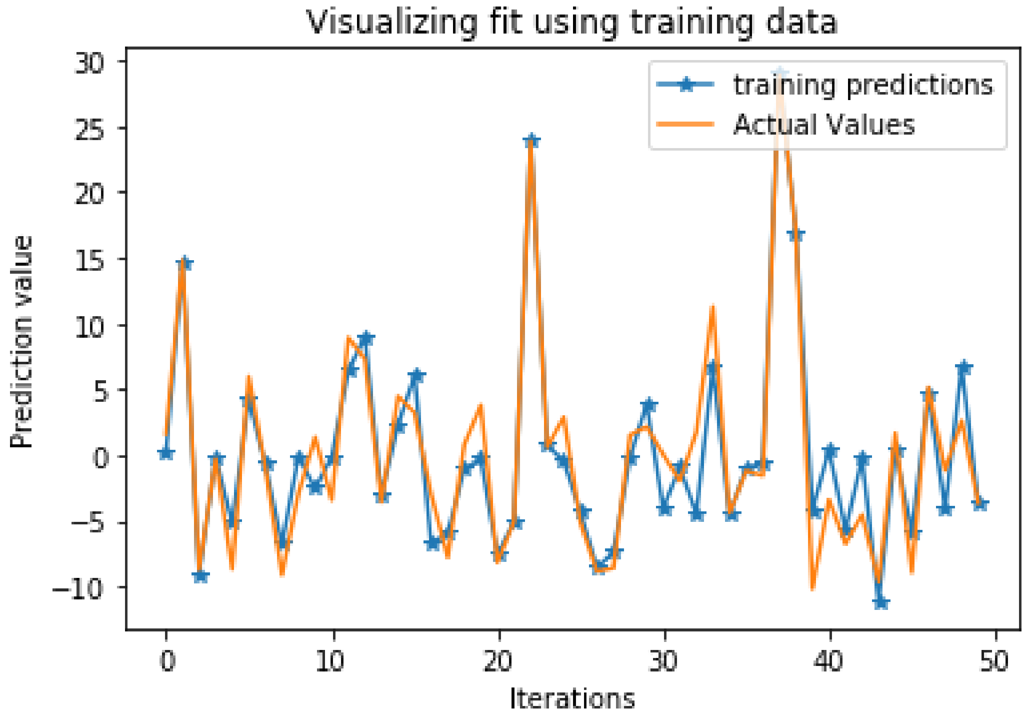
Comp551 : Assignment 1

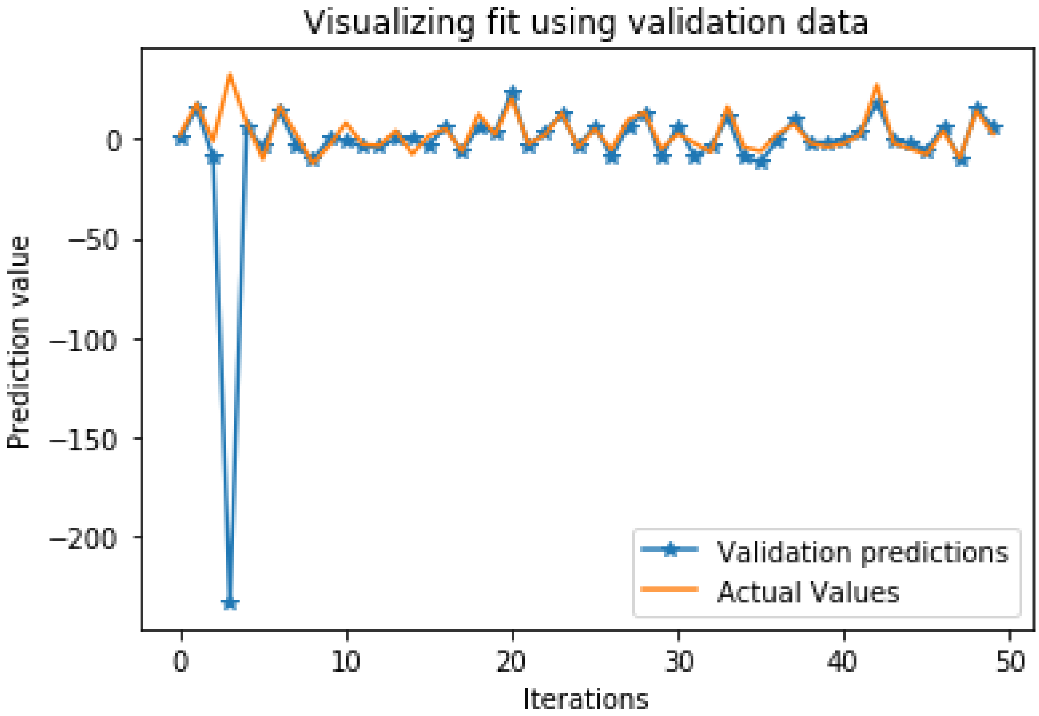
Anand Kamat

ID# 260773313

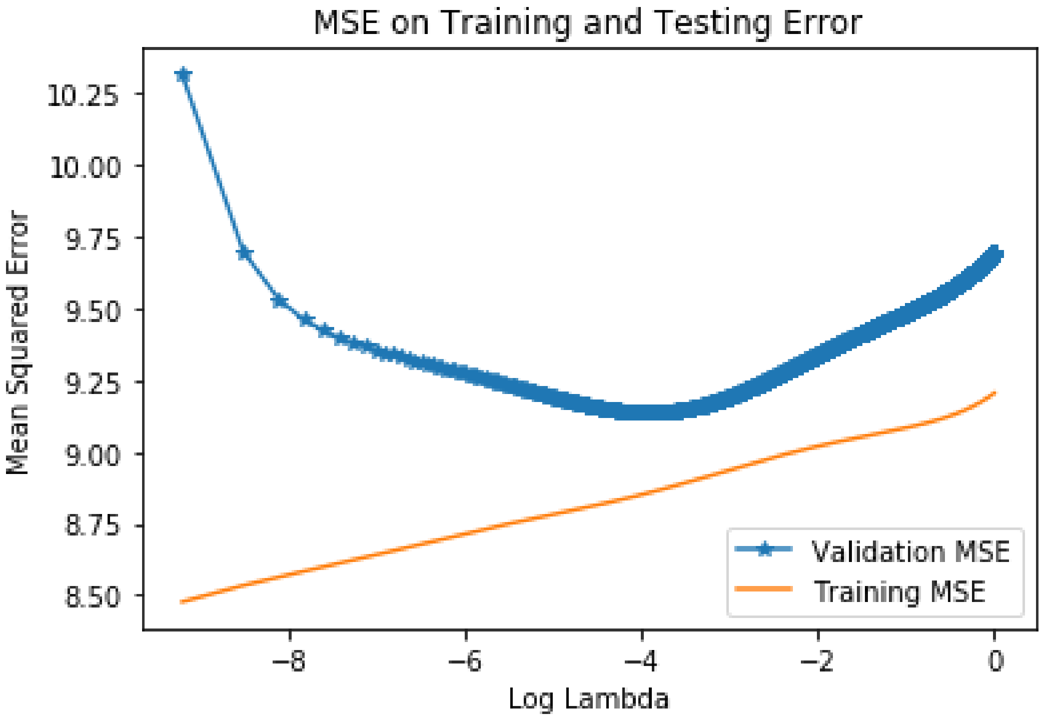
Question 1:

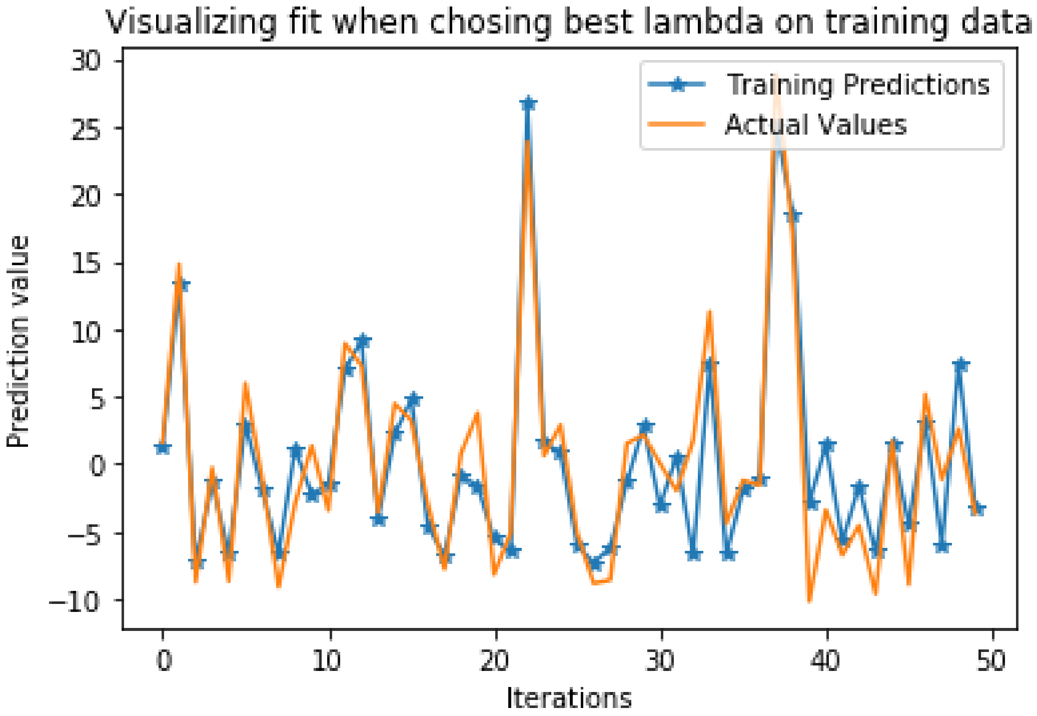
1. The following graph represents the training fit for the data provided. We can observe the predicted values are quite close to the actual values given in the data which is why the mean squared error for the training error is only 6.47474257381. The validation fit is shown in the graph that follows. We observe besides one huge fluctuation, the data is also closely fit. However due to the big jump in the data point the validation error comes to be significantly bigger than the training data (with MSE of 1422.05475784).

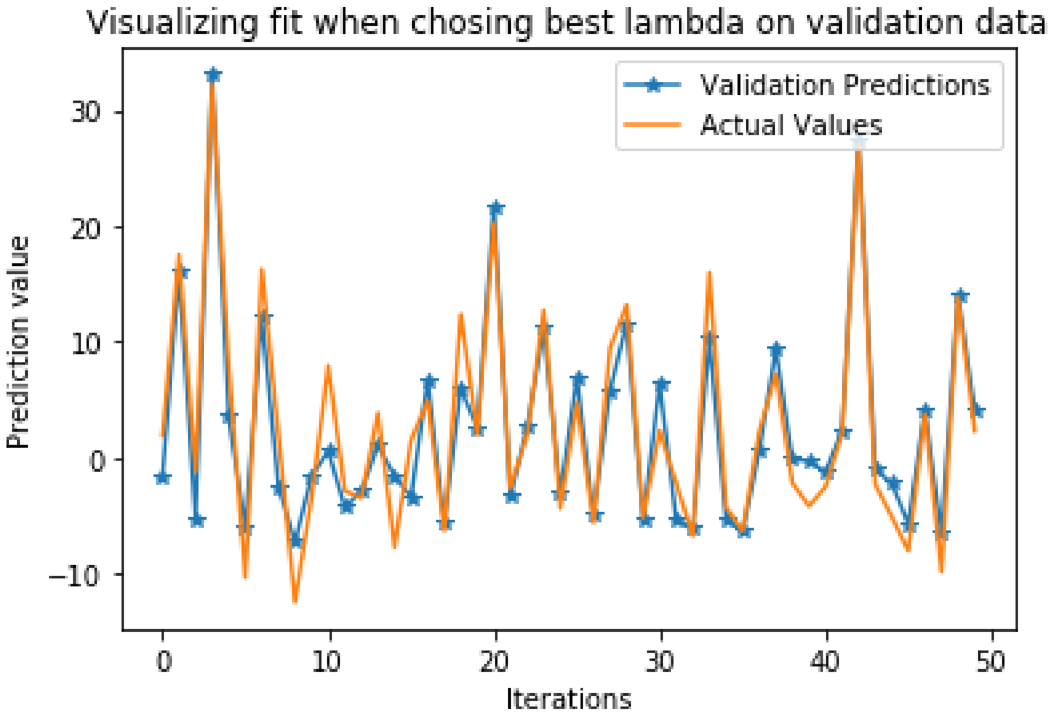




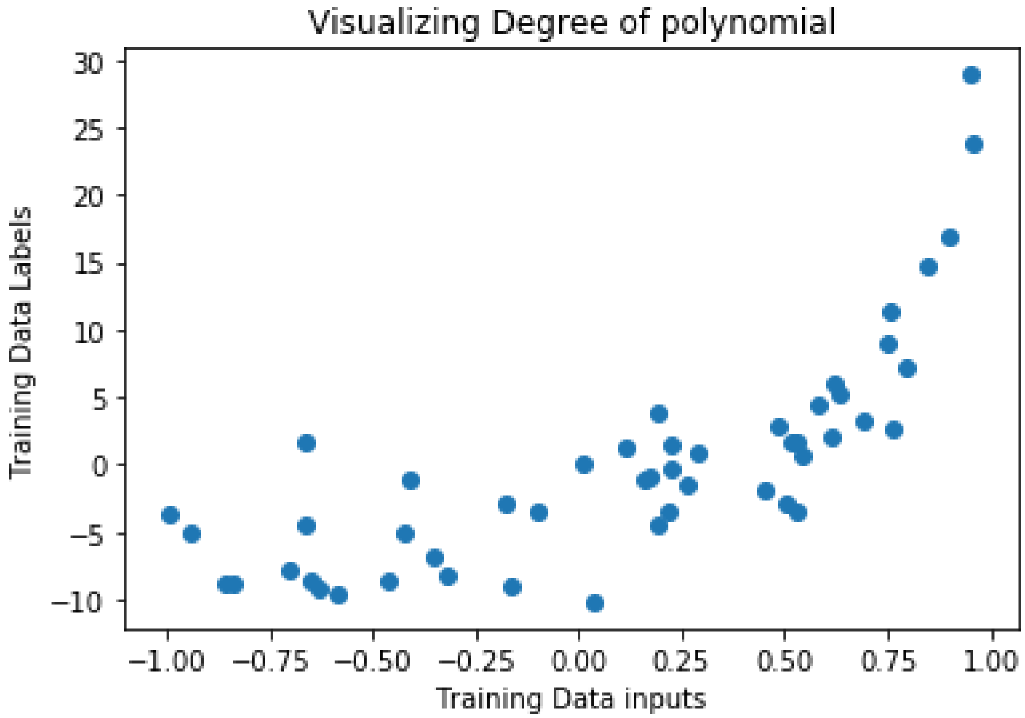
2. Lambda was varied from 0 to one with an increment of size 0.0001. The plot below depicts the Mean squared error changing as lambda changes (log of lambda is considered as in the lecture slides). As expected the training error continuously drops with increasing in lambda value(log lambda in the graph) and the validation error first drops and then starts increasing again. The best value of lambda for this data is 0.0197 which gives testing mean squared error of 10.7323010053. The following two graphs depict the fit of the training predictions and the validation predictions when the value of lambda is 0.0197. We can observe the validation fit to be much better compared to the previous plots.


MSE on Training and Testing Error
Mean Squared Error
10.25
10.00
9.75
9.50
9.25
9.00
8.75
8.50
−8
−6
−4
−2
0
Log Lambda
Validation MSE
Training MSE


Visualizing fit when chosing best lambda on training data
Prediction value
30
25
20
15
10
5
0
−5
−10
0
10
20
30
40
50
Iterations
Training Predictions
Actual Values

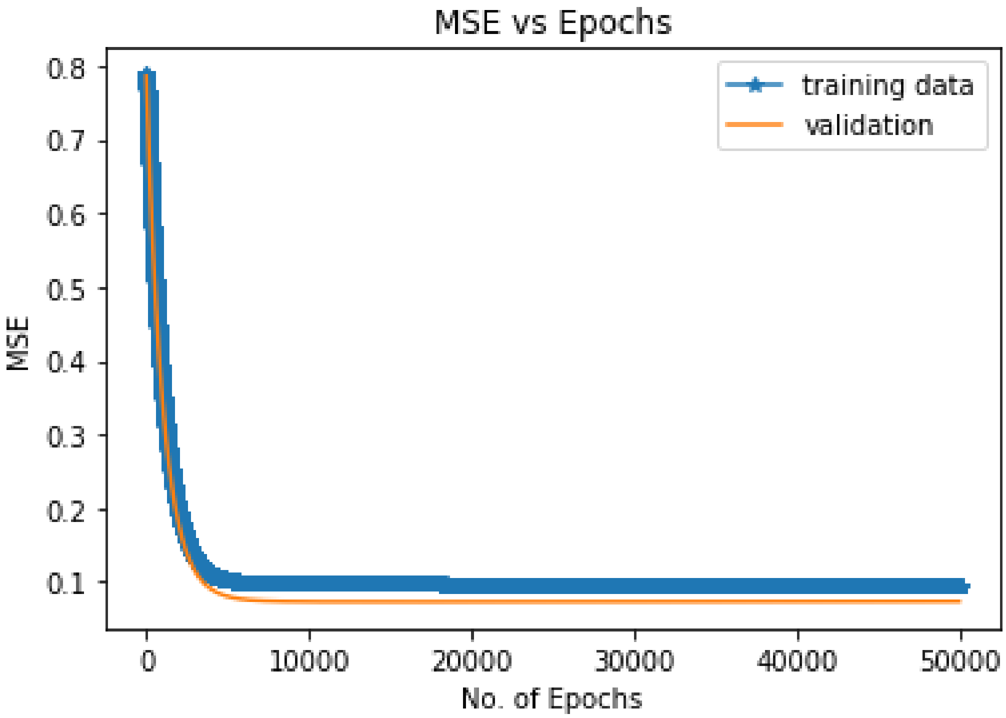


3. Accurately predicting the degree of polynomial on such a dataset isn't a direct process. The following scatter plot demonstrates the data points on the graph. Based on the number of noticeable undulations the line would be having, I would guess the degree to be 8 or 9. The degree of the polynomial is definitely less than 20 as deduced by the regularization process. It would prove to be difficult to guess the degree of the polynomial from the above graphs as despite regularization there is significant error and uncertainty in the predictions.



Question 2

1. The following graph shows the Mean Squared Error (MSE) for the training and validation data as the number of epochs increases. The initial values of the parameters were 3 and 4 and the algorithm was set to run 50,000 epochs. We can observe a very similar trend for both training MSE and validation MSE.

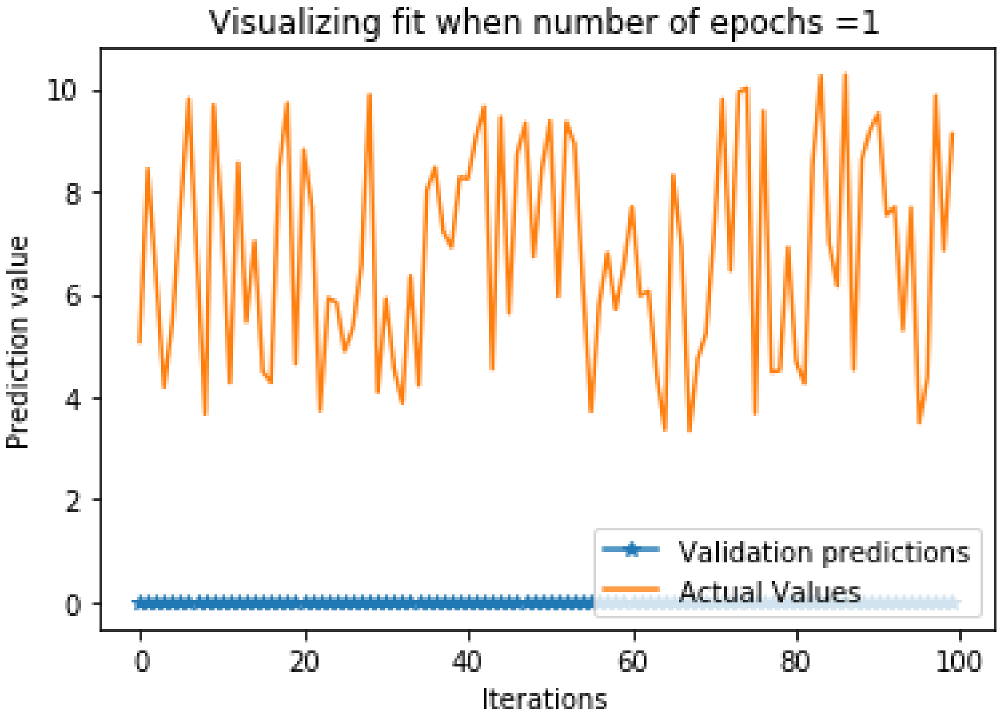


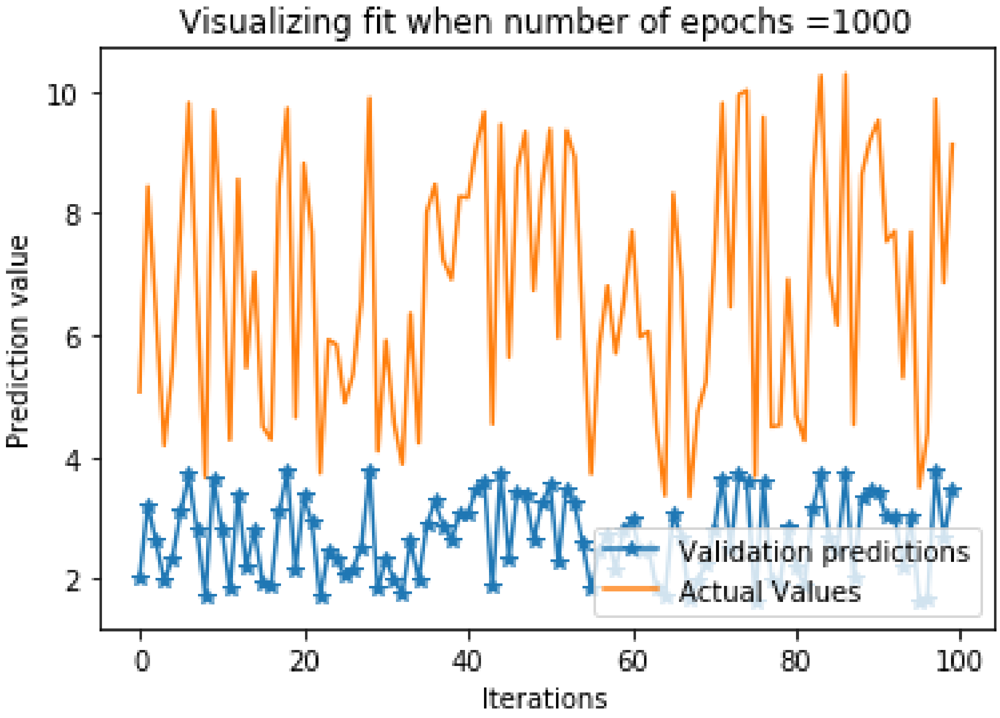
2. Gradient descent was carried out in these following step sizes and the following validation error was achieved:

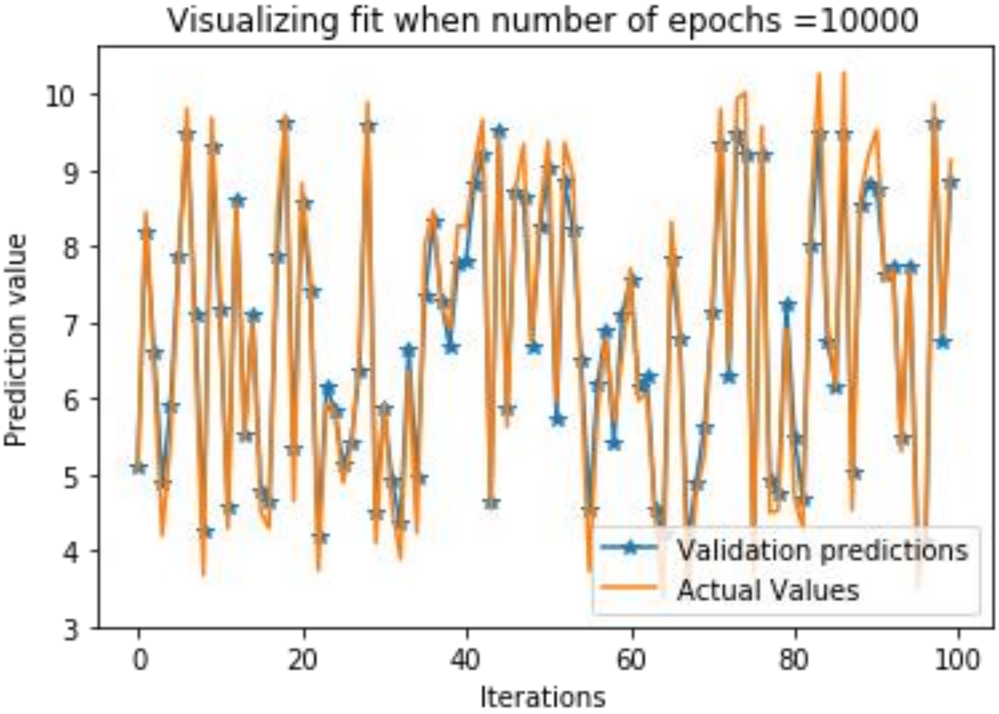
Step Size	**Training Error**	**Validation Error**
1e-6	0.0955856192636	0.0734708365076
1e-5	0.0955063571238	0.0740694476135
1e-7	0.101653524692	0.0801727105788

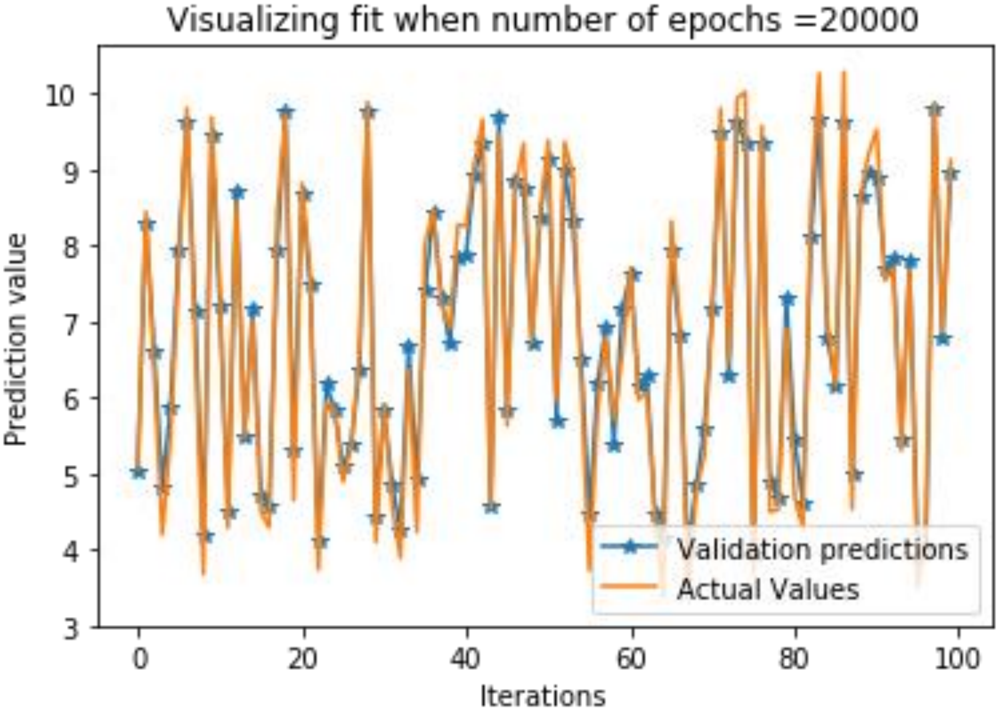
The least MSE was achieved when the learning rate was 1e-6 when the validation error was 0.0734708365076. The testing MSE when learning rate was 1e-6 is 0.0734708365076.

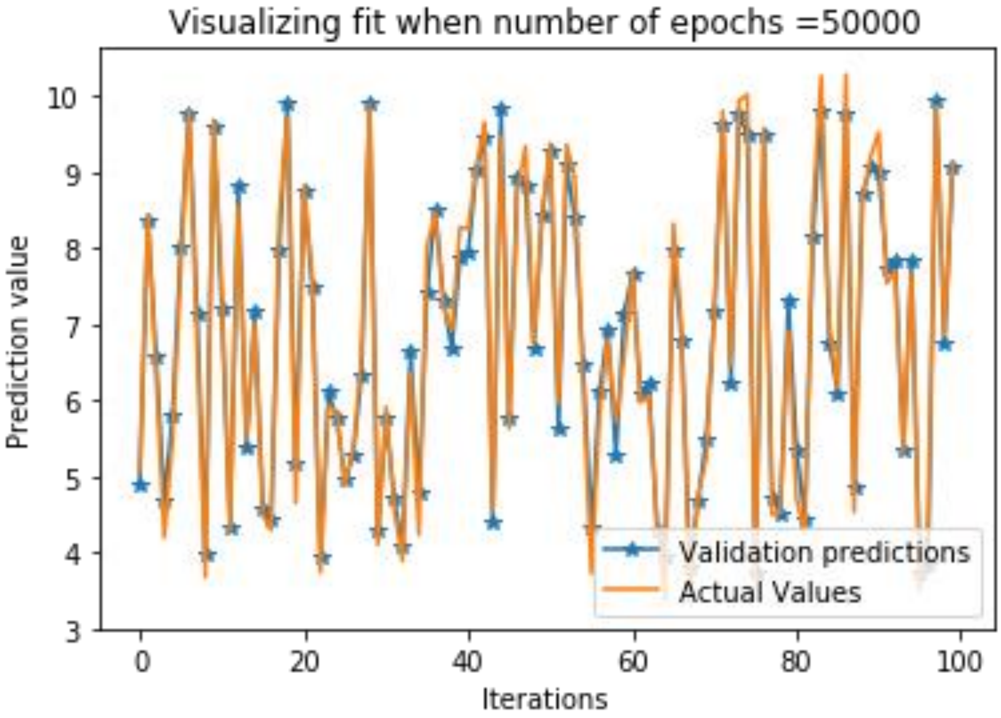
3. Below are five visualizations at epoch 1, epoch 1000, epoch 10000, epoch 30000 and epoch 50000 respectively.


Visualizing fit when number of epochs =1
Prediction value
0
2
4
6
8
10
0
20
40
60
80
100
Iterations
Validation predictions
Actual Values


Visualizing fit when number of epochs =1000
Prediction value
2
4
6
8
10
0
20
40
60
80
100
Iterations
Validation predictions
Actual Values


Visualizing fit when number of epochs =10000
Prediction value
10
9
8
7
6
5
4
3
0
20
40
60
80
100
Iterations
Validation predictions
Actual Values


Visualizing fit when number of epochs =20000
Prediction value
10
9
8
7
6
5
4
3
0
20
40
60
80
100
Iterations
Validation predictions
Actual Values



Question 3

1. Sample mean seems like a logically reasonable choice to fill in the missing values given the limited data available for this question. There are several other values one can use instead of mean, like median. My preference would be towards removing the rows containing missing data points. This would fairly maintain the bias and noise the dataset originally contained in the dataset. However, I did not remove the rows from the dataset as it would shrink the dataset a lot.
2. The five-fold cross-validation MSE on the testing data without regularization is 0.223086110832.
3. The five-fold cross-validation MSE on the testing data with regularization is given in the csv file labelled **MSE with regularization.csv**. The value of Lambda giving the lowest MSE is 1.9 when the average test MSE is 0.0184443222005.

   Ridge Regression of L2 regularization isn't radical enough to drastically reduce the weights to almost zero although its likely given a high regularization coefficient. Feature selection is more prominent in L1 regularization or Lasso Regression where some of the weights may be forced to reduce to zero. Since in none of the instances the parameters drew close to zero, we cannot definitively state which features can be set to zero. L1 regularization might be more suited here.