Assignment 2

By-

Drishti Mamtani(20165A70574H)

Shreyansh Jain(2016B1A30947H)

Shreya Saxena(2016A4PS0367H)

**Part A - Gradient Descent Method**

**Training**

Degree =1

Mean Sum of Squared Error=338.32

R^2 = 0.0267

RMSE =18.393

Degree =2

Mean Sum of Squared Error=326.32

R^2 = 0.0612

RMSE =18.064

Degree =3

Mean Sum of Squared Error=294.86

R^2 = 0.1517

RMSE =17.171

Degree =4

Mean Sum of Squared Error=277.03

R^2 = 0.203

RMSE =16.644

Degree =5

Mean Sum of Squared Error=261.25

R^2 = 0.2484

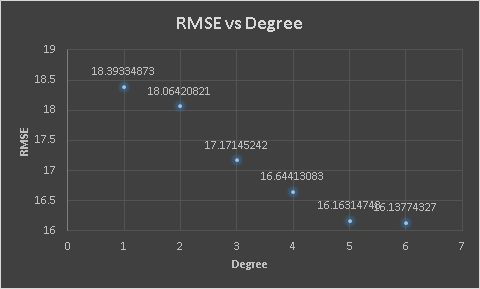
RMSE = 16.163

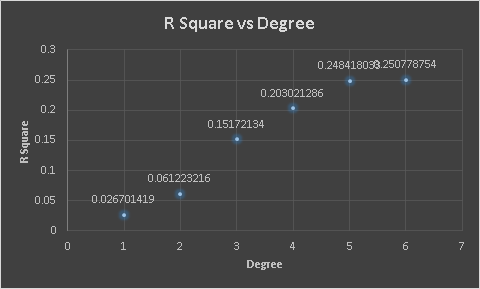
Degree =6

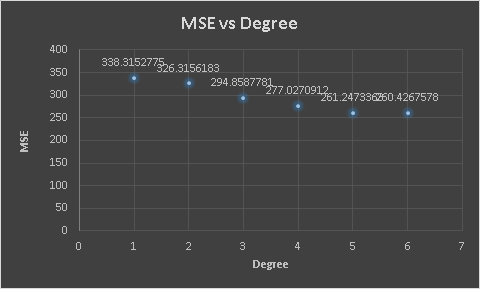
Mean Sum of Squared Error=260.23

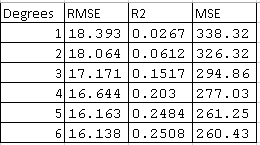
R^2 = 0.2508

RMSE =16.138

****

****

****

****

**Testing**

Degree =1

Mean Sum of Squared Error=528.64

R^2 = 0.0278

RMSE =22.99

Degree =2

Mean Sum of Squared Error=399.38

R^2 = 0.0258

RMSE =19.98

Degree =3

Mean Sum of Squared Error=328.72

R^2 = 0.1478

RMSE =18.13

Degree =4

Mean Sum of Squared Error=275.69

R^2 = 0.19942

RMSE =16.604

Degree =5

Mean Sum of Squared Error=260.43

R^2 = 0.2437

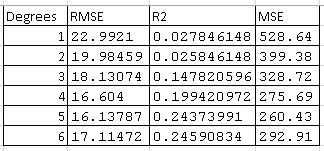
RMSE = 16.1378

Degree =6

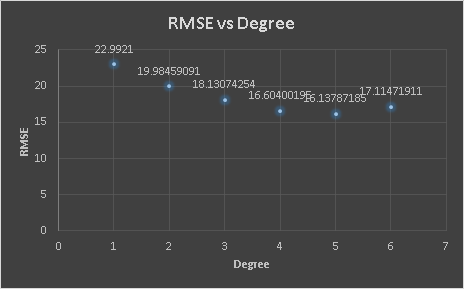
Mean Sum of Squared Error=229.91

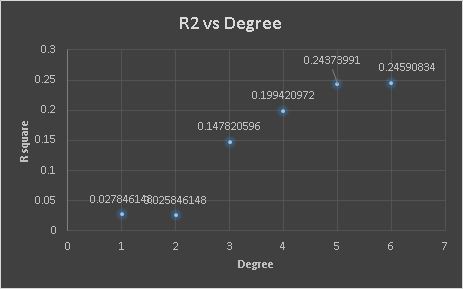
R^2 = 0.24590

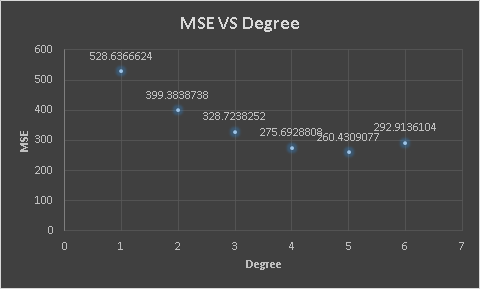
RMSE =17.114



|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |







Part B:

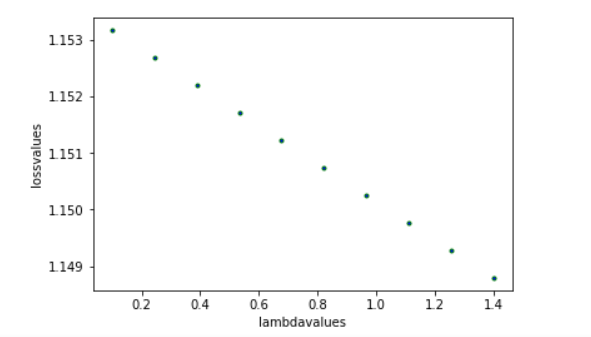
* Sum of Square of Error (or SSE) measures the total deviation of the predicted values from the target values. Hence, lower the SSE, better is the regression model
* As shown from the training data RMSE and MSE values are lowest for the training data and very high for a polynomial of the degree of 1 and 2. This implies that polynomial of degree 6 overfit the data and polynomials of degree 1 and 2 which give very high MSE values underfit the data.
* Now when the models are run for testing data the error (MSE) increases which is obvious because it is a new set of data given to the model. But as we can see the error obtained for a polynomial of degree 6 is more than the error for a polynomial of degree 5. This validates the fact that Polynomial of degree overfits the data.
* The model which gives the least error (MSE) in testing data is polynomial of degree 5.
* Hence the model which best fits the data is polynomial of degree 5, model which underfits the data is polynomial of degree 1 and 2 and the model which overfits the data is polynomial of degree.
* Regularization obtained on degree 6 would decrease the error hence we go for Part C.

**Part C - Gradient Descent Method along with regularization**

**L1 Regularization**

For a polynomial of degree 6, there are 28 terms and 7 ws. Initially, all were taken as 10 and the regularization coefficient was varied from 0.1 to 1.4. The dataset was divided into 3 parts Training Data,Cross-Validation Data and Testing Data. Cross-Validation Data was used to find the best Regularization coefficient and for L2 regularization the best Lambda value came out to be 0.1. Hence the trend is that as the value of Lambda increases Validation Loss increases. This is true because as the w’s are more the regularization term would be more but more weight has to be given in minimizing the error term.

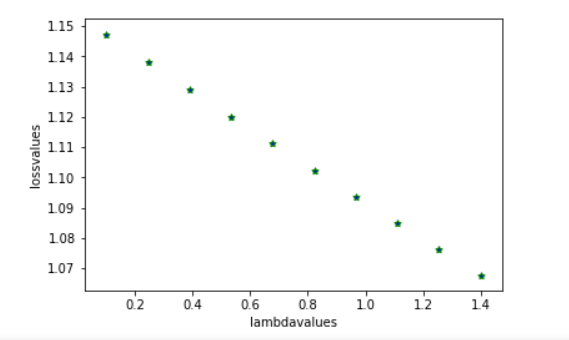
The results displayed are for Testing Data



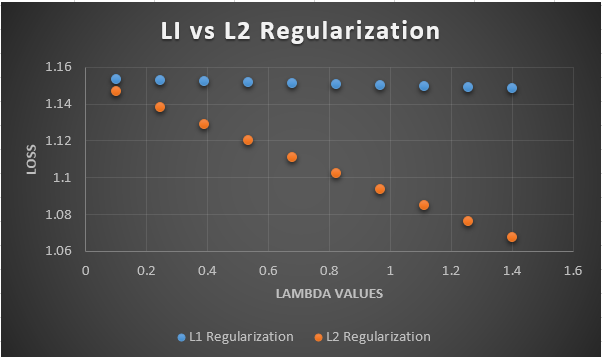
RMSE:12.96834946211578

**L2 Regularization**

For a polynomial of degree 6, there are 28 terms and 7 ws. Initially, all were taken as 10 and the regularization coefficient was varied from 0.1 to 1.4. The dataset was divided into 3 parts Training Data, Cross-Validation Data, and Testing Data. Cross-Validation Data was used to find the best Regularization coefficient and for L2 regularization the best Lambda value came out to be 0.1. Hence the trend is that as the value of Lambda increases Validation Loss increases. This is true because as the w’s are more the regularization term would be more but more weight has to be given in minimizing the error term.



RMSE=15.60282736375703



The above graph shows a significant difference in the validation loss of L1 and L2, this occurred due to an insufficient number of iterations(20)(more iterations could not be performed due to the high computation times). If the number of iterations is increased both iterations might lead to similar results.

After Regularization, the RMSE obtained here is lesser than what was obtained by any model of degree lesser than or equal to 6 without Regularization in Part A. This is because Regularization restricts the limits of coefficients of the Regression equation, hence reduces overfitting. Therefore the RMSE values are the least for the Regularized model of the polynomial of degree 6.