Assignment 2

Data Preprocessing -

1)Drop the first column of ‘id’.

2)Randomly Split the data into training(70%) and testing(30%) for cross-validation.

3)Normalize the training data by applying standardization to make mean as 0 and standard deviation as 1.

4) Use the training data mean and standard deviation to standardize the testing data.

5)Expand the data set to include new columns according to the respective degrees .

PROCESS FLOW

Apply gradient descent for degrees 1 to 5 . And apply gradient descent with regularization to degree 6.

1.For training data:

Comparison of various approaches:

Approach applied on Training Data	Iterations	Learning Rate	Stopping Criteria	R^2	Error	RMSE	SSE	Regularization coefficient
Gradient Descent on 1 degree	**30**	**1e-6**	**1e-4**	**0.0259**	**148252.534**	**0.987**	**296505.068**	**-**
Gradient Descent on 2 degree	**20**	**1e-6**	**1e-5**	**0.060**	**143013.4**	**0.969**	**286026.72**	**-**
Gradient Descent on 3 degree	**600**	**1e-7**	**1e-5**	**0.1447**	**130180.931**	**0.9248**	**260361.862**	**-**
Gradient Descent on 4 degree	**2000**	**1e-8**	**1e-6**	**0.1736**	**125767.4**	**0.9090**	**251534.85**	**-**
Gradient Descent on 5 degree	**3500**	**1e-8**	**1e-6**	**0.2145**	**119555.0319**	**0.8862**	**239110.063**	**-**
L1-Regularization on 6 degree	**500**	**1e-9**	**1e-4**	**0.0867**	**139007.968**	**0.9556**	**278015.936**	**1e-16**
L2-Regularization on 6 degree	**500**	**1e-9**	**1e-4**	**0.0867**	**139007.968**	**0.9556**	**278015.936**	**1e-16**

Comments on Regularization Coefficients :

For L1 regularization on 6 degree polynomial, we get λ=1e-16

For L2 regularization on 6 degree polynomial, we get λ =1e-16

2.For testing data:

Comparison between various approaches:

Approach tested on Testing Data	R^2	Error	RMSE	SSE	Regularization coefficient
Gradient Descent on 1 degree	0.0275	64077.26	0.99111	128154.52	-
Gradient Descent on 2 degree	0.056	62221.95	0.976	124443.911	-
Gradient Descent on 3 degree	0.1471	56196.93	0.9281	112393.86	-
Gradient Descent on 4 degree	0.17508	54353.012	0.9128	108706.024	-
Gradient Descent on 5 degree	0.2152	51704.131	0.8902	103408.263	-
L1-Regularization on 6 degree	0.0876	60113.866	0.9599	120227.733	1e-16
L2-Regularization on 6 degree	0.0876	60113.866	0.9599	120227.733	1e-16

Weights corresponding to polynomial of different degrees

1. **Degree 2**

Number of weights = 6

[0.082, 0.0252, -0.147, -0.142, 0.076, -0.039]

2. Degree 3

Number of weights = 10

[0.0425, 0.3405, 0.067, -0.234, 0.123, 0.098, -0.2405 ,-0.19813185 , 0.37579568 ,-0.10670022]

3. Degree 4

Number of weights = 15

[-0.02430904 ,0.21842771, -0.00355097, -0.1379215, 0.2858239 0.17015734, -0.15485221, -0.12422135, 0.15726159, 0.04443806, -0.02917758, -0.04225718, 0.09770351, -0.02514362, -0.1744744]

4. Degree 5

Number of weights = 21

[-0.05969687, 0.30120738, -0.05297262 ,-0.15302335
0.40056281 , 0.24213786 ,-0.21583168 ,-0.22862247
0.2044249, 0.09984751, -0.03440302, -0.07330723
,0.12420305, -0.07912063, -0.21939714 ,-0.00521156
0.04575243, 0.10410096 ,-0.01419789, -0.21923004
, 0.08915316]

5. Degree 6

Number of weights = 28

L2

[0.00689789 , 0.01166721, -0.0008426, -0.00516704, 0.01930913,
0.00544683, 0.00311111, -0.00724565, 0.00793199 , 0.00687061,
-0.00998526, 0.01787344 , 0.0009727, -0.00043947, -0.00334592,
-0.01093142, -0.02153433, -0.00047727, 0.00024188, -0.001468 ,
0.00400684 ,-0.00457427 , 0.0100237, 0.00418119, -0.0127856,
-0.00446561, -0.0146811, -0.00460315]

L1

[0.00689789 , 0.01166721, -0.0008426, -0.00516704, 0.01930913,
0.00544683, 0.00311111, -0.00724565, 0.00793199 , 0.00687061,
-0.00998526, 0.01787344 , 0.0009727, -0.00043947, -0.00334592,
-0.01093142, -0.02153433, -0.00047727, 0.00024188, -0.001468 ,
0.00400684 ,-0.00457427 , 0.0100237, 0.00418119, -0.0127856,
-0.00446561, -0.0146811, -0.00460315]

Formulas Used:

$H = X.W^T + b$

Where X is the modified data and W is the weight of respective degrees

H=Hypothesis = Standardized Predicted Altitude
Predicted Altitude = H*std(training_Y) + mean(training_Y)
b - constant term
$\lambda =$ *Regularization Coefficient*

In case of Gradient descent without regularization :
$Error = 0.5 * [(H - Y)^T . (H - Y)]$

In case of Gradient descent with L1 regularization :

$$Error = 0.5 * [(H - Y)^T . (H - Y)] + 0.5 * \lambda * \sum_{i=1}^{3} |w_i|$$

In case of Gradient descent with L2 regularization :

$$\text{Error} = 0.5 * [(H - Y)^T . (H - Y)] + 0.5 * \lambda * \sum_{i=1}^{3} (w_i)^2$$

$$\mathbf{R^2 =}\ 1 - \frac{SSE}{SST}$$

Where

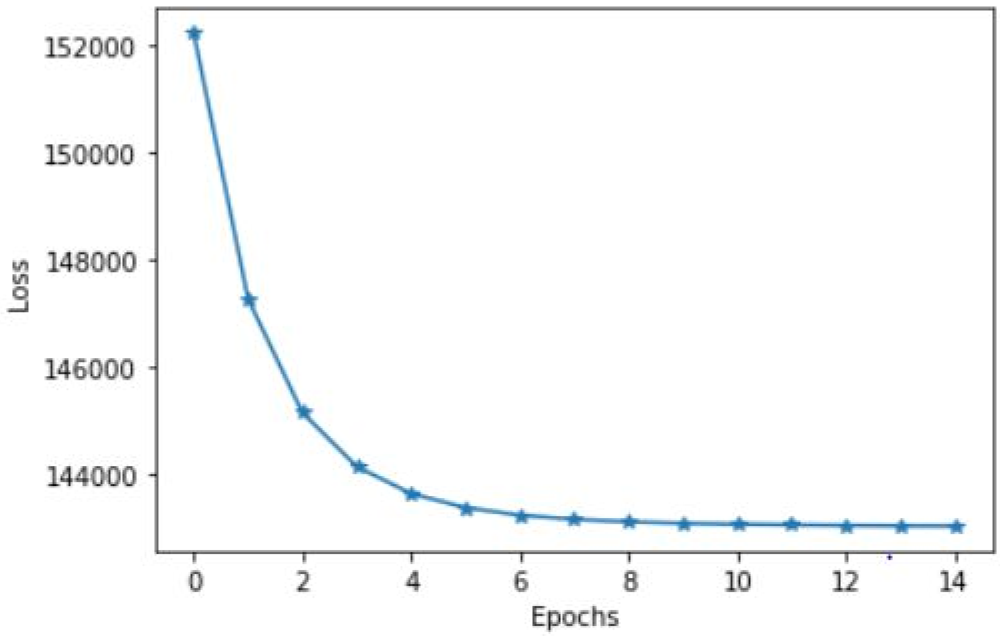
SSE = 2 * Error

$$SST = \sum_{\textit{all data points}} (Y - \bar{Y})^2$$

SST = 304411 (fixed for all the approaches for training data)

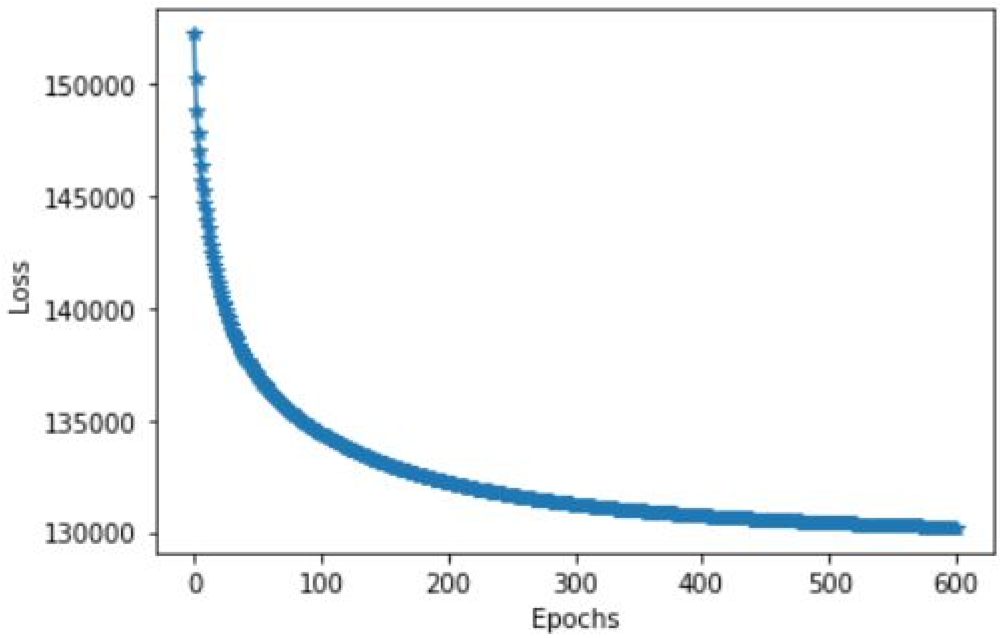
2 degree graph:

Loss vs Epochs -



3 degree graph:

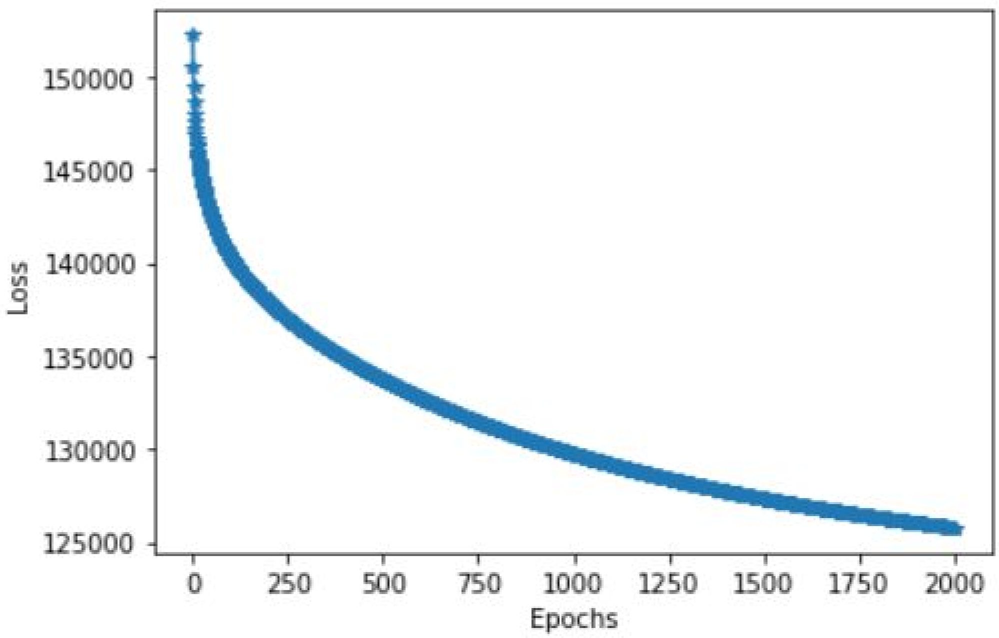
Loss vs Epochs -



1e-07

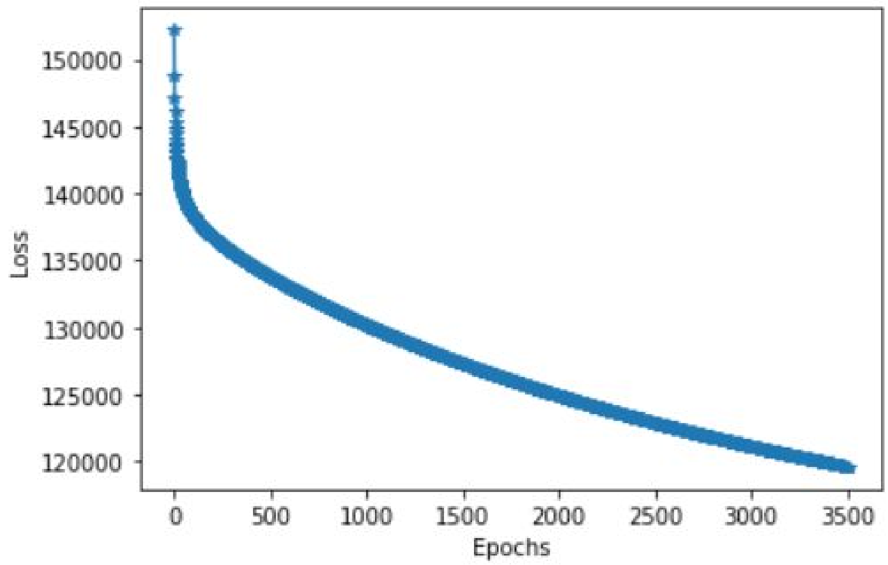
4 degree graph:

Loss vs Epochs -



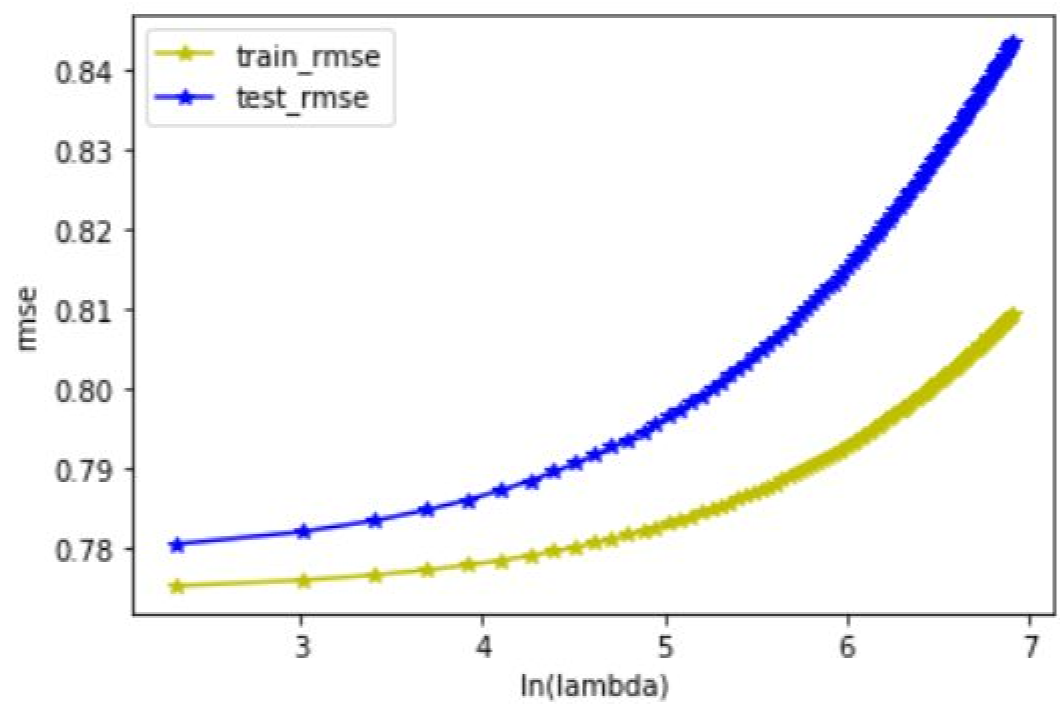
5 degree graph:

Loss vs Epochs -



6 degree graph (L2 regularization):

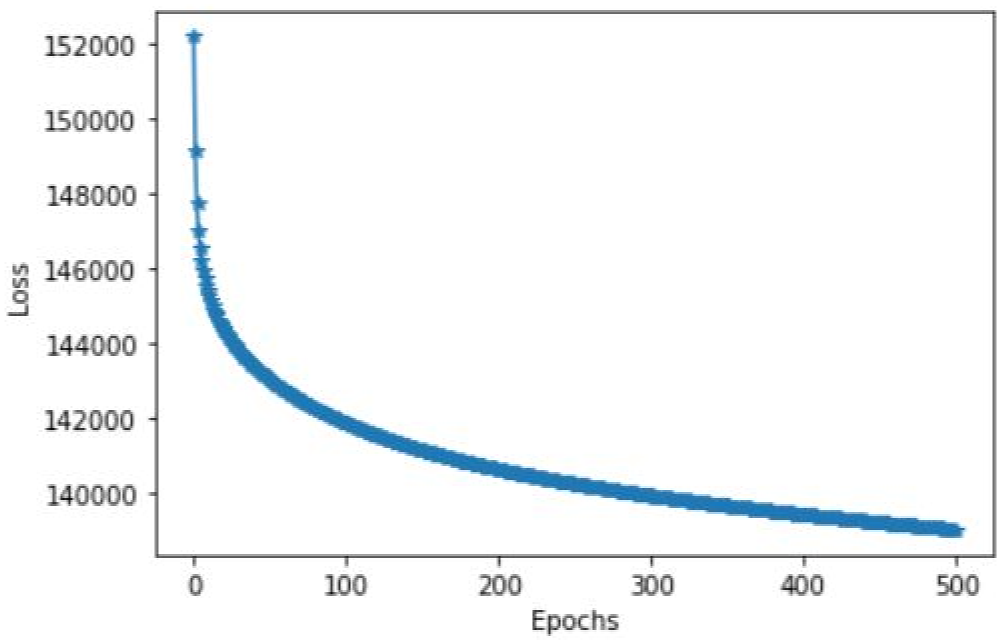
Rmse vs ln(lambda) -



<u>Conclusion</u>

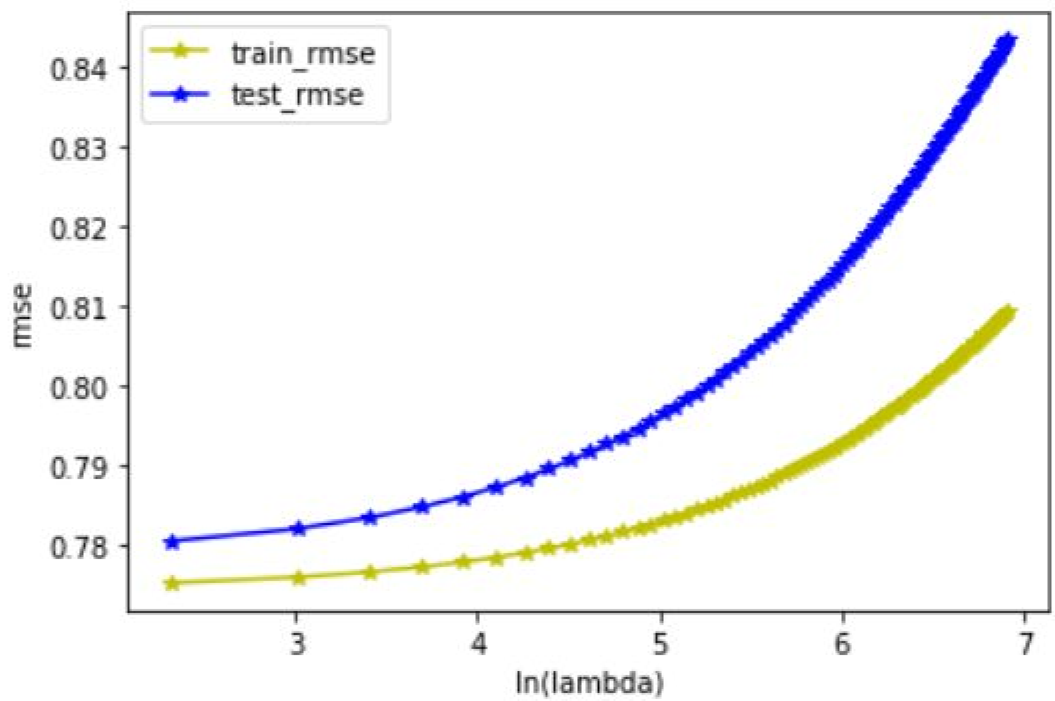
As there is no overfitting we take $\lambda = 0$. Also, at $\lambda = 0$ there is low bias and low variance between training and testing data .

Loss vs Epochs -



6 degree graph (L1 regularization) :

Rmse vs ln(lambda) -



<u>Conclusion</u>

As there is no overfitting we take $\lambda = 0$. Also, at $\lambda = 0$ there is low bias and low variance between training and testing data .

Loss vs Epochs -