Foundations of Data Science (CS F320)
Assignment - 2

TEAM MEMBERS:

1. **Shreeya Nelakar 2017A7PS0093H**
2. **Prateek Agarwal 2017A7PS0074H**
3. **Shriya Choudhary 2017AAPS0409H**

1. Dataset Information:

The dataset used on which our model was fit was the 3-d Road Network, where the latitude, longitude and altitude values were recorded from 434874 different points. Models were built using linear regression based on different types of descent algorithms.

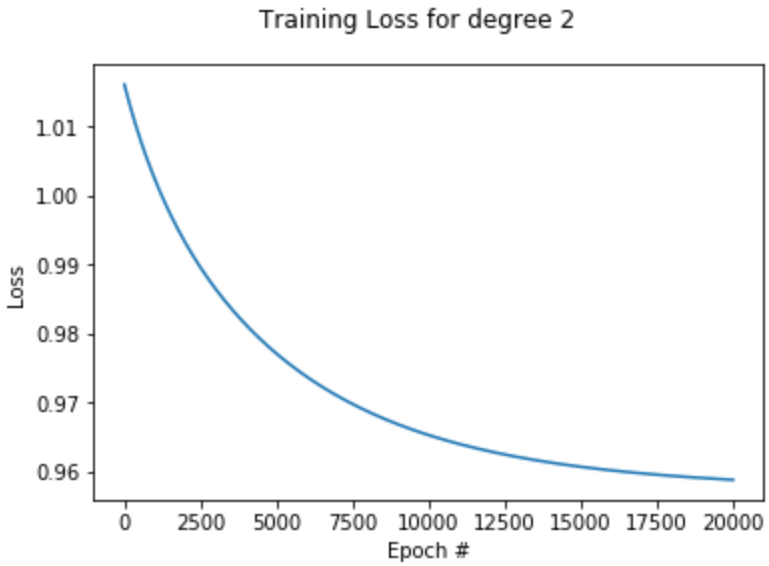
Attributes:

1. Latitude
2. Longitude
3. Altitude

Dataset characteristics	Sequential	**Number of Instances:**	434874
Attribute Characteristics:	Real	**Number of ATtributes:**	3

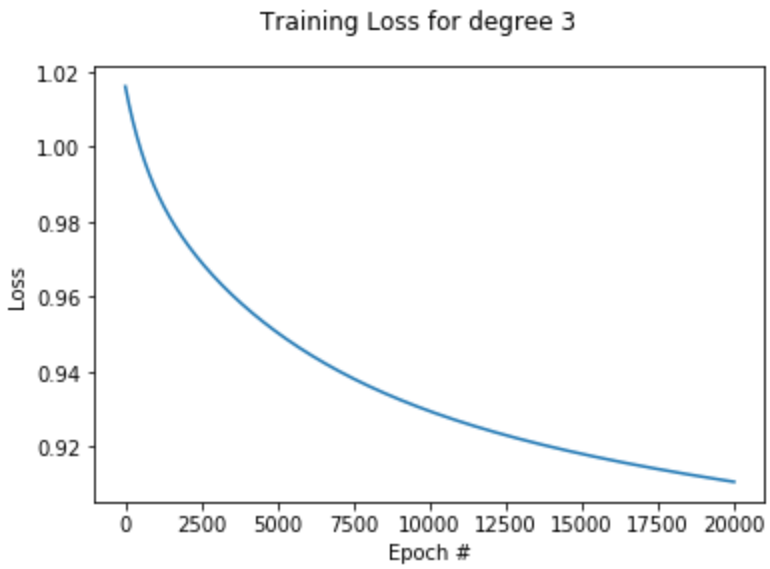
2 Degree:

1. Initialization for the weights: [0,0,0]
2. Learning rate: 0.0001
3. Stopping criteria:Iteration number =20000
4. RMSE value for Degree 2 is for training Data : 0.9791608053464746
5. R Sq value for Degree 2 is for training Data : 0.05625613953956077
6. RMSE value for Degree 2 is for test Data : 0.8590821666010684
7. R Sq value for Degree 2 is for test Data : 0.06177193998267594



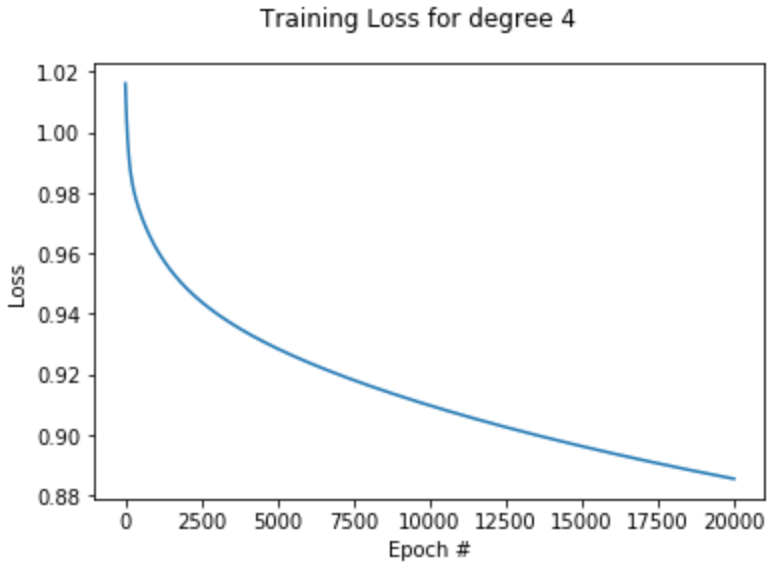
3 Degree:

1. Initialization for the weights: [0,0,0]
2. Learning rate: 0.0001
3. Stopping criteria:Iteration number =20000
4. RMSE value for Degree 3 is for training Data : 0.9542647775613271
5. R Sq value for Degree 3 is for training Data : 0.10362810976655412
6. RMSE value for Degree 3 is for test Data : 0.8246377411214945
7. R Sq value for Degree 3 is for test Data : 0.1353999620642098



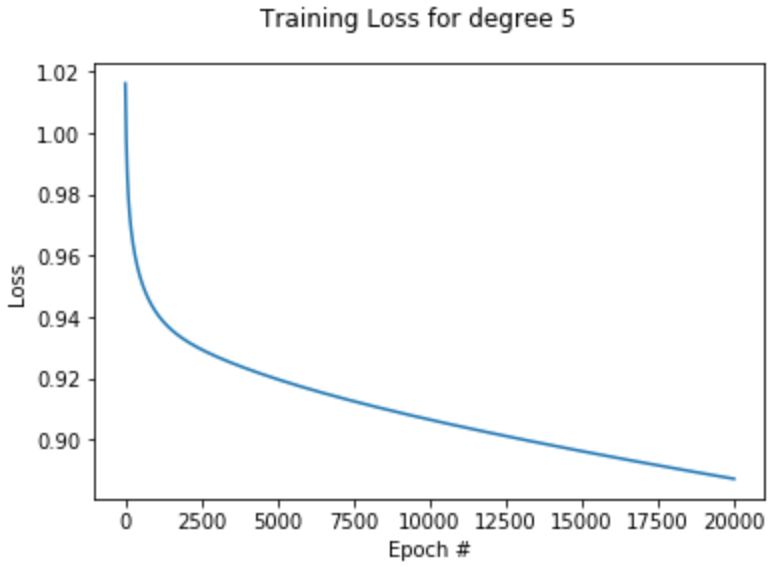
4 Degree:

1. Initialization for the weights: [0,0,0]
2. Learning rate: 0.0001
3. Stopping criteria:Iteration number =20000
4. RMSE value for Degree 4 is for training Data :0.9410256031230492
5. R Sq value for Degree 4 is for training Data : 0.12831665323825414
6. RMSE value for Degree 4 is for test Data : 0.8154343647919028
7. R Sq value for Degree 4 is for test Data : 0.15446975717295952



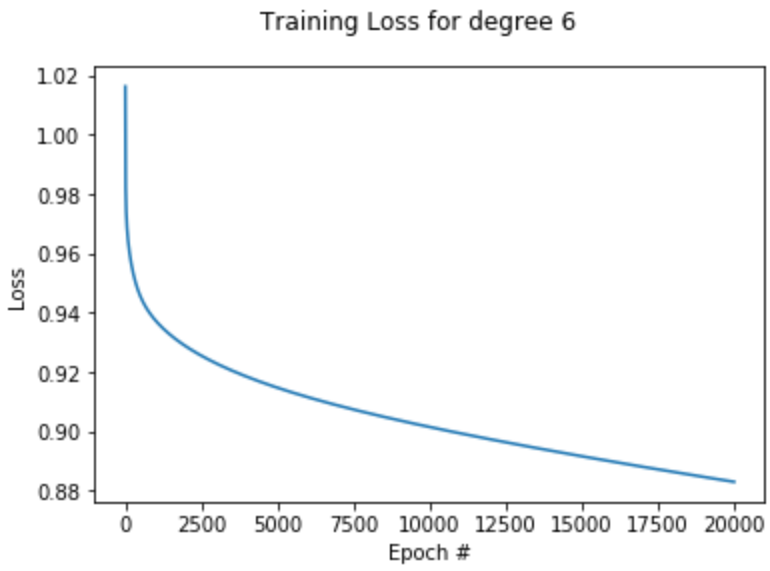
5 Degree:

1. Initialization for the weights: [0,0,0]
2. Learning rate: 0.0001
3. Stopping criteria:Iteration number =20000
4. RMSE value for Degree 5 is for training Data: 0.9418529979470314
5. R Sq value for Degree 5 is for training Data : 0.12677002881070076
6. RMSE value for Degree 5 is for test Data : 0.8180624235100478
7. R Sq value for Degree 5 is for test Data : 0.1488643621585548



6 Degree:

1. Initialization for the weights: [0,0,0]
2. Learning rate: 0.0001
3. Stopping criteria:Iteration number =20000
4. RMSE value for Degree 6 is for training Data: 0.9396766855641743
5. R Sq value for Degree 6 is for training Data: 0.13078564866691222
6. RMSE value for Degree 6 is for test Data : 0.817743336512959
7. R Sq value for Degree 6 is for test Data : 0.14935735623194213

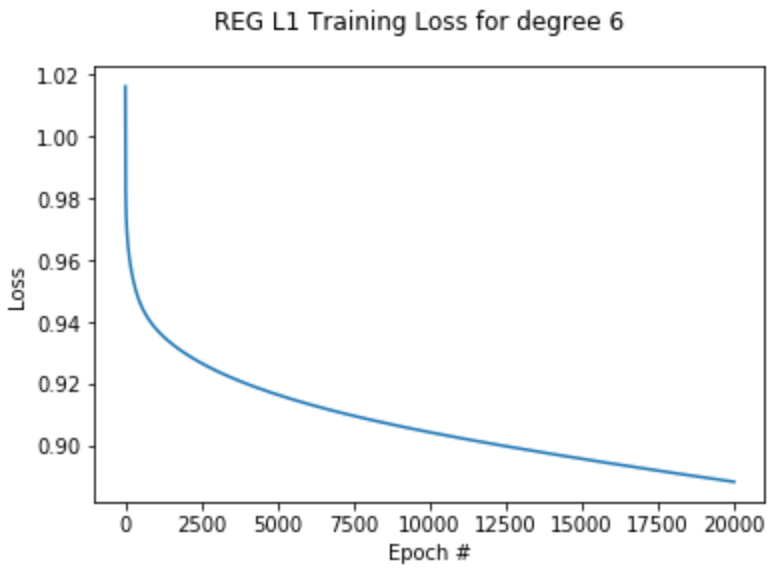


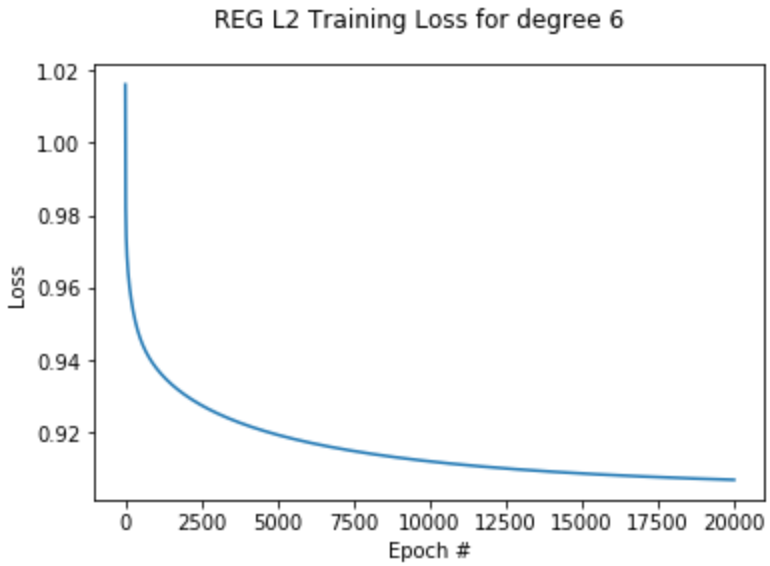
Comparison:

1. 6th Degree with Regularization:

Regularization	Training MSE	Testing MSE
L1 (0.0001)	0.94245	0.8191488
L2 (0.0001)	0.9523000	0.823556

Parameters	Training RMSE	Testing RMSE
With Regularization:	0.94245	0.8191488
Without Regularization:	0.939676	0.817743





Observations:

1. There is no overfitting for degree 6 polynomial regression, hence there isn't much variation in RMSE after implementation of regularization on the model.
2. Regularization sets constraints on the coefficients in such a way that it shrinks the coefficient estimates towards zero. Overfitting can mostly be seen in nonlinear models.