ASSIGNMENT -1

Rohan Panda - 2017A3PS0487H
Rayyan Shaikh - 2017A3PS0351H

Introduction:

In the following assignment, linear models were built to fit a certain dataset. 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. Namely:

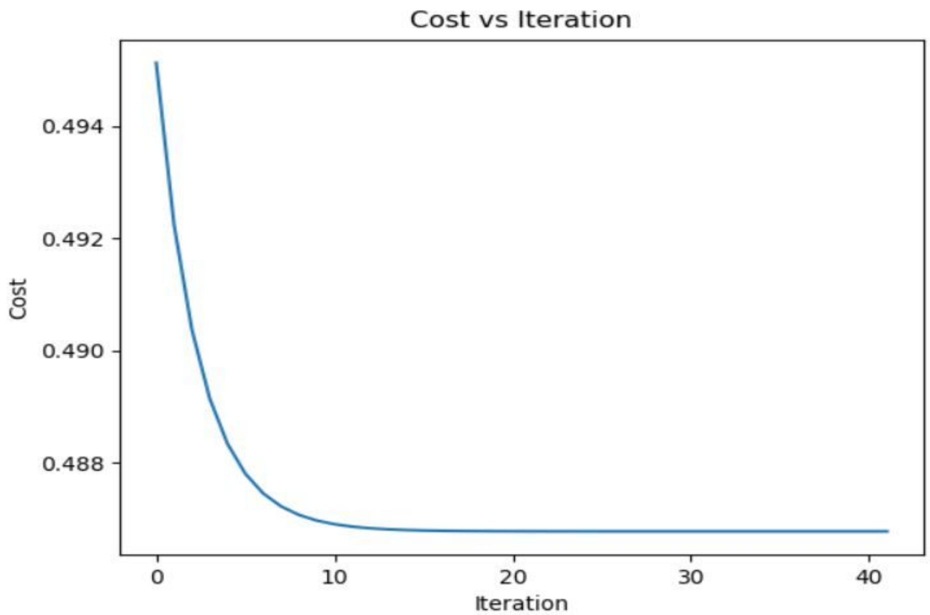
1. Batch Gradient Descent
2. Stochastic Gradient Descent
3. Gradient Descent with L1/L2 norm
4. Normal Equations method

Since the data consisted of values holding high magnitudes, the data was first normalized such that it has a mean = 0 and standard deviation = 1. The performance of all the models were accounted for on the basis of R2 and RMSE values.

A) Batch Gradient Descent:

The following method was used to fit a regression model on the complete data. The model met the given stopping criteria after 41 iterations hence the graph is plotted for the same. The early stopping of the model's descent can be based on the normalization of the data.

Graph



Results:

```
W values  [[ 1.07112712e-14]
 [ 1.50355015e-01]
 [-1.90600964e-01]]
rsquare: [0.02644573]
RMSE 0.6976913398079841
```

No. of iterations taken: 1000

Learning Rate(Alpha): 1e-6

R2: 0.026445

RMSE: 0.697691

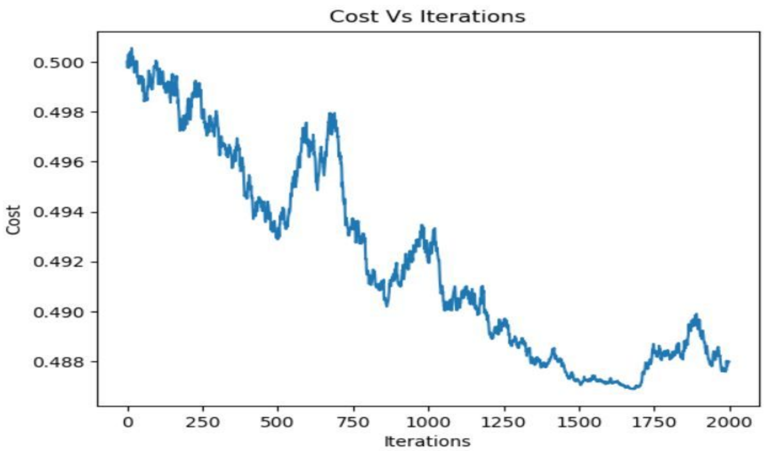
Stopping condition : epsilon < 2e-10

Where “epsilon” is the difference between two consecutive cost values.

B) **Stochastic Gradient Descent:**

Following batch gradient descent, stochastic gradient descent algorithm was studied. The algorithm is found to converge at a faster speed as each iteration takes very less computation time as compared to the previous method. The graph of the descent is seen to be noisy owing to the method in which the descent works. The descent was stopped when the R2 and RMSE values were comparatively significant to that of part A.

Graph:



Results:

```
W values  [[ 0.0423648 ]
 [ 0.11955695]
 [-0.17524659]]
rsquare: [0.02304659]
RMSE 0.6985665374719897
```

No. of iterations taken: 2000

Learning Rate(Alpha): 1e-6

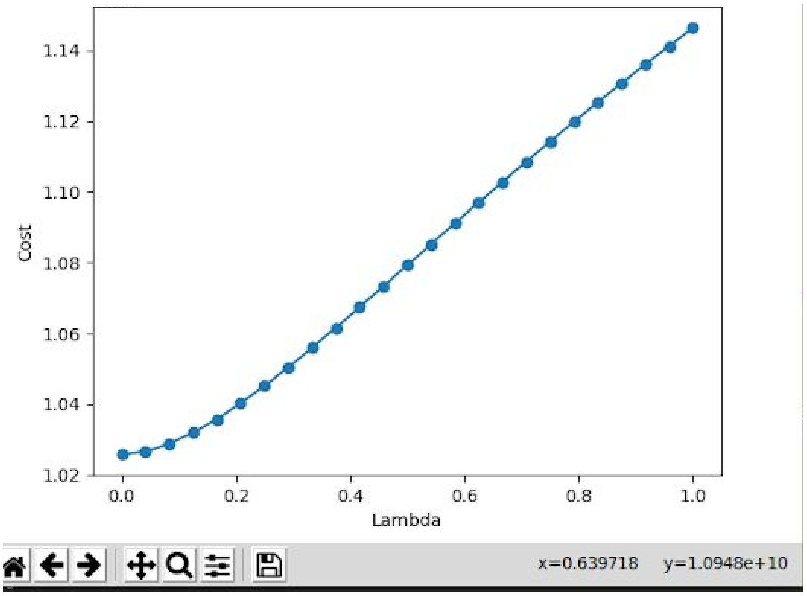
R-Squared: 0.023046

RMSE: 0.698566

C) <u>L2 Norm Regularization:</u>

Norm regularization techniques prove to be vital in tackling overfitting related issues. However, from our observations it is seen that the models do not overfit on the given data and increasing the norm parameters causes the cost function to increase as the weights tend to become very small in magnitude.

Graph: Validation Loss Vs Regularization Coefficient(Lambda)



Results:

```
W values  [[ 6.80296353e-15]
 [ 1.50354635e-01]
 [-1.90600558e-01]]
rsquare: [0.02644561]
RMSE 0.697691360932417
```

No. of iterations taken: 2000

Learning Rate(Alpha): 1e-6

Regularization Constant(Lambda): 0.01

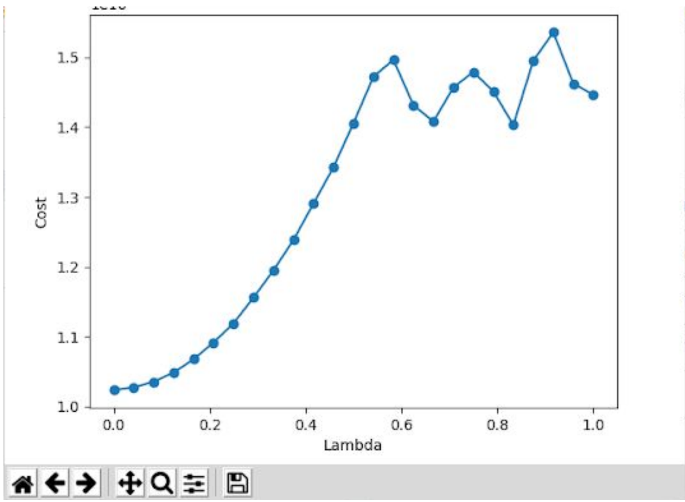
R-Squared: 0.026445

RMSE: 0.697691

L1 Norm Regularization:

This model is very similar to L2 regularized model as here too we constraint the model. Here the regularization factor is the sum of absolute values of the weights.

Graph: Validation Loss Vs Regularization Coefficient(Lambda)



Results:

```
W values  [[-0.00082023]
 [ 0.1482041 ]
 [-0.19075178]]
rsquare: [0.02629604]
RMSE 0.6975397990561742
```

No. of iterations taken: 2000

Learning Rate(Alpha): 1e-6

Regularization Constant(Lambda): 0.01

Test R-Squared: 0.026296

Test RMSE: 0.697539

D) **<u>Normal:</u>**

This is the most accurate but the most computationally expensive method for very large datasets with lots of features. The process uses various matrix manipulation techniques that solve the normal equations. This method is the most accurate as it gives us the exact minimum rather than approximations like the other methods do, and hence the minimum error compared to all the other methods.

Results:

```
rsquare: [0.02645359]
RMSE 0.6976913395883109
[[ 6.98613907e-15]
 [ 1.50381617e-01]
 [-1.90627566e-01]]
```

Test R-Squared: 0.026453

Test RMSE: 0.697691

Comparison Of the Models

Model	R-Square	RMSE
Batch Gradient Descent	0.026445	0.697691
Stochastic Gradient Descent	0.023046	0.698566
L2 Norm Regularization	0.026445	0.697691
L1 Norm Regularization	0.026296	0.697539
Normal	0.026453	0.697691

As we see all the models perform reasonably well and have close values of R2 and RMSE. As noted, Normal equations provide us with the best results as they do not depend on the descent algorithm and rather solve for the weights of the model from solving the equations.

L1 and L2 regularization both perform in the same fashion by increasing the RMSE value as the model is not overfitting the data, these regularizations cause the weights to become very small which causes the high error. Hence L1 and L2 regularization would not be needed in this dataset.