# **Assignment 1**

By-

Drishti Mamtani(20165A70574H)

Shreyansh Jain(2016B1A30947H)

Shreya Saxena(2016A4PS0367H)

**Part A : Gradient Descent Algorithm**

The weights were initialised by selecting three values from standard normal distribution randomly.

Number of iterations = 4000 is taken as the stopping criteria

Learning rate =1e-2

Loss for train data:

Mean Sum of Squared Error=338.32

R2 : 0.0269

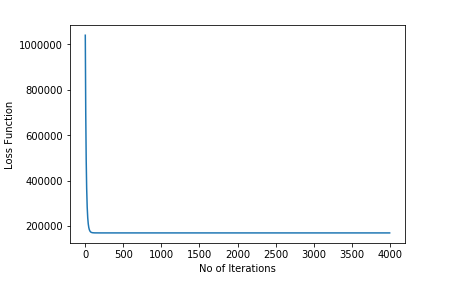
RMSE : 22.985

Loss for test data :

Mean Sum of Squared Error= 528.64

R2 : 0.0287

RMSE :18.393



**Part B : Stochastic Gradient Descent Algorithm**

The model was run for 500 iterations, Training loss was recorded for every 20 iterations, As expected, it be seen in the graph the training losses reduce as the number of iterations increase and the step sizes also reduce as we approach the optimal coefficients.

MSE: 0.5026659970441093

RMSE: 21.708989419557238

R2 :0.02585

initial w0 : -0.8021728386486427

initial w1 : -0.4488778077207725

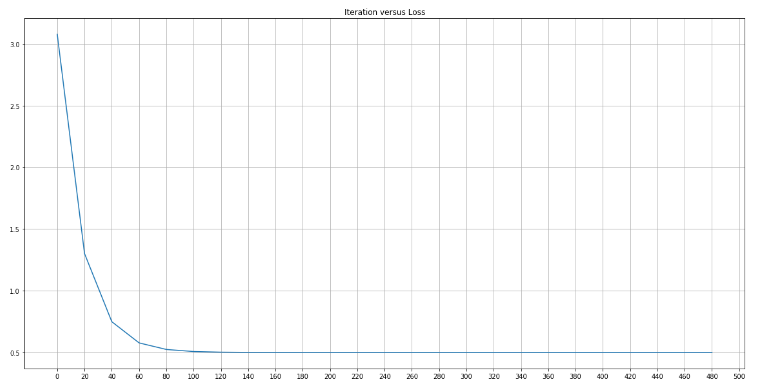
initial w2 : -1.1059350760083153

final w0 : -0.04721818152926071

final w1: 0.2990740926760206

final w2: -0.27615087584477704

In the graph below, X-axis represents number of Iterations while Y-axis represents Loss

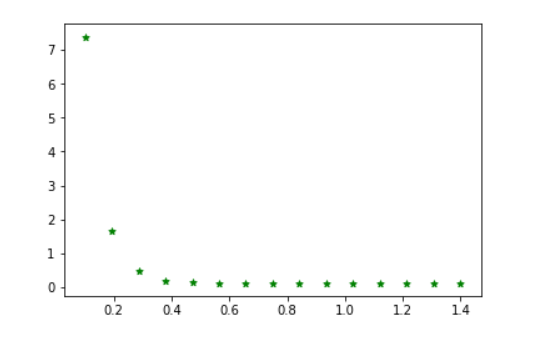


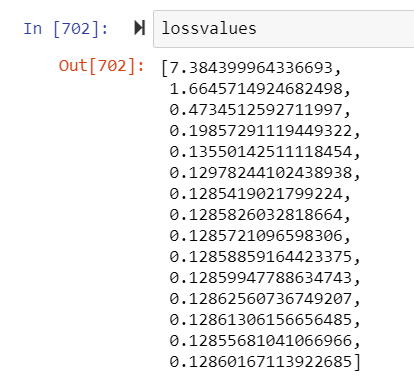
**Part C - Gradient Descent Method along with regularization**

**L1 Regularization**

Initially, w0,w1, and w2 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 1.4. Hence the trend is that as the value of Lambda increases Validation Loss decreases. This means more focus is on reducing the coefficients.

In the graph below, X-axis represents Lambda values while Y-axis represents Validation Loss



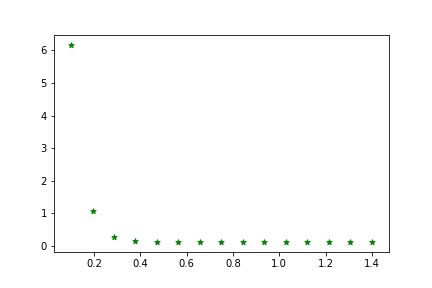


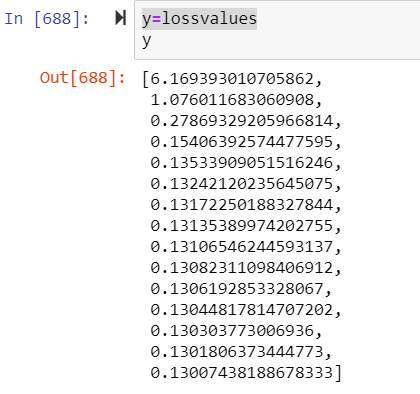
This shows that loss is least for **lambda=1.3**. The same is depicted via a graph. RMSE value came out to be 19.99988759438357. R2  value=0.0224798

**L2 Regularization**

Here also the same procedure was followed like LI Regularization.

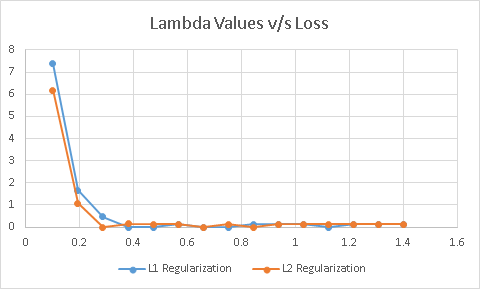
Initially, w0,w1, and w2 all were taken as 10 and the regularization coefficient was varied from 0.1 to 1.4. The dataset was divided into 3 parts Training Data, Cross-Validation Data, and Testing Data. Cross-Validation Data was used to find the best Regularization coefficient and for L2 regularization the best Lambda value came out to be 1.4. Hence the trend is that as the value of Lambda increases Validation Loss decreases. This means more focus is on reducing the coefficients

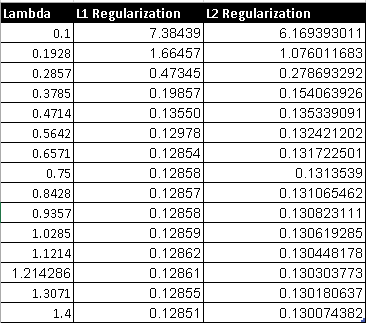




This shows that loss is least for lambda=1.4. The same is depicted via a graph. RMSE value came out to be 19.993942151311575. R2  value=0.0232

The key difference between these two is the penalty term.





As we know, the losses by both the Regularization techniques should be similar, This is thus verified by the above Results wherein when both the regularization techniques are applied for the same Regularization coefficient values the validation losses come to be similar.

**Note:** This is only possible when the model is run for a sufficient number of iterations.

**Part D : Normal Method**

The normal method gives RMSE value of 18.315739045043244 and R2 value of 0.0300021.

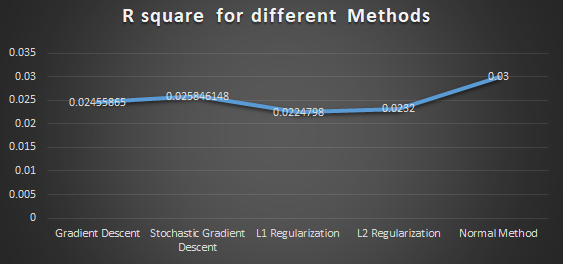
When the dataset is small, normal is considered as the best method as it gives the exact answers and not the approximations. But here the dataset is large and hence all other methods are supposed to give decent approximations.

Weights obtained are [681.34825732 4.52127649 -12.31770487]

# **Analysis**

**R2  for Different Models**

|  |  |
| --- | --- |
| **Method** | **R square** |
| Gradient Descent | 0.02456 |
| Stochastic Gradient Descent | 0.02585 |
| L1 Regularization | 0.02248 |
| L2 Regularization | 0.0232 |
| Normal Method | 0.03 |

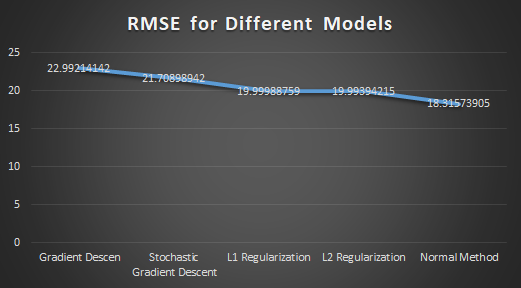


From the above graph we can clearly observe the following things -

1. As expected the R2 value of the Stochastic Model is greater than normal Gradient Descent.
2. L1 and L2 Regularization have similar R2 values.
3. The R2 value signify the variation described by the model.

|  |  |
| --- | --- |
| **Method** | **RMSE** |
| Gradient Descent | 22.9921 |
| Stochastic Gradient Descent | 21.709 |
| L1 Regularization | 19.9999 |
| L2 Regularization | 19.9939 |
| Normal Method | 18.3157 |

**RMSE for Different Models**



From the above graph, we observe the following properties-

1. Stochastic Descent has lesser RMSE as compared to the Gradient Descent method, thus has more accurate results.
2. L1 and L2 Regularizations have very similar RMSE which verifies the theory.
3. Normal Method has the least RMSE, this is because where other methods predict approximate values, the Normal method shows exact values .
4. Gradient Descent,Stochastic Gradient Descent,L1 Regularization, L2 Regularization tend to move near the approximate solution but don’t reach the stationary point. Whereas Normal method gives us the stationary point.
5. Also as there is not much difference in order of RMSE values for all methods, hence all methods tend to give us good models. This is only possible because the dataset is very large.