Assignment 1

BY

Ravisanker E
2017A7PS0433H

S Ankit
2017A7PS0297H

CONTENTS

Part A

Gradient Descent degree 1

Before performing the gradient descent, the given data was standard normalized and spliced into 80 percent training and 20 percent testing set respectively.

Degree 1 polynomial used was:

$$y = w_0 + w_1 x_1 + w_2 x_2$$

X1 and x2 values are normalized in this equation.

The weights were initialized with random values between 0-1

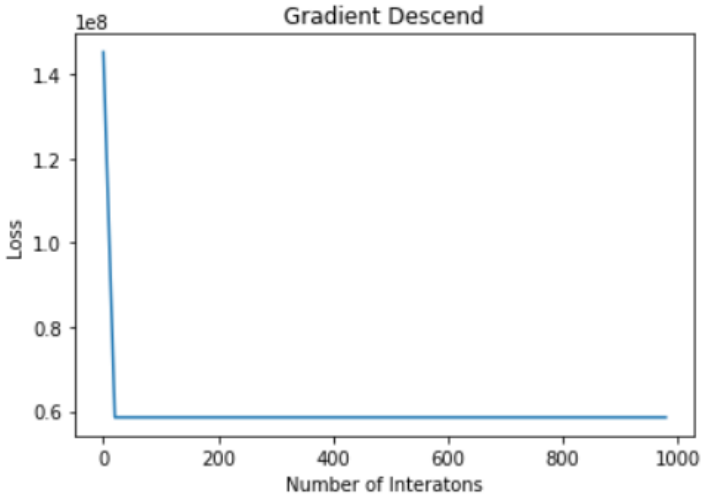
The learning rate α was taken as 3×10^{-6}

Model created after 2000 iterations of gradient descent

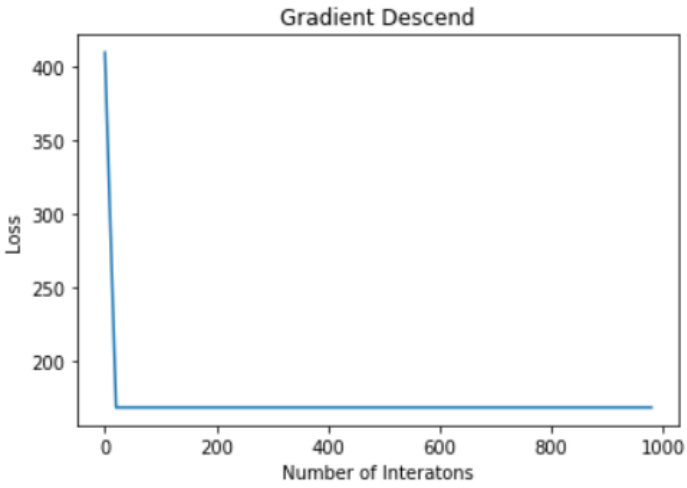
$$y = 22.17800962 + 2.79440781 x_1 - 3.5682518 x_2$$

With mean squared error = 169.27

Loss vs number of iterations for every 20 iterations on the training set.



Loss function considered is half sum of Squares



Loss function considered is Mean Squared Error

On the testing set

RMSE = 18.4698

R2 = 0.02540

Part B

Stochastic Gradient Descent degree 1

The weights were initialized with random values between 0-1

The learning rate α was taken as 3×10^{-6}

The model obtained was almost the same as that of gradient descent

$$y = 22.1704988 + 2.75735116x_1 - 3.53390022x_2$$

Half sum of squares = 58541239.2574

Mean Squared error = 168.27127

On the testing set

RMSE = 18.4698

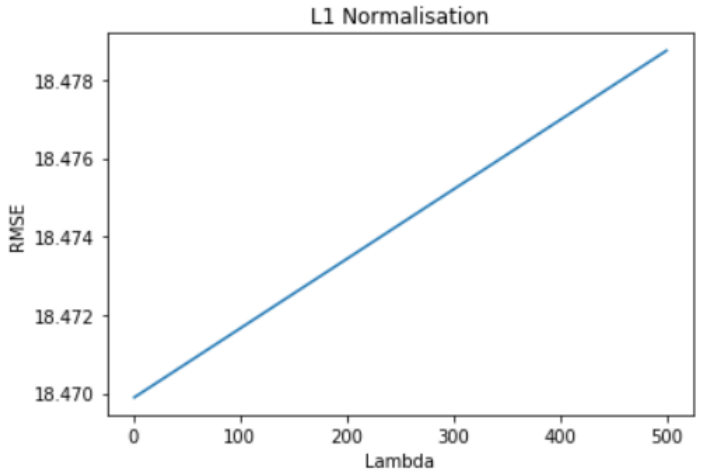
R2 = 0.02540

Part C

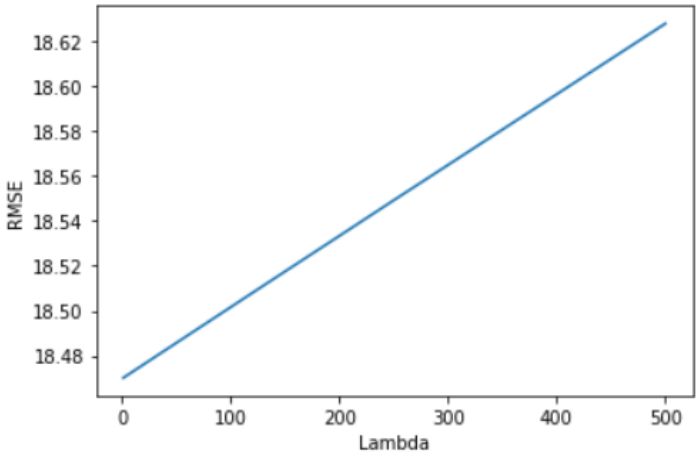
L1 and L2 regularisation.

To compute an optimum value for λ a binary search was performed between 0 and 1000 for both L1 and L2 regularization and simultaneously models were developed for the given λ and the RMSE for the validation set was calculated.

Below are the plots of RMSE on validation set vs λ.



L2 Normalisation.



In both L1 and L2 normalisation it was observed that λ was tending towards zero for lower values of RMSE.

No difference was observed for λ calculated for L1 and L2, λ

Part D

Normal Equations

Using normal equations, the model obtained is as follows:

$$y = 22.17800962 + 2.79440781x_1 - 3.5682518x_2$$

On the Validation set

RMSE = 18.4698

R2 = 0.02540

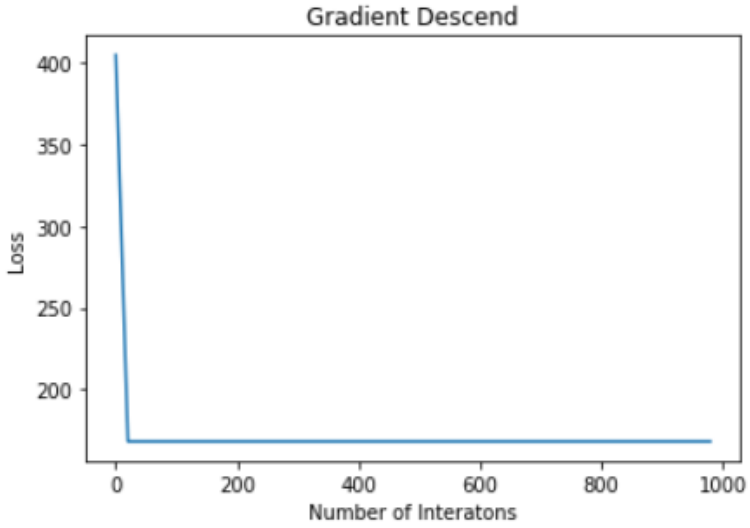
Comparison of the models developed:

Model	RMSE on Validation Set	R2 on Validation Set
Gradient Descent degree 1	18.4698	0.02540
Stochastic Gradient Descent	18.4698	0.02540
L1 Regularisation $\lambda=0$	18.4698	0.02540
L2 Regularisation $\lambda=0$	18.4698	0.02540
Normal Equations	18.4698	0.02540

Inferences:

- For all the models RMSE and R^2 were perfectly matching with the ones computed using Normal equations. For the provided datasets under the conditions specified all models provide similar performance though it was expected for the model formed from the normal equations to perform slightly better than the rest.
- We observe that $\lambda=0$ for both L1 and L2 regularisation. Both the models give the same validation metrics (R^2 and RMSE) as the model obtained from Normal Equations, so we infer that the linear model does not overfit the dataset.
- R^2 validation metric comes out as 0.02540 suggests that the Linear model is not best suited to the provided dataset

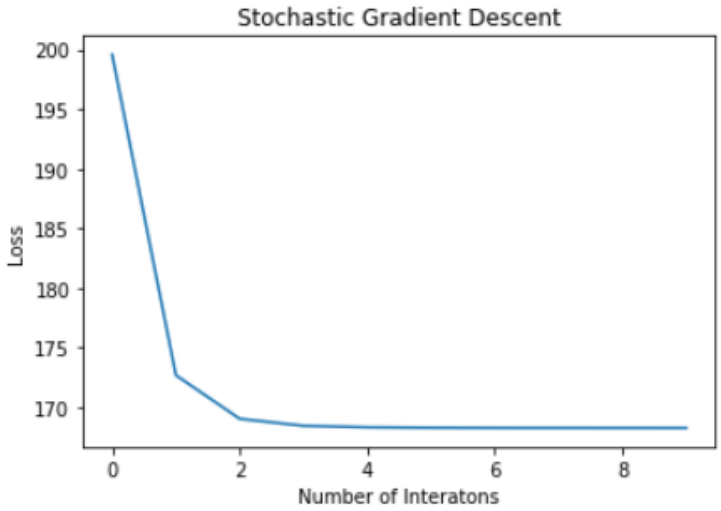
Screenshots of results:




```
Model:
 [[22.17800962]
 [ 2.79440781]
 [-3.5682518 ]]

On the Test Data

 RMSE:  18.46987496271496
 R2:  0.025401222950555913
```




```
Model:
 [[22.17050229]
 [ 2.75795041]
 [-3.53449896]]
sum of squares: 58541232.85610572
Mean squared error: 168.271254379461

On the Test Data

 RMSE:  18.46987810963661
 R2:  0.025400890843678137
```

```
Part-D
Normal Equations:
Model:
 [[22.17800962]
 [ 2.79440781]
 [-3.5682518 ]]

On the Test Data

 RMSE:  18.46987496271496
 R2:  0.025401222950555913
```