SMAI
ASSIGNMENT 7
PRAKASH NATH JHA
2018201013

Question 1:

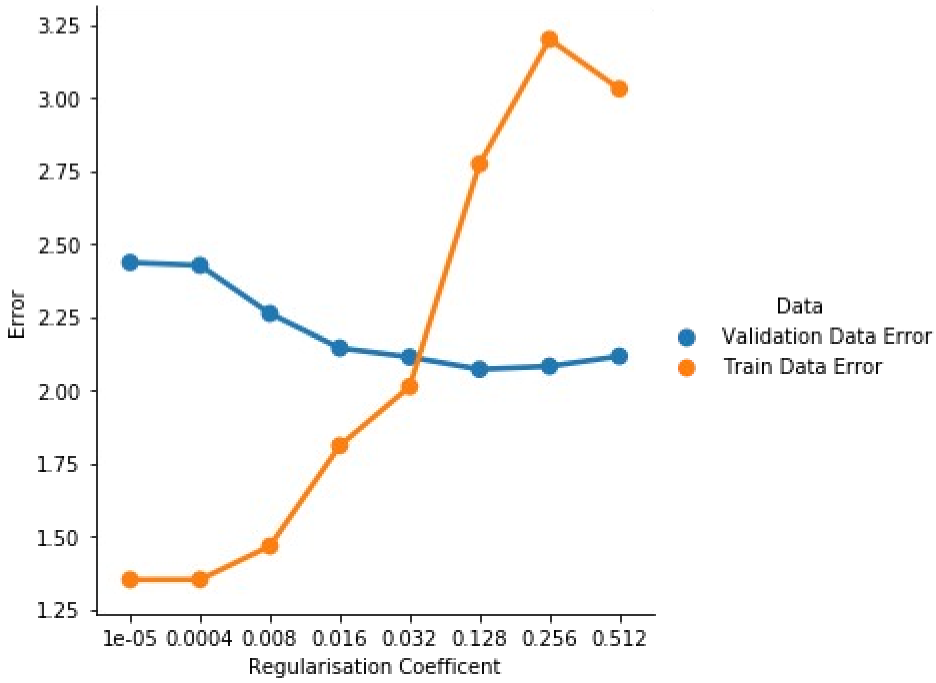
Linear regression using Gradient Descent with L1 regularization

Hyper Parameters:

Learning Rate: 0.01
Error Tolerance: 1e-09
Iterations: 10000
Regularization Constant: 0.01
Type of regression used: L1 (Lasso Regression)

Result:

R2 Score: 0.7732511125845698
Mean Square Error: 2.4279934952760516
Mean Absolute Error: 11.896316219195295
Mean Percentage Error: -2.016526990373679



From the above graph we can observe that with small value of regularization constant model suffers from high variance and with increasing value of regularization constant model suffers from high bias.The optimal value of regularization is around 0.32

Question 2:

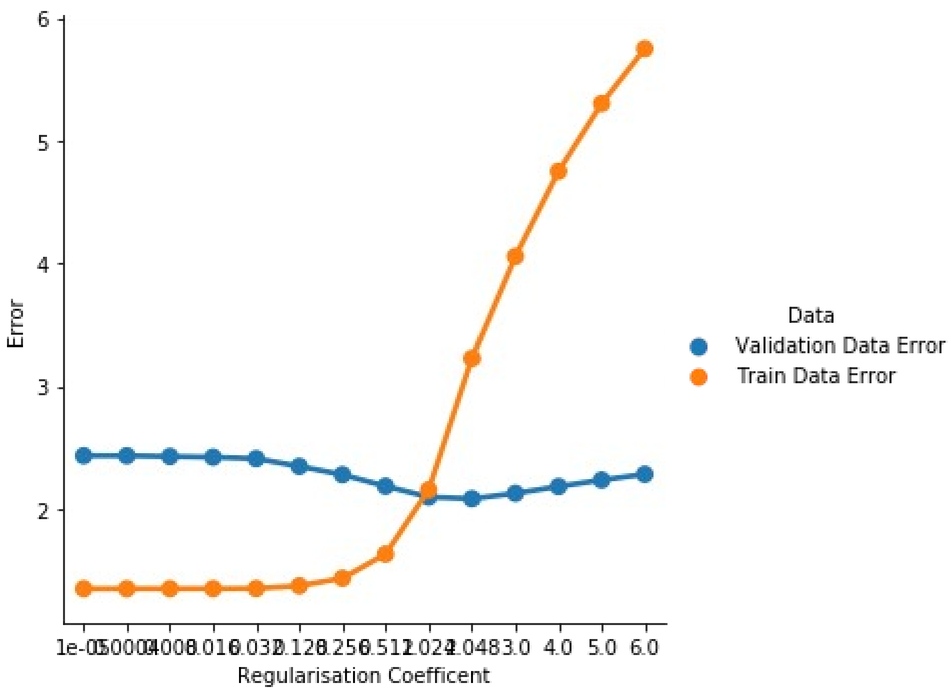
Linear regression using Gradient Descent with L2 regularization

Hyper Parameters:

Learning Rate: 0.01
Error Tolerance: 1e-09
Iterations: 10000
Regularization Constant: 0.01
Type of regression used: L2 (Ridge Regression)

Result:

R2 Score: 0.7732955450687018
Mean Square Error: 0.0032450195796822905
Mean Absolute Error: 0.04199650727761043
Mean Percentage Error: 0.0009783869770144547



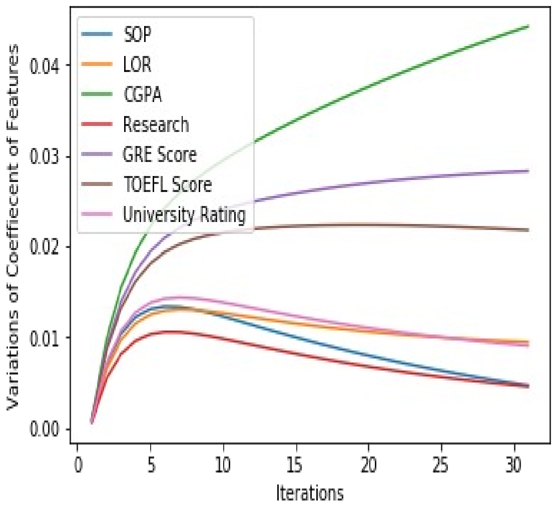
From the above graph we can observe that with increasing value of regularization coefficent model performs poorly on train dataset itself which is a clear sign of high bias. With low value of regularization the model overfits the training data and performs badly on validation data. The optimal value of regularization is around 0.512 to 1.024.

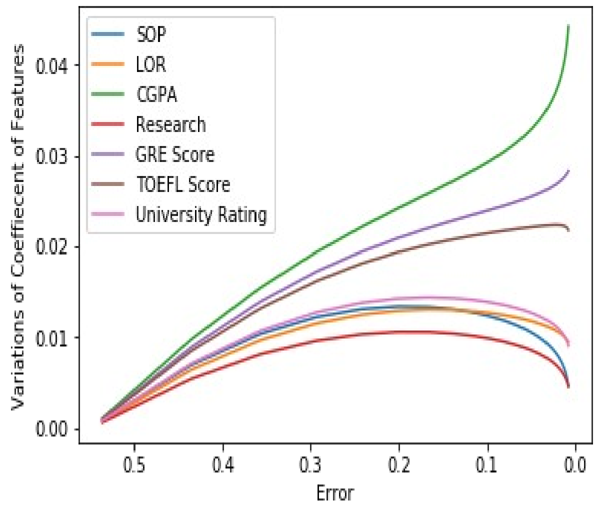
Question 3:

From the plots presented in above questions (1,2), we can see that with increasing value of λ model performs worse on train dataset which is a clear indication of underfitting i.e. having high bias in the model. Moreover we can also observe that with smaller value of λ we can see that model perform very well on train data but performs poorly in validation data which is a clear indication of overfitting on training dataset where model is also trying to fit the inherent noise in training dataset and hence its not generalising well on unseen dataset. Our goal is to minimize both bias and variance and by choosing appropriate value of λ we can perform a sort of tradeoff such that both minimized at the same time.

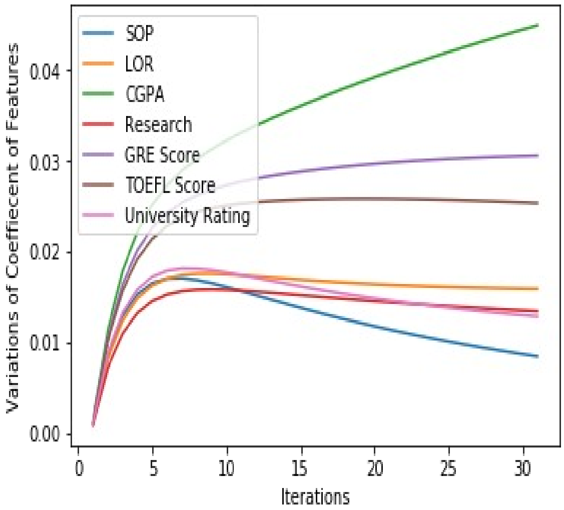
Question 4:

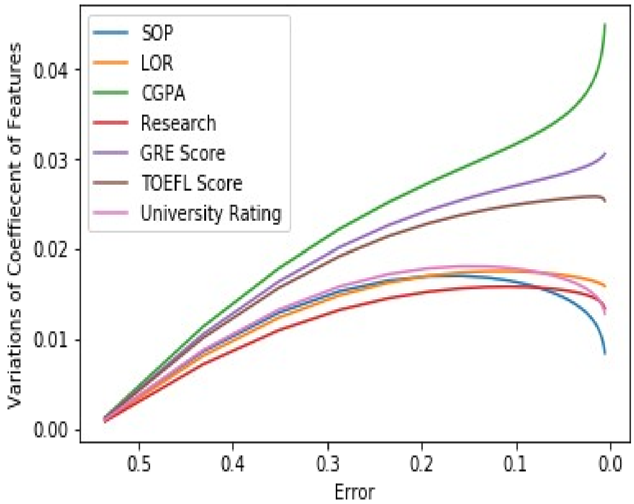
Using L1 regularization:





Using L2 regularization:





From above plots we can observe that with L1 regularization parameters/weights of features tends to more towards 0 as compared to L2 regularization. Both methods produce similar results in terms of feature relationship with the output variable.

Question 5:

K fold cross validation:

Hyper Parameters:

Learning Rate: 0.01
Error Tolerance: 1e-09
Iterations: 10000
Regularization Constant: 0.01
Type of regression used: L2 (Ridge Regression)

Result:

R2 Score: 0.7721875651575957
Mean Square Error: 2.4354535937995396
Mean Absolute Error: 11.910530934913696
Mean Percentage Error: -2.017220194771312

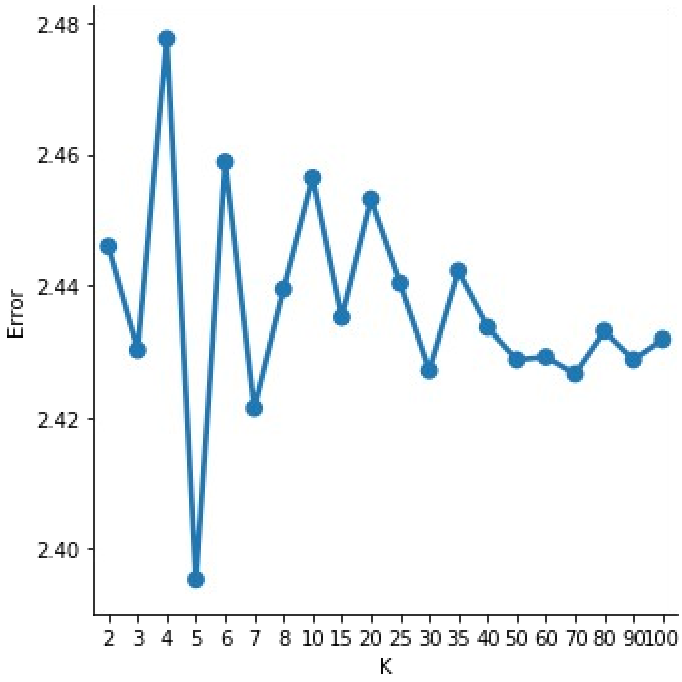
Leave one out cross validation:

Hyper Parameters:

Learning Rate: 0.01
Error Tolerance: 1e-09
Iterations: 10000
Regularization Constant: 0.01
Type of regression used: L2 (Ridge Regression)

Result:

R2 Score: 0.7732509240139638
Mean Square Error: 2.427991675757996
Mean Absolute Error: 11.896311051271303
Mean Percentage Error: -2.016552284751225



We can observe from the above graph that error is minimum at k = 5 and at at all other values of k error is more or less same.