CS725 Assignment-1 Report

IIT Bombay

Submitted By:

1. Akash Kumar (213050020)
2. Prince Gogayan (213050022)
3. Ayush Kumar Tripathi (213050040)
4. Ashish Verma (213050058)
5. Manoj Kumar Maurya (213050067)

1. Solving Linear Regression:

We are given two methods to compute mean squared error (MSE) by Implementing the (A) closed form/analytical solution and (B) Gradient Descent iterative solution

The following MSE losses on the development set are achieved on computation: -

a) Analytical Solution:

37426.934074995406

b) Gradient Descent Solution:

42152.7367114286

2. Gradient descent stopping criteria:

Here we are using two constraints to stopping the gradient descent iteration for convergence

i) Predefined number steps:

We are taking 1,00,000 steps to iterate

ii) and error is less than hyper-parameter(β):

when the difference of two successive error is less than hyper-parameter, we will stop our iterations.

NOTE: Here value of hyper-parameter is **β=0.8**

Early Stopping Criteria:

Early stopping is a form of regularization used to avoid **overfitting** when training a learner with an iterative method, such as gradient descent

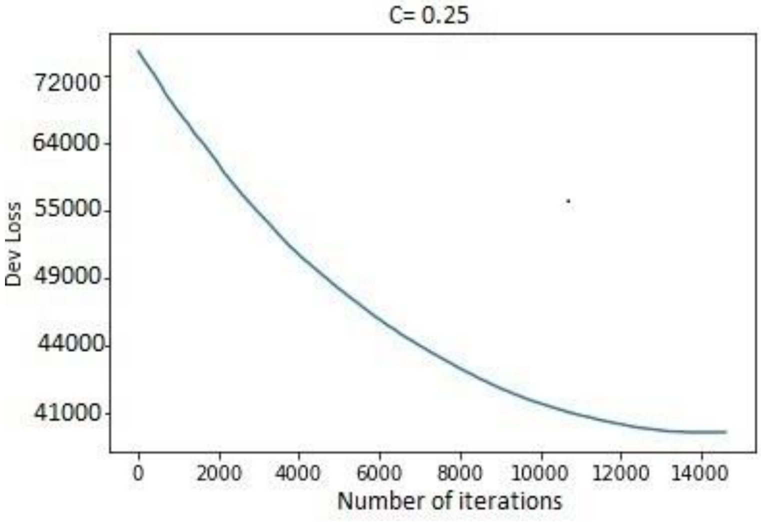
MSE losses on development set instances with and without the use of early stopping criteria:

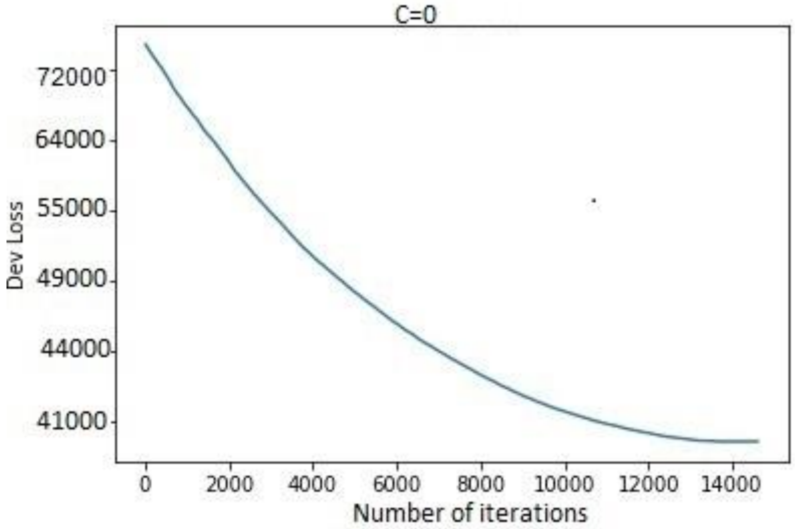
Without early stopping: 42152.7367114286

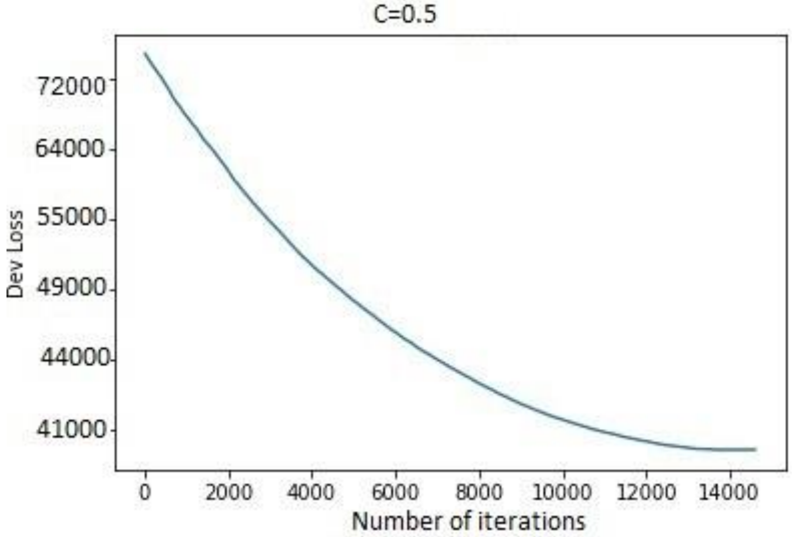
With early stopping: 41825.3042313232

3. Effect of regularization:

Regularization helps in reducing over-fitting by shrinking the weights of the features and getting rid of high degree polynomial features in the model




C=0
72000
64000
55000
49000
44000
41000
Dev Loss
0
2000
4000
6000
8000
10000
12000
14000
Number of iterations


C=0.5
72000
64000
55000
49000
44000
41000
Dev Loss
0
2000
4000
6000
8000
10000
12000
14000
Number of iterations

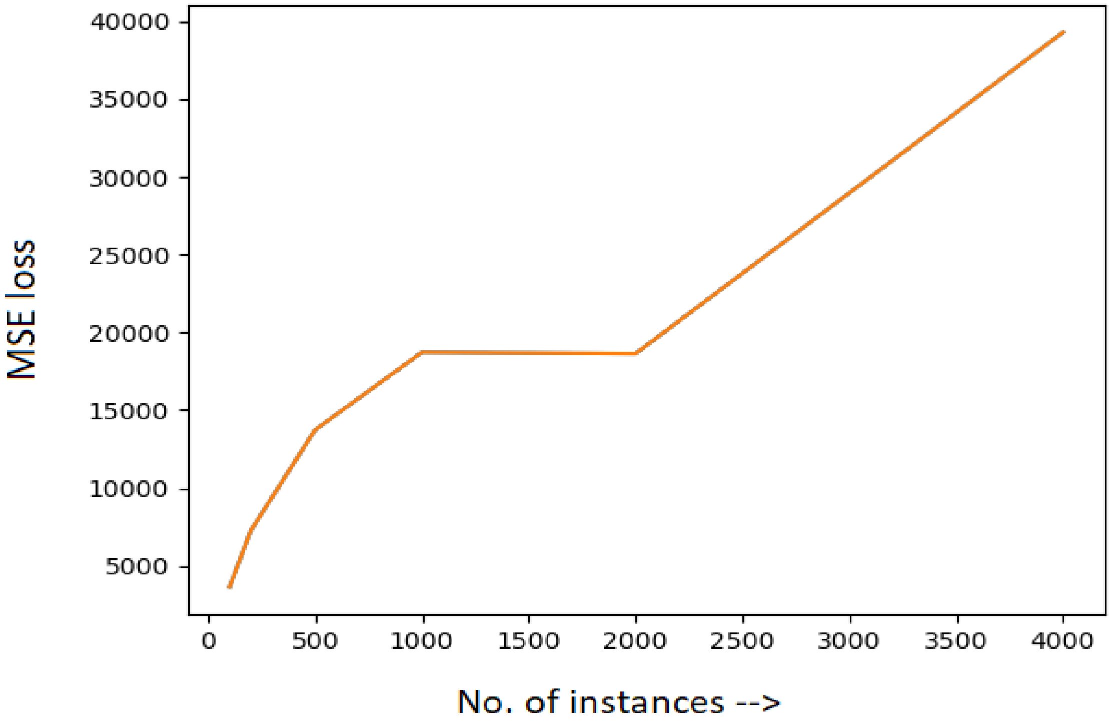
4. Basis Functions:

basis1(): Polynomial basis function; MSE on dev = 42343.32738

basis2(): basis function of type

$x_i^2 + \frac{const.}{1+\left(\frac{1}{\ln 10}\right)^{x_i}}$; MSE on development set = 44354.92764

5. Training Plots:



6. Feature Importance:

Least important field → instrument, version

Most important field → scan, brightness, track, confidence, bright_t31

Reason:

‘instrument', ‘version’ fields are the least important field because its value stays constant throughout the data-set. Therefore, we can safely say that this field does not affect the target field('frp') at all.

'scan' field is the most relevant field in the dataset because it has the highest positive correlation with the target field('frp'). Similarly, brightness, track, confidence, bright_t31 have high correlation with frp.

7. Climb the Leaderboard:

Feature selection: Removed unimportant/least important features by screening the results of correlation factors, successfully reducing the list of the input parameters to only relevant parameters.

Scaling used: Z-score scaling is used to normalized the features.