Programming Assignment Report

A Programming Assignment Report
Submitted in partial fulfillment
of the course requirements
Of CS725
Foundations of Machine Learning

by

Ayush Singh Chauhan, (Mtech RA, 213059001)
Karnika Shivhare, (PhD, 204050010)
Krishnakant Bhatt, (MS, 21Q050016)

Instructor
Prof. Preethi Jyothi
Department of Computer Science and Engineering
Indian Institute of Technology Bombay
Mumbai 400076 (India)
09 September 2021

Table of Contents:

1. Problem statement

2. Solving Linear Regression

3. Gradient descent stopping criteria

4. Effect of regularization.

5. Basis Functions

6. Training plots

7. Feature Importance

8. Climb the leaderboard

Problem Statement:

Forest fires have become an increasingly common occurrence across the world. Most recently, raging forest fires in Turkey caused huge devastation. In this assignment, we worked with a dataset with fourteen different attributes describing fires in Australia. Given these features, the task was to predict a real valued `frp` that refers to "Fire Radiative Power" using Linear regression Model with closed form solution and gradient descent based approaches.

1.Solving Linear Regression:

The L2 regularized Linear Regression has been implemented using the following two approaches:

(a) Closed form / Analytical Solution:

MSE loss on the development set = **37260.37**

(b) Iterative solution using mini-batch gradient descent:

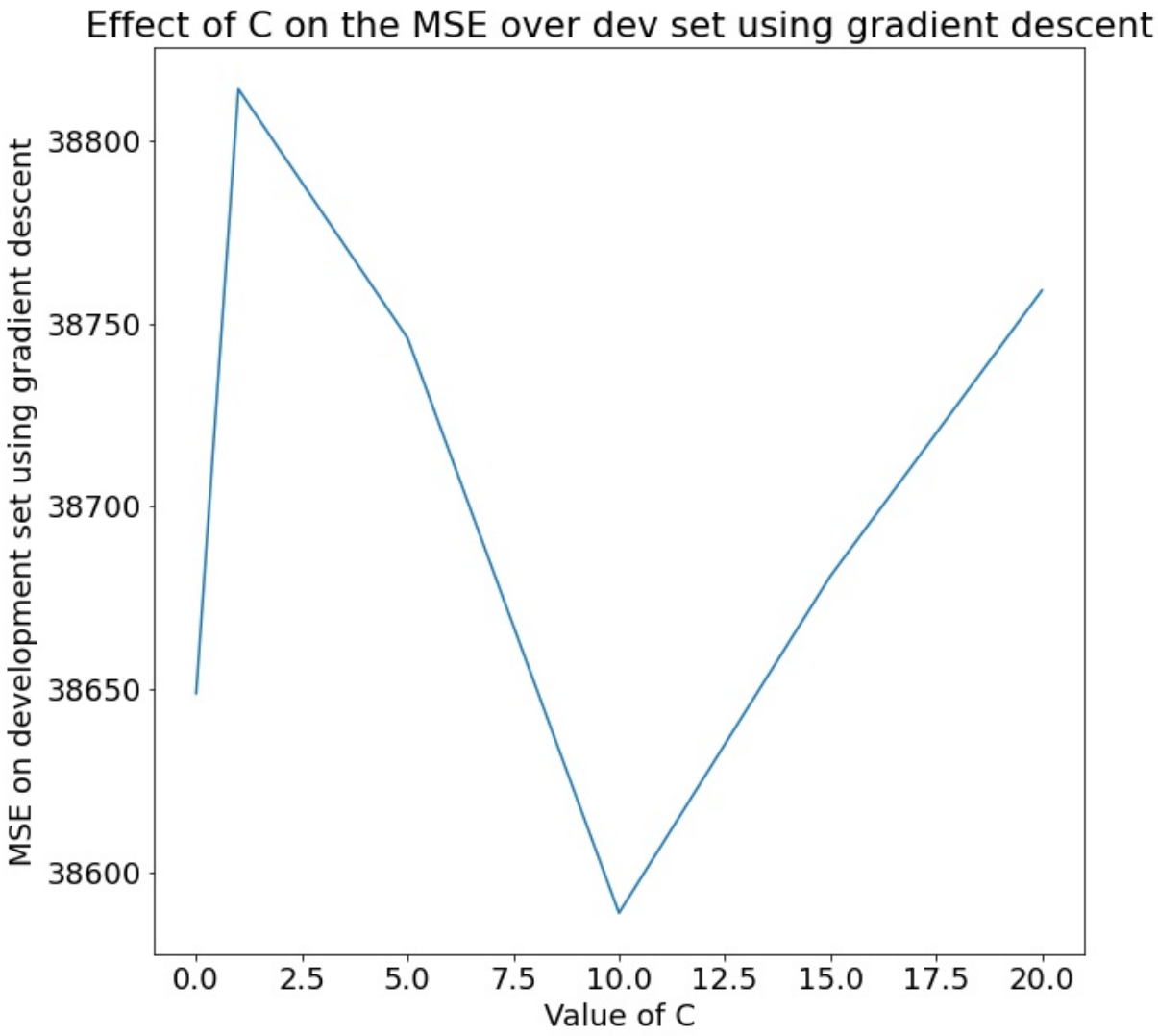
MSE loss on the development set = **73509.31**

2. Gradient Descent Stopping Criteria:

The stopping criteria we used for convergence in gradient descent is the difference between the MSE losses on the development set in two consecutive iterations.
If the difference between the MSE losses on the development set in two consecutive iterations is less than **10** then we declare the convergence by stop the further Iterations of gradient descent.

3. Effect of Regularization:

Here is the plot of effect of value of C in gradient descent on the MSE for development set.
The values of C used are { 0,1,5,10,15,20 }. The corresponding MSE's can be seen Here in the plot.



4. Basis Functions:

We have implemented two different basis functions which are applied to our input features with the L2-regularized model and optimized using gradient descent.

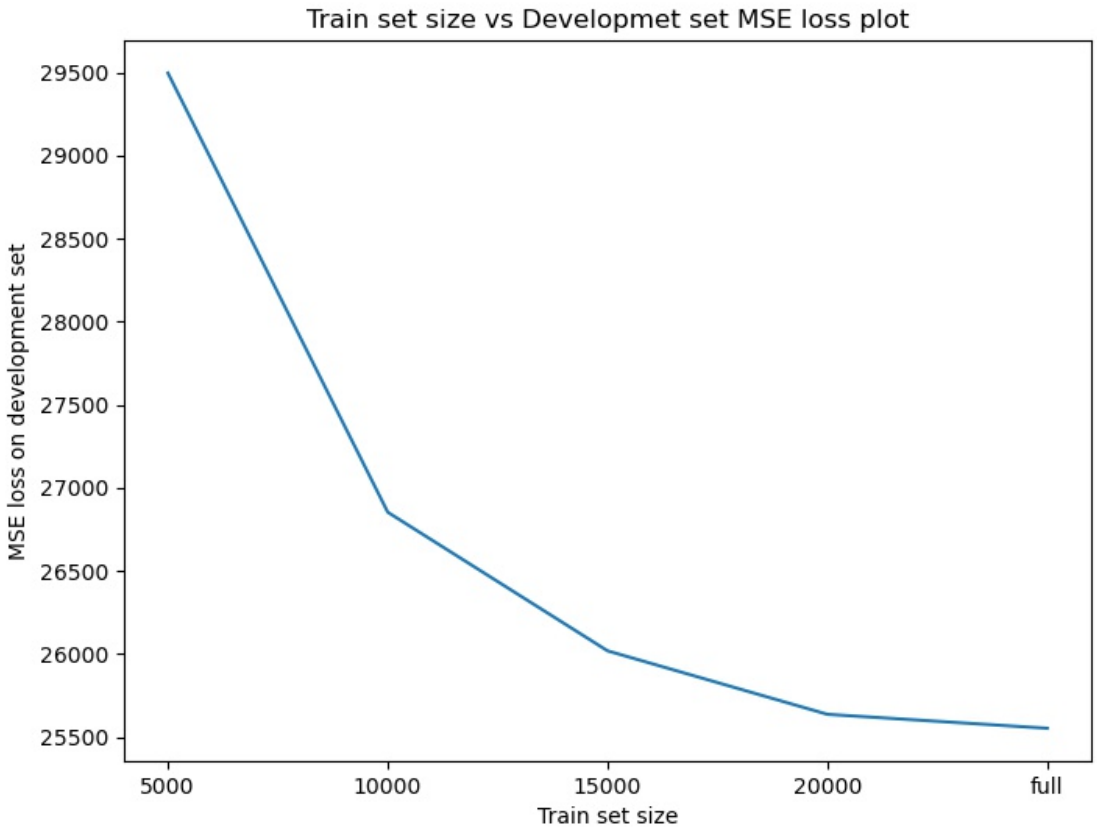
MSE loss on development set using the basis function 1 = **34214.36**

MSE loss on development set using the basis function 2 = **24733.72**

Reason behind choosing these functions:
As we have observed while the data analysis that the features ‘brightness’, ‘bright_t31’, ‘confidence’, ‘scan’, ‘track’ are the features which are having the most predictive power among all (i.e. the target variable is most sensitive to these features). That’s why We chose these functions in order to increase the weight of these features while training the model.

5.Training Plots:

Below is the plot which represents the MSE losses on the development set when we train the model by taking the first 5000 instances, first 10000 instances, first 15000 instances, first 20000 instances, all instances from the train set respectively.



6. Feature Importance:

During the data analysis, we have observed some patterns related to how much the the target variable “frp” depends on a particular feature and how much the target variable Changes when a particular feature changes. Based on those observations we have found out the following points.

Most Important feature for the prediction of “frp” : “brightness”
Least important feature for the prediction of “frp” : “instrument”

-> How we identified these two features:-

Identifying the least important feature was straightforward. The feature “instrument” was having the same value “MODIS” across all the instances in the training set. So this can be easily figured out that this feature is not providing any important information for predicting the target “frp”. Similarly we can say that the feature “version” can also be tagged as least important.

On the other hand for predicting the most important feature first we took one feature at a time to predict the target and observed the pattern in the MSE. Then we analysed the weights assigned by the model to the features model is trained using all the features. Also we used some standard methods using libraries like scikit-learn, which helps to get the importance measures of different features.

7. Climbing the leaderboard:

For improving our scores on the kaggle leaderboard we tried the following approaches.

1. We incorporated new features using some basis functions on the features from the training set.
2. We identified the features which are not affecting the target much, and removed those features.
3. We identified the features which are affecting the target the most and increased their contribution in training the model.
4. Tuned the hyperparameters in order to achieve better results.