Project 2 : Learning to Rank using Linear Regression

CS-574 – Introduction to Machine Learning

Name : Dhiren Patel
UBitName : dhirenbh
UBit number : 50170084

1 Problem Introduction

In this project, We are required to use linear regression technique to predict target variable from given input vector data. Here 2 techniques are used, Maximum likelihood solution and stochastic gradient method. These techniques are used on 2 dataset, Synthetic dataset and LEarning TO Rank (LETOR) dataset.

2 Reading data into proper format from QueryLevelNorm.txt

I have imported all data into 1 column by below command.

```
fileimport = importdata('Querylevelnorm.txt',' ');
```

I have got all columns from it into proper format using below commands.

```
c=strsplit(num2str(cell2mat(fileimport.textdata(i,j))),':');
```

3 Maximum Likelihood Method

3.1 Real (LETOR) DataSet

I have fetched input data into x_real[69623,46] from QueryLevelNorm.txt and target values in t_real[69623,1].

From which 55698 rows (80 %) are training set , next 6962 rows (10%) are validation set, others are testing set.

How have I set hyper parameters :

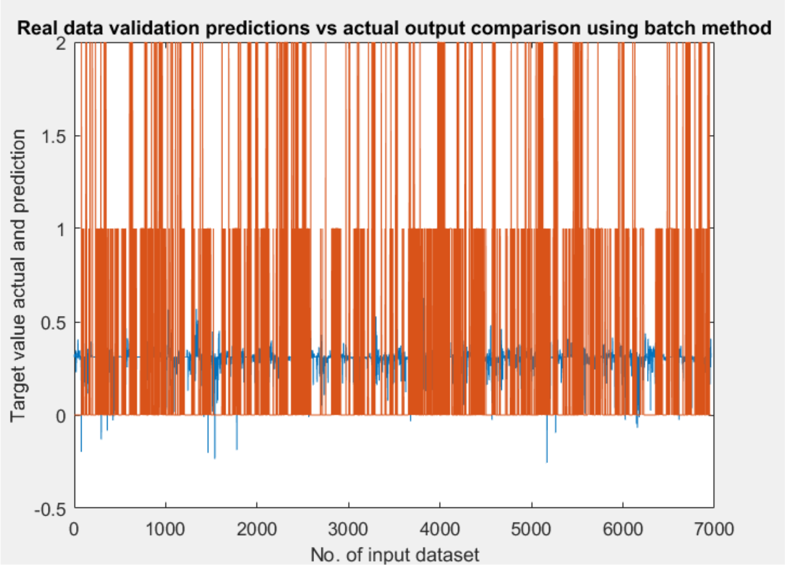
1) I have checked through all possible M values from 1 to 50. For each value of M, I have checked for all values of lambda between 0 to 0.5.
2) For dividing training set into clusters, I have used k-mean clustering technique. Then I'm taking mean of each cluster for each basis function. For Sigma, I have taken as diagonal matrix of all complete training set and same sigma is being used for all basis functions.
3) For each value of M and lambda, I found training error rms and validation error rms. Finally I have chosen M and lambda with smallest validation rms error.

   Below data can very as k-mean will divide data into different clusters each time.
   M1 = 29 (vary in 25 to 30)
   Lambda1 = 0.1

trainPer1 = 0.5619
validPer1 = 0.5510

Below Graph shows variation between actual output and predictions on validation dataset.

Orange line → Actual Output
Blue line → Predictions done by batch method



3.2 Synthetic DataSet

Input data is into x[2000,10] and target values in t[2000,1].

Training dataset size : 1600 x 10 (80 %)
Validation dataset size : 200 (10 %)
Testing dataset size : 200 (rest 10 %)

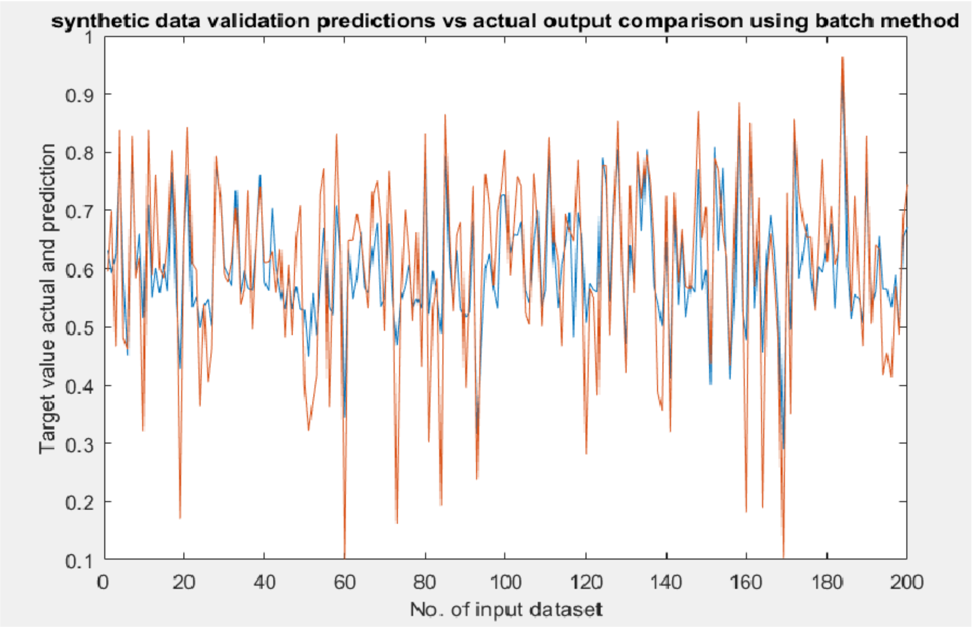
How have I set hyper parameters :

1) I have checked through all possible M values from 1 to 100. For each value of M, I have checked for all values of lambda between 0 to 5 with 0.1 interval. (As this won't take lots of time like real dataset)
2) Clustering, mean and sigma methods are same as real dataset methods.
3) For each value of M and lambda, I found training error rms and validation error rms. Finally I have chosen M and lambda with smallest validation rms error.

Below data can very as k-mean will divide data into different clusters each time.
M2 = will vary between 35 to 60
Lambda2 = 0.01
trainPer2 = 0.0928
validPer2 = 0.0961

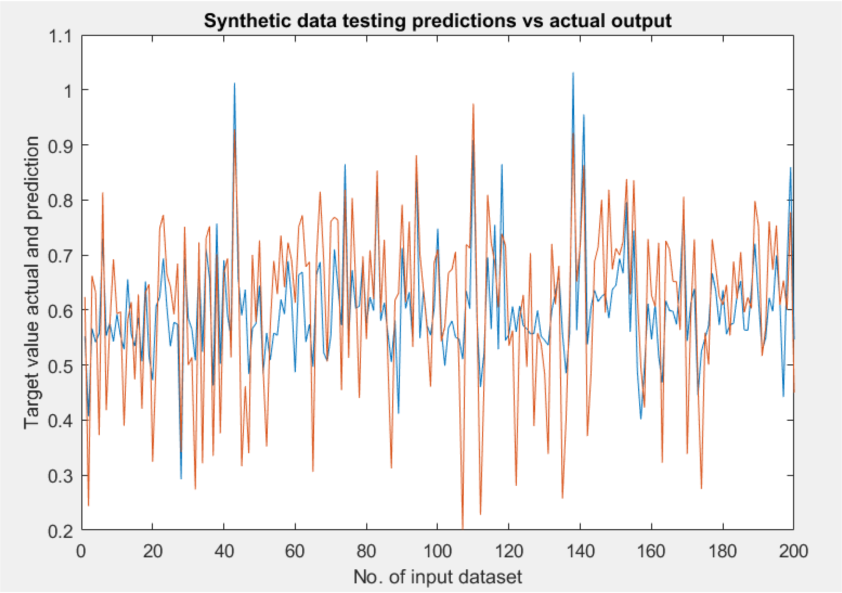
Below Graph shows variation between actual output and predictions on validation dataset.

Orange line → Actual Output
Blue line → Predictions done by batch method



Below Graph shows variation between actual output and predictions on testing dataset.

Orange line → Actual Output
Blue line → Predictions done by batch method



So, As we can see how our prediction varies with actual output and avoids overfitting too.

4 Stochastic Gradient Method

4.1 Real (LETOR) Dataset

I'm using same Hyper Parameters which I got from batch method.

Here is exact method how I performed it :

1) I processed all data 1 by 1 row wise and kept updating hyperparameters.
2) After processing all training set, I tested this model on validation set. If error is more then I repeated step 1.
3) I kept repeating step 1 and step 2 until I didn't get model with less error as batch method.

Parameter Learning

I used Bold Driver Method for setting adaptive learning rate.

1) If current error will be more than previous error that means you missed optimal point. So I reduced learning rate by 2.
2) If current error will be less than previous error then increase learning rate by 5%.

Stopping Criteria

I set precision as 0.000001.

If in stochastic gradient method, Gradient of error is less then precision while following above step 1 to step 3 then I stopped further processing.

Final ERMS from validation data is 0.80.

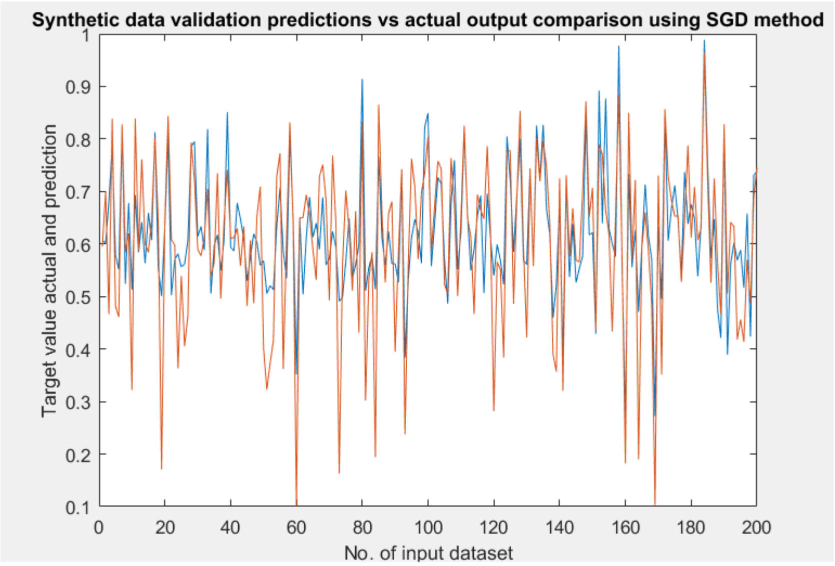
4.2 Synthetic Dataset

Method is same as above.

Validation ERMS is 0.100 obtained.

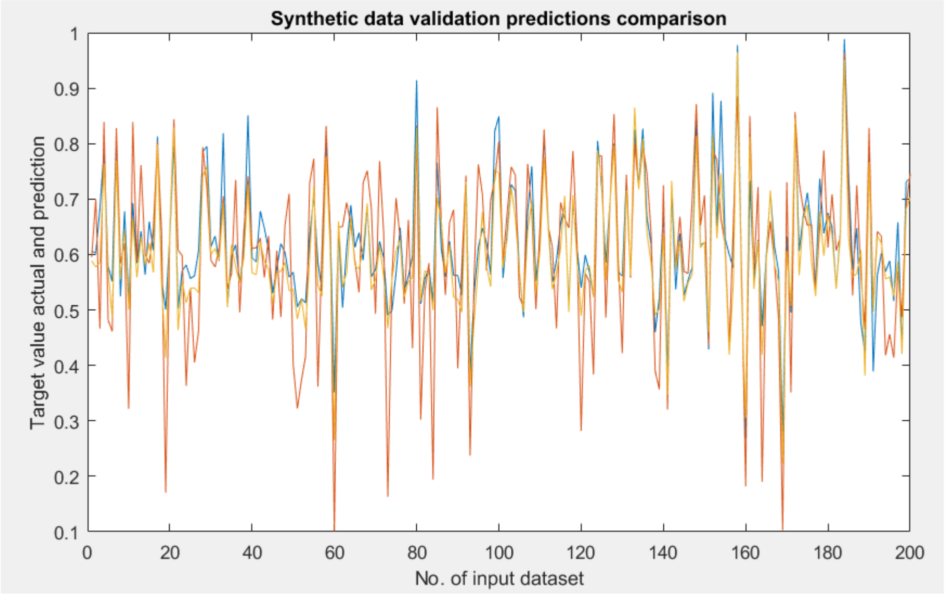
Below Graph shows variation between actual output and predictions on validation dataset.

Orange line → Actual Output
Blue line → Predictions done by SGD method



5 Comparison of Maximum Likelihood method and SGD method

Below graph shows comparison of predictions achieved through batch and SGD method.



Orange line → Actual output
Yellow line → Predictions using batch method
Blue line → Predictions using SGD method

6 Relationship between Hyper Parameters

With increase in M, there is increase in lambda too for better performance.