Machine Learning Report of Pro1.1

Haorui Dong
UBIT number: 50291149
Department of Computer Science and Engineering
University at Buffalo
haoruido@buffalo.edu

Abstract

In this project, we use machine learning to solve a problem called the Learning to Rank (LeToR) problem. We should first train a linear regression model using a closed-form solution. Then we need to train a linear regression model using stochastic gradient descent. We modified some hyper-parameters in order to find what values those hyper-parameters should take so as to give better performance.

1. The original result

First I run the sample code to get the original result. I change some print to see the number and plot of accuracy.

```
─────────Gradient Descent Solution─────────
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training    = 0.54964
E_rms Validation = 0.53846
E_rms Testing     = 0.62372
Accuracy Testing     = 70.30599
Accuracy Training     = 74.52198
Accuracy validation     = 75.25136
```
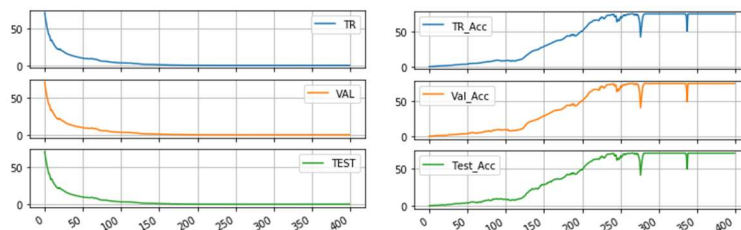
Figure1: SGD result

Figure2: Erms of SGD   Figure3: Accuracy of SGD

This result is under the condition that: M = 10, C_Lambda = 0.03.

The Testing Accuracy got 70.3. The E_rms of testing got 0.062372.

2. Change the value of C_Lambda

I change the value of C_Lambda from 0.03 to 0.9 to see how it influence the result.

```
─────────Gradient Descent Solution─────────
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training    = 0.55008
E_rms Validation = 0.53925
E_rms Testing     = 0.62408
Accuracy Testing     = 70.29162
Accuracy Training     = 74.52917
Accuracy validation     = 75.22264
```
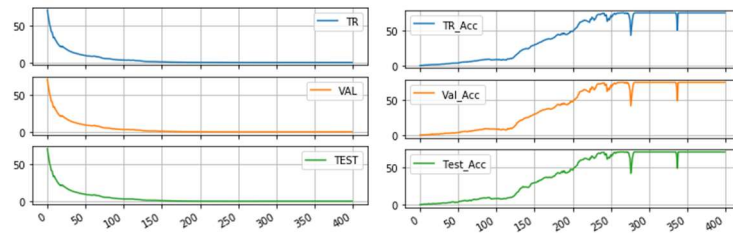
Figure4: Result of 0.9 C_Lambda

Figure5: Erms of 0.9 C_Lambda   Figure6: Accuracy of 0.9 C_Lambda

Then I change C_Lambda to 0.001.

```
————————Gradient Descent Solution————————
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training   = 0.54962
E_rms Validation = 0.53849
E_rms Testing    = 0.6236
Accuracy Testing     = 70.32036
Accuracy Training    = 74.52198
Accuracy validation  = 75.25136
```
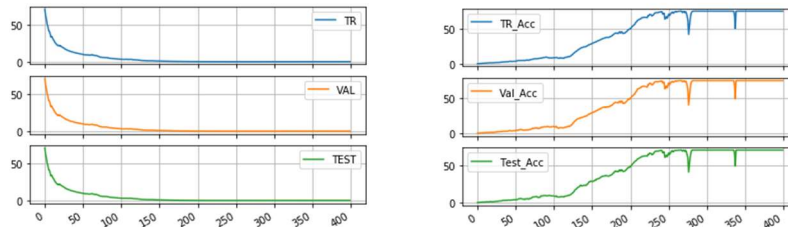
Figure7: Result of C_Lambda = 0.001

Figure8: Erms of C_Lambda = 0.001   Figure9: Accuracy of C_Lambda = 0.01

It doesn't change the result a lot.

3. Change the value of M

M determines how many cluster will be separated. So I change the M from 10 to 2 and 50 to see the change of result.

```
————————Gradient Descent Solution————————
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training   = 0.56868
E_rms Validation = 0.55485
E_rms Testing    = 0.65091
Accuracy Testing     = 70.23416
Accuracy Training    = 74.52198
Accuracy validation  = 75.17955
```
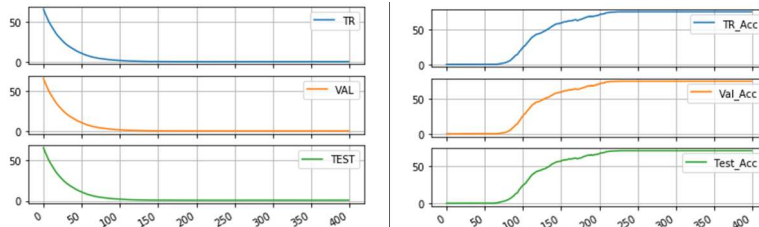
Figure10: Result of M = 2

Figure11: Erms of M = 2   Figure12: Accuracy of M = 2

The curve got more smooth.

```
——————————Gradient Descent Solution——————————
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training    = 0.54108
E_rms Validation = 0.53714
E_rms Testing     = 0.61612
Accuracy Testing      = 70.26289
Accuracy Training     = 74.58303
Accuracy validation   = 75.19391
```
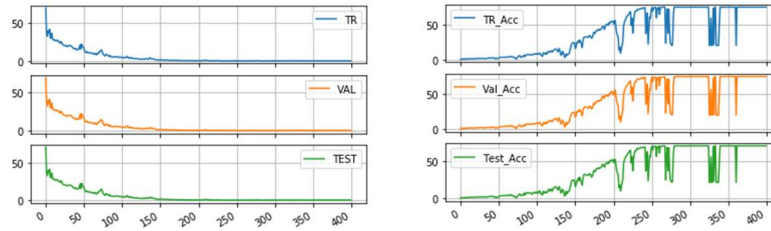
Figure13: Result of M = 50

Figure14: Erms of M = 50   Figure15: Accuracy of M = 50

The final result seems same as the original one. But the shapes of curves are changed a lot.

4. Change some other parameters

In the part of Grandient Descent, I keep the learning rate and change the La to 10.

```
——————————Gradient Descent Solution——————————
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training    = 0.54218
E_rms Validation = 0.5369
E_rms Testing     = 0.61764
Accuracy Testing      = 70.29162
Accuracy Training     = 74.53276
Accuracy validation   = 75.17955
```
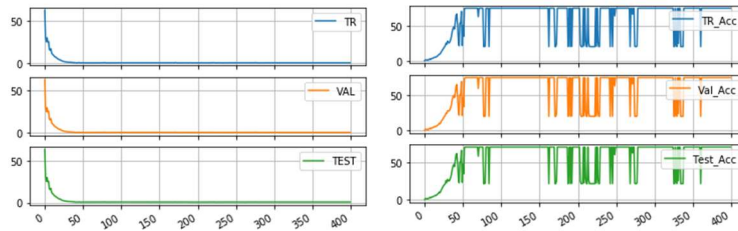
Figure16: Result of La = 10

Figure17: Erms of La=10   Figure18: Accuracy of La=10

We found that the results do not change a lot, but it changed to final value in a higher speed.

Then I change the La to a very small value such as 0.1.

```
——————————Gradient Descent Solution——————————
M = 15
Lambda  = 0.0001
eta=0.01
E_rms Training    = 22.11086
E_rms Validation = 21.98863
E_rms Testing     = 22.38816
Accuracy Testing      = 2.0974
Accuracy Training     = 1.88693
Accuracy validation   = 2.16892
```

Figure19: Result of La = 0.1

Obviously, it had a huge influence to result. And we got a very bad result, so I will not put the plot of curve.

I keep the La and change the Learning rate to 0.05 from 0.1. And the program cannot run with math range error.