Factorization-Based Data Modeling
Practical Work 3

CHEN Hang, FAN Zheng

January 2019

1 Complete the file dsgd_mf_template.cpp

See the codes in dsgd_mf_template.cpp.

2 Set the rank to 10 and the step size to 0.00001

To compile:

```
mpicxx dsgd_mf_template.cpp -Wall -I/usr/local/include
-L/usr/local/lib  -lgsl -lgslcblas -lm -o dsgd_mf
```

To run:

```
mpirun -n 4 ./dsgd_mf 3883 6040 10 20 0.00001 1
```

We get the information from the terminal:

```
Process3: Initial data loaded
Process2: Initial data loaded
Process0: Initial data loaded
Process1: Initial data loaded
Iteration: 0
Iteration: 0
Iteration: 0
Iteration: 0
Iteration: 10
Iteration: 10
Iteration: 10
Iteration: 10
*586
*586
*581
*586
```

3 Compute the RMSE by using the code compute_rmse.cpp and plot the RMSE in Matlab by using plot_rmse.m

To compile:

```
mpicxx compute_rmse.cpp -Wall -I/usr/local/include
-L/usr/local/lib  -lgsl -lgslcblas -lm -o compute_rmse
```

To run:

```
mpirun -n 10 ./compute_rmse 3883 6040 10 20 4
```

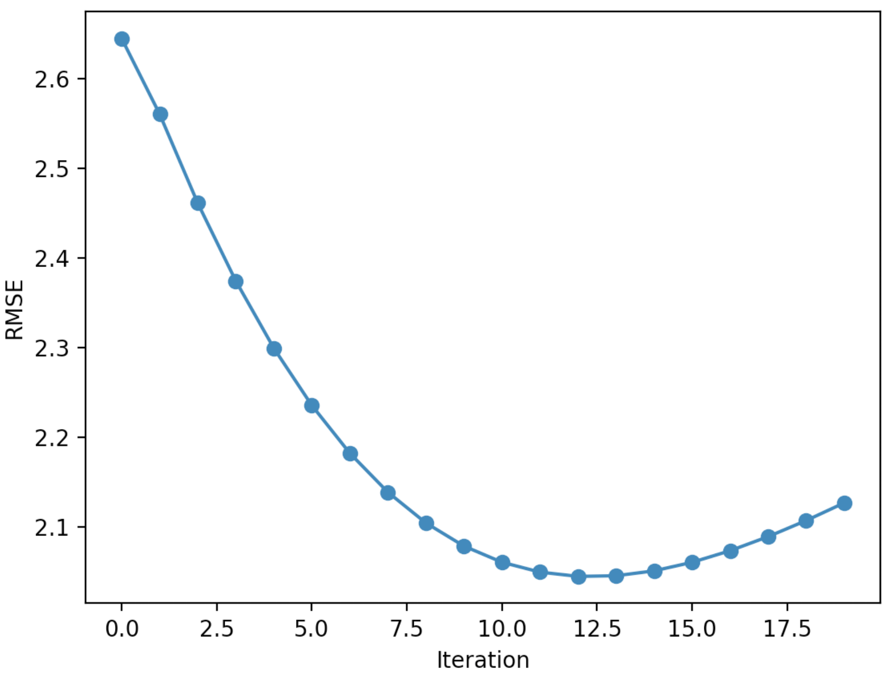
We get the information from the terminal:

```
Data loaded
Data loaded
Data loaded
Data loaded
Data loaded
Data loaded
Data loaded
Data loaded
Data loaded
Data loaded
Iteration: 10
Iteration: 4
Iteration: 6
Iteration: 12
Iteration: 2
Iteration: 18
Iteration: 16
Iteration: 8
Iteration: 0
Iteration: 14
Iteration: 11
Iteration: 13
Iteration: 3
Iteration: 5
Iteration: 7
Iteration: 15
Iteration: 19
Iteration: 17
Iteration: 9
Iteration: 1
```

Then we need to plot the RMSE(in Python).

```
python3 plot_rmse.py
```

We get the result:

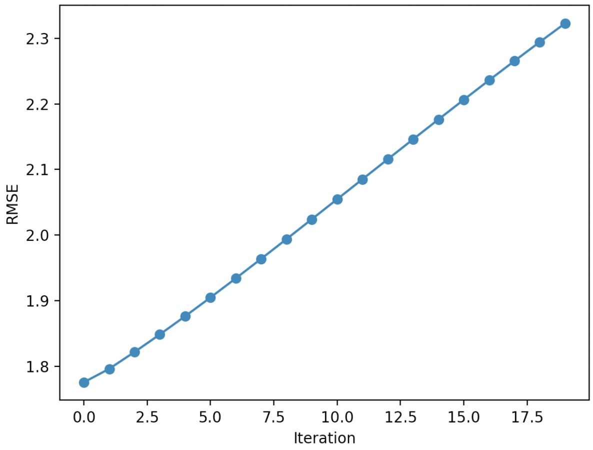


The RMSE is for checking if the function is converged or not, if the variation of value of RMSE is small, we can know that the function is converged.

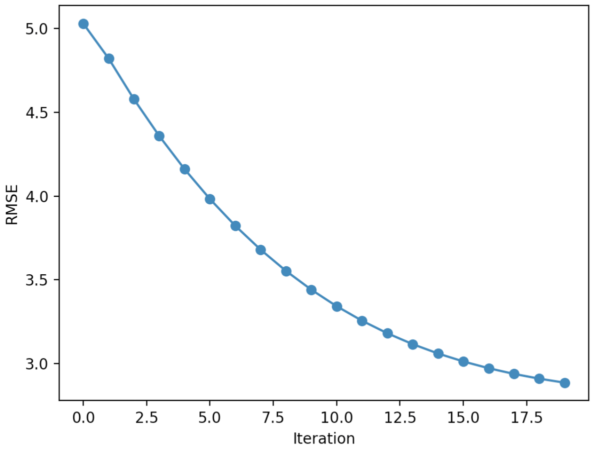
4 Play with the rank and the step-size

Comparing different ranks

When k = 5, step-size = 0.00001:



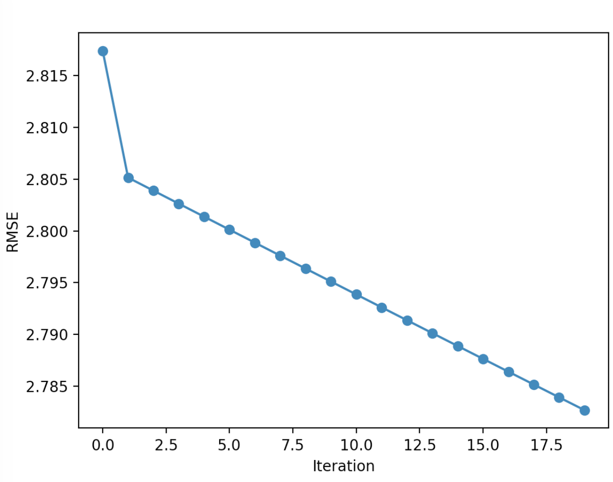
When k = 15, step-size = 0.00001:



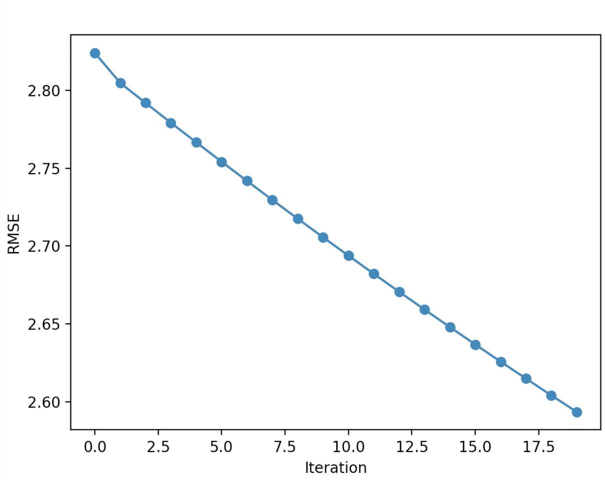
The rank k controls the dimension of the matrix W and H, it represents the features of the matrix. When k is too small(k = 5), it means that there are less features and we can't have a good result, the RMSE is going to be high and can't converge. And when k is large(k = 15), it means that we have more features, as we can see from the figure, the RMSE begins at a big value 5.0, then it converges very slowly, it has not converged after 20 iterations. What's more, when the k is too large, it maybe cause overfitting. So for the chose of k, it can nigher be too small nor too large.

Comparing different step-size

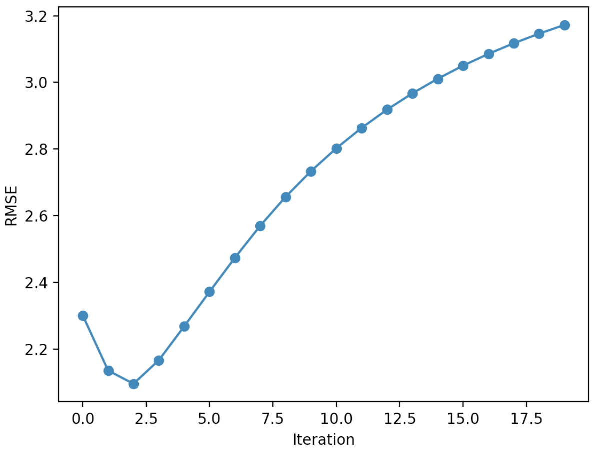
When k = 10, step-size = 0.0000001:



When k = 10, step-size = 0.000001:



When k = 10, step-size = 0.00005:



We can notice that when we have the original step-size = 0.00001, when the number of iteration is 13, the value of RMSE is the best, and when the number of iteration is 20, the value of RMSE increases to about 2.15. When we set the step size = 0.0000001 and the step size = 0.000001, we can notice that when the number of iteration arrives to 20, the results of the 2 models are worse than the original graph, but we can find these 2 graphs continue descending because the learning rate is much smaller than the original value of learning rate, so the speed of descending is slower than before.

When we set the value of step size equal to 0.00005 which is larger than the original value, we can notice that the lowest value of the graph is corresponding to the iteration = 2, the result is about 2.05, so when the step-size becomes larger, the curve descends quickly, but when the iteration is 20, the value of RMSE is pretty high.

So for the chose of step-size, it can nigher be too small nor too large, because when the step-size is larger, the curve can converge quickly, but may be it can miss the best value, when the step-size is too small, the curve converges too slowly, so when all of the iteration finish, the curve doesn't converge to the lowest value.