1. First let’s look at the ratings.csv file. The first few columns of ratings.csv is as shown in the picture below:

########################################################################

A picture containing clock, object, meter

Description automatically generated

########################################################################

1. MAE, Mean Absolute Error measures the average magnitude of the errors in a set of predictions

RMSE, Root Mean Squared Error is a quadratic scoring that measures the average magnitude of the error.

RMSE will always be larger or equal to the MAE. The greater the difference between them, the greater the variance in the individual errors in the sample. Both measures can range from 0 to infinity. Lower the values of MAE and RMSE, the better.

1. RMSE and MSE for Probabilistic Matrix Factorization (PMF) is as shown below:

########################################################################

Output:

Text

Description automatically generated

RMSE and MAE for five-fold PMF:

{'test\_mae': array([0.76888102, 0.77694119, 0.78394215, 0.78330379, 0.78947437]),

‘test\_rmse': array([0.99355912, 1.00720332, 1.01273677, 1.01490639, 1.02196274]),

‘fit\_time': (3.1727535724639893, 3.1059951782226562, 3.1052229404449463, 3.1319563388824463, 3.128542900085449),

'test\_time': (0.144212007522583, 0.1401071548461914, 0.12485003471374512, 0.1422138214111328, 0.12299180030822754)}

########################################################################

RMSE and MSE for User Based Collaborative Filtering is as shown below:

########################################################################

Output:

Text

Description automatically generated

RMSE and MAE for five-fold User Based Collaborative Filtering:

{'test\_mae': array([0.7141557 , 0.7064661 , 0.70278552, 0.70853913, 0.70401333]),

'test\_rmse': array([0.93101601, 0.91979979, 0.92004312, 0.92286234, 0.92019265]),

'fit\_time': (0.2722916603088379, 0.276334285736084, 0.28095579147338867, 0.2749972343444824, 0.2758452892303467), 'test\_time': (1.305346965789795, 1.2526752948760986, 1.2655532360076904, 1.2794365882873535, 1.250403881072998)}

########################################################################

RMSE and MSE for Item Based Collaborative Filtering is as shown below:

########################################################################

Output:

Text

Description automatically generated

{'test\_mae': array([0.69668777, 0.6834048 , 0.69497176, 0.68137891, 0.68707677]),

'test\_rmse': array([0.91582721, 0.89923181, 0.91691483, 0.89652777, 0.90592884]),

'fit\_time': (7.029903888702393, 6.840514659881592, 6.929088354110718, 6.81726861000061, 6.9107584953308105), 'test\_time': (4.7717673778533936, 4.765731334686279, 4.694500923156738, 4.674083471298218, 4.624952793121338)}

########################################################################

1. The average mean value of PMF, User based and Item Based collaborative filtering:

########################################################################

Output:







########################################################################

Lower MAE and RMSE indicates better fit. Hence, based on the average MAE and RMSE values, Item based Collaborative Filtering has a better fit.

1. The following are the graphs for MSE and RMSE over five-fold cross validation for Item and User based collaborative filtering with MSD, Pearson and Cosine distance. The dashed line represents the mean value for the corresponding value.

1) User-based Collaborative filtering using MSD:

########################################################################

Output:

A large clock mounted to the side

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

########################################################################

2) User-based Collaborative filtering using Pearson similarity:

########################################################################

Output:

A clock mounted to the side

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

########################################################################

3) User-based Collaborative filtering with Cosine similarity:

########################################################################

Output:

A large clock mounted to the side

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

########################################################################

4) Item based Collaborative Filtering with MSD:

########################################################################

Output:

A close up of a clock

Description automatically generated

Chart, bar chart

Description automatically generated

##############################################################################

5) Item based Collaborative Filtering with Pearson similarity:

##############################################################################

Output:

A large clock mounted to the side

Description automatically generated

Chart, bar chart

Description automatically generated

##############################################################################

6) Item based Collaborative Filtering with Cosine similarity:

##############################################################################

Output:

A picture containing clock, object, ball, green

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

##############################################################################

Among User based Collaborative filtering with MSD, Pearson and Cosine similarity, it was noticed that Mean Squared Distance (MSD) had the lowest mean RMSE and MAE.

Among Item based Collaborative filtering with MSD, Pearson and Cosine similarity, once again, it was noticed that Mean Squared Distance (MSD) had the lowest mean RMSE and MAE.

**The results of user-based and item-based CF are consistent with each other for all similarity metrics.**

1. Since MSD gave us the smallest RMSE and MAE for both user-based and item-based collaborative filtering, I decided to use this similarity metric for tasks f and g. K vs Error graphs were draw for K 1 to 100.
2. User based Collaborative Filtering using MSD:

##############################################################################

Output:

Chart, line chart

Description automatically generated

############################################################################

1. Item based Collaborative Filtering using MSD:

############################################################################

Output:

Chart, line chart

Description automatically generated

############################################################################

1. The best K was the K value that gives us the lowest RMSE and MAE value. For User based CF, the lowest RMSE value was 0.91741 and it was obtained when K = 60

########################################################################

Output:

The minimum RMSE value in User based Collaborative Filtering using MSD is: 0.9174160874647195

The K value that gives the minimum RMSE in User based Collaborative Filtering using MSD is: 60



############################################################################

Similarly, for Item based CF, the lowest RMSE value was 0.908932 and it was obtained when K = 93

############################################################################

Output:

The minimum RMSE value in Item based Collaborative Filtering using MSD is: 0.9089329378274155

The K value that gives the minimum RMSE in Item based Collaborative Filtering using MSD is: 93



#####################################################################

**The best K is not the same for User-based and Item based collaborative filtering.**

GitHub Link:

Assignment\_06

https://github.com/monicabernard/CAP-5610\_Machine-Learning.git