INFORMATION RETRIEVAL (CS F469) ASSIGNMENT 3

Implementation of Recommender System Algorithms

-Tarun Raheja

-2015A7PS0106H

This assignment attempts to compare various techniques used in implementing Recommender Systems on the basis of their errors using Root Mean Square Error, Precision on top K and Spearman Rank Correlation.

DATASET:

The MovieLens dataset which has 100k reviews is used.

MAJOR DATA STRUCTURES USED

- The code is for the most part written in the main recom.py file. It is clearly partitioned into three sections for CF, SVD and CUR for ease of readablilty.
- The main data structures used are numpy and sparse matrices.

PACKAGES USED

- Numpy, Scipy, SciKitlearn and pandas were used.

BRIEF NOTES ON CODE EXECUTION:

- The main python file supplied is to be executed.
- It sequentially executes CF, SVD and CUR.
- After each execution, it computes the RMSE, Top K error and Spearman Rank Correlation values and displays them.

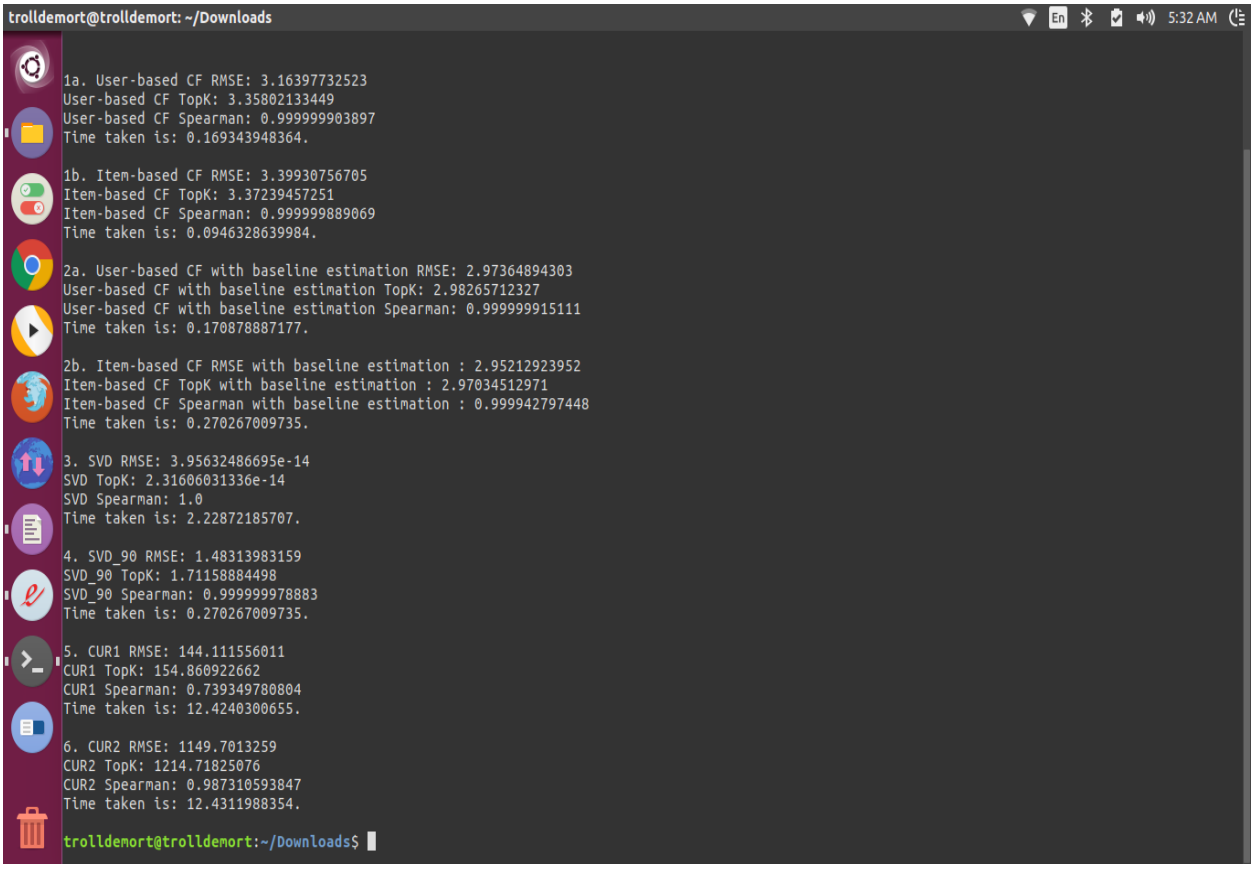
DOCUMENTATION

The code is written without classes, and most of the functionality is implemented within functions. For more clarification all the functions are listed below:

- For CF:
    - Predict( ) : This function returns a matrix of predicted values based on user input and existing ratings. It is central to the calculations of CF.
    - Rmse( ) : It calculates the Root Mean Square Error.
    - Sper( ): It calculates the Spearman Correlation Coefficient.
- For SVD:
    - Svd( ) : This fairly large method takes the training matrix as input and decomposes it.
    - Svd_90( ) : This method takes training matrix as input and decomposes it following the correct rules which 90%decomp requires.
- For CUR:
    - Cur( ): This is a function that helps calculate Cur decomposition.
    - Cur_2: This function helps calculate Cur decomposition with no replacement.

TABLE

TECHNIQUE	RMSE	SPEARMAN	TOP K	TIME(s)
UU CF	3.1639	0.9999	3.3580	0.1693
UU-BF CF	3.3993	0.9999	3.3723	0.0946
II CF	2.9736	0.9999	2.9826	0.1708
II-BF CF	2.9521	0.9999	2.9703	0.2702
SVD	3.9563e-14	1.0000	2.3160e-14	2.2287
SVD(90)	1.4831	0.9999	1.7115	0.2702
CUR	144.1115	0.7393	154.8609	12.4240
CUR(90)	1149.701	0.9873	1214.7182	12.4311


trolldemort@trolldemort: ~/Downloads
5:32 AM
1a. User-based CF RMSE: 3.16397732523
User-based CF TopK: 3.35802133449
User-based CF Spearman: 0.999999903897
Time taken is: 0.169343948364.
1b. Item-based CF RMSE: 3.39930756705
Item-based CF TopK: 3.37239457251
Item-based CF Spearman: 0.999999889069
Time taken is: 0.0946328639984.
2a. User-based CF with baseline estimation RMSE: 2.97364894303
User-based CF with baseline estimation TopK: 2.98265712327
User-based CF with baseline estimation Spearman: 0.999999915111
Time taken is: 0.170878887177.
2b. Item-based CF RMSE with baseline estimation : 2.95212923952
Item-based CF TopK with baseline estimation : 2.97034512971
Item-based CF Spearman with baseline estimation : 0.999942797448
Time taken is: 0.270267009735.
3. SVD RMSE: 3.95632486695e-14
SVD TopK: 2.31606031336e-14
SVD Spearman: 1.0
Time taken is: 2.22872185707.
4. SVD_90 RMSE: 1.48313983159
SVD_90 TopK: 1.71158884498
SVD_90 Spearman: 0.999999978883
Time taken is: 0.270267009735.
5. CUR1 RMSE: 144.111556011
CUR1 TopK: 154.860922662
CUR1 Spearman: 0.739349780804
Time taken is: 12.4240300655.
6. CUR2 RMSE: 1149.7013259
CUR2 TopK: 1214.71825076
CUR2 Spearman: 0.987310593847
Time taken is: 12.4311988354.
trolldemort@trolldemort:~/Downloads$