Puja Sonawane
UCF ID: 5075654

Cloud Computing: Assignment 2

(1 point) Please write a brief report (e.g., 1 page) to describe (1) how to run your applications, (2) how you tune the accuracy of each approach, (3) comparing accuracy of the two approaches.

How to Run the Program:
I have submitted the .ipynb file which runs on Jupyter notebook. For reference I have submitted .py file as well.
Please run the code cell by cell in Jupyter notebook. There are two separate files for K-means and ALS. Import the files into Jupyter notebook and run each cell. You might need to change the path of the data file. I have kept datafiles as well for reference in assignment2 folder.

Comparing Accuracy of two approaches:
K means gives RMSE of 0.93 whereas ALS performs better with RMSE of 0.87
ALS is better choice for recommendation system in this case.

Recommendation System with K-means

(6 points) Please design a movie recommendation system using K-means. In K-means, please enumerate the value of K and find the best result according to RMSE.

To tune the accuracy of K-means approach, I tried range of K values. The best K is selected depending on minimum of SSE.
Best K is with least SSE which is 29.

```
least_ssee
#best K is 29
```

```
[[502123.6624947026, 25],
 [498955.0994429273, 26],
 [489590.6635881172, 27],
 [502650.16439045937, 28],
 [472141.6583442273, 29],
 [515729.22612895304, 30],
 [435086.11919278424, 31],
 [483245.95672973094, 32],
 [426129.84960044944, 33],
 [433239.09027276083, 34]]
```

Puja Sonawane
UCF ID: 5075654

The following is the best RMSE value obtained after parameter tuning is, 0.93

```
rmse.real
```

```
0.9303716850954153
```

Recommendation System with ALS

Design a workable ALS code. In ALS, please find the best ranks, lambdas, and numIters according to RMSE.

After tuning the hyper-parameter, the best performing values are as follow:

```
model.bestModel.rank
```

```
5
```

```
final_model._java_obj.parent().getMaxIter()
```

```
10
```

```
final_model._java_obj.parent().getRegParam()
```

```
0.01
```

Rank: 5, Max Iteration: 10, lambda/Reg parameter: 0.01
The best RMSE obtained with parameter tuning is 0.87

```
print(rmse)
```

```
0.8748213132371326
```