Joshua Harkness

COP 6526

Assignment 02

**Run the program:**

Clone the git repo at this url: <https://github.com/darkhark/COP6526.git>

This assignment is located in COP6526/assignment02/assignment02.py

Each of the current tasks are set up to run as long as the file is run and all the libraries in the import section are installed.

If you’d like to see the silhouette plot used to determine the optimal k for kmeans, uncomment the function on line 174.

**Tune the accuracy:**

For this project, RMSE was used to determine the best model. For the kmeans portion, the RMSE value can be changed by changing the k value. This can be done on line 175.

For ALS, simply running the runTask2 method with the given parameters will provide many results for RMSE. Different values can be tested by editing any of the three lists that are parameters on line 218.

**Comparison:**

The optimal k for kmeans was determined to be 6, with an RMSE of 0.9892157893894079.

The optimal parameters for ALS were: rank = 100, lambdas/regParams = .1, and numIters = 15. These parameters resulted in an RMSE of 0.859185841025245.

The resulting RMSE values show how ALS is a better predictor for values than kmeans because the RMSE value is less.