Rohan Kamat
013759252
CMPE-256

Recommender Systems PR-1

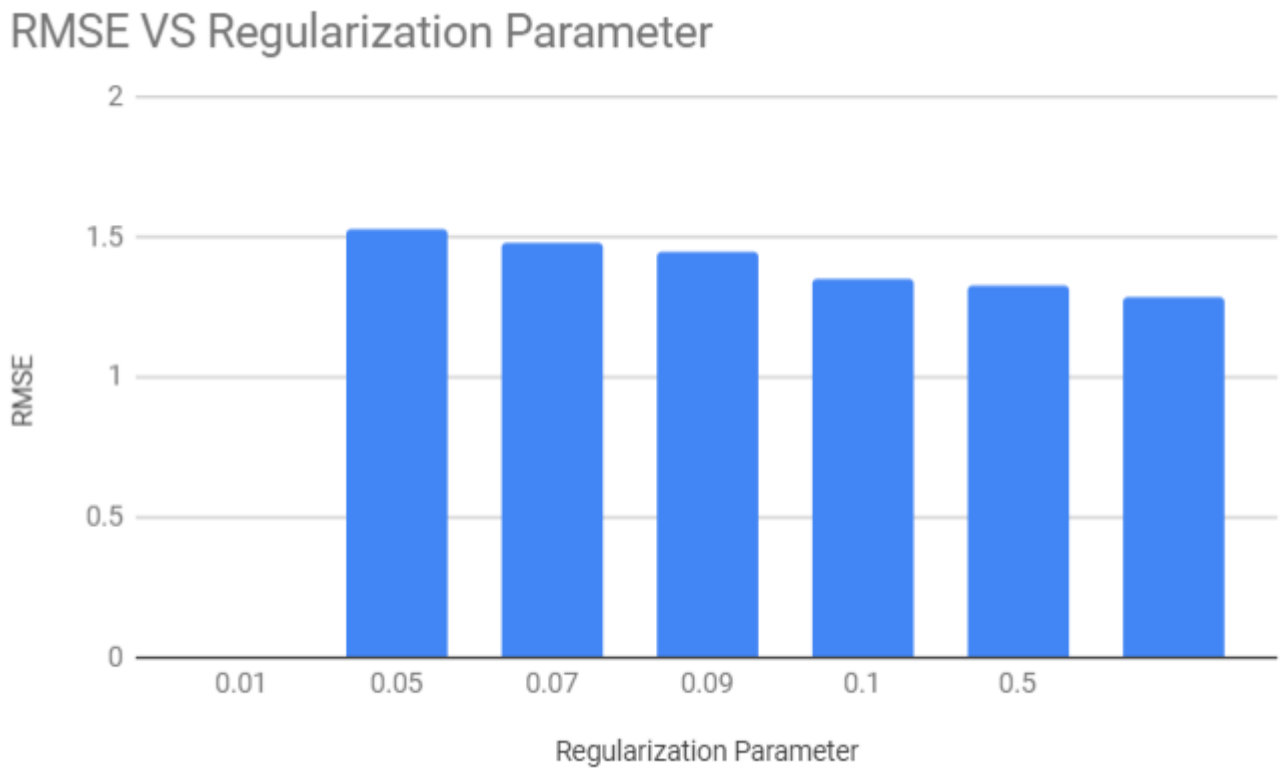
Collaborative Filtering is used by many e-commerce giants, to recommend users new products and services.There are 2 types of collaborative filtering item-item and user-user.The approach ALS(alternative least square) uses user-user collaborative filtering mechanism.In this assignment many techniques such as matrix factorization, regressionEvaluator.SVD approach was tried but produced imbalanced results so discarded.

Approach 1:The first approach was using a ALS along with matrix factorization.The result was an rmse score of 1.33.The rank and number of iterations were kept as 3.The issue was that the testing data had values which were not present in the training set.The ALS algorithms omitted the missing values and produced 3 less results respectively.On changing the parameters the results worsened.

Approach 2: After multiple attempts, a cold start mechanism was used to declare the values as missing values as 'nan'.This method used RegressionEvaluator mechanism.The train and test data were imported as rdds from arguments and split using tab(space).Row function was used to map user, movie and ratings respectively.Map function of rdd was to map the user and movie from the test to the predicted ratings.The output file had to have only ratings.Another rdd was used to store the ratings from prediction variable using lambda function.The output file had 'nan' values which were replaced by the mean of the rating.This is related to real life

scenario when a new user has not rated a product.The changing parameters have been plotted on the graph below.

Special Instructions: Spark-submit path/to/python file name path/to/train data path/to/test data



RMSE	RANK
1.297	37