Spark Project Modeling part walkthrough

1. Split data into training, testing, and validation with proportion of 6:2:2
2. Tune the following three parameters: First set iterations = 10,

Then pass in a parameter list to both Rank and Lambda

Rank = [4, 6, 8, 10]

Lambda = [0.005,0.01, 0.05, 0.1, 0.2]

1. After the first round training, evaluate the result, finding the best performing model contains rank = 10. Lambda = 0.2.
2. During the second round training, I set values of (rank, lambda) around (10, 0.2). Meanwhile I train everything with iteration 15.

The best performing result contains rank = 8, lambda = 0.2

The mean square error on validation set is 0.924

1. Finally test the model on testing set, the mean square error rate is 0.918