Movie Recommendation System

Kregg Jackson

Business Context

- This project was built for a website designed to make personalized movie recommendations to the user based off their ratings of movies
- When done correctly the system will support and improve the quality of the decisions users make while searching for movies

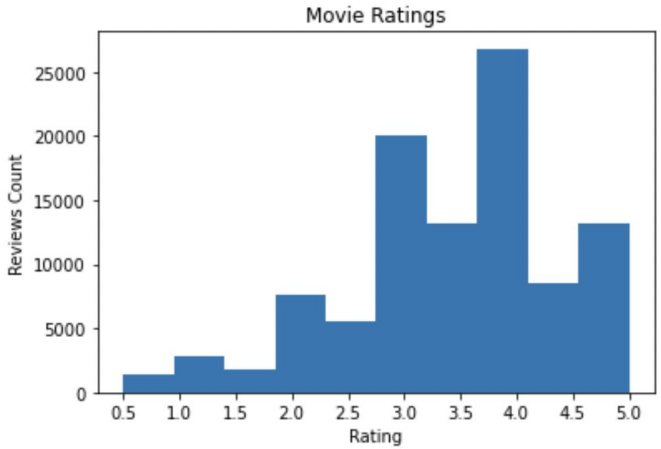
Movie Review Data

The movies dataframe has 9,700 rows of three columns :

- ‘movieId’
- ‘title’
- ‘genre’

The ratings column originally had over 100,000 rows of reviews with four columns of:

- ‘userId’
- ‘movieId’
- ‘rating’
- ‘timestamp’



Process Steps

My next steps were to:

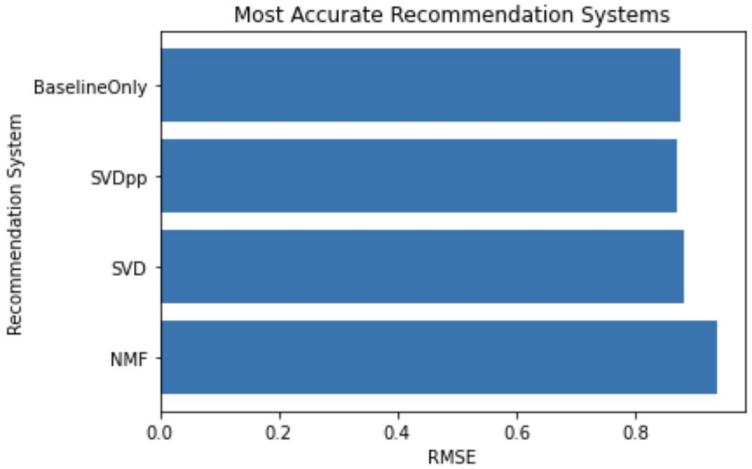
- Build and tune the recommendation systems
- Find the best parameters for the lowest Root means squared error (RMSE)
- Obtain a prediction for a specific user for a particular item
- Create a function that will return the top 5 movie recommendations for a user

The best performing systems were:

- SVD ++
- SVD
- BaselineOnly

Results

- The SVD system was able to predict rating within 0.88 points on average
- The BaselineOnly recommendation system predicted rating within 0.87 points
- The best performing system was the SVD ++, and was able to predict user ratings within 0.86 points on average



Conclusive Evaluation

- The best performing recommendation system is the SVD ++ iteration
- SVD ++ predicts user preference within 1 point of the actual rating
- The movie suggestion website will be able to take users previous rating and recommend movies the user has not seen

Future Improvements

- Have recommendation system not only suggest singular movies but movie genres
- Running gridsearch on SVD++ was very taxing on my machine's CPU with more time I could test more parameters to see if I could lower RMSE
- Build a function to accept new user data

Thank You

Email: kreggthegoat@gmail.com

Github: @kreggthegoat