Movielens Recommendation system

Hany Awadalla

May 29, 2019

Movielens Recommendation system Project

The goal of this project is to built a recommendation system with different method to achieve the best fiting module algorithm that gives us the least Root Mean Square Error (RMSE) so that we will Predict values of other dataset.

- The data that we are using its 10M set of the original data set, we will have edx data set as the (training set) and validation data set to validate our prediction.

here is a sample of how edx data set looks like.

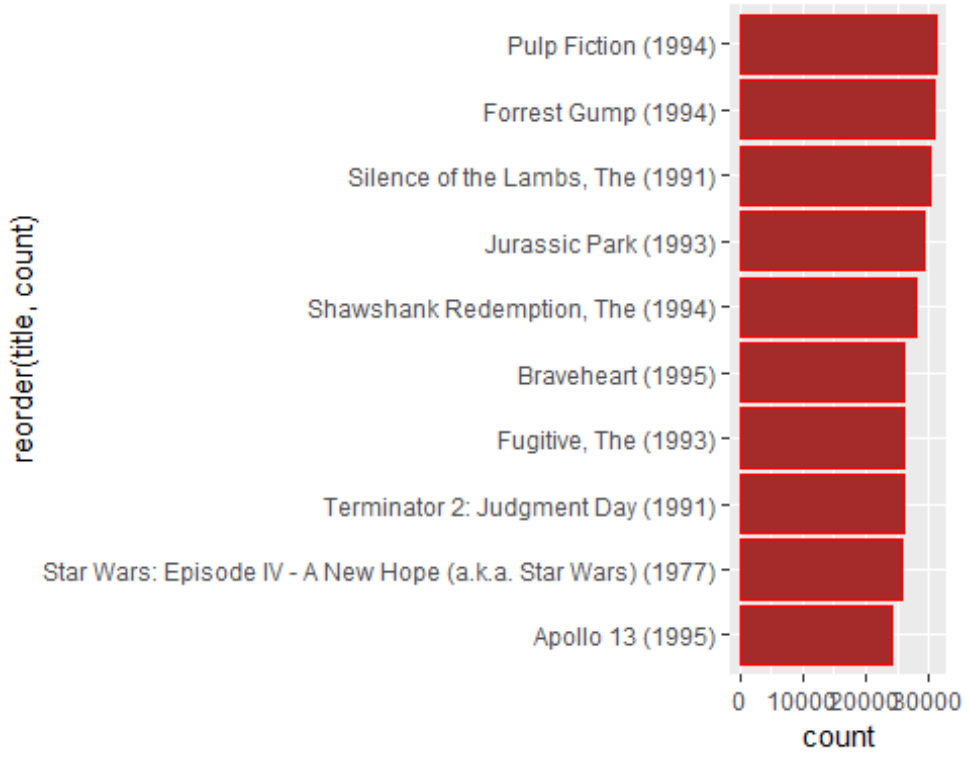
	userId	movieId	rating	timestamp	title	genres
1	1	122	5	838985046	Boomerang (1992)	Comedy|Romance
2	1	185	5	838983525	Net, The (1995)	Action|Crime|Thriller
4	1	292	5	838983421	Outbreak (1995)	Action|Drama|Sci-Fi|Thriller
5	1	316	5	838983392	Stargate (1994)	Action|Adventure|Sci-Fi
6	1	329	5	838983392	Star Trek: Generations (1994)	Action|Adventure|Drama|Sci-Fi
7	1	355	5	838984474	Flintstones, The (1994)	Children|Comedy|Fantasy

I will use different methods to develop the algorithm using the edx set. For a final test of the algorithm i will predict movie ratings in the validation set as if they were unknown. RMSE will be used to evaluate how close is the predictions are to the true values in the validation set.

The edx set consists of 9000055 rows and 6 columns and

- 69878 userID.
- 10677 movies.
- the data set has no missing values.

Here is graph for the top 10 movies according to ratings.



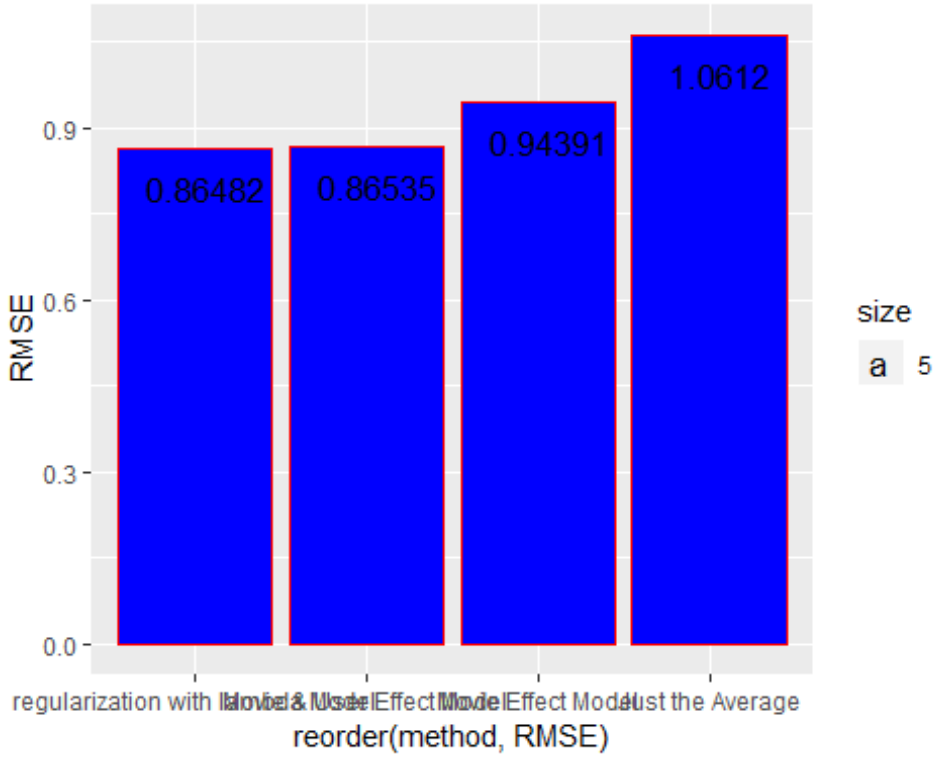
check the best fitting module

the methods that im gonna use are:

- Use just the average.
- use Movie Effect on the Model.
- use Movie and User Effect on the Model.
- use Regularization method on the Model for best fit.

Here is the results for our models.

method	RMSE
Just the Average	1.0612018
Movie Effect Model	0.9439087
Movie & User Effect Model	0.8653488
regularization with lambda Model	0.8648170



Summarization

According to the last model we could achieve RMSE of 0.86482, Which we could achieve our target.