MovieLens Project Submission

Nir Levy, January 2022

Summary

This report describes my work on the MovieLens project, within the Capstone course of Harvardx's Data Science Professional Certificate.
The goal of the project was to predict the ratings that will be given to a set of movies by unknown users, based on the ratings that were given by previous users. After downloading, cleaning and exploring the data (section 2), I checked the performance of several prediction models on the training set (section 3).
The Funk Singular Value Decomposition model (SVDF) performed the best, but it took a long time to run, so I decided to use the Popular model which performed
reasonably well and ran quickly. I applied the Popular model to the test set (section 4) and received an RMSE of 0.856 My main learning is that it is important to be aware of the trade-off between accuracy and speed. (section 5) Although the SVDF model was the most accurate, due to my limited processing power it was impractical. A less accurate model such as the 'Popular' one was much faster and still provided reasonable accuracy.

Structure of the report

The report is structured as follows: the first section describes the goal of the analysis. The second section presents
some exploratory analysis of the data, and the third section presents some prediction models that I checked out using the training set, before choosing the one that I used for predicting the test set scores (the 'Popular' model). The fourth section presents the results of applying this model on the test set.
Finally, the fifth section presents the conclusion.

Table of contents

1. Introduction
2. Downloading and exploring the data
3. Choosing a prediction model
4. Applying the model to the test set
5. Conclusion

1. Introduction

The assignment

The analysis

[Key steps that were performed]

```
5+3
```

```
## [1] 8
```