Recommender System & Latent Space Analysis Workshop

Week 10 - Session 1

Recommender System Programming

In this workshop, you will learn how matrix factorization method is applied to recommender systems. Matrix factorization is basically the breaking down of one matrix in a product of multiple matrices, and among several methods, singular value decomposition (SVD) is known to be useful for making recommendations. For the following parts, you can build a basic recommender system with a dataset of users' movie ratings. Follow the instructions below and include all results in the report.

Data You will be using one of the MovieLens datasets, which were collected by GroupLens Research at the University of Minnesota. This dataset (ml-latest-small) describes 5-star rating from MovieLens, a movie recommendation service. It contains 100,836 ratings across 9,742 movies. These data were created by 610 users between March 29, 1996 and September 24, 2018. For more information, refer to the README file or visit `https://grouplens.org/datasets/movielens/`.

Task Use the "w10_RecSys_WS_starter.ipynb" file to explore the latent factor matrix factorization method.

1. Load the movies.csv and ratings.csv and report the subset of the datasets.
2. For a simple data analysis, plot the distribution of movie ratings.
3. Generate user-movie rating matrix where the row represents userId and the column indicates movieId. Once the matrix is generated, fill null values with 0s. Report the final matrix.
4. Perform SVD on the matrix (Set the number of latent factors to 2) and report the shape of output matrices, u, s, and vt. Briefly describe each matrix.
5. Plot users and movies in the same latent space. Report your observations from the plot.
6. Calculate RMSEs using different numbers of latent factors. Briefly describe the result.

7. Using the best number of latent features, recommend top 10 movies for a specific user and briefly describe the recommendations.