Homework 3.1: Collaborative Filtering (50 points)

Machine Learning

- Read the paper Empirical Analysis of Predictive Algorithms for Collaborative Filtering linked on the course web page. You need to read up to Section 2.1, and are encouraged to read further if you have time.
- The dataset we will be using is a subset of the movie ratings data from the Netflix Prize. You need to download it from the course web page. It contains a training set, a test set, a movies file, a dataset description file, and a README file. The training and test sets are both subsets of the Netflix training data. You will use the ratings provided in the training set to predict those in the test set. You will compare your predictions with the actual ratings provided in the test set. The evaluation metrics you need to use are the Mean Absolute Error and the Root Mean Squared Error. The dataset description file further describes the dataset, and will help you get started. The README file is from the original set of Netflix files, and has been included to comply with the terms of use for this data.
- Implement the collaborative filtering algorithm described in Section 2.1 of the paper (Equations 1 and 2; ignore Section 2.1.2) for making the predictions. You may program in C, C++, Java or Python.
- What to Turn in.
  In a single zip file include the following:
  - Your code.
  - PDF Report describing your evaluation.