CS 340 Data Mining, Fall 2018

Homework 2: Recommender Systems

Due Nov 27, 11:59pm

Submit by the **blackboard system**

**Data:**

Please download a part of Netflix data attached in the blackboard system.

The zip file includes 1) *m\_u\_ratings.txt*; and 2) *selected\_films.txt*.

The *selected\_films.txt* lists out the title and release year of 2000 films, which are largely rated.

The *m\_u\_ratings.txt* is the rating records given by active users (who rated a lot of movies) to these 2000 films. Each line of this file has a format: [movie\_ID user\_ID Rate].

**Code:** Please write your own code for predicting users’ ratings.

**Tasks**:

1. **Rating prediction** (5pts)

Randomly mask 10% of ratings given by some users to some movies. Apply Collaborative Filtering method to predict these ratings. Measure the prediction quality by **MAE.**

1. What is average MAE from 10-fold cross validation when using User-based collaborative filtering?
2. What is average MAE from 10-fold cross validation when using Item-based collaborative filtering?
3. **Item Recommendation** (8pts)

Randomly choose 100 users. Mask randomly 10% of their ratings. Predict their ratings and recommend them the movies with positive predicted ratings (ratings >=4). Measure the recommendation quality by **precision@K, recall@K, MAP@K,** and **nDCG@K** (averaged on 100 users)**.**

1. What is performance of User-based collaborative filtering?
2. What is performance of Item-based collaborative filtering?