

CSE 255, Winter 2015: Homework 6

Instructions

Please submit your solution **at the beginning of the next lecture (February 23)** or outside of CSE 4102 beforehand. Please complete homework **individually**. I will accept homework as late as **Wednesday morning** in 4102 in light of the assignment being due on Monday.

Download the “Video Game Reviews” data from the course webpage:

http://jmcauley.ucsd.edu/cse255/data/homework6_7.tar.gz

This data is equivalent to the data provided for Assignment 2, with three crucial differences:

1. It is for a different category (Video Games)
2. It is smaller (1/10th the size)
3. Labels for the test data **have been provided** (‘labeled_*.txt’)

These homework exercises are intended to help you get started on potential solutions to Assignment 2.

Tasks

1. Using the *training* data (‘train.json.gz’) fit a simple predictor of the form

$$\text{rating}(\text{user}, \text{item}) \simeq \alpha.$$

Report α and the MSE of your predictor against the test data (‘labeled_Rating.txt’) (1 mark).

2. Fit a predictor of the form

$$\text{rating}(\text{user}, \text{item}) \simeq \alpha + \beta_{\text{user}} + \beta_{\text{item}},$$

by fitting the mean and the two bias terms as described in the lecture notes (with the regularization parameter $\lambda = 1$). Report the item bias $\beta_{I1102776733}$ and the user bias $\beta_{U566105319}$, and the MSE of the predictor against the test data (1 mark).

3. The following experiments should be conducted on the *training data* (‘train.json.gz’):
 - (a) Compute the Jaccard similarity between the users ‘U229891973’ and ‘U622491081’ in terms of the sets of items they have reviewed in the training set (1 mark).
 - (b) Find the user (or users) with the highest Jaccard similarity to ‘U622491081’ (if there are multiple users with the same maximum similarity, list all of them) (1 mark).
 - (c) Suppose we want to recommend a video game to the user ‘U639726733’. Possible candidates might include video games with very high Jaccard similarity (in terms of users who purchased them) to games that the user has already purchased. For each item that user ‘U639726733’ has reviewed, find the item with the highest Jaccard similarity out of all items in the training data. Which one of these might you recommend to the user and why (1 mark)?