Programming Assignment 1: Instructions

Help

When you're ready to submit your solution, go to the assignments list.

Overview

This assignment will explore non-personalized recommendations. You will be given a program stub and a data set (in .csv format) where columns represent user number, a user name/key, a movie ID, and rating. You will use these to write a program that makes basic, non-personalized recommendations.

NOTE: Actual data will not be posted until approximately September 12 (after data is collected and processed from Written assignment 0). This is a small toy dataset.

Download Sample Data — This is a comma-separated values file, with the user number, user name/key, movie ID, and rating, in that order.

Download Java Stub (be sure you can successfully compile and run before making changes)

Notes

This assignment requires you to write the code needed to parse the ratings file. It is up to you how you do this (including whether you skip ahead and use LensKit data structures or simply build your own matrix). It is particularly important to make sure you can distinguish between rated and non-rated cells in your matrix.

Deliverables

There are 2 deliverables for this assignment. Each deliverable represents a different analysis of the data provided to you. For each deliverable, you will submit a list of the top 5 movies that occur with movies x, y, and z; where x, y, and z will be uniquely provided to you before submission time (on or after September 12). Do this for each of the two association formulas described in class:

1. Simple: (x and y) / x
2. Advanced: ((x and y) / x) / ((!x and y) / !x)

Output Format

For each formula, your output should be as CSV file (a file of comma-separated values) defined as follows: Each file will have three rows (one for each movie you're computing associations for). Each row will have the movie-id of the movie assigned to you, followed by five pairs "movie id,predicted-score", from first to last, showing the top-five associated movies using that formula. A sample will be provided below when the real data set is available. Note: you will be graded on both choosing the right movies and getting correct scores (within a rounding factor); you should provide at least two decimal places precision on your predicted scores.

Examples

Example solutions and instructions on finding your test data will be posted approximately September 12