

Assigned:
May 17, 2025

Homework 5.0

Due:
May 26, 2025

Please complete the assigned problems to the best of your abilities. Ensure that the work you do is entirely your own, external resources are only used as permitted by the instructor, and all allowed sources are given proper credit for non-original content.

1. Practicum Problems

These problems will primarily reference the lecture materials and the examples given in class using Python. It is suggested that a Jupyter/IPython notebook be used for the programmatic components.

1.1 Problem 1

Load the Movielens 100k dataset (ml-100k.zip) into Python using Pandas data frames. Convert the ratings data into a utility matrix representation and find the 10 most similar users for user 1 based on the cosine similarity of the centered user ratings data. Based on the average of the ratings for item 508 from similar users, what is the expected rating for this item for user 1?

Top 10 similar users to user 1: [773, 868, 592, 880, 429, 276, 916, 222, 457,

8]

Expected rating of user 1 for item 508: 4.20

1.2 Problem 2

Load the Movielens 100k dataset (ml-100k.zip) into Python using Pandas data frames. Build a user profile on centered data (by user rating) for both users 200 and 15, and calculate the cosine similarity and distance between the user's preferences and the item/movie 95. Which user would a recommender system suggest this movie to?

User 15 similarity: 0.0000, distance: 1.0000

User 200 similarity: 0.0768, distance: 0.9232

-

END