

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?

Data processing: Automatically download and load Movielens 100k dataset, construct user item rating matrix

Data preprocessing: Centralize user ratings to improve the accuracy of similarity calculation

Similarity calculation: Use cosine similarity to find other users who are most similar to the target user

Rating prediction: Predicting the target user's rating for a specific item based on similar user ratings

Here are the results

```
Target users: 1
Target item(ID: 508): People vs. Larry Flynt, The (1996)
```

```
The 10 most similar users:
user's ID: 773.0, similarity: 0.2048
user's ID: 868.0, similarity: 0.2023
user's ID: 592.0, similarity: 0.1966
user's ID: 880.0, similarity: 0.1958
user's ID: 429.0, similarity: 0.1907
user's ID: 276.0, similarity: 0.1875
user's ID: 916.0, similarity: 0.1864
user's ID: 222.0, similarity: 0.1824
user's ID: 457.0, similarity: 0.1823
user's ID: 8.0, similarity: 0.1809
```

```
Predicting Users 1 rating for item 508 to be: 7.81
```

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?

Item 95 feature vector shape: (1, 19)

User 200 profile shape: (1, 19)

User 15 profile shape: (1, 19)

Analyzing Movie (ID: 95): Aladdin (1992)

User 200:

Cosine Similarity to Movie 95: -0.2652

Cosine Distance to Movie 95: 1.2652

User 15:

Cosine Similarity to Movie 95: -0.3259

Cosine Distance to Movie 95: 1.3259

Recommendation: Movie 95 should be recommended to User 200

Reason: User 200 has a higher similarity score to the movie's profile.

-

END