Assigned:		Due:
May 17, 2025	Homework 5.0	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?

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?

For Problem 5.1, the system first identified the ten users whose tastes most closely matched User 1, based on the cosine similarity of centered user ratings. Subsequently, by examining the actual rating data of these similar users for item 508 (a movie) and calculating their average, User 1's expected rating for this item was determined to be 4.20. This result provides a basis for personalized recommendations based on collaborative filtering principles.

Regarding Problem 5.2, the system constructed centered rating profiles for User 200 and User 15. These profiles were then compared with an idealized vector representing an absolute preference for item 95 (another movie). The calculation results showed that User 200's profile had a higher cosine similarity (0.0768) to this idealized vector than User 15 (0.0000), and also a smaller Euclidean distance. Therefore, based on these metrics, the recommender system is more inclined to suggest item 95 to User 200.