Due Date: January 3, 2022 at 23:59.

The Task:

You received a file containing data of items that users consumed (e.g., purchased, watched, etc.). The task is to build a model that predicts which future items a user is likely to consume next. The test files contain rows with user ID and two items. One item was consumed by the user and the other item was not consumed by the user. Your goal is to distinguish between the two items.

There are two test files: in the first file (RandomTest.csv), the "negative" items were sampled uniformly from all the list of all items that were not consumed by the user. In the second file (PopularityTest.csv), the "negative" item was sampled from the list of items that were not consumed by the user according to items popularity distribution (the probably for an item to be consumed).

Your goal is to fill both test files as follows:

<userId>, <item1>,<item2>,<bitClassification>

where <userId>, <item1>,<item2> are the user ID, and the IDs of the first and second items respectively as given to you. The last column <bitClassification>, is a binary classification marking your prediction as follows: '0' if you predict that the first item was the item that was liked by the user and '1' if you predict that the second item was liked by the user.

Build models for each task and follow these steps:

1) Write the objective of the model and explain its reasoning.
2) Are you using negative examples? If so, what scheme have been employed to sample the negatives?
3) Write the update step for each parameter.
4) Write a pseudo code for the algorithm.
5) What hyper-parameters do you need to tune and how?
6) Do you need a validation set? If so, how would you create it? How would you check for convergence?
7) How would you train the last \ best model?
8) Implement your model. Explain the main work items you had to take.
9) What is the Precision @K)K=1, K=10, K=50) and the MPR of your model?
   Note:
   Hit Ratio at *K* is a fraction of the test items that were ranked in the top *K*.
   MPR is similar to its definition in class.
10) Submit the test result files according to the instructions above. Also note these additional instructions:
    a. The name of the files should be made from the student ID numbers separated with an underline.
       i. random_<id1>_<id2>_<id3>.csv
       ii. popularity_<id1>_<id2>_<id3>.csv
    b. The files content should be CSV in the same order as the test file you received.
    The report and the results should be emailed to: amitshreiber@mail.tau.ac.il
11) You also need to submit your python code along with the report.