Homework 3: Neural Collaborative Filtering

Form:	Jupyter notebook file including images and text explanation
Language:	English
Requirements:	The report should be clear, readable and include all code documented
Submission:	.ipynb file via Moodle. The file name should include the students' ids
Contact:	
Deadline for submission:	**February 7, 2021**

Students will form teams of two people each, and submit a single homework for each team. The same score for the homework will be given to each member of the team.

Submit your solution in the form of an Jupyter notebook file (with extension ipynb). Images of graphs or tables should be submitted as part of the notebook itself. The code used to answer the questions should be included, runnable and documented in the notebook. Python 3.6 should be used.

The goal of this homework is to let you understand the concept of recommendations based on implicit data which is very common in real life, and learn how 'Deep neural networks' components can be used to implement a collaborative filtering and hybrid approach recommenders. Implementation example is presented in the NeuralCollaborativeFiltering_Implicit notebook in Moodle.

Submission: Submission of the homework will be done via Moodle by uploading a Jupyter notebook file containing code, plots and explanations. The homework needs to be entirely in English. The deadline for submission of Homework 1 is set to February 7, 2021 end of day Israel.

We will use a dataset based on the MovieLens 1M rating dataset after some pre-processing to adapt it to an implicit feedback use case scenario. You can download the dataset used by this implementation of the paper Neural Collaborative Filtering or from the NeuralCollaborativeFiltering_implicit notebook in Moodle.

Question 1: Dataset preparation (10 points)

a. This implementation contains one file for training: ml-1m.train.rating and two files for testing: ml-1m.test.rating; ml-1m.test.negative. Explain the role and structure of each file and how it was created from the original MovieLens 1M rating dataset.
b. Explain how the training dataset is created.
c. Explain how the test dataset is created.

Question 2: Neural Collaborative filtering (50 points)

a. Build the following four models using the neural collaborative filtering approach: Matrix Factorization (MF), Multi layer perceptron (MLP), Generalized Matrix Factorization (GMF) and NeuroMatrixFactorization (NMF).

b. Train and evaluate the recommendations accuracy of three models: MF or GMF, MLP and NMF. Compare the learning curve and recommendations accuracy using NDCG and MRR metrics with cutoff values of 5 and 10. Discuss the comparison.
c. How the values of MRR and NDCG differ from the results you got in the previous exercises which implemented learning from explicit data. What is the difference in preparing the dataset for evaluation.
d. How will you represent an item and measure item similarity for the NeuMF model?

Question 3: Loss function (40 points)

a. One of the enhancements presented in the Neural Collaborative Filtering paper is the usage of probabilistic activation function (the sigmoid) and binary cross entropy loss function. Select one of the models you implemented in question 2 and change the loss function to a Mean Squared Error and the activation function of the last layer to RELU. Train the model and evaluate it in a similar way to what you did in question 2. Compare the results and discuss.

Good luck