Homework 2: Content Based Recommendations

Form:	Zip file including Jupyter notebook file and images.
Language:	English
Requirements:	The report should be clear, readable and include all code documented
Submission:	Zip file via Moodle. The file name should include the students' ids
Contact:	
Deadline for submission:	**January 10, 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 PNG or JPG files. 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 details of content based recommender systems and its pros\cons comparing to other recommender system approaches. Implementation examples are presented in the Movielens_content_meta and Movielens_content_text notebooks in Moodle.

Submission: Submission of the homework will be done via Moodle by uploading a zip file containing a Jupyter notbook and images. The homework needs to be entirely in English. The deadline for submission of Homework 1 is set to January 10, 2021 end of day Israel.

We will use the same dataset we used for exercise #1 MovieLens 100K rating dataset.

Split your dataset to train and test. You can use the built-in splits included in the dataset.

Question 1: Item Similarity (40 points)

a. Build a movie profile vector based on the item features of your choice. Suggested options: item genres, item title.
b. Build a function which provides the 5 most similar items to a specific item. Item to item similarity is calculated using the cosine similarity metric.
c. Use the above function to find most similar items for 2 items from the dataset. Explain your results.
d. Use the matrix factorization item representation you built in exercise 1 to find the most similar items for the same 2 items you used above. Compare between the results.

Question 2: Content based recommendation (60 points)

a. Build a function which recommends 5 most relevant items to a user.
b. Test your recommender system on 2 users. Explain your results.

    c.   Use the MRR metric to evaluate your recommender system on the test set. Use a cutoff value of 5.

    d.   Compare the results of the content based recommender system to the matrix factorization recommender system using the same train and test set. Use the MRR metric for the comparison.

    e.   Discuss the advantages and disadvantages of matrix factorization vs. content based recommender systems (refer to accuracy, complexity, explainaiblity and new user\new items) and when will you recommend using each approach.

Good luck