

# Home Assignment 1

This assignment consists of 2 theoretical and 3 practical tasks. The description of the theoretical tasks can be found in the current document. The descriptions and templates for the practical tasks can be found in the corresponding ipynb files. Your task is to complete the code and answer questions where necessary. Please don't remove any of the code we provide.

You must submit one pdf file with the answers to the theoretical problems and three ipynb files with the solutions to the practical problems. Please use LaTeX to type the answers to the theoretical questions. For the practical tasks, develop a solution yourself and provide a reproducible and readable code. Your code should be reproduced without any mistakes with "Run all" mode. You may use `evaluation.py` and `dataprep.py` modules from previous seminars.

You will receive a total of 100 points for this assignment, plus some extra points if you make some improvements to the code or method (please discuss this with your instructor). If you hand in the assignment the day after the deadline, you will be penalised 50 points, and every day after that you will lose 1 point.

## 1 Problem 1 (5 pts)

### Theoretical Task

It is often stated, that pure content-based recommendation models provide very low level of personalization to users.

Prove this claim using a standard regression-based formulation for the case when a single global model is learned in the form:

$$r \approx \theta x + \epsilon,$$

where vector  $x$  encodes some features of both users and items (e.g., user attributes and item characteristics), and  $\theta$  are the corresponding learnable weights of the regression model. Recall that personalization task is formulated as the ranking problem of the top- $n$  best-matching items:

$$\text{toprec}(u, n) = \arg \max_i^n r_{ui},$$

where  $r_{ui}$  is the relevance score assigned by the model to item  $i$  for user  $u$ .

## 2 Problem 2 (5 pts)

### Theoretical Task

You're given the matrix of interactions between 3 users and 6 items:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Is it possible to build a personalized recommendation model with this data? Explain your answer.

## 3 Problem 3 (20 pts)

### Content-based models

`hw1_content_based.ipynb`

In this problem you need to train simple content-based model for each user individually in order to achieve some level of personalization. Thus your model may be considered as ensemble of the personal models.

## 4 Problem 4 (30 pts)

### KNN models

`hw1_knn.ipynb`

Implement two variants of user-based KNN for the top-n recommendations task when:

- similarity matrix is symmetric,
- similarity matrix is asymmetric

Recall, there's no reason for implementing row-wise weighting scheme in user-based KNN. So choose the weighting scheme wisely.

Test your solution against both weak and strong generalization. Follow the "most-recent-item" sampling strategy for constructing holdout. Report evaluation metrics, compare the models, and analyse the results.

## 5 Problem 5 (40 pts)

### Matrix factorization models

`hw1_matrix_factorization.ipynb`

Your task is to implement federated collaborative filtering model for privacy-preserving personalized recommendation system, compare it with one or two

baseline models, tune hyper-parameters and to examine, how the weights and loss value changes during training.