

## Session 8: Recommender systems

### Exercise List, Fall 2023

---

#### Basic comprehension questions.

Check that you can answer them before proceeding. Not for credit.

1. True or false or criticize: Content-based recommender systems require longer customization time than Collaborative Filtering approaches in new application scenarios.
  2. Explain the cold start problem in recommender systems. Provide a general definition and explain how it concerns recommendation methods.
  3. True or false or criticize: Matrix factorization methods based on SVD can be used to solve the cold start problem in Collaborative Filtering recommendation systems.
- 

**Exercises for credit.** Solving three of these exercises (not solved by the instructors in class) suffice for full credit for this assignment.

#### Exercise 1

Suppose a recommender system that has to estimate  $r(a, s)$ , i.e. the preference of user  $a$  for a certain item  $s$  based on the preferences of other users from a set  $U$  over the same product. Suppose that  $r(a, s)$  is estimated with  $pred(a, s)$ , that is defined as a simple mean over  $U$ , namely

$$pred(a, s) = \frac{1}{|U|} \sum_{b \in U} r(b, s).$$

Analyze the limitations of the formula above and explain how it could be improved, proposing better formulae for estimating  $pred(a, s)$ . Suppose that  $U$  is given (cannot be changed) and that the scores of user  $a$  or users in  $U$  on other items are available. Explain the rationale behind your choices.

### Exercise 2

Discuss ethical issues involved in real applications of recommender system techniques.

### Exercise 3

We have designed a sophisticated recommender system based on collaborative filtering and we would like to evaluate its performance by comparing it against a simple system that does not employ collaborative filtering. We wish to make sure that the sophistication of our system is worth.

1. What system would you use for comparison?
2. Why?

### Exercise 4

A team of former students of this course is developing an online social network where users can follow other users (but a user cannot follow him/herself). Formally, the social network is represented by a binary adjacency matrix  $A = \{a_{ij}\}$  that indicates who follows who, i.e.  $a_{ij} = 1$  if user  $i$  follows user  $j$  ( $a_{ij} = 0$  otherwise). The system applies collaborative filtering on the adjacency matrix to make recommendations of users to follow based on  $k$ -nearest neighbours, user-to-user similarity and normalized Hamming distance as similarity score. Given the following adjacency matrix

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

1. Calculate the normalized Hamming distance of user 3 to all other users (including him/her self and all the columns of the adjacency matrix). You have to fill the following table

User	Hamming distance	Normalized Hamming distance
1		
2		
3		
4		
5		

**Use that matrix to find the nearest neighbours in the remainder of the exercise.**

2. Predict the potential interest of user 3 in following user 2 with  $k = 1$  in two ways: (a) majority vote and (b) taking into account the average rating of the users and their similarity. For each procedure, indicate if you would recommend user 2 to user 3.
3. Same as before with  $k = 2$ .
4. To sum up, fill the following summary table on the predicted interest of user 3 in following user 2.

	$k = 1$	$k = 2$
Identity of nearest users		
Predicted interest with (a)		
Recommendation with (a)		
Predicted interest with (b)		
Recommendation with (b)		

### Exercise 5

Same as in Exercise 4 but on the interest of user 3 in following users 1 and 5.

### Exercise 6

Same as in Exercise 4 but with item-to-item similarity and prediction of the interest of user 3 in following users 1, 2 and 5.

### Exercise 7

Users have been expressing their degree of preference for items on a scale from 1 to 6. The ratings of two users on 7 items are the following:

user	item 1	item 2	item 3	item 4	item 5	item 6	item 7
1	1	4	5	1	3	6	5
2	2	3	4	4	4	5	3

We define the similarity between two users as the Pearson correlation between their ratings.

1. What is the degree of similarity between user 1 and user 2?

2. Show the ratings of another user  $x$  whose similarity with respect to user 1 is maximized? Give the vector of ratings of  $x$ , the value of the correlation and justify why it is maximum.
3. Show the ratings of another user  $y$  whose similarity with respect to user 1 is minimized? Give the vector of ratings of  $y$ , the value of the correlation and justify why it is minimum.
4. Suppose a user  $z$  that has been rating items by rolling a fair die. What would be the similarity between  $z$  and user 1?
5. Is the similarity between user 1 and 2 significant?