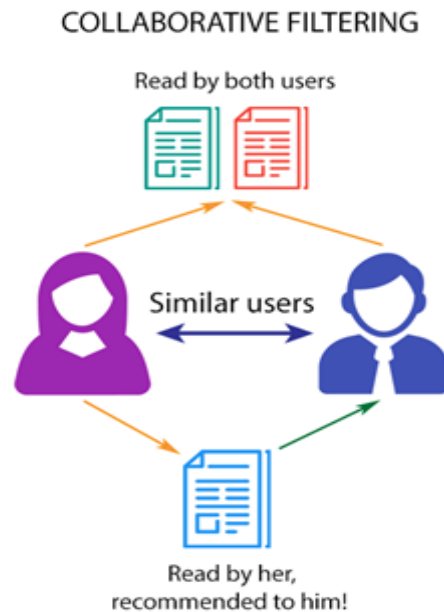


Recommendation Systems

(Collaborative Filtering)

A Recommendation System aims at suggesting optimal recommendation of items to various users. Collaborative Filtering approach is based on the ratings of the users for various items in a domain.



1. Give a formal description of this application in terms of Task, Experience, and Performance. (Make it a well posed problem)
2. A dataset is provided consisting of the ratings of various products from Amazon users. Write a function $Utility_matrix(data)$ that converts the uploaded dataset 'amazon_rating' into an Utility matrix where columns represent items and rows represent users and the values represent rating.

For instance, consider the following matrix

Item	User	Rating
9707716436	A3KP1BUNRQY69J	3
9707716436	A3P7REQQEXHATA	1
9707716436	A39XKVLWEHYCI1	1
9707716436	A340KNBKZ86MKN	5
9707716436	A207HOQVQ3F552	5
9707716371	AT9HSLVQB7OFT	1
9707716371	A39D5S9WPOY5XR	5
9707716371	AUBJWP9VI3FUZ	3

The given matrix should be converted to Utility matrix as follows. Item 1 corresponds to Item 9707716436 and Item 2 corresponds to 9707716371. Similarly, for users

User	Item 1	Item 2
A3KP1BUNRQY69J	3	
A3P7REOQEXHATA	1	
A39XKVLWEHYCI1	1	
A340KNBKZ86MKN	5	
A207HOQVQ3F552	5	
AT9HSLVQB7OFT		1
A39D5S9WPOY5XR		5
AUBJWP9VI3FUZ		3

3. Write a function $Normalize(U)$ to normalize the ratings of the users for the items.
4. Write a function $PearsonCorr(x,y)$ to find the similarity between the ratings of the items rated by both users x and y .

S_{xy} = items rated by both users x and y

$$sim(x, y) = \frac{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)(r_{ys} - \bar{r}_y)}{\sqrt{\sum_{s \in S_{xy}} (r_{xs} - \bar{r}_x)^2} \sqrt{\sum_{s \in S_{xy}} (r_{ys} - \bar{r}_y)^2}}$$

5. Write a function $NearestNbrs(U, q, k)$ that takes the normalized utility matrix 'U', the query 'q' as the rating vector of a user, and finds the best 'k' neighbours from 'U' based on the similarity metric *Pearson Correlation coefficient*
6. Write a function $PredictRating(x, s)$ that predicts the rating of user x for item i based on the following formula

$$r_{xi} = \frac{\sum_{y \in N} s_{xy} \cdot r_{yi}}{\sum_{y \in N} s_{xy}}$$

where $s_{xy} = sim(x, y)$ as given in Q.4

7. **Extra credit** : Evaluate the Rating based on RMSE. A part of the utility matrix can be considered as Test Data as shown in the following figure.

Utility Matrix R : Evaluation

