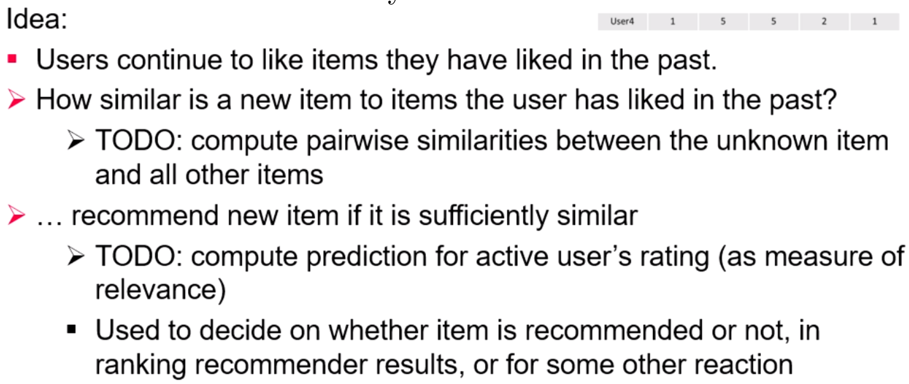
Overview

- asssuming items x and y are similar
  - similarity score needs to be above threshhold
- if A liked x also recommend y



User4	1	5	5	2	1

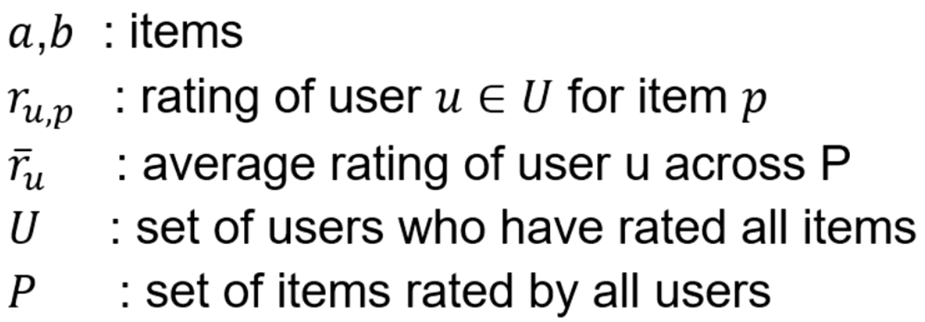
Idea:

- Users continue to like items they have liked in the past.
- How similar is a new item to items the user has liked in the past?
  - TODO: compute pairwise similarities between the unknown item and all other items
- … recommend new item if it is sufficiently similar
  - TODO: compute prediction for active user's rating (as measure of relevance)
  - Used to decide on whether item is recommended or not, in ranking recommender results, or for some other reaction

-
- interests might change over time

Similarity Score

- cosine similarity of item vectors
  - does not account for different user rating tendencies
    - some easily 10/10, some 8/10 at max
- cosine similarity of centered item vectors
  - normalize user ratings by each user's average rating value



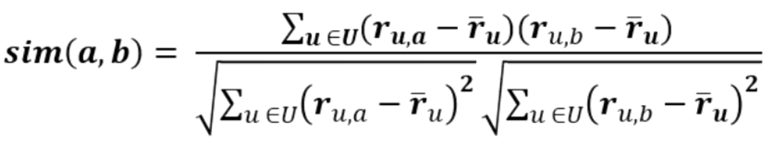
a, b : items

$r_{u,p}$: rating of user $u \in U$ for item p

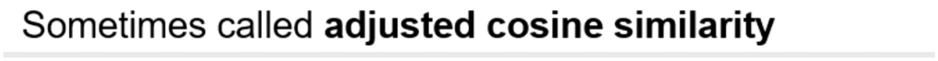
$\bar{r}_u$: average rating of user u across P

U : set of users who have rated all items

P : set of items rated by all users

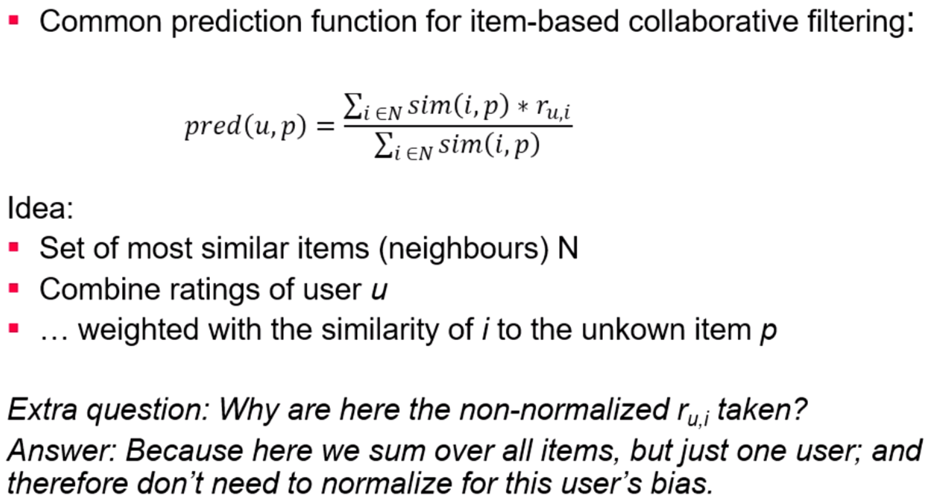


$$sim(a,b) = \frac{\sum_{u \in U}(r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U}(r_{u,a} - \bar{r}_u)^2}\sqrt{\sum_{u \in U}(r_{u,b} - \bar{r}_u)^2}}$$

- 

  Sometimes called **adjusted cosine similarity**

Prediction

- 

  - Common prediction function for item-based collaborative filtering:

  $$pred(u,p) = \frac{\sum_{i \in N} sim(i,p) * r_{u,i}}{\sum_{i \in N} sim(i,p)}$$

  Idea:

  - Set of most similar items (neighbours) N
  - Combine ratings of user *u*
  - … weighted with the similarity of *i* to the unkown item *p*

  Extra question: Why are here the non-normalized $r_{u,i}$ taken?
  Answer: Because here we sum over all items, but just one user; and therefore don't need to normalize for this user's bias.

- fine tuning via
  - more similarity if users agree on controversial items
    - ∗ controversial if high variance in ratings
    - ∗ more weight of those items
  - more weight to ratings of similar users