Cosine Similarity
Item Based Predictions

77B Recommender Systems
Item Based Collaborative Filtering

- In previous lectures we looked at similarities between users (user based collab. Filt.)
  Now we look at similarities between items (item based collab. filtering)

- Find the nearest neighbors of a particular item \( i \) by some measure of similarity.

- The similarity can depend on predefined features (genre of a movie) \( \rightarrow \) content based filtering or on the rating behavior of people for this item (item based collab. Filt.)

- Average the ratings of these nearest neighbors for some item to predict the rating for a new item
Cosine Similarity Measure

\[ CSM(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} (x_i)^2} \sqrt{\sum_{i=1}^{n} (y_i)^2}} \]

- Note that we are now interested in similarity between items, so the sum is over users
- If there many people who rated the items similarly then the cosine is large
- If \( x_i \) is approximately equal to \(-y_i \) then it is large and negative.
- Two vectors \( x, y \) that are orthogonal have similarity 0
- We do NOT subtract the mean rating!
- We divide by the scale (standard deviation) so that a wider range of values has no influence.
Prediction

\[ x_{iu} = \frac{\sum_{j=1}^{I} csm(i, j) \times y_{ju}}{\sum_{j=1}^{I} csm(i, j)} \]

- We assume we have all ratings (relax later)
- csm is the similarity between two items.
- We weight item’s contributions with their similarity
- Divide by total weight so that sum of weights = 1