Neighborhood methods

Friday, 16 April 2021 08:46

Interaction matrix

$[r_{ui}]$

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

5 users

6 items

Interaction vectors as features

Users

$r_0 = [1,1,0,0,1,0]$ $r_1 = [0,0,1,1,0,0]$

$r_2 = [1,0,0,1,1,1]$ $r_3 = [0,0,1,0,0,1]$

$r_4 = [0,1,1,0,1,1]$

Items

$i_0 = [1,0,1,0,0]$ $i_1 = [1,0,0,0,1]$

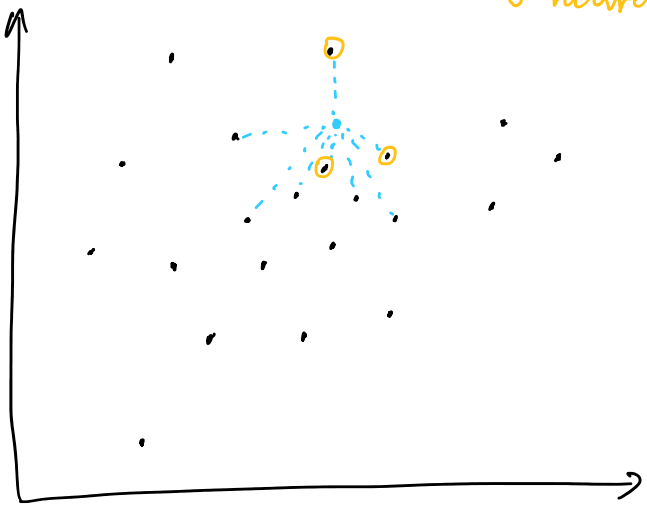
$i_2 = [0,1,0,1,1]$ $i_3 = [0,1,1,0,0]$

$i_4 = [1,0,1,0,1]$ $i_5 = [0,0,1,1,1]$

These features can directly be treated as input for an ML algorithm

Neighborhood

3 nearest neighbors



User-based neighborhood methods

Voting by neighbors

1. Find k-nearest neighbors to the active user.
2. Identify items those neighbors interacted with.
3. For every such item calculate its score as an average neighbor vote weighted with similarities.
4. Recommend items with the highest score the active user has not interacted with.

Example

neighbor 1: similarity = 0.9 items = {item 1, item 3}
neighbor 2: similarity = 0.7 items = {item 2}
neighbor 3: similarity = 0.4 items = {item 2, item 3}

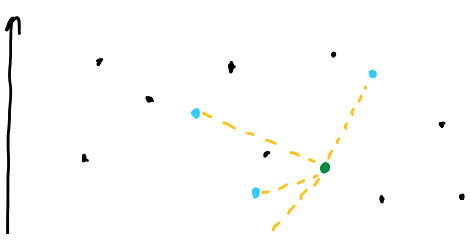
$$\text{item 1 score} = \frac{0.9}{0.9+0.7+0.4} = 0.45$$

$$\text{item 2 score} = \frac{0.7+0.4}{0.9+0.7+0.4} = 0.55$$

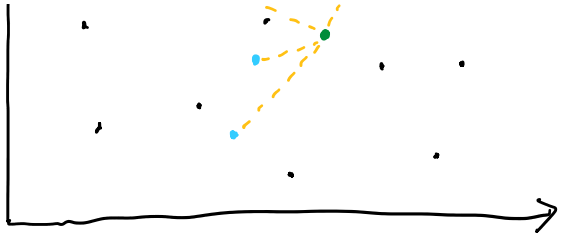
$$\text{item 3 score} = \frac{0.9+0.4}{0.9+0.7+0.4} = 0.65$$

Item-based neighborhood methods

1. For every item the active user has not interacted with find its overall similarity (sum of similarities) to items the active user has interacted with.
2. Recommend items with the highest score.



- items already bought
- scored item
- other items
- ---- summed similarities

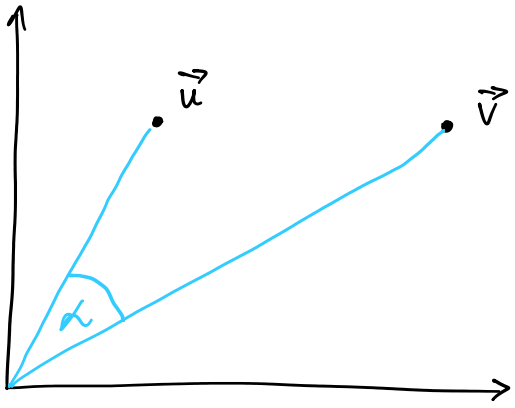


- - - - summed similarities

How to find neighbors?

Similarity measures

Cosine similarity



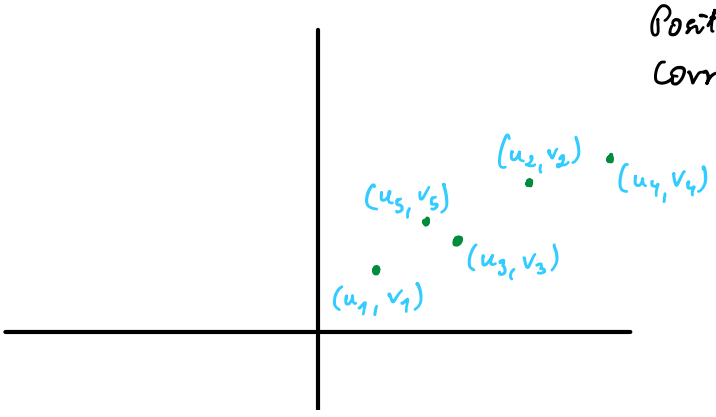
$$sim(u, v) = Cos(u, v) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|}$$

Especially useful for ratings

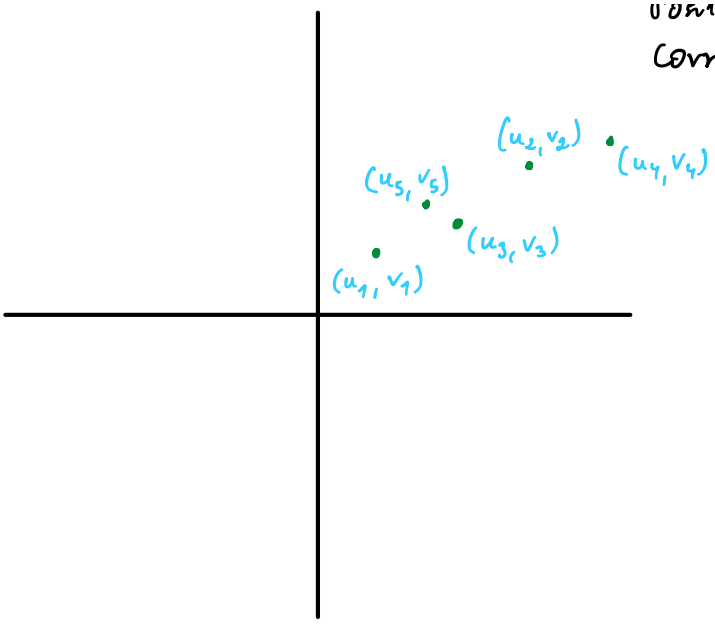
Pearson similarity (Pearson correlation)

$u = (u_1, u_2, \dots, u_m)$ - user 1 interactions (or ratings)

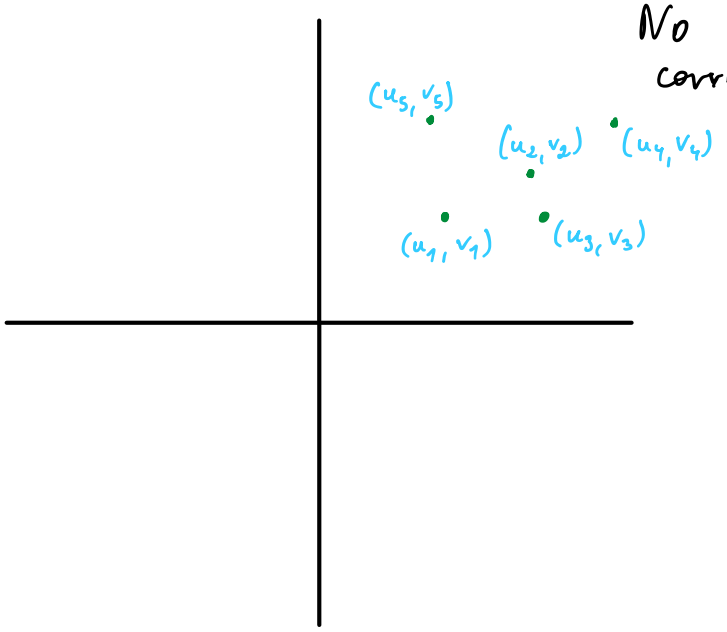
$v = (v_1, v_2, \dots, v_m)$ - user 2 interactions (or ratings)



Positively correlated



No or little correlation



Negatively correlated