

RS ASSIGNMENT 01

Bilal Ahmed Khan

20K-0183

Section : B

QUESTION NO. 01

Part a:-

$$\Rightarrow r(U_1, U_2) = \frac{(3-3.33) * (2-3)}{\sqrt{(-0.33)^2 * (-1)^2}}$$

$$\boxed{r(U_1, U_2) = 1}$$

$\Rightarrow r(U_1, U_3)$: Not possible

$$\Rightarrow r(U_1, U_4) = \frac{(4-3.33)(3-3.33)}{\sqrt{(4-3.33)^2 * (3-3.33)^2}} = -1$$

$$\boxed{r(U_1, U_4) = -1}$$

$$r(U_1, U_5) = \frac{(3-3.33) \times (3-3.75) + (3-3.33) \times (4-3.75)}{(4-3.33)(4-3.75)}$$

$$\sqrt{(0.33^2 + 0.33^2 + 0.67^2) \times (0.75^2 + 0.25^2 + 0.25^2)}$$

$$r(U_1, U_5) = 0.491$$

For $K=2$, the nearest neighbors of user 1 are user 2 & user 5

Part B

Collaborative user-based filtering for user 1 & item 3

Most similar users (User 2 & 5) ~~are~~ haven't rated item 03

Thus we must choose a neighbor which has

- 1) Be closest in vector space
- 2) Have a raw rating for the item we wish to predict the rating for.

Thus selecting user 04

$$\text{rating}(U_1, I_3) = 3.33 + \frac{0.67(-1)}{(-1)}$$

$$\text{rating}(U_1, I_3) = 4$$

Part c) Item Based Collaborative filtering for

i) $R(U_3, I_2)$

$$r(I_1, I_2) = \frac{(0.25)(-0.75)}{\sqrt{0.25^2 + 0.75^2}} = -1$$

$$r(I_1, I_2) = -1$$

$$r(I_3, I_2) = 0$$

$$r(I_4, I_2) = \frac{(0.33 \times -0.33) + (-0.75 \times 0.25)}{\sqrt{(0.33^2 + 0.75^2) + (0.33^2 + 0.25^2)}}$$

$$r(I_4, I_2) = -0.232$$

$$r(I_5, I_2) = \frac{(0.33 \times 0.6) + (-0.75 \times 0.25)}{\sqrt{0.75^2 + 0.33^2} \times \sqrt{0.6^2 + 0.25^2}}$$

$$r(I_5, I_2) = -0.697$$

Item 04 & Item 05 are the most similar to Item 03 but none of them have rated item 02, Thus we calculate

~~$R(I_3, I_2)$~~

a/c to the similarity of item 05

$$R(U_3, I_2) = 3 + \frac{0(1)}{(1)}$$

$$R(U_3, I_2) = 3$$

ii) $R(U_4, I_4)$

$$r(I_1, I_4) = \frac{1(-1) + (0.25)0.25}{\sqrt{(1+0.25^2) \times (1^2+0.25^2)}}$$

$$r(I_1, I_4) = -0.88$$

$$r(I_2, I_4) = \frac{-0.33(-0.33) + 0.25(-0.75)}{\sqrt{(0.33^2+0.75^2) \times (0.33^2+0.25^2)}}$$

$$r(I_2, I_4) = -0.23$$

$$r(I_5, I_4) = \frac{-0.33(0.67) + 0.25(0.25)}{\sqrt{(0.67^2+0.25^2) \times (0.33^2+0.25^2)}}$$

$$r(I_5, I_4) = -0.54$$

For $K=2$, I_5 & I_2 are the nearest neighbors

For rating prediction

$$R(U_4, I_4) = 3.33 + \frac{(-0.33)(-0.88) + (-0.33)(-0.54)}{|-0.88| + |-0.54|}$$

$$R(U_4, I_4) = 3.66 \approx 4$$

Q1: part d)

Mode measures the central tendency of data while mean takes all the values of the user into account.

Using a mode ~~and~~ centered function would converge our ratings towards the central tendency of the data.

Therefore for general data we tend to use mean centered function unless the ratings are very extreme in which case we would opt for a mode centered function.