20k-0212
BSCS-6B
Date: M T W T F S S

Recommender Systems Assignment 1

1)

	item 1	2	3	4	5	Mean
user 1	?	-0.33	?	-0.33	0.67	3.33
2	1	?	?	-1	?	3
3	?	?	0	?	?	3
4	-0.33	?	0.67	?	-0.33	3.33
5	0.25	-0.75	?	0.25	0.25	3.75

a) Pearson correlation b/w:

U_1 and U_2 = $\frac{(-0.33)(-1)}{\sqrt{(-0.33)^2}\sqrt{(-1)^2}}$ = 1

U_1 and U_3 = cannot be calculated

U_1 and U_4 = $\frac{0.67 \times -0.33}{0.67 \times 0.33}$ = -1

U_1 and U_5 = $\frac{(-0.33)(-0.75) + (-0.33)(0.25) + (0.67)(0.25)}{\sqrt{(0.33)^2 + (0.33)^2 + (0.67)^2}\sqrt{(0.75)^2 + (0.25)^2 + (0.25)^2}}$ = 0.491

The two nearest neighbours ~~are~~ of U_1 are U_2 and U_5

b) Since both nearest neighbours of U_1 have not rated item 3, we will consider the next nearest neighbours which in this case will become U_3 and U_4. Since there is no value for a pearson correlation with U_3, we will only consider U_4.

Date: M T W T F S S

$3.33 + \frac{0.67 \times 0.491}{|0.491|}$: ~~0.67 ≅ 1 rating~~ 2.66 ≅ 3

c) Since U_3 has no items rated besides item 3, it is not possible to calculate item similarity. Hence, we will first calculate the ratings of $U_3 I_1$, $U_3 I_2$, $U_3 I_4$ and $U_3 I_4$ using user based similarity measure. This method is known as cascading.

Similarity b/w

U_3 and U_1

first calculate similarity of I_2 with all others.

User	item 1	2	3	4	5	Mean
User 1	?	~~-0.25 -0.33~~	~~-0.25 -0.33~~	~~-0.25 -0.33~~	~~0.75 -0.67~~	3.25
2	1	?	?	-1	?	3
3	?	?	0	?	?	3
4	-0.33	?	0.67	?	-0.33	3.33
5	0.25	-0.75	?	0.25	0.25	3.75

I_2 and I_1: $\frac{0.25 \times -0.75}{(0.25)^2 * 0.75} = -1$

I_2 and I_3: $\frac{-0.25 \times -0.25}{-0.25 \times -0.25} = 1$

I_2 and I_4: $\frac{-0.25 \times -0.25 + -0.75 \times 0.25}{\sqrt{(0.25)^2 + (0.75)^2} \times \sqrt{(0.25)^2 + (0.25)^2}} = -0.2$

Date:
M T W T F S S

I_2 and I_5 : $\frac{-0.25 \times -0.75 + -0.75 \times 0.75}{\sqrt{(0.25)^2 + (0.75)^2}\sqrt{(0.75)^2 + 0.25^2}} = 0$

Since there does not exist a rating for the nearest neighbours, ~~I_3, we will consider~~ of any other items, we will calculate it according to the similarity of item 3. Our results will be as follows:

$3 + \left(\frac{0 \times 1}{1}\right) = 3$

$R(U_4, I_4)$

Similarity b/w

I_1 and I_4 : $\frac{1 \times -1 + 0.25 \times 0.25}{\sqrt{1 + 0.25^2} \times \sqrt{1 + 0.25^2}} = \cancel{-0.036}\ -0.88$

I_3 and I_5 : $\frac{-0.33 \times 0.67 + 0.25 \times 0.25}{\sqrt{0.33^2 + 0.25^2} \times \sqrt{0.67^2 + 0.25^2}} = \cancel{-0.35}\ -0.53$

I_2 and I_4 : $\frac{-0.33 \times -0.33 + -0.75 \times 0.25}{\sqrt{0.33^2 + 0.75^2} \times \sqrt{0.33^2 + 0.25^2}} = -0.23$

nearest neighbours : I_2 and I_5

$3.33 + \frac{-0.33 \times -0.88 + -0.33 \times -0.54}{|-0.88| + |-0.54|} = 3.66 \simeq 4$

Date:
M T W T F S S

d) Mode is a measure of central tendency that represents the most repeated value in a dataset whereas mean is the average of all values. Therefore, using mean or mode for prediction can lead to very different results.

Using mode can be useful where users give extreme ratings, since mode is not affected by anomalies of or values on the far edges of the spectrum

However mean is generally better than mode since it takes into account all the values and also captures subtle differences in user preferences. It is more accurate in providing estimates of user preference.