(Item-based collaborative filtering) Suppose that we have the user-item rating data as follows:

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Alice	**5**	**2**	**3**	**3**		**?**
User 1	2		4		4	1
User 2	3	1	3		1	2
User 3	4	2	3	1		1
User 4	3	3	2	1	3	1
User 5		3		1	2	
User 6	4	3		3	3	2
User 7		5		1	5	1

(1) Calculate the item similarities using adjusted cosine similarity between

- Item 6 and (Item 1 ~ Item 4)
- Item 5 and (Item 1 ~ Item 4)

$R_{Alice} = 3.25$

$R_{User1} = 2.75$

$R_{User2} = 2$

$R_{User3} = 2.2$

$R_{User4} = 2.17$

$R_{User5} = 2$

$R_{User6} = 3$

$R_{User7} = 3$

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Alice	1.75	-1.25	-0.25	-0.25		
User 1	-0.75		1.25		1.25	-1.75
User 2	1.00	-1.00	1.00		-1.00	0.00
User 3	1.80	-0.20	0.80	-1.20		-1.20
User 4	0.83	0.83	-0.17	-1.17	0.83	-1.17
User 5		1.00		-1.00	0.00	
User 6	1.00	0.00		0.00	0.00	-1.00
User 7		2.00		-2.00	2.00	-2.00

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}}.$$

sim(6,1) = -0.42

sim(6,2) = -0.70

sim(6,3) = -0.68

sim(6,4) = 0.93

sim(5,1) = -0.38

sim(5,2) = 0.92

sim(5,3) = 0.15

sim(5,4) = -0.91

(2) Suppose that we want to recommend an item for Alice among Item 5 and Item 6. Which item will you recommend for Alice using item-based recommendation? Provide the reason for your choice (Use Cosine similarity; show the formula and complete the calculation. You can use your smart phone calculator).

Use the top-2 similar items for N in the following formula:

$$P_{u,i} = \frac{\sum_{\text{all similar items, N}}(s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, N}}(|s_{i,N}|)}$$

$$P_{Alice,5} = \frac{sim(5,2) \times 2 + sim(5,3) \times 3}{|sim(5,2)| + |sim(5,3)|} = 2.14$$

$$P_{Alice,6} = \frac{sim(6,4) \times 3 + sim(6,1) \times 5}{|sim(6,4)| + |sim(6,1)|} = 1.13$$

Suppose that we have a table of user-item ratings

	Book1	Book2	Book3	Book4
Alice	1	2	5	?
George	5	?	1	?
Mary	?	?	4	3
Tom	1	1	5	4

For those unrated items, the following table shows the scores estimated by our recommender system and the actual scores obtained by asking the users to rate them:

	Estimated	Actual
$r_{Alice,Book4}$	3.5	4
$r_{George,Book2}$	4	4
$r_{George,Book4}$	4	1
$r_{Mary,Book1}$	1	2
$r_{Mary,Book2}$	2	1

Calculate the RMSE and MAE respectively.

$$\mathrm{RMSE} = \sqrt{\frac{1}{|\mathscr{T}|} \sum_{(u,i)\in\mathscr{T}} (\hat{r}_{ui} - r_{ui})^2}$$

Rmse = 1.5

$$\mathrm{MAE} = \sqrt{\frac{1}{|\mathscr{T}|} \sum_{(u,i)\in\mathscr{T}} |\hat{r}_{ui} - r_{ui}|}$$

mae = 0.45

The following table lists a set of evaluation metrics and metric descriptions. Draw lines that connect the metric names with the correct metric descriptions.

● Sparsity (5)	◆ The system cannot draw any inferences for users or items about which it has not yet gathered sufficient information
● Precision (3)	◆ The average deviation between computed recommendation scores and actual rating values for all evaluated users and all items in their testing sets
● Recall (6)	◆ The proportion of relevant instances that are retrieved
● Accuracy (4)	◆ The proportion of true results (both true positives and true negatives) among the total number of cases examined
● Coldstart (1)	◆ The ratio of empty and total entries in the user-item matrix
● MAE (2)	◆ The proportion of retrieved instances that are relevant

(Precision) For an active user, two recommender systems return the following estimated scores for 8 unrated (=not purchased yet) items.

	I1	I2	I3	I4	I5	I6	I7	I8
System 1	5	4	1	4	2	2	3	2
System 2	5	2	4	4	3	1	1	1

The actual purchases and the rate for those items by the active user are

	I1	I2	I3	I4	I5	I6	I7	I8
Purchase	O	O		O		O		
Rate (=gain)	5	4		3		2		

Let say we recommend the active user the top-4 items in the order of decreasing (estimated) scores. Among those two recommender systems, which one makes better recommendation in terms of precision?

System1: 3/4

System2: 2/4

(Recall) Which one makes better recommendation in terms of recall?

System1: 3/4

System2: 2/4

(DCG) Which one makes better recommendation in terms of cumulative discounted gain (DCG)?

$$\mathrm{DCG} = \frac{1}{N}\sum_{u=1}^{N}\sum_{j=1}^{J}\frac{g_{ui_j}}{\max(1,\log_b j)}$$

System1: 5 + 4 / $\log_2 2$ + 3 / $\log_2 3$ + 2 / $\log_2 5$ =

System2: 5 + 3 / $\log_2 2$ + 4 / $\log_2 5$ + 2 / $\log_2 6$ =

(Coverage – Gini index) Among 100 users recommended by two recommender systems, the following number of users have actually bought the items:

	I1	I2	I3	I4	I5
System 1	80	10	5	5	0
System 2	50	25	15	5	5

Compute Gini index of the sales by each system. Which system show higher diversity in sales?

$$G = \frac{1}{n-1}\sum_{j=1}^{n}(2j-n-1)p(i_j)$$

System1: 1/(5-1) * ((2*1-5-1)*0 + (2*2-5-1)*0.05 + (2*3-5-1)*0.05 + (2*4-5-1)*0.1 + (2*5-5-1)*0.8) = 0.825

System2: 1/(5-1)* ((2*1-5-1)*0.05 + (2*2-5-1)*0.05 + (2*3-5-1)*0.15 +(2*4-5-1)*0.25+(2*5-5-1)*0.5) = 0.55