1. Exercise Sheet: Recommender Systems May 3, 2018

Exercise 1: Collaborative-filtering Recommender Systems

a) Given the following table:

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

The cosine similarities between Alice and all the other users are the following:

$$sim(Alice, User1) = 0.9762$$
$$sim(Alice, User2) = 0.9939$$
$$sim(Alice, User3) = 0.9080$$
$$sim(Alice, User4) = 0.7962$$

The predictions computed from the prediction function are the following:

$pred(Alice, Item4) = \frac{13}{4} + \frac{1}{|0.9762+0.9939|} \cdot \left(0.9939 \cdot \left(3 - \frac{20}{6}\right) + 0.9762 \cdot \left(3 - \frac{13}{6}\right)\right)$
$pred(Alice, Item4) = 3.4947$

$pred(Alice, Item6) = \frac{13}{4} + \frac{1}{|0.9762+0.9939|} \cdot \left(0.9939 \cdot \left(5 - \frac{20}{6}\right) + 0.9762 \cdot \left(3 - \frac{13}{6}\right)\right)$
$pred(Alice, Item6) = 4.5037$

Exercise 2: Similarity metrics

a) $Proof$:

We have the following cosine similarity formula:

$$sim(X, Y) = \frac{|X| \cdot |Y|}{\sqrt{X^2} \cdot \sqrt{Y^2}}$$

We now replace X by bX and Y by dY:

$$sim(X,Y) = \frac{|bX| \cdot |dY|}{\sqrt{(bX)^2} \cdot \sqrt{(dY)^2}}$$

By equivalence:

$$sim(X,Y) = \frac{|b| \cdot |X| \cdot |d| \cdot |Y|}{\sqrt{(b)^2} \cdot \sqrt{(X)^2} \cdot \sqrt{(d)^2} \cdot \sqrt{(Y)^2}}$$
$$sim(X,Y) = \frac{|b| \cdot |X| \cdot |d| \cdot |Y|}{|b| \cdot \sqrt{(X)^2} \cdot |d| \cdot \sqrt{(Y)^2}}$$

By eliminating b and d we get:

$$sim(X,Y) = \frac{|X| \cdot |Y|}{\sqrt{X^2} \cdot \sqrt{Y^2}}$$

Thus, the cosine similarity is invariant to the scale of variables.

b) $Proof$: We have the following Pearson similarity formula:

$$sim(X,Y) = \frac{\sum\left(X-\bar{X}\right) \cdot \left(Y-\bar{Y}\right)}{\sqrt{\sum\left(X-\bar{X}\right)^2 \sum\left(Y-\bar{Y}\right)^2}}$$

We then replace

$$a + bX for X$$
$$c + dY for Y$$

Remark that the previous mean values will become:

$$a + b\bar{X} for \bar{X}$$
$$c + d\bar{X} for \bar{X}$$

By equivalence we get:

$$sim(X,Y) = \frac{\sum\left(a+bX-a-b\bar{X}\right) \cdot \left(c+dY-c-d\bar{Y}\right)}{\sqrt{\sum\left(a+bX-a-b\bar{X}\right)^2 \sum\left(c+dY-c-d\bar{Y}\right)^2}}$$
$$sim(X,Y) = \frac{\sum\left(bX-b\bar{X}\right) \cdot \left(dY-d\bar{Y}\right)}{\sqrt{\sum\left(bX-b\bar{X}\right)^2 \sum\left(dY-d\bar{Y}\right)^2}}$$
$$sim(X,Y) = \frac{\sum(bd) \cdot \left(X-\bar{X}\right) \cdot \left(Y-\bar{Y}\right)}{\sqrt{\sum(b)^2 \cdot \left(X-\bar{X}\right)^2 \sum(d)^2 \cdot \left(Y-\bar{Y}\right)^2}}$$

b and d are constant thefore:

$$sim(X,Y) = \frac{(bd) \cdot \sum\left(X-\bar{X}\right) \cdot \left(Y-\bar{Y}\right)}{\sqrt{(bd)^2 \cdot \sum\left(X-\bar{X}\right)^2 \sum\left(Y-\bar{Y}\right)^2}}$$
$$sim(X,Y) = \frac{(bd) \cdot \sum\left(X-\bar{X}\right) \cdot \left(Y-\bar{Y}\right)}{(bd) \cdot \sqrt{\sum\left(X-\bar{X}\right)^2 \sum\left(Y-\bar{Y}\right)^2}}$$
$$sim(X,Y) = \frac{\sum\left(X-\bar{X}\right) \cdot \left(Y-\bar{Y}\right)}{\sqrt{\sum\left(X-\bar{X}\right)^2 \sum\left(Y-\bar{Y}\right)^2}}$$

Thus, the pearson similarity is invariant to separate changes of location and scale of the variables.

Exercise 3: Recommendations based on association rules

The mean adjusted rating matrix R′ from R:

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	mean
Alice	1.75	-0.25	0.75	?	-2.25	?	3.25
User 1	0.84	-1.16	-0.16	0.84	-1.16	0.84	2.16
User 2	0.66	-0.33	0.66	-0.33	-2.33	1.66	3.33
User 3	0.17	0.17	-1.83	2.17	-1.83	1.17	2.83
User 4	-1.5	2.5	2.5	-0.5	-1.5	-1.5	2.5

The transformed mean-adjusted utility matrix:

Table 1: My caption

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Alice	1	0	1	?	0	?
User 1	1	0	0	1	0	1
User 2	1	0	1	0	0	1
User 3	1	1	0	1	0	1
User 4	0	1	1	0	0	0

Support Calculations:

$$support(Item1) = \frac{3}{4}$$
$$support(Item2) = \frac{1}{2}$$
$$support(Item3) = \frac{1}{2}$$
$$support(Item4) = \frac{1}{2}$$
$$support(Item5) = 0$$
$$support(Item6) = \frac{3}{4}$$
$$support(Item1, Item4) = \frac{1}{2}$$
$$support(Item1, Item6) = \frac{3}{4}$$
$$support(Item1, Item4, team6) = \frac{1}{2}$$

Confidence Calculations:

$$confidence(Item1 \Rightarrow Item4) = \frac{support(Item1,Item4)}{support(Item1)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$
$$confidence(Item1 \Rightarrow Item6) = \frac{support(Item1,Item6)}{support(Item1)} = \frac{\frac{3}{4}}{\frac{3}{4}} = 1$$
$$confidence(Item1, Item4 \Rightarrow Item6) = \frac{support(Item1,Item4,Item6)}{support(Item1,Item4)} = \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{2}{3}$$
$$confidence(Item1, Item6 \Rightarrow Item4) = \frac{support(Item1,Item6,Item4)}{support(Item1,Item6)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$$

Frequent Item sets: (all sets greater or equal to the support treshhold)

$$[Item1], [Item2], [Item3], [Item4], [Item6], [Item1, Item4], [Item1, Item6], [Item1, Item4, Item6]$$

Inference Rules: (all inference rules greater or equal to the confidence treshhold)

$$[Item1 \Rightarrow Item4], [Item1 \Rightarrow Item6], [Item1, Item4 \Rightarrow Item6], [Item1, Item6 \Rightarrow Item4]$$

This two association rules will most likely generate a recommendation because of their high confidence value:

$$[Item1 \Rightarrow Item6], [Item1, Item6 \Rightarrow Item4]$$

The system will use the first inference rule in order to recommend Item 6 because its confidence has the highest confidence value given the initial state. Then the system will most likely recommend item 4 as a second step because it has the highest confidence given the state where the system already recommended item 6 for Alice.

Exercise 4: Recommender systems, Probabilistic Approach

a) Calculation of priors for Item 2 and Item 4

	Item 2	Item 4
1	P(Item2 = 1) = 1/4	P(Item4= 1) = 0
2	P(Item2 = 2) = 0	P(Item4= 2) = 1/4
3	P(Item2 = 3) = 1/2	P(Item4= 3) = 1/2
4	P(Item2 = 4) = 0	P(Item4= 4) = 0
5	P(Item2 = 5) = 1/4	P(Item4= 5) = 1/4

b) Calculation of class-conditional probabilities for Alice‘s ratings X= Item1 = 5, Item 3= 4:

	1	2	3	4	5
Item1	$P(Item1 = 5 \mid Item2 = 1) = 0$	$P(Item1 = 5 \mid Item2 = 2) = 0$	$P(Item1 = 5 \mid Item2 = 3) = 1/2$	$P(Item1 = 5 \mid Item2 = 4) = 0$	$P(Item1 = 5 \mid Item2 = 5) = 0$
Item3	$P(Item3 = 4 \mid Item2 = 1) = 0$	$P(Item3 = 4 \mid Item2 = 2) = 0$	$P(Item3 = 4 \mid Item2 = 3) = 1/2$	$P(Item3 = 4 \mid Item2 = 4) = 0$	$P(Item3 = 4 \mid Item2 = 5) = 0$

	1	2	3	4	5
Item1	$P(Item1 = 5 \mid Item4 = 1) = 0$	$P(Item1 = 5 \mid Item4 = 2) = 0$	$P(Item1 = 5 \mid Item4 = 3) = 0$	$P(Item1 = 5 \mid Item4 = 4) = 0$	$P(Item1 = 5 \mid Item4 = 5) = 1/2$
Item3	$P(Item3 = 4 \mid Item4 = 1) = 0$	$P(Item3 = 4 \mid Item4 = 2) = 0$	$P(Item3 = 4 \mid Item4 = 3) = 1/2$	$P(Item3 = 4 \mid Item4 = 4) = 0$	$P(Item3 = 4 \mid Item4 = 5) = 0$

b) Posterior probabilities of Alice's ratings:

	1	2	3	4	5
Item2	P(Item2 = 1 | X) = 0	P(Item2 = 2 | X) = 0	P(Item2 = 3 | X) = 1/8	P(Item2 = 4 | X) = 0	P(Item2 = 5 | X) = 0
Item4	P(Item4 = 1 | X) = 0	P(Item4 = 2 | X) = 0	P(Item4 = 3 | X) = 0	P(Item4 = 4 | X) = 0	P(Item4 = 5 | X) = 0

d) Posterior probabilities of Alice's ratings:
The only value we could derive is 1/8 thefore we would recommend Item 2 and assign it a rating of 3.