Lecture 11

Collaborative Filtering

COMP3278

Introduction to Database Management Systems

Dr. Ping Luo
Email : pluo@cs.hku.hk

Department of Computer Science, The University of Hong Kong

Making Prediction

We are going to predict Bob's rating(s). We call him the **Active user**.

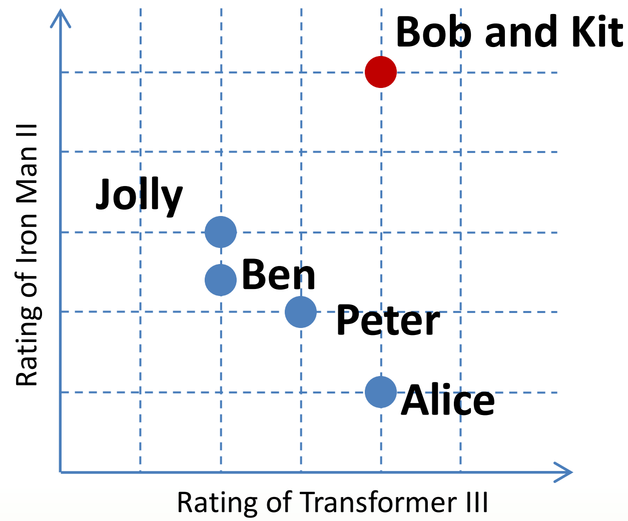
	Transformer III	Avengers	Iron Man II	The Amazing Spiderman
Bob	4	?	5	5
Alice	4	2	1	
Peter	3		2	4
Kit	4	4		
Jolly	2	1	3	5

Considering the above ratings given by the site users, what is the predicted rating of **Bob** to the film **"Avengers"**?

1. Measuring similarity

- There are various methods to represent the similarity of two users in memory-based CF.
  - **1) Vector Cosine-based similarity**
  - **2) Correlation-based similarity**

1. Vector-cosine similarity



	Transformer III	Iron Man II
Bob	4	5
Alice	4	1
Peter	3	2
Jolly	2	3
Kit	4	5
Ben	2	2.5

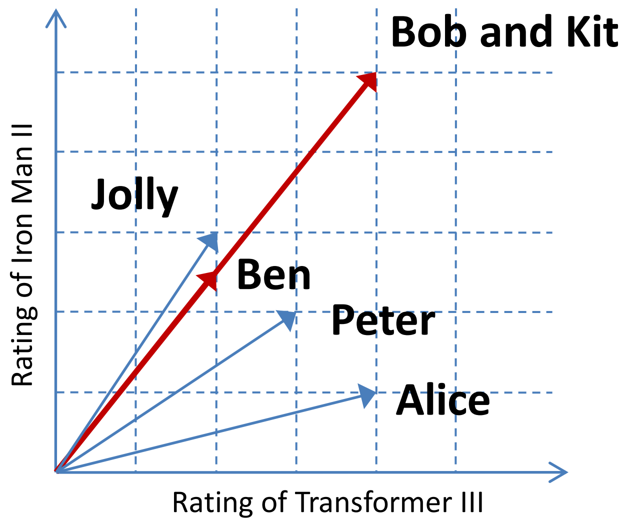
Think about it…

We would like to have a similarity value that reflects things like:

- As **Bob** and **Kit** have exactly the same ratings, they should have very high similarity.
- **Scaling:** As the scale of the ratings that **Bob** and **Ben** gave are the same i.e., 4:5 (0.8), and 2:2.5 (0.8) , they should have high similarity

How will you model the users (represent them mathematically)?

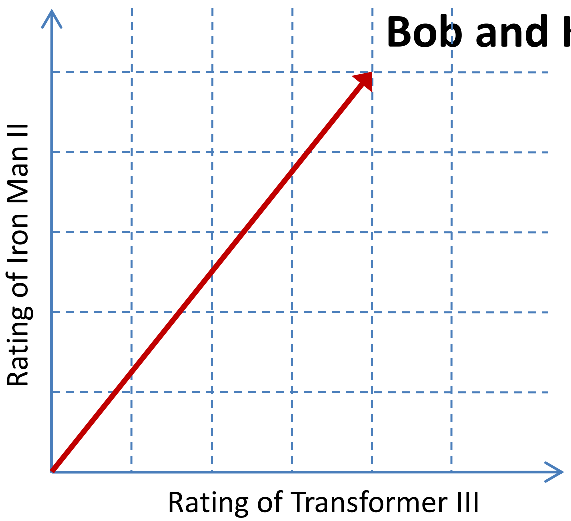
1. Vector-cosine similarity



	Transformer III	Iron Man II
Bob	4	5
Alice	4	1
Peter	3	2
Jolly	2	3
Kit	4	5
Ben	2	2.5

- **User modeling:** Each user is modeled as a "vector" of ratings.
- **Similarity measure:** The inner angle formed by the two users' vectors.
  - **Rationale** - If the inner angle of the vectors of two users is smaller, then they are more similar.

Example



	Transformer III	Iron Man II
Bob	4	5
Alice	4	0
Peter	3	2
Jolly	0	3
Kit	4	5
Ben	2	2.5

Similarity		similarity
Bob	Kit	1

$$\cos\theta = \frac{r_{u,x}r_{v,x} + r_{u,y}r_{v,y}}{|\vec{u}||\vec{v}|}$$

$$|\vec{u}| = \sqrt{r_{u,x}^2 + r_{u,y}^2} \qquad |\vec{v}| = \sqrt{r_{v,x}^2 + r_{v,y}^2}$$

Similarity between Bob and Kit

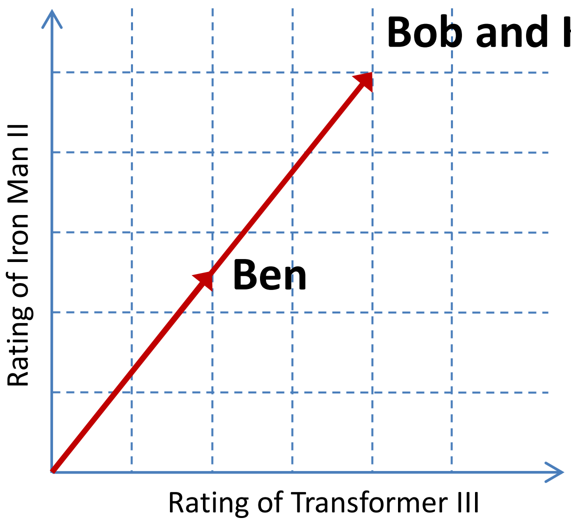
(4 × 4 + 5 × 5) / [sqrt(4^2+5^2) × sqrt(4^2+5^2)]

= (16 + 25) / [sqrt(41) × sqrt(41)]

= 1

Since similarity between Bob and Kit calculated in this way is 1, they are “the same” in this similarity model.

Example



	Transformer III	Iron Man II
Bob	4	5
Alice	4	0
Peter	3	2
Jolly	0	3
Kit	4	5
Ben	2	2.5

Similarity		similarity
Bob	Kit	1
Bob	Ben	1

$$\cos\theta = \frac{r_{u,x}r_{v,x} + r_{u,y}r_{v,y}}{|\vec{u}||\vec{v}|}$$

$$|\vec{u}| = \sqrt{r_{u,x}^2 + r_{u,y}^2} \qquad |\vec{v}| = \sqrt{r_{v,x}^2 + r_{v,y}^2}$$

Similarity between Bob and Ben

$(4 \times 2 + 5 \times 2.5)/[\text{sqrt}(4^2+5^2) \times \text{sqrt}(2^2+2.5^2)]$

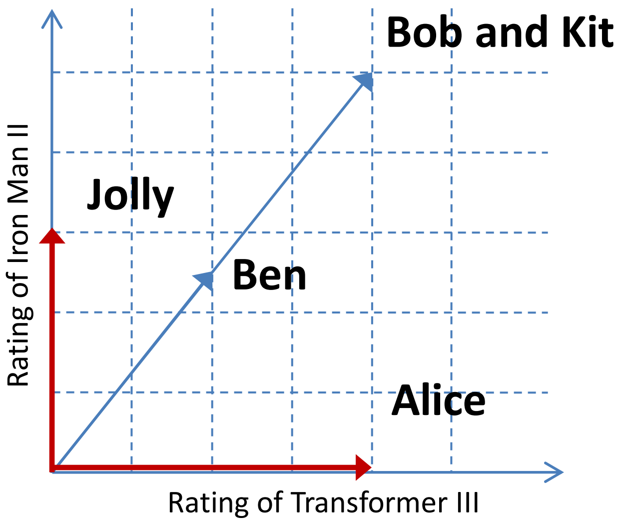
$= (8 + 12.5) / [\text{sqrt}(41) \times \text{sqrt}(10.25)]$

$= 20.5 / 20.5$

$= 1$

Bob and Ben are "the same" in this similarity model.
This shows that this model handles different scale of ratings by different users.

Example



	Transformer III	Iron Man II
Bob	4	5
Alice	4	0
Peter	3	2
Jolly	0	3
Kit	4	5
Ben	2	2.5

Similarity		similarity
Bob	Kit	1
Bob	Ben	1
Jolly	Alice	0

$$\cos\theta = \frac{r_{u,x}r_{v,x} + r_{u,y}r_{v,y}}{|\vec{u}||\vec{v}|}$$

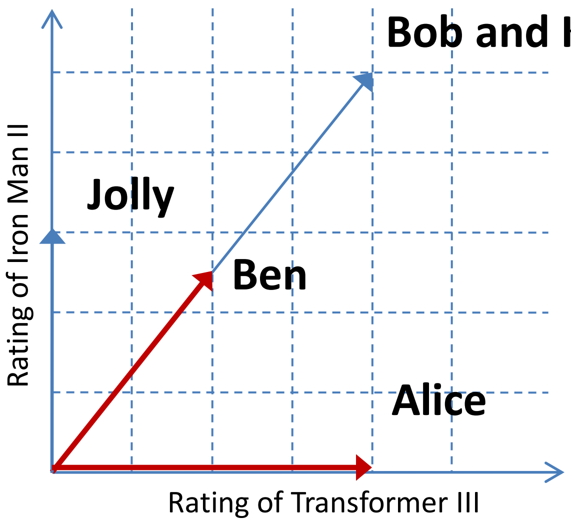
$$|\vec{u}| = \sqrt{r_{u,x}^2 + r_{u,y}^2} \qquad |\vec{v}| = \sqrt{r_{v,x}^2 + r_{v,y}^2}$$

Similarity between Ben and Alice

$(0 \times 4 + 3 \times 0)$ / [$\mathrm{sqrt}(0^2+3^2) \times \mathrm{sqrt}(4^2+0^2)$]
= 0

Since similarity between Jolly and Alice calculated in this way is 0, they are not similar .

Example



	Transformer III	Iron Man II
Bob	4	5
Alice	4	0
Peter	3	2
Jolly	0	3
Kit	4	5
Ben	2	2.5

Similarity		similarity
Bob	Kit	1
Bob	Ben	1
Jolly	Alice	0
Ben	Alice	0.62

$$\cos\theta = \frac{r_{u,x}r_{v,x} + r_{u,y}r_{v,y}}{|\vec{u}||\vec{v}|}$$

$$|\vec{u}| = \sqrt{r_{u,x}^2 + r_{u,y}^2} \qquad |\vec{v}| = \sqrt{r_{v,x}^2 + r_{v,y}^2}$$

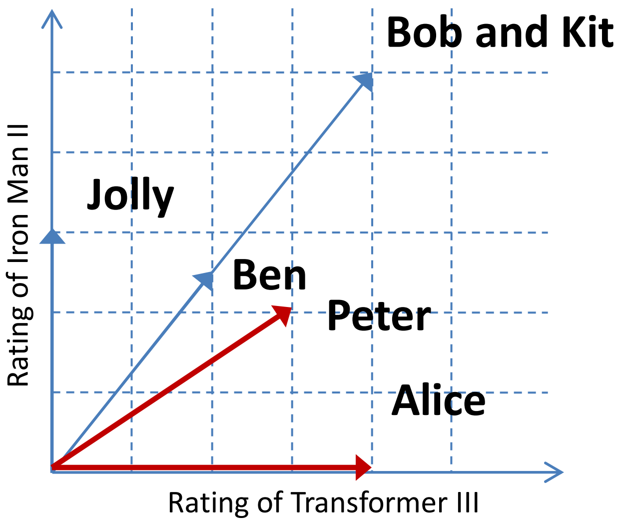
Similarity between Ben and Alice

$(2 \times 4 + 2.5 \times 0)$ / [$\text{sqrt}(2^2+2.5^2) \times \text{sqrt}(4^2+0^2)$]

= 8 / [sqrt(10.25) × sqrt(16)]

= 0.62

Example



	Transformer III	Iron Man II
Bob	4	5
Alice	4	0
Peter	3	2
Jolly	0	3
Kit	4	5
Ben	2	2.5

Similarity		similarity
Bob	Kit	1
Bob	Ben	1
Jolly	Alice	0
Ben	Alice	0.62
Peter	Alice	0.83

$$\cos\theta = \frac{r_{u,x} r_{v,x} + r_{u,y} r_{v,y}}{|\vec{u}||\vec{v}|}$$

$$|\vec{u}| = \sqrt{r_{u,x}^2 + r_{u,y}^2} \qquad |\vec{v}| = \sqrt{r_{v,x}^2 + r_{v,y}^2}$$

Similarity between Ben and Alice

$(2 \times 4 + 2.5 \times 0) / [\ sqrt(2^2+2.5^2) \times sqrt(4^2+0^2)\ ]$

$= 8 / [sqrt(10.25) \times sqrt(16)\ ]$

$= 0.62$

Similarity between Peter and Alice

$(3 \times 4 + 2 \times 0) / [\ sqrt(3^2+2^2) \times sqrt(4^2+0^2)\ ]$

$= 12 / [sqrt(13) \times sqrt(16)\ ]$

$= 0.83$

Comparing Ben and Peter with Alice, Peter is more similar to Alice than Ben.

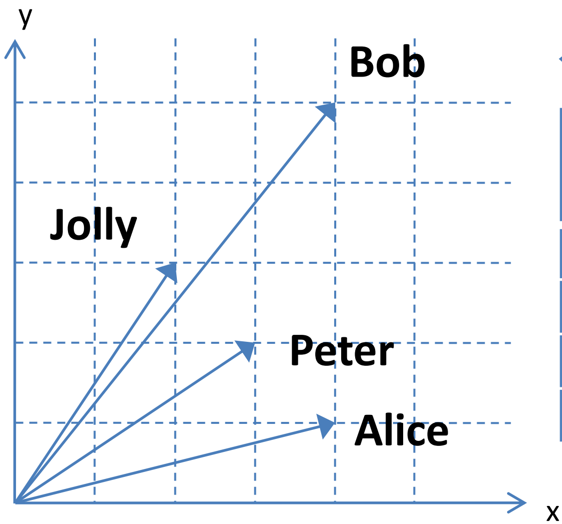
2. Correlation-based similarity

- Some users may have their rating **shifted** by certain amount.
  - E.g., In fact, Bob and Jolly’s are very similar except Jolly’s rating is always Bob’s rating minus 2 (shifting).
- Cosine similarity does not reflect negative correlation.
  - E.g., Bob and Peter’s ratings are negative correlated in the sense that when Bob’s rating is low, Peter’s rating is high, vice versa.

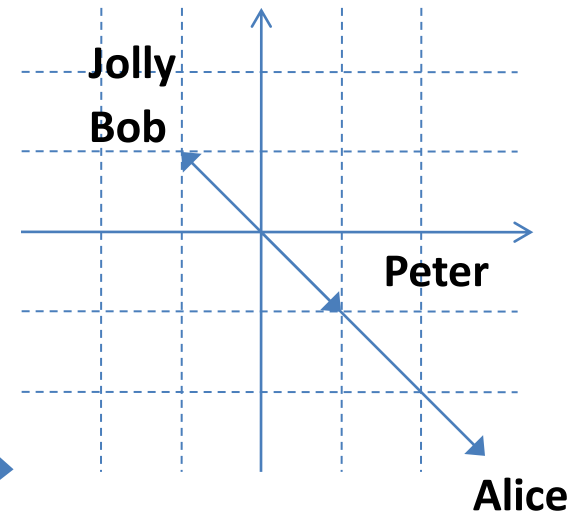
Think about it…
How can we tackle the shifting problem?

	Transformer III	Iron Man II
Bob	4	5
Alice	4	1
Peter	3	2
Jolly	2	3

Correlation-based similarity



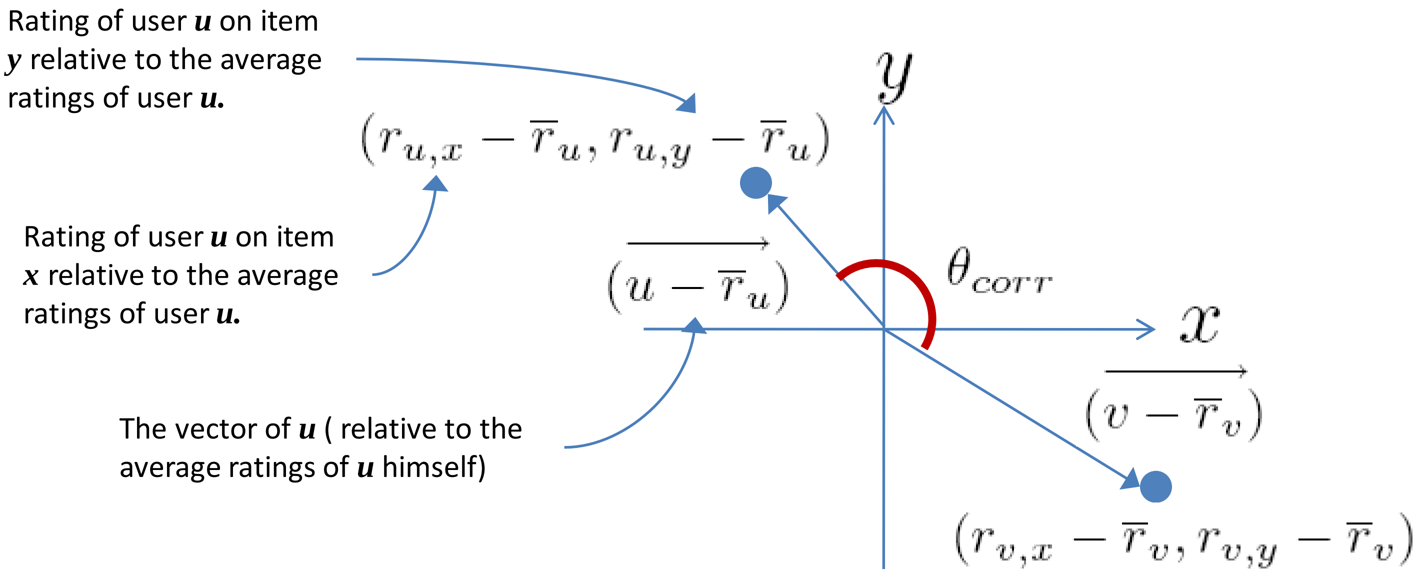
	x	y	Average (x,y)	x – Average (x,y)	y – Average (x,y)
Bob	4	5	4.5	-0.5	0.5
Alice	4	1	2.5	1.5	-1.5
Peter	3	2	2.5	0.5	-0.5
Jolly	2	3	2.5	-0.5	0.5



Idea: Instead of working on the absolute ratings, let's look at the **variation** w.r.t. the **average rating** of a user.

- Bob's rating on *x and y are* **-0.5, +0.5** **from his average**,
- Jolly's rating on *x and y are* **-0.5, +0.5** **from her average**
- Bob and Jolly are exactly the same in this sense.

Correlation-based similarity

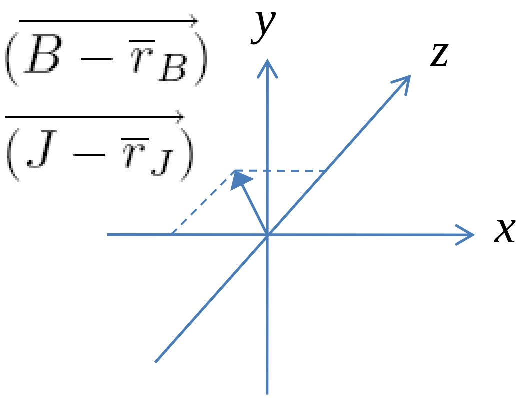


- We use $\cos \theta_{corr}$ to model the correlation-based similarity between two users.

Before we derive the formula of calculating the similarity, I want to understand the meaning of the similarity value. What does it represents?

Correlation-based similarity

	x	y	z	Average (x,y,z)	x-Average (x,y,z)	y-Average (x,y,z)	z-Average (x,y,z)
Bob	4	5	6	5	**-1**	**0**	**1**
Alice	4	1	4	3	1	-2	1
Peter	3	2	1	2	1	0	-1
Jolly	2	3	4	3	**-1**	**0**	**1**

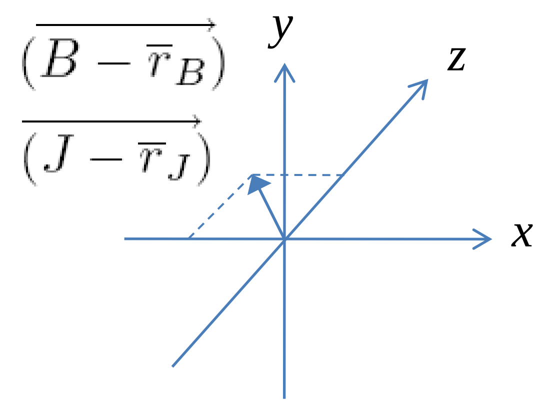


Now, let's look at the data with 3 films (3 dimensions) for easier explanation.
Take Bob and Jolly as an example. What will be the correlation $\cos\theta_{corr}$?

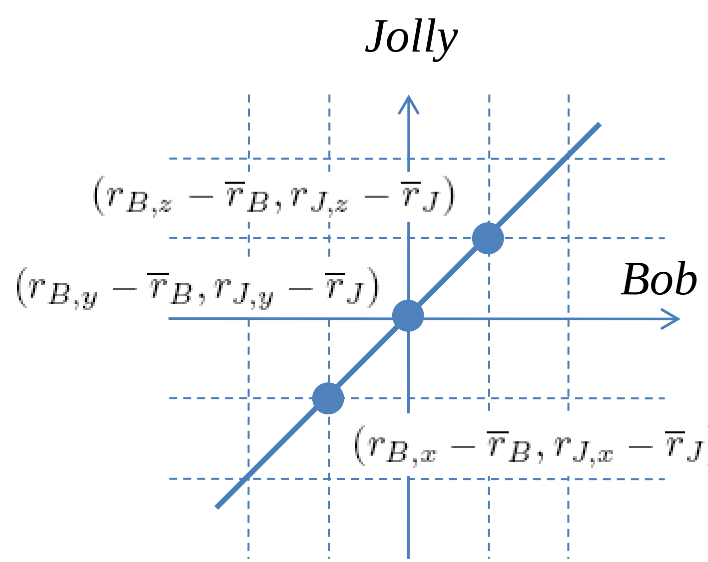
Since the two vectors overlap, the angle is 0 and cos(0) is **1**.
What does the correlation value of 1 imply?

Correlation-based similarity

	x	y	z	Average (x,y,z)	x-Average (x,y,z)	y-Average (x,y,z)	z-Average (x,y,z)
Bob	4	5	6	5	**-1**	**0**	**1**
Alice	4	1	4	3	1	-2	1
Peter	3	2	1	2	1	0	-1
Jolly	2	3	4	3	**-1**	**0**	**1**

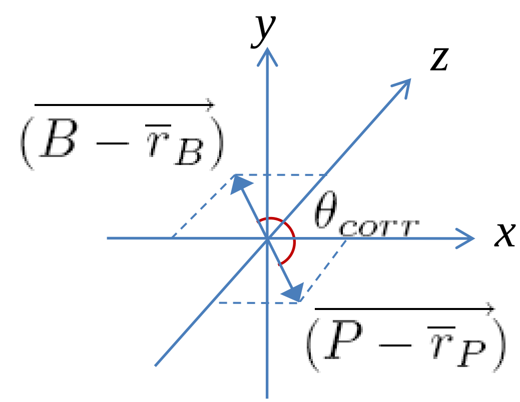


It means Bob and Jolly's ratings (minus their corresponding average) follow **perfect positive correlation**!!!
If Bob's rating (relative to his avg rating) is larger, Jolly rating (relative to her avg rating) is also larger.



Correlation-based similarity

	x	y	z	Average (x,y,z)	x-Average (x,y,z)	y-Average (x,y,z)	z-Average (x,y,z)
Bob	4	5	6	5	**-1**	**0**	**1**
Alice	4	1	4	3	1	-2	1
Peter	3	2	1	2	**1**	**0**	**-1**
Jolly	2	3	4	3	-1	0	1



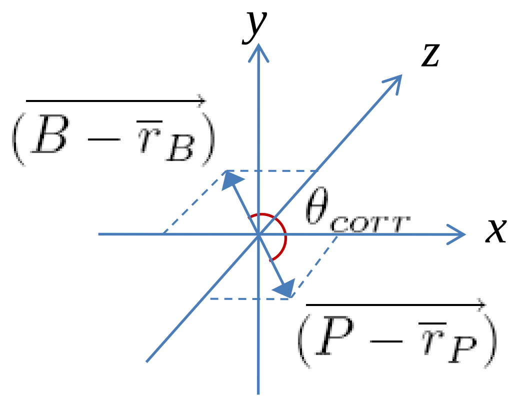
Now lets look at Bob and Peter's similarity. What will be the correlation $\cos\theta_{corr}$?

Since the two vectors pointing exactly in opposite direction, the angle is 180 and cos(180) is **-1**.

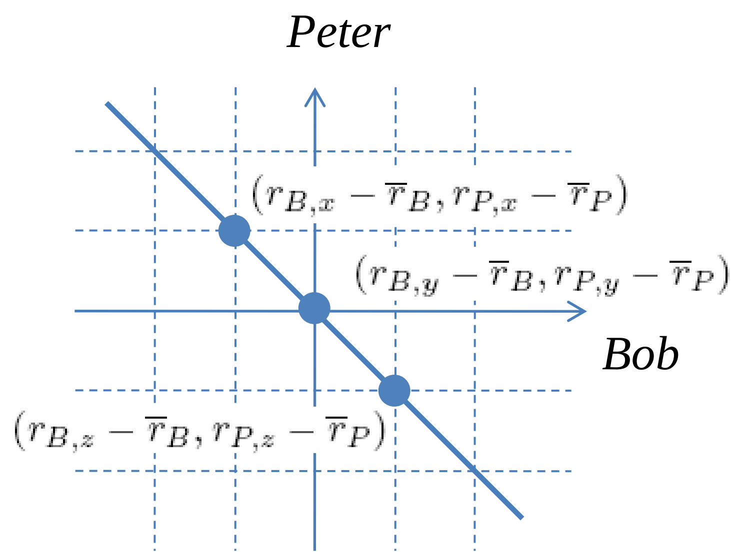
What does the correlation value of -1 imply?

Correlation-based similarity

	x	y	z	Average (x,y,z)	x-Average (x,y,z)	y-Average (x,y,z)	z-Average (x,y,z)
Bob	4	5	6	5	**-1**	**0**	**1**
Alice	4	1	4	3	1	-2	1
Peter	3	2	1	2	**1**	**0**	**-1**
Jolly	2	3	4	3	-1	0	1

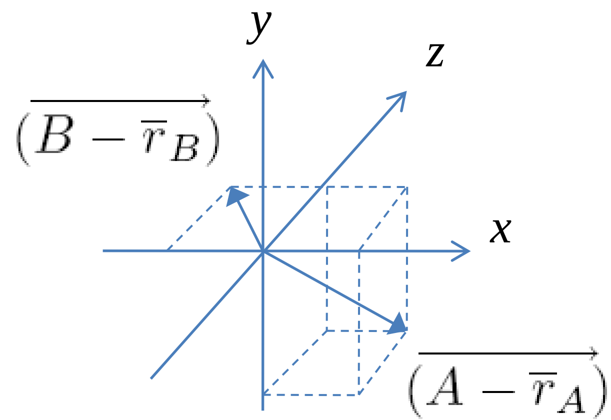


It means Bob and Peter's ratings (minus their corresponding average) forms **perfect negative correlation**!!!
if Bob's rating (relative to his avg rating) is larger, Peter's rating (relative to his avg rating) is lower, vice versa.

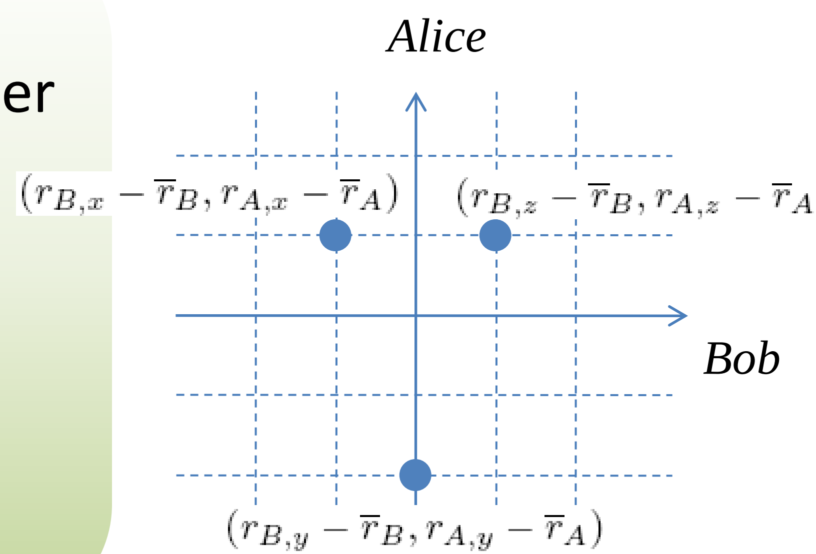


Correlation-based similarity

	x	y	z	Average (x,y,z)	x-Average (x,y,z)	y-Average (x,y,z)	z-Average (x,y,z)
Bob	4	5	6	5	**-1**	**0**	**1**
Alice	4	1	4	3	**1**	**-2**	**1**
Peter	3	2	1	2	1	0	-1
Jolly	2	3	4	3	-1	0	1



Now take Bob and Alice's similarity as an example. The inner angle is 90 degree (we will show how to calculate shortly) and $\cos \theta_{corr}$ equals cos(90) = **0**. This reflects the Bob and Alice 's rating **does not show any linear correlations**.



Correlation-based similarity

- Note that the correlation reflects the extend to which two variables **linearly related** with each other.

Perfect positive correlation (Like Bob and Jolly)

Perfect negative correlation (Like Bob and Peter)

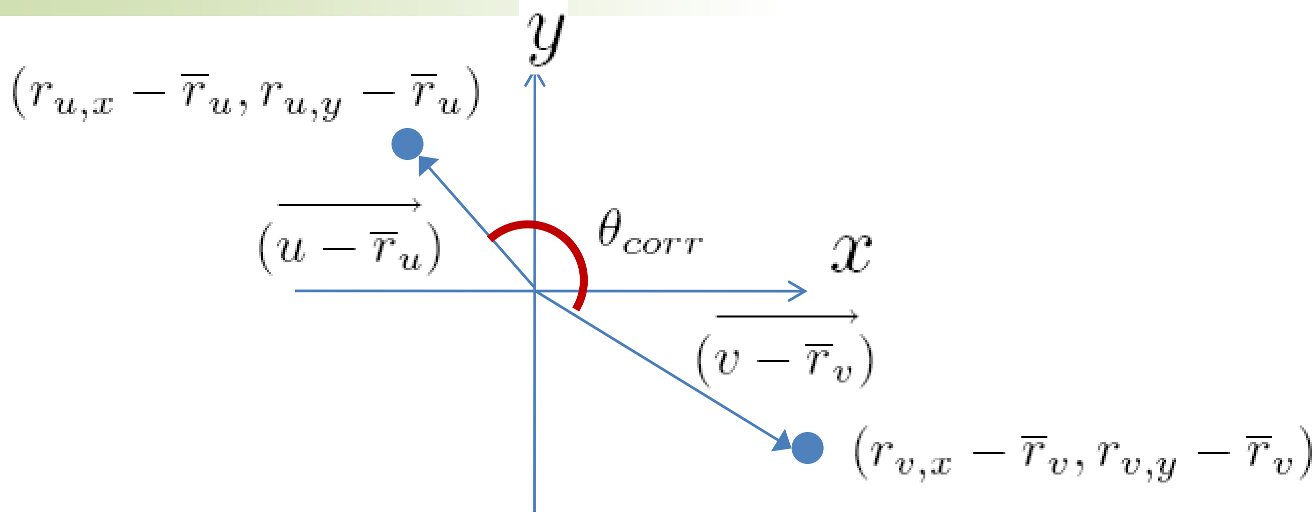
No linear correlation (Like Bob and Alice)

- The correlation value is **not the slope** of that relationship.

They are all **perfect correlations** (Except the middle one) although they have different slopes.

- Note that the middle case has correlation undefined because there is no variation w.r.t. the mean value for the y-dimension.

Correlation-based similarity

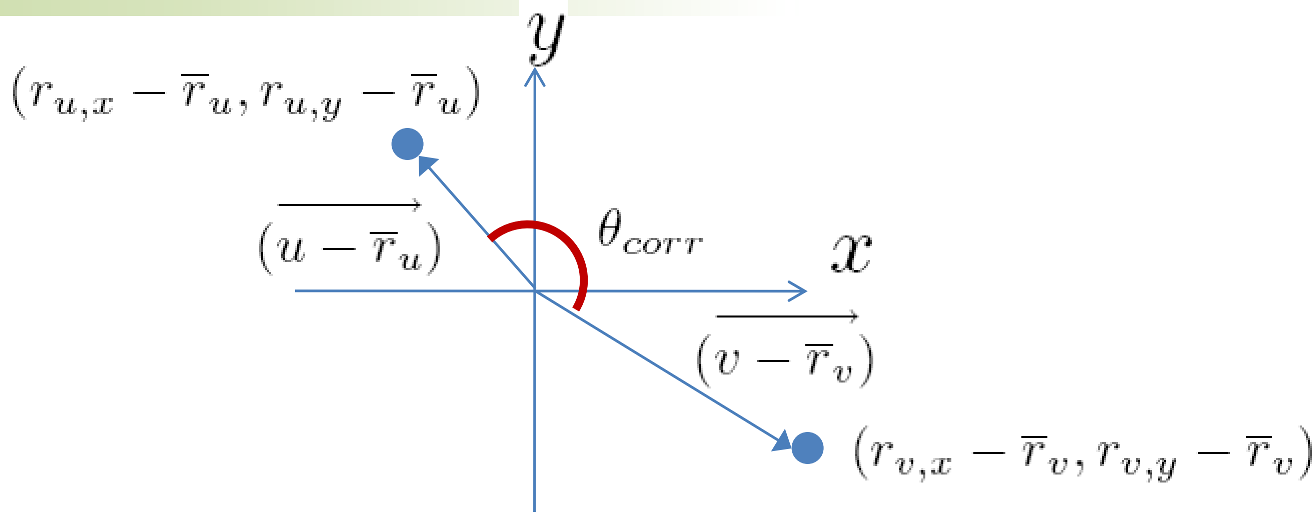


$$corr(u, v) = \cos \theta_{corr} = \frac{\overrightarrow{(u - \overline{r}_u)} \cdot \overrightarrow{(v - \overline{r}_v)}}{|\overrightarrow{(u - \overline{r}_u)}||\overrightarrow{(v - \overline{r}_v)}|}$$

Now let's derive the formula of the similarity measure.

We reuse the proof for the vector-cosine similarity, but the vectors change from u, v to $\overrightarrow{(u - \overline{r}_u)}$ and $\overrightarrow{(v - \overline{r}_v)}$.

Correlation-based similarity



$$corr(u,v) = \cos\theta_{corr} = \frac{\overrightarrow{(u - \overline{r}_u)} \cdot \overrightarrow{(v - \overline{r}_v)}}{|\overrightarrow{(u - \overline{r}_u)}||\overrightarrow{(v - \overline{r}_v)}|}$$

Note that ***I*** **is the set of co-rated items** of user ***u*** and ***v***. (so that we won't have null value fit into this formula)

$$= \frac{\sum_{i \in I}(r_{u,i} - \overline{r}_u)(r_{v,i} - \overline{r}_v)}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r}_u)^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r}_v)^2}}.$$

These are the average on the **co-rated items**.

DONE

Correlation-based similarity

- The correlation is actually called the **Pearson coefficient**, a well known measure of correlation between two random variables ($\boldsymbol{u}$ and $\boldsymbol{v}$).
- In collaborative filtering, we can use the Pearson coefficient as the **weight** in the prediction model.

2. Computing prediction

- **Target:** To calculate a prediction $P_{a,i}$ for the user ***a***'s rating on item ***i*** (a film which ***a*** has not rated).

Average rating that user ***a*** gave

$$P_{a,i} = \overline{r}_a +$$

Weighted average
of all other users' ratings on the item ***i*** (relative to the corresponding user's average rating).

2. Computing prediction

Average rating that user ***a*** gave

For each user rated item ***i***

Weighted

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} \text{[Relative ratings of each of the user on item } i \text{] ?}}{\text{?}}$$

Relative ratings of each of the user on item ***i***

?

Average → **?**

2. Computing prediction

Insight: If the weighting with user ***u*** is 1 (**perfect positive correlation**), and if user ***u*** rate the item ***i*** as +3 from the his average. Then we predict user ***a*** will also rate +3 from ***a***'s average.

Average rating that user ***a*** gave

For each user rated item ***i***

$$P_{a,i} = \overline{r}_a + \frac{\sum_{u \in U} [\text{Relative ratings of each of the user on item } i] \cdot [\text{Weighting (the similarity with user } a)]}{\text{Sum of weightings over all users rated on item } i}$$

Relative ratings of each of the user on item ***i*** · **Weighting** (the similarity with user ***a***)

Sum of weightings over all users rated on item ***i***

Why multiply with the weighting for each item in the nominator?

Insight: If the weighting with user ***u*** is -1 (**perfect negative correlation**), and if user ***u*** rate the item ***i*** as +3 from the his average. Then we predict user ***a*** will rate -3 from ***a***'s average.

2. Computing prediction

Average rating that user ***a*** gave

For each user rated item ***i***

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot \text{Weighting (the similarity with user } a\text{)}}{\text{Sum of weightings over all users rated on item } i}$$

Weighting (the similarity with user ***a***)

Sum of weightings over all users rated on item ***i***

- Relative rating is the user's rating on item ***i*** minus the user's rating over all items he/she rated (handle shifting).

2. Computing prediction

Average rating that user ***a*** gave

For each user rated item ***i***

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\quad}$$

Sum of weightings over all users rated on item ***i***

- Let the similarity between two users ***a*** and ***u*** be $w_{a,u}$
- Note that $w_{a,u}$ can be **Cosine similarity** or **Pearson correlation** coefficient.

2. Computing prediction

Average rating that user ***a*** gave

For each user rated item ***i***

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

- Since the weighting **can be negative value** (e.g., -1 for perfect negative correlation), we take the absolute value when getting the sum of the weight.

Example

	1. Transformer III	2. Avengers	3. Iron Man II	4. The Amazing Spiderman
1. Bob	4	**?**	5	5
2. Alice	4	2	1	
3. Peter	3		2	4
4. Kit	4	4		
5. Jolly	2	1	3	5

- What is the predicted rating of Bob on “Avengers”?
  - Users who have rated Avengers are “Alice”, “Kit” and “Jolly”. **(i.e., U = {2,4,5})**

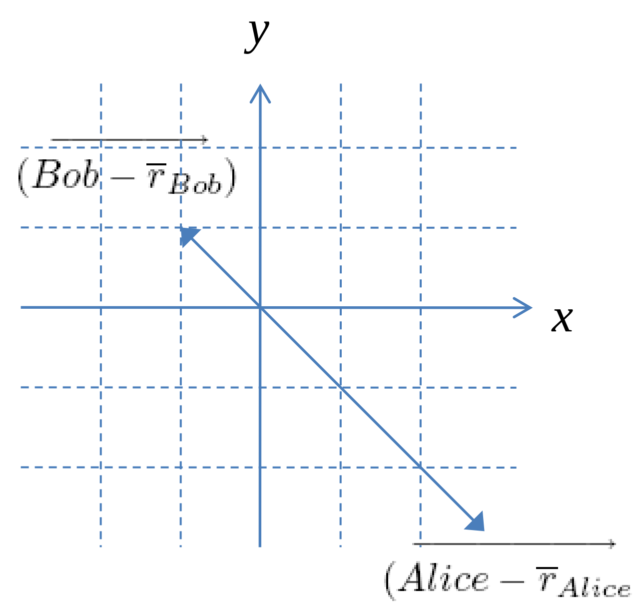
What are the **weights** between **Bob & Alice ($w_{1,2}$), Bob & Kit ($w_{1,4}$), Bob & Jolly ($w_{1,5}$)** respectively?
(Assume that we use **Pearson Correlation**)

$w_{1,2}$

	1. Transformer III	2. Avengers	3. Iron Man II	4. The Amazing Spiderman
1. Bob	4	?	5	5
2. Alice	4	2	1	
3. Peter	3		2	4
4. Kit	4	4		
5. Jolly	2	1	3	5

Co-rated items : **1. Transformer III** and **3. Iron Man II**

	1. Transformer III	3. Iron Man II	Average (1,3)	Film 1 – Average (1,3)	Film 3 – Average (1,3)
1. Bob	4	5	4.5	-0.5	0.5
2. Alice	4	1	2.5	1.5	-1.5



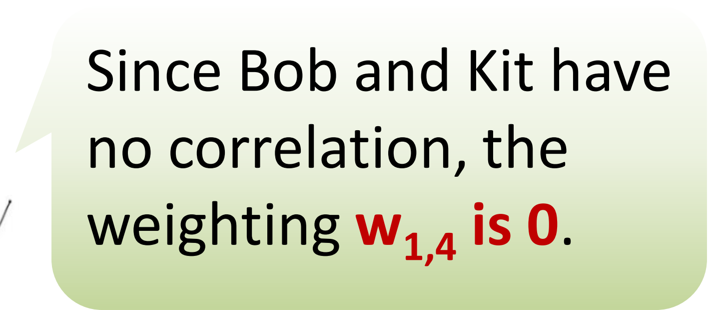
Since the two vectors are having opposite direction, their angle is 180 and cos(180) = -1. Therefore **$w_{1,2}$ is -1**.

$w_{1,4}$

	1. Transformer III	2. Avengers	3. Iron Man II	4. The Amazing Spiderman
1. Bob	4	?	5	5
2. Alice	4	2	1	
3. Peter	3		2	4
4. Kit	4	4		
5. Jolly	2	1	3	5

Co-rated item : **1. Transformer III** only, no correlations can be derived.

	1. Transformer III	Average (1)	Film 1 – Average (1)
1. Bob	4	4	0
4. Kit	4	4	0



Since Bob and Kit have no correlation, the weighting $w_{1,4}$ **is 0**.

$w_{1,5}$

I is the set of **co-rated items**, for Bob and Jolly, the co-rated items are (**1.** Transformer III, **3.** Iron Man II, and **4.** The Amazing Spiderman)

$$corr(u,v) = \frac{\sum_{i \in I}(r_{u,i} - \overline{r}_u)(r_{v,i} - \overline{r}_v)}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r}_u)^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r}_v)^2}}.$$

$$w_{1,5} = \frac{\sum_{i \in \{1,3,4\}}(r_{1,i} - \overline{r}_1)(r_{5,i} - \overline{r}_5)}{\sqrt{\sum_{i \in \{1,3,4\}}(r_{1,i} - \overline{r}_1)^2}\sqrt{\sum_{i \in \{1,3,4\}}(r_{5,i} - \overline{r}_5)^2}}$$

$$= \frac{\sum_{i \in \{1,3,4\}}(r_{1,i} - \frac{14}{3})(r_{5,i} - \frac{10}{3})}{\sqrt{\sum_{i \in \{1,3,4\}}(r_{1,i} - \frac{14}{3})^2}\sqrt{\sum_{i \in \{1,3,4\}}(r_{5,i} - \frac{10}{3})^2}}$$

$$= 0.756.$$

The averages $\overline{r}_1$ and $\overline{r}_5$ are the average over the **co-rated items** (1,3,4).

The Pearson correlation is **0.756**, which mean a **positive correlation** (If Alice's rating is larger, then Bob's rating should be relatively larger, vice versa).

	1. Transformer III	2. Avengers	3. Iron Man II	4. The Amazing Spiderman
1. Bob	4		5	5
5. Jolly	2	1	3	5

What is $P_{1,2}$?

	1. Transformer III	2. Avengers	3. Iron Man II	4. Amazing Spiderman
1. Bob	4	?	5	5
2. Alice	4	2	1	
3. Peter	3		2	4
4. Kit	4	4		
5. Jolly	2	1	3	5

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

$$P_{1,2} = \bar{r}_1 + \frac{\sum_u (r_{u,2} - \bar{r}_u) \cdot w_{1,u}}{\sum_u |w_{1,u}|}$$

$$= \bar{r}_1 + \frac{(r_{2,2} - \bar{r}_2) w_{1,2} + (r_{4,2} - \bar{r}_4) w_{1,4} + (r_{5,2} - \bar{r}_5) w_{1,5}}{|w_{1,2}| + |w_{1,4}| + |w_{1,5}|}$$

$$= 4.67 + \frac{(2 - 2.5)(-1) + (4 - 4)0 + (1 - 3.33)0.756}{1 + 0 + 0.756}$$

$$= 3.95.$$

Done!

Therefore, the predicted rating of Bob on "Avenger" is **3.95**.

Example

$$corr(u,v) = \frac{\sum_{i \in I}(r_{u,i} - \overline{r}_u)(r_{v,i} - \overline{r}_v)}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r}_u)^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r}_v)^2}}.$$

Note that ***I*** **is the set of co-rated items** of user ***u*** and ***v***.
(so that we won't have null value fit into this formula)

	Item1	Item2	Item3	Item4	Item5	Average of 1,2,3,4	Item1	Item2	Item3	Item4	Similarity with Alice
Alice	5	3	4	4	?	**4**	+1	-1	0	0	**/**
User1	3	1	2	3	3	**2.25**	+0.75	-1.25	-0.25	+0.75	**0.85**
User2	4	3	4	3	5	**3.5**	+0.5	-0.5	+0.5	-0.5	**0.7**
User3	3	3	1	5	4	**3**	0	0	-2	+2	
User4	1	5	5	2	1	**3.25**	-2.25	+1.75	+1.75	-1.25	

- $w_{Alice,User1} = \dfrac{[1*0.75 + (-1)*(-1.25) + 0*(-0.25) + 0*0.75]}{sqrt(1^2 + 1^2 + 0^2 + 0^2) * sqrt(0.75^2 + (-1.25)^2 + (-0.25)^2 + 0.75^2)}$

  $= \dfrac{0.75 + 1.25}{sqrt(2) * sqrt(2.75)}$

  = 0.85

Determine the rating of Alice on Item5. Use correlation-based similarity in the prediction model.

- $w_{Alice,User2} = \dfrac{[1*0.5 + (-1)*(-0.5) + 0*0.5 + 0*(-0.5)]}{sqrt(1^2 + 1^2 + 0^2 + 0^2) * sqrt(0.5^2 + (-0.5)^2 + 0.5^2 + (-0.5)^2)}$

  $= \dfrac{0.5 + 0.5}{sqrt(2) * sqrt(1)}$

  = **0.7**

Example

	Item1	Item2	Item3	Item4	Item5	Average of 1,2,3,4	Item1	Item2	Item3	Item4	Similarity with Alice
Alice	5	3	4	4	?	**4**	+1	-1	0	0	**/**
User1	3	1	2	3	3	**2.25**	+0.75	-1.25	-0.25	+0.75	**0.85**
User2	4	3	4	3	5	**3.5**	+0.5	-0.5	+0.5	-0.5	**0.7**
User3	3	3	1	5	4	**3**	0	0	-2	+2	**0**
User4	1	5	5	2	1	**3.25**	-2.25	+1.75	+1.75	-1.25	**-0.79**

- $$w_{Alice,3} = \frac{[1*0 + (-1)*0 + 0*(-2) + 0*2]}{\text{sqrt}(1^2 + 1^2 + 0^2 + 0^2) * \text{sqrt}(0^2 + 0^2 + (-2)^2 + 2^2)}$$

  $$= \frac{0}{\text{sqrt}(2) * \text{sqrt}(8)}$$

  $$= \mathbf{0}$$

- $$w_{Alice,4} = \frac{[1*(-2.25) + (-1)*1.75 + 0*1.75 + 0*(-1.25)]}{\text{sqrt}(1^2 + 1^2 + 0^2 + 0^2) * \text{sqrt}((-2.25)^2 + 1.75^2 + 1.75^2 + (-1.25)^2)}$$

  $$= \frac{-2.25 - 1.75}{\text{sqrt}(2) * \text{sqrt}(12.75)}$$

  $$= \mathbf{-0.79}$$

Example

For each user rated item i

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

	Item1	Item2	Item3	Item4	Item5	Average of 1,2,3,4	Item1	Item2	Item3	Item4	Similarity with Alice
Alice	5	3	4	4	?	**4**	+1	-1	0	0	**/**
User1	3	1	2	3	3	**2.25**	+0.75	-1.25	-0.25	+0.75	**0.85**
User2	4	3	4	3	5	**3.5**	+0.5	-0.5	+0.5	-0.5	**0.7**
User3	3	3	1	5	4	**3**	0	0	-2	+2	**0**
User4	1	5	5	2	1	**3.25**	-2.25	+1.75	+1.75	-1.25	**-0.79**

- $$P_{Alice,item5} = 4 + \frac{(3 - 2.25)w_{Alice,2} + (5 - 3.5)w_{Alice,3} + (4 - 3)w_{Alice,4} + (1 - 3.25)w_{Alice,5}}{|w_{Alice,2}| + |w_{Alice,3}| + |w_{Alice,4}| + |w_{Alice,5}|}$$

$$= 4 + \frac{(0.75)0.85 + (2.5)0.7 + (1)0 + (-2.25)(-0.79)}{|0.85| + |0.7| + |0| + |-0.79|}$$

$$= 4 + \frac{4.165}{2.34}$$

$$= 5.78$$

End