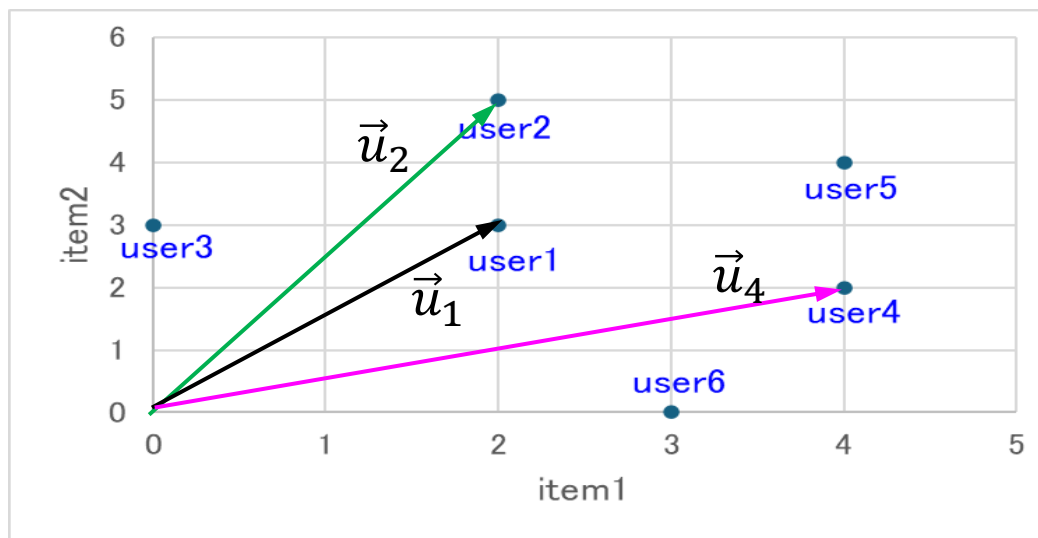


	item1	item2
user1	2	3
user2	2	5
user3	0	3
user4	4	2
user5	4	4
user6	3	0

コサイン類似度



$$\vec{u}_1 = (2,3) \quad \vec{u}_2 = (2,5) \quad \vec{u}_4 = (4,2)$$

内積は、

$$\vec{u}_1 \cdot \vec{u}_2 = 2 \times 2 + 3 \times 5 = 19$$

$$\vec{u}_1 \cdot \vec{u}_4 = 2 \times 4 + 3 \times 2 = 14$$

$$\cos(u_1, u_2) = \frac{\vec{u}_1 \cdot \vec{u}_2}{|\vec{u}_1| |\vec{u}_2|} = \frac{19}{\sqrt{2^2 + 3^2} \times \sqrt{2^2 + 5^2}}$$

$$= \frac{19}{\sqrt{13} \times \sqrt{29}} \doteq 0.978$$

$$\cos(u_1, u_4) = \frac{\vec{u}_1 \cdot \vec{u}_4}{|\vec{u}_1| |\vec{u}_4|} = \frac{14}{\sqrt{2^2 + 3^2} \times \sqrt{4^2 + 2^2}}$$

$$= \frac{14}{\sqrt{13} \times \sqrt{20}} \doteq 0.808$$

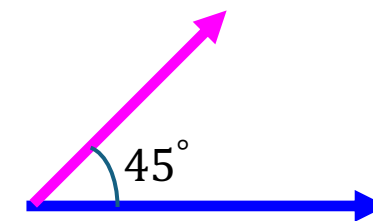
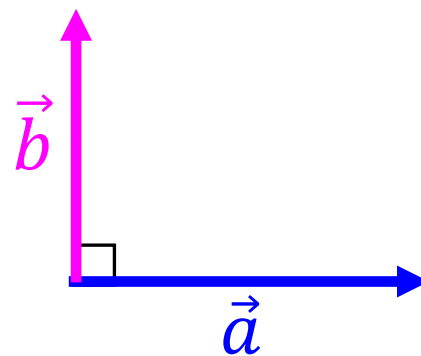
$$\cos(u_1, u_2) > \cos(u_1, u_4)$$

u_1 と u_2 の関係の方が u_1 と u_4 よりも近い(相関あり)

相関係数 $\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}$

$$\vec{a} \cdot \vec{b} = |\vec{a}| |\vec{b}| \cos \theta \quad \text{より}$$

$$\cos 90^\circ = 0 \quad \cos 45^\circ = 0.707 \quad \cos 0^\circ = 1$$



itemが3個以上の場合

$$\vec{a} = (a_1, a_2, a_3, \dots, a_n) \quad \vec{b} = (b_1, b_2, b_3, \dots, b_n)$$

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} = \frac{\sum_{k=1}^n a_k b_k}{\sqrt{\sum_{k=1}^n a_k^2} \sqrt{\sum_{k=1}^n b_k^2}}$$

真逆の評価の場合

	item1	item2	item3	item4	item5
user1	1	2	3	4	5
user2	5	4	3	2	1

↓ 平均値3を差引く

	item1	item2	item3	item4	item5
user1	-2	-1	0	1	2
user2	2	1	0	-1	-2

$$\cos \theta = \frac{1 \times 5 + 2 \times 4 + 3 \times 3 + 4 \times 2 + 5 \times 1}{\sqrt{1^2 + 2^2 + 3^2 + 4^2 + 5^2} \sqrt{5^2 + 4^2 + 3^2 + 2^2 + 1^2}} \doteq 0.636$$

相関がそこそこある？ 違和感あり

$$\cos \theta = \frac{(-2) \times 2 + (-1) \times 1 + 0 \times 0 + 1 \times (-1) + 2 \times (-2)}{\sqrt{(-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2} \sqrt{2^2 + 1^2 + 0^2 + (-1)^2 + (-2)^2}} = -1$$

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} = \frac{\sum_{k=1}^n (a_k - \bar{a})(b_k - \bar{b})}{\sqrt{\sum_{k=1}^n (a_k - \bar{a})^2} \sqrt{\sum_{k=1}^n (b_k - \bar{b})^2}}$$

← 一般式はこちらを使用します

	item1	item2	item3	item4	平均値
user1	2	3	—	5	3.333
user2	2	5	—	5	4.000
user3	—	3	4	4	3.667
user4	4	2	3	—	3.000

	user1	user2	user3	user4
user1	1.000	0.756	0.614	-0.514
user2	0.756	1.000	-0.316	-0.949
user3	0.614	-0.316	1.000	0.894
user4	-0.514	-0.949	0.894	1.000

$$\cos(u_1, u_2) = \frac{(2 - 3.333) \times (2 - 4) + (3 - 3.333)(5 - 4) + (5 - 3.333)(5 - 4)}{\sqrt{(2 - 3.333)^2 + (3 - 3.333)^2 + (5 - 3.333)^2} \sqrt{(2 - 4)^2 + (5 - 4)^2 + (5 - 4)^2}} = \frac{4}{2.160 \times 2.449} = 0.756$$

同様にしてuser間の相関係数を算出

user3と他のuserとの類似性は以下の順

$\cos(u_3, u_4) > \cos(u_3, u_1) > \cos(u_3, u_2)$ の順に類似性あり → user1とuser4を用いてitem1の評価点を推測する

	item1	平均値	評価点 - 平均値	user3との類似度
user1	2 $r_{1,1}$	3.333 \bar{u}_1	-1.333 $r_{1,1} - \bar{u}_1$	0.614 $\cos(u_3, u_1)$
user3	— $r_{3,1}$	3.667 \bar{u}_3	—	—
user4	4 $r_{4,1}$	3.000 \bar{u}_4	1.000 $r_{4,1} - \bar{u}_4$	0.894 $\cos(u_3, u_4)$

$$\hat{r}_{3,1} = \bar{u}_3 + \overbrace{(r_{1,1} - \bar{u}_1) \times \cos(u_3, u_1)}^{\text{重み付け評価値}} + \underbrace{(r_{4,1} - \bar{u}_4) \times \cos(u_3, u_4)}_{\text{重み付け評価値}}$$

$$= 3.667 + (-1.333) \times 0.614 + (1.000) \times 0.894$$

$$= 3.743$$

協調フィルタリングによる予測例

	item1	item2	item3	item4
user1	2	3	4	5
user2	2	5	5	5
user3	2	3	4	4
user4	4	2	3	3

行列因子分解の実行例

	item1	item2	item3	item4
user1	2.00	3.00	4.05	5.00
user2	2.00	5.00	5.00	5.00
user3	1.60	3.00	4.00	4.00
user4	4.00	2.00	3.00	3.00

コサイン類似法による予測例

	item1	item2	item3	item4
user1	2	3	4	5
user2	2	5	5	5
user3	3.743	3	4	4
user4	4	2	3	3.180