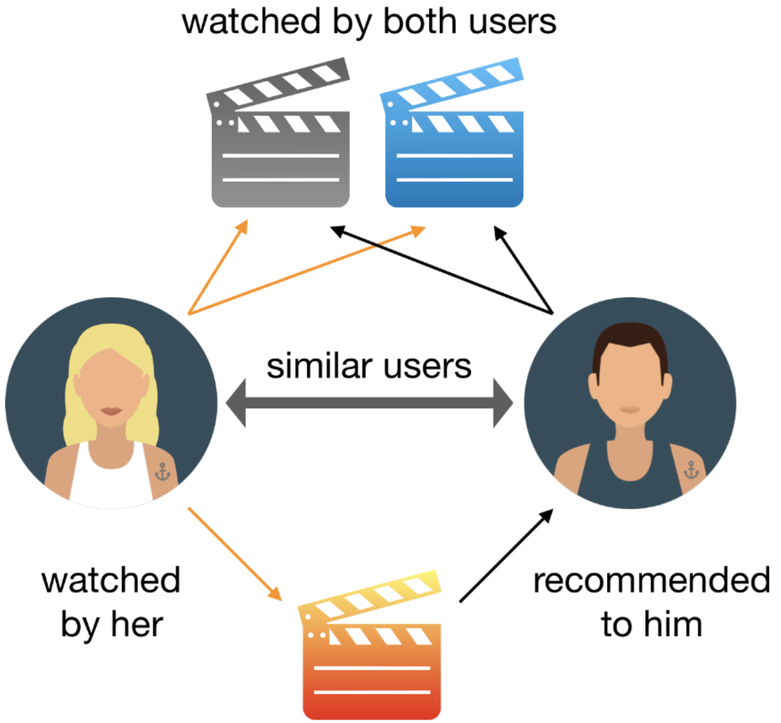
Collaborative Filtering

Collaborative Filtering

- A method to recommend items based on the **behavior of many users**.
- Two main types:
  - **User-Based Collaborative Filtering**
  - **Item-Based Collaborative Filtering**
- Works on the idea: *"People who agreed in the past will agree again."*

User-Based Collaborative Filtering

- Recommends items to a user based on the preferences of **similar users**.
- **Users with similar tastes** are identified using rating patterns.
- Example: If User A and User B liked the same movies, and B liked another movie that A hasn't seen, recommend it to A.



User Based Collaborative Filtering

Term	Definition
Users	Individuals using the system.
Items	Products, movies, books, etc., being recommended.
Rating Matrix	A user-item matrix where each entry represents a user's rating for an item.
Neighborhood	A set of users similar to the active user (user for whom recommendations are made).

Steps in User Based CF

1.Create User-Item Rating Matrix

2.Find similar users using similarity measures (e.g., Pearson, Cosine)

3.Select top-N similar users (neighbors)

4.Predict rating for the target item using weighted average of neighbors' ratings

5.Recommend items with highest predicted ratings

Similarity Metrics

- Cosine Similarity Formula:

$$\text{sim}(u, v) = \frac{\vec{R}_u \cdot \vec{R}_v}{||\vec{R}_u|| \cdot ||\vec{R}_v||}$$

- Pearson Correlation Formula:

$$\text{sim}(u, v) = \frac{\sum_i (R_{u,i} - \bar{R}_u)(R_{v,i} - \bar{R}_v)}{\sqrt{\sum_i (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_i (R_{v,i} - \bar{R}_v)^2}}$$

Where:

- $R_{u,i}$: rating of user u on item i
- $\bar{R}_u$: average rating of user u

Range of Pearson correlation:-1 to +1

+1 indicates perfect positive linear relation

-1 indicates perfect negative linear relation

Rating Prediction Formula

$$\hat{R}_{u,i} = \bar{R}_u + \frac{\sum_v \text{sim}(u,v) \cdot (R_{v,i} - \bar{R}_v)}{\sum |\text{sim}(u,v)|}$$

Where:

- $\hat{R}_{u,i}$ = predicted rating of user *u* for item *i*
- $\bar{R}_u, \bar{R}_v$ = average ratings of users *u* and *v*
- $R_{v,i}$ = rating given by similar user *v* to item *i*

Pearson Correlation Calculation

Movie	User A (u)	User B (v)
M1	4	5
M2	3	1
M3	5	4
M4	4	4

🔢 Step 1: Compute Average Ratings

- $\bar{R}_u = \frac{4+3+5+4}{4} = 4.0$
- $\bar{R}_v = \frac{5+1+4+4}{4} = 3.5$

Movie	Ru	Rv	$R_u - \bar{R}_u$	$R_v - \bar{R}_v$	Product
M1	4	5	0	1.5	0
M2	3	1	-1	-2.5	2.5
M3	5	4	1	0.5	0.5
M4	4	4	0	0.5	0

- **Numerator**:

$$\sum (R_u - \bar{R}_u)(R_v - \bar{R}_v) = 0 + 2.5 + 0.5 + 0 = 3.0$$

Denominator:

- For user u:

$$\sqrt{(-1)^2 + 0^2 + 1^2 + 0^2} = \sqrt{2} \approx 1.414$$

- For user v:

$$\sqrt{(1.5)^2 + (-2.5)^2 + (0.5)^2 + (0.5)^2} = \sqrt{2.25 + 6.25 + 0.25 + 0.25} = \sqrt{9.0} = 3.0$$

Movie	Ru	Rv	$R_u - \bar{R}_u$	$R_v - \bar{R}_v$	Product
M1	4	5	0	1.5	0
M2	3	1	-1	-2.5	2.5
M3	5	4	1	0.5	0.5
M4	4	4	0	0.5	0

✅ Final Calculation

$$\text{sim}(u, v) = \frac{3.0}{1.414 \cdot 3.0} = \frac{3.0}{4.242} \approx 0.707$$

✅ Final Answer:

$$\boxed{\text{Pearson similarity between User A and User B} \approx 0.707}$$

Rating predictions

$$\hat{R}_{u,i} = \bar{R}_u + \frac{\sum_v \text{sim}(u,v) \cdot (R_{v,i} - \bar{R}_v)}{\sum_v |\text{sim}(u,v)|}$$

From the earlier example:

- $\bar{R}_u = 4.0$ (User A's average)
- $\bar{R}_v = 3.5$ (User B's average)
- $\text{sim}(u,v) = 0.707$
- $R_{v,i} = 3$ (User B's rating on M5)

Movie	R_u	R_v	$R_u - \bar{R}_u$	$R_v - \bar{R}_v$	Product
M1	4	5	0	1.5	0
M2	3	1	-1	-2.5	2.5
M3	5	4	1	0.5	0.5
M4	4	4	0	0.5	0
M5	??	3			

$$\hat{R}_{u,\text{M5}} = 4.0 + \frac{0.707 \cdot (3 - 3.5)}{0.707} = 4.0 + \frac{0.707 \cdot (-0.5)}{0.707} = 4.0 - 0.5 = \boxed{3.5}$$

Example

User-Based Collaborative Filtering (UBCF) with **5 users** and **5 movies**, where we:

1. Create a user-item rating matrix
2. Calculate similarities
3. Choose 2 nearest neighbors
4. Predict ratings for unseen movies
5. Recommend the best movie

User-Item rating matrix

Movie/User	U1	U2	U3	U4	U5
M1	5	3	4	4	1
M2	3	1	2	3	5
M3	4	2	5	2	–
M4	–	3	4	–	2
M5	–	–	5	3	1

Calculate Similarities

Movie	UserU1	U2	U3	U4	U5
M1	5	3	4	4	1
M2	3	1	2	3	5
M3	4	2	5	2	–
M4	–	3	4	–	2
M5	–	–	5	3	1

To recommend movie to user 1,

We need to find similarity between User 1 and Other users

Similarity Calculations

Movie	U1	U3
M1	5	4
M2	3	2
M3	4	5

- $\bar{R}_{U1} = \frac{5+3+4}{3} = 4.0$
- $\bar{R}_{U3} = \frac{4+2+5}{3} = 3.67$

Now calculate:

$$\text{sim}(U1, U3) = \frac{(5-4)(4-3.67) + (3-4)(2-3.67) + (4-4)(5-3.67)}{\sqrt{(1)^2 + (-1)^2 + (0)^2} \cdot \sqrt{(0.33)^2 + (-1.67)^2 + (1.33)^2}}$$

$$= \frac{1 \cdot 0.33 + (-1) \cdot (-1.67) + 0 \cdot 1.33}{\sqrt{1+1+0} \cdot \sqrt{0.1089 + 2.7889 + 1.7689}} = \frac{0.33 + 1.67 + 0}{\sqrt{2} \cdot \sqrt{4.6667}} \approx \frac{2.0}{1.41 \cdot 2.16} \approx \frac{2.0}{3.05} \approx 0.656$$

> Similarly, calculate sim(U1, U2), sim(U1, U4), sim(U1, U5) (skip exact math here for brevity):

Calculate Similarity of User 1 with other users

- **sim(U1, U2) ≈ 0.4**
- **sim(U1, U3) ≈ 0.66**
- **sim(U1, U4) ≈ 0.3**
- **sim(U1, U5) ≈ -0.8** (opposite taste)

✅ Step 3: Choose Top 2 Similar Users

Top neighbors for **U1** are:

- **U3 (sim = 0.66)**
- **U2 (sim = 0.4)**

Rating Prediction

$$\hat{R}_{u,i} = \bar{R}_u + \frac{\sum_v \text{sim}(u,v) \cdot (R_{v,i} - \bar{R}_v)}{\sum |\text{sim}(u,v)|}$$

🔷 **Predict $\hat{R}_{U1,M4}$**

Only U2 and U3 rated M4:

- $R_{U2,M4} = 3, \bar{R}_{U2} = (3+1+2+3)/4 = 2.25$
- $R_{U3,M4} = 4, \bar{R}_{U3} = (4+2+5+4+5)/5 = 4.0$

Apply formula:

$$\hat{R}_{U1,M4} = \bar{R}_{U1} + \frac{0.4 \cdot (3 - 2.25) + 0.66 \cdot (4 - 4)}{0.4 + 0.66} = 4.0 + \frac{0.4 \cdot 0.75}{1.06} \approx 4.0 + 0.283 \approx \boxed{4.28}$$

Rating Prediction

$$\hat{R}_{u,i} = \bar{R}_u + \frac{\sum_v \text{sim}(u,v) \cdot (R_{v,i} - \bar{R}_v)}{\sum |\text{sim}(u,v)|}$$

🔷 **Predict $\hat{R}_{U1,M5}$**

Only U3 rated M5 = 5, and $\bar{R}_{U3} = 4.0$

$$\hat{R}_{U1,M5} = 4.0 + \frac{0.66 \cdot (5 - 4)}{0.66} = 4.0 + 1.0 = \boxed{5.0}$$

Final User-Item Matrix

Movie	UserU1	U2	U3	U4	U5
M1	5	3	4	4	1
M2	3	1	2	3	5
M3	4	2	5	2	–
M4	4.28	3	4	–	2
M5	5.0	–	5	3	1

✅ Step 5: Recommend the Best Movie

U1 hasn't rated M4 and M5. We predicted:

- **M4 → 4.28**
- **M5 → 5.0**

🎉 Recommend: Movie M5 to User U1

Item Based Collaborative Filtering

Item-Based Collaborative Filtering (IBCF) recommends items to a user based on **how similar items are** to those the user has already **rated** or interacted with.

"People who rated Item A highly also rated Item B highly — so if you liked A, you'll probably like B."

It’s like:

- You rated **Avengers** 5 stars.
- Other users who rated **Avengers** also rated **Iron Man** 5 stars.
- So we recommend **Iron Man** to you — based on **similar rating behavior**, not content.

Note: Unlike content-based filtering, we do not use item features to find similarities. Instead, we use user ratings of the items.

Steps of item-based CF

Step 1: Build Item Similarity Matrix

Compare every item with every other item using **user ratings**.

Example: How similar are *Movie A* and *Movie B* based on how users rated both?

Step 2: Look at Items the User Has Rated

Find the items that the target user has already rated.

Example: User U1 rated *Movie A = 5*, *Movie C = 4*

Step 3: Find Similar Items

For each item the user hasn't rated, find how **similar** it is to the ones the user **has** rated.

Example: How similar is *Movie B* to *Movie A* and *Movie C*?

Step 4: Predict Ratings for Unseen Items

Estimate how the user would rate unseen items using the similarity scores and the user's own ratings.

Predicted rating for *Movie B* = weighted average of similarities × user's ratings

Step 5: Recommend Top-N Items

Recommend the **top N items** with the **highest predicted scores**.

Recommend the ones the user is most likely to enjoy!

Example

User / Movie	M1	M2	M3	M4	M5
U1	5	3	4	?	?
U2	3	1	2	3	3
U3	4	2	5	4	?
U4	4	3	2	?	1
U5	1	5	2	1	2

We want to **recommend either M4 or M5** to **User 1**, based on items they have already rated: **M1 = 5, M2 = 3, M3 = 4**.

Compute Pearson Correlation Between M4 and (M1, M2, M3)

sim(M4, M1)

User	M4	M1
U2	3	3
U3	4	4
U5	1	1

Pearson Formula for Items *i* and *j*:

$$\text{sim}(i,j) = \frac{\sum_u (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_u (R_{u,i} - \bar{R}_i)^2} \cdot \sqrt{\sum_u (R_{u,j} - \bar{R}_j)^2}}$$

📊 Step 1: Mean Ratings

- Mean for **M4** = (3 + 4 + 1) / 3 = **2.67**
- Mean for **M1** = (3 + 4 + 1) / 3 = **2.67**

🧮 Step 2: Pearson Numerator

$$\sum (R_{u,M4} - \bar{R}_{M4})(R_{u,M1} - \bar{R}_{M1})$$

- U2: (3 – 2.67)(3 – 2.67) = (0.33)(0.33) ≈ 0.1089
- U3: (4 – 2.67)(4 – 2.67) = (1.33)(1.33) ≈ 1.7689
- U5: (1 – 2.67)(1 – 2.67) = (–1.67)(–1.67) ≈ 2.7889

Numerator = 0.1089 + 1.7689 + 2.7889 = **4.6667**

Compute Pearson Correlation Between M4 and M1

User	M4	M1
U2	3	3
U3	4	4
U5	1	1

Pearson Formula for Items *i* and *j*:

$$\text{sim}(i,j) = \frac{\sum_u (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_u (R_{u,i} - \bar{R}_i)^2} \cdot \sqrt{\sum_u (R_{u,j} - \bar{R}_j)^2}}$$

📊 Step 1: Mean Ratings

- Mean for **M4** = (3 + 4 + 1) / 3 = **2.67**
- Mean for **M1** = (3 + 4 + 1) / 3 = **2.67**

📐 Step 3: Pearson Denominator

Same as:

$$\sqrt{\sum (R_{u,M4} - \bar{R}_{M4})^2} \cdot \sqrt{\sum (R_{u,M1} - \bar{R}_{M1})^2}$$

Since the differences are the same for M4 and M1, we'll square the same terms twice:

$$\sqrt{0.1089 + 1.7689 + 2.7889} \cdot \sqrt{0.1089 + 1.7689 + 2.7889} = \sqrt{4.6667} \cdot \sqrt{4.6667} = 4.6667$$

Pearson Formula for Items *i* and *j*:

$$\text{sim}(i, j) = \frac{\sum_u (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_u (R_{u,i} - \bar{R}_i)^2} \cdot \sqrt{\sum_u (R_{u,j} - \bar{R}_j)^2}}$$

✅ Final Pearson Correlation

$$\text{sim}(\text{M4}, \text{M1}) = \frac{4.6667}{4.6667} = 1.0$$

sim(M4 and M2)

User	M4	M2
U2	3	1
U3	4	2
U5	1	5

- Mean M4 = (3+4+1)/3 = 2.67
- Mean M2 = (1+2+5)/3 = 2.67

$$\text{Numerator} = (3-2.67)(1-2.67) + (4-2.67)(2-2.67) + (1-2.67)(5-2.67)$$

$$= (0.33)(-1.67) + (1.33)(-0.67) + (-1.67)(2.33)$$

$$\approx -5.332$$

$$\text{Denominator} = \sqrt{(0.33)^2 + (1.33)^2 + (-1.67)^2} \cdot \sqrt{(-1.67)^2 + (-0.67)^2 + (2.33)^2} =$$

$$= \sqrt{4.6667} \cdot \sqrt{8.6667} \approx 2.16 \times 2.94 \approx 6.36$$

$$\text{sim}(\text{M4}, \text{M2}) \approx \frac{-5.332}{6.36} \approx -0.838$$

Similarly, calculate similarity between m4 and m3

$$\text{sim}(\text{M4}, \text{M3}) = \frac{1.5}{1.5} = 1.0$$

Rating prediction formula

$$\hat{R}_{u,i} = \bar{R}_i + \frac{\sum_{j \in N} \text{sim}(i,j) \cdot (R_{u,j} - \bar{R}_j)}{\sum_{j \in N} |\text{sim}(i,j)|}$$

Where:

- $\hat{R}_{u,i}$ is the predicted rating of user u on item i
- $\bar{R}_i$ is the **average rating** of item i (not of user u!)
- $R_{u,j}$ is the rating user u gave to item j
- $\bar{R}_j$ is the **mean rating of item** j (since we're comparing items, not users)
- $\text{sim}(i,j)$ is the Pearson similarity between item i and j
- N is the set of items that user u has rated and are similar to i

Rating prediction formula

$$\hat{R}_{u,i} = \bar{R}_i + \frac{\sum_{j \in N} \text{sim}(i,j) \cdot (R_{u,j} - \bar{R}_j)}{\sum_{j \in N} |\text{sim}(i,j)|}$$

User 1 ratings:

- M1 = 5 → (5 – 3.0 = **+2.0**)
- M2 = 3 → (3 – 3.4 = **–0.4**)
- M3 = 4 → (4 – 3.6 = **+0.4**)

Use similarities:

- sim(M4, M1) = 1.0
- sim(M4, M2) = –0.838
- sim(M4, M3) = 1.0

Now:

$$\hat{R}_{U1,M4} = \bar{R}_{M4} + \frac{1.0 \cdot 2.0 + (-0.838) \cdot (-0.4) + 1.0 \cdot 0.4}{|1.0| + |-0.838| + |1.0|}$$

$$= 2.5 + \frac{2.0 + 0.3352 + 0.4}{2.838} = 2.5 + \frac{2.7352}{2.838} \approx 2.5 + 0.964 = \boxed{3.46}$$

🎯 Similarly, Predict M5 for U1?

You can repeat with:

- Similarities: sim(M5, M1), sim(M5, M2), sim(M5, M3)
- Predict rating like above

- Lets assume, rating of m5 by user 1 is 4.5

Rating by User 1

M4	3.46
M5	4.5 (recommend)

End of Lecture