**✅ Similarity Function Variants – Purpose and Formulas**

**📘 Introduction**

In recommender systems, **similarity functions** help to **compare users or items** to find which are alike.  
The system uses this comparison to **recommend items** based on **similar users’ preferences** or **similar item features**.

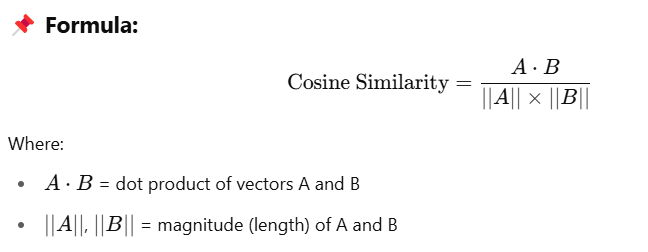
🔁 **Purpose:** To measure how close two users or two items are, based on their **ratings** or **attributes**.

There are several similarity functions used in **content-based** and **collaborative filtering** methods.

**🧮 1. Cosine Similarity**

**📌 Purpose:**

Measures the **angle** between two vectors (user or item profiles). It checks how similar their **directions** are, not their values.

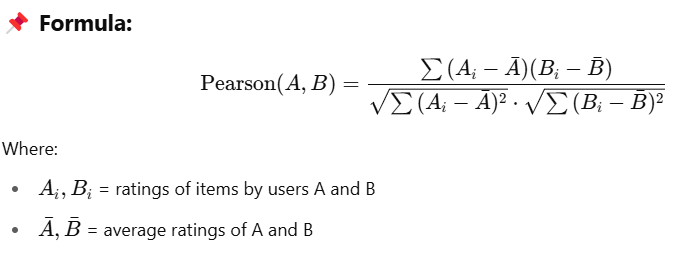
****

**📌 Use Case:**

* Used in **text-based systems**, item profiling, and **content-based** recommenders.
* Best when only **pattern** of interest matters, not rating values.

**📊 2. Pearson Correlation Coefficient**

**📌 Purpose:**

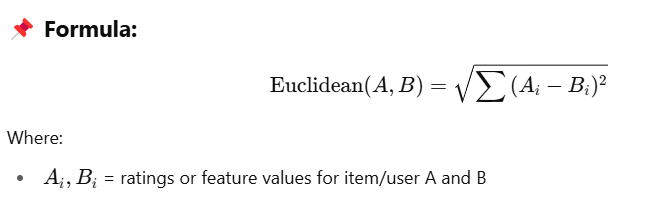
****Measures how **strongly two users or items are linearly related**, based on their ratings.

**📌 Use Case:**

* Used in **user-based collaborative filtering**.
* Accounts for **individual rating behavior** (like lenient or strict raters).

**🧮 3. Euclidean Distance**

**📌 Purpose:**

****Measures the **straight-line distance** between two users or items in multi-dimensional space.

**📌 Use Case:**

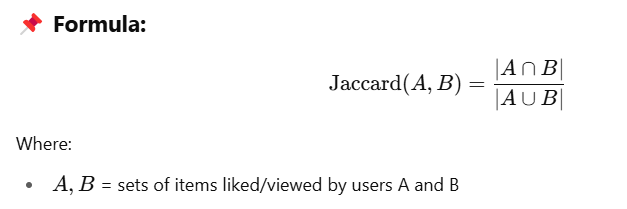
* Used when **absolute difference** in ratings matters.
* Common in **item-based filtering**.

🔻 **Note:** Smaller distance = more similar

**💡 4. Jaccard Similarity**

**📌 Purpose:**

Measures similarity between two **sets** – useful when users interact with items (like purchased, viewed, liked) instead of giving ratings.

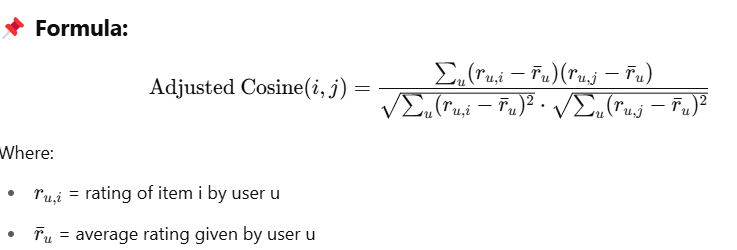
****

**📌 Use Case:**

* Works for **binary data** like clicks or purchases.
* Suitable for **implicit feedback** systems.

**🎯 5. Adjusted Cosine Similarity**

**📌 Purpose:**

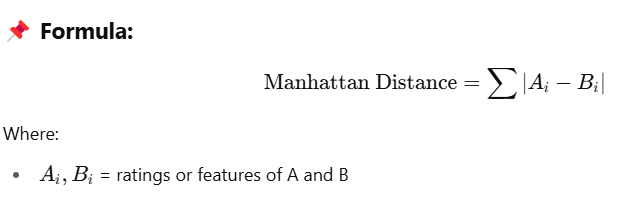
****Improves cosine similarity by removing **user bias** in ratings.

**📌 Use Case:**

* Used in **item-based collaborative filtering**.
* Adjusts for differences in user rating scale.

**🔍 6. Manhattan Distance (L1 Norm)**

**📌 Purpose:**

****Measures total absolute difference between feature values.

**📌 Use Case:**

* Less sensitive to outliers than Euclidean.
* Used in **simple item comparisons**.

**📊 Comparison of Similarity Measures**

| **Similarity Function** | **Handles Ratings** | **Handles Binary** | **Handles User Bias** | **Common Use** |
| --- | --- | --- | --- | --- |
| Cosine Similarity | ✅ | ❌ | ❌ | Text-based, CBRS |
| Pearson Correlation | ✅ | ❌ | ✅ | CF (User-based) |
| Euclidean Distance | ✅ | ❌ | ❌ | CF (Item-based) |
| Jaccard Similarity | ❌ | ✅ | ❌ | Implicit feedback |
| Adjusted Cosine Similarity | ✅ | ❌ | ✅ | CF (Item-based) |
| Manhattan Distance | ✅ | ❌ | ❌ | General use |

**📝 Conclusion**

Similarity functions are at the heart of recommender systems.  
They help to **identify related users or items**, making it possible to generate **personalized recommendations**.  
Choosing the **right similarity function** depends on the type of data (ratings, binary, text) and the algorithm used (content-based or collaborative filtering).