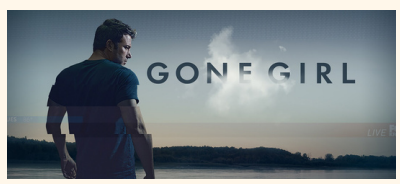
How Euclidean Distance can measure similarity between users

Movies	David	Eva	Frank
 Inception	5	4	2
 The Matrix	5	5	1
 Inter-stellar	4	4	2
 Seven	2	2	5
 Gone Girl	1	2	4
 Fight Club	2	1	5

- As can be visually seen by the ratings given by David, Eva and Frank: David and Eva likes Sci-Fi movies, while Frank likes Thriller movies
- We will see how Euclidean distance clearly captures similarity between user based on the rating information

Calculating Euclidean Distance

1. Euclidean Distance Between  David and  Eva:

For David and Eva, the ratings are as follows:

- David: (5, 5, 4, 2, 1, 2)
- Eva: (4, 5, 4, 2, 2, 1)

The Euclidean distance d_{DE} is calculated as:

$$d_{DE} = \sqrt{(5-4)^2 + (5-5)^2 + (4-4)^2 + (2-2)^2 + (1-2)^2 + (2-1)^2}$$

$$d_{DE} = \sqrt{1+0+0+0+1+1}$$

$$d_{DE} = \sqrt{3} \approx 1.73$$

2. Euclidean Distance Between  David and  Frank:

For David and Frank, the ratings are as follows:

- David: (5, 5, 4, 2, 1, 2)
- Frank: (2, 1, 2, 5, 4, 5)

The Euclidean distance d_{DF} is calculated as:

$$d_{DF} = \sqrt{(5-2)^2 + (5-1)^2 + (4-2)^2 + (2-5)^2 + (1-4)^2 + (2-5)^2}$$

$$d_{DF} = \sqrt{9+16+4+9+9+9}$$

$$d_{DF} = \sqrt{56} \approx 7.48$$

in DineshLal

Calculating Euclidean Distance

3. Euclidean Distance Between  Eva and  Frank:

For Eva and Frank, the ratings are as follows:

- Eva: (4, 5, 4, 2, 2, 1)
- Frank: (2, 1, 2, 5, 4, 5)

The Euclidean distance d_{EF} is calculated as:

$$d_{EF} = \sqrt{(4-2)^2 + (5-1)^2 + (4-2)^2 + (2-5)^2 + (2-4)^2 + (1-5)^2}$$

$$d_{EF} = \sqrt{4 + 16 + 4 + 9 + 4 + 16}$$

$$d_{EF} = \sqrt{53} \approx 7.28$$

All Distances Summarized:

User Pair	Formula	Distance
David-Eva	$\sqrt{(5-4)^2 + (5-5)^2 + (4-4)^2 + (2-2)^2 + (1-2)^2 + (2-1)^2}$	1.73
	$\sqrt{1 + 0 + 0 + 0 + 1 + 1}$	
	$\sqrt{3} \approx 1.73$	
David-Frank	$\sqrt{(5-2)^2 + (5-1)^2 + (4-2)^2 + (2-5)^2 + (1-4)^2 + (2-5)^2}$	7.48
	$\sqrt{9 + 16 + 4 + 9 + 9 + 9}$	
	$\sqrt{56} \approx 7.48$	
Eva-Frank	$\sqrt{(4-2)^2 + (5-1)^2 + (4-2)^2 + (2-5)^2 + (2-4)^2 + (1-5)^2}$	7.28
	$\sqrt{4 + 16 + 4 + 9 + 4 + 16}$	
	$\sqrt{53} \approx 7.28$	

in DineshLal

Interpretation of Results

User Pair	Euclidean Distance	Interpretation
David-Eva 	1.73	David and Eva have very similar tastes in movies.
David-Frank 	7.48	David and Frank have very different tastes in movies.
Eva-Frank 	7.28	Eva and Frank have quite different tastes in movies.

DineshLal

Python Implementation

```
import numpy as np

# Ratings by users
ratings = {
    'David': np.array([5, 5, 4, 2, 1, 2]),
    'Eva': np.array([4, 5, 4, 2, 2, 1]),
    'Frank': np.array([2, 1, 2, 5, 4, 5])
}

# Function to calculate Euclidean distance
def euclidean_distance(vec1, vec2):
    return np.sqrt(np.sum((vec1 - vec2) ** 2))

# Distances between users
distance_DE = euclidean_distance(ratings['David'], ratings['Eva'])
distance_DF = euclidean_distance(ratings['David'], ratings['Frank'])
distance_EF = euclidean_distance(ratings['Eva'], ratings['Frank'])

print(f"Euclidean Distance between David and Eva: {distance_DE}")
print(f"Euclidean Distance between David and Frank: {distance_DF}")
print(f"Euclidean Distance between Eva and Frank: {distance_EF}")
```

Output

```
Euclidean Distance between David and Eva: 1.7320508075688772
Euclidean Distance between David and Frank: 7.483314773547883
Euclidean Distance between Eva and Frank: 7.280109889280518
```

DineshLal

Conclusion

- By using Euclidean distance, we can measure the similarity between different users based on their movie ratings.
- In this example, David and Eva have very similar tastes in movies, while Frank's tastes are quite different.
- This measure helps recommendation systems to identify users with similar tastes and make personalized suggestions.
- Understanding these distances allows recommendation systems to provide more accurate and relevant recommendations to users.

Mathematical Definition of Euclidean Distance

Euclidean distance is a measure of the straight-line distance between two points in Euclidean space.

In a two-dimensional space, the Euclidean distance between two points p and q, with coordinates (p1, p2) and (q1, q2) respectively, is given by the following formula:

$$d(p, q) = \sqrt{((p1 - q1)^2 + (p2 - q2)^2)}$$

This formula can be extended to n-dimensional space, where the Euclidean distance between points p and q with coordinates (p1, p2, ..., pn) and (q1, q2, ..., qn) is:

$$d(p, q) = \sqrt{((p1 - q1)^2 + (p2 - q2)^2 + ... + (pn - qn)^2)}$$

in DineshLal

Benefits of Using Euclidean Distance

1. Simplicity
Euclidean distance is simple to understand and implement. It provides an intuitive measure of similarity based on geometric distance.

2. Scalability
Calculating Euclidean distance is computationally efficient, making it suitable for large-scale recommendation systems.

3. Versatility
Euclidean distance can be applied to various types of data, including ratings, purchase history, and user preferences.

Real-World Applications

1. E-commerce
Online retailers use Euclidean distance to recommend products similar to those a user has previously purchased or viewed.

2. Streaming Services
Platforms like Netflix and Spotify use Euclidean distance to recommend movies, TV shows, and music based on user preferences and viewing/listening history.

3. Social Networks
Social media platforms recommend friends or connections by measuring the similarity between user profiles, interests, and activities.

in DineshLal

Special Thanks to ChatGPT and Gemini for Content support

Sources - Key Scores Public Information

DineshLal