

Recommending Products Based on the Nearest Neighbors Model



Swetha Kolalapudi

CO-FOUNDER, LOONYCORN

www.loonycorn.com

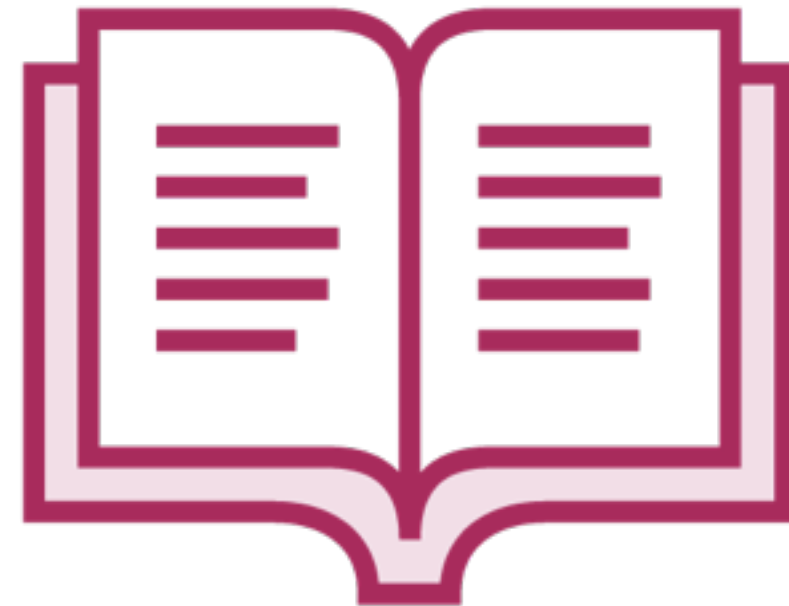
Overview

Understand the nearest neighbors model for collaborative filtering

Measure similarity of users using distance metrics

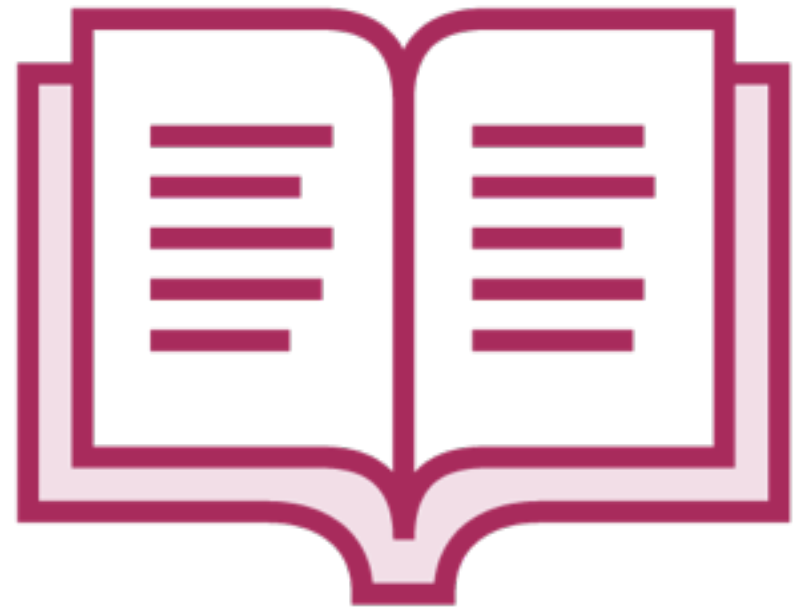
Find the top 10 book recommendations for a user

Book Crossing



A network of book lovers

Finding Book Recommendations



You know what books a user already likes

- User ratings database

Find the top N books to recommend to that user

Recommendation Algorithms

Content based filtering



Find products with
“similar” attributes

Collaborative filtering



Find products liked
by “similar” users

Association rules learning



Find “complementary”
products

Collaborative filtering



Find products liked by “similar” users

Rating Matrix

Users

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	-
U ₂	3	2	-	-	5
U ₃	-	2	-	5	4
U ₄	-	-	4	-	-
U ₅	1	-	-	-	-
U ₆	3	4	-	-	5

Products

Collaborative filtering



Find products liked by “similar” users

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	-
U ₂	3	2	-	-	5
U ₃	-	2	-	5	4
U ₄	-	-	4	-	-
U ₅	1	-	-	-	-
U ₆	3	4	-	-	5

Collaborative Filtering Techniques

Nearest Neighbors Model



Use the ratings of
“most similar” users

Latent Factor Analysis



Solve for underlying
factors that drive the
ratings

Understanding the Nearest Neighbors Model

The rating data
is represented
using a matrix

Rating Matrix

Users

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

Books

Ratings are on a
scale of 1-5

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

User 1 has read
books 1, 2, 5

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

User 1 has not
read books 3, 4

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

Estimate the ratings
for unrated books

Sort them in
descending order

Pick the top 10

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

How do we
estimate user 1's
rating for book 3?

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

An option:
Weighted average of
ratings given by the
“most similar” users

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

User 2 is “more similar” to user 1

User 2 gets a higher weight

Rating Matrix

	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

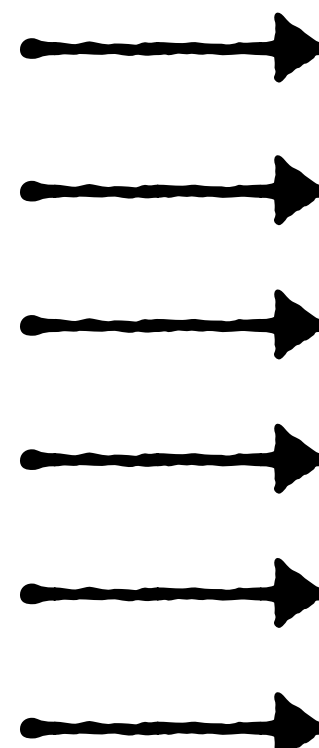
User 5 is “less similar” to user 1

User 5 gets a lower weight

Rating Matrix

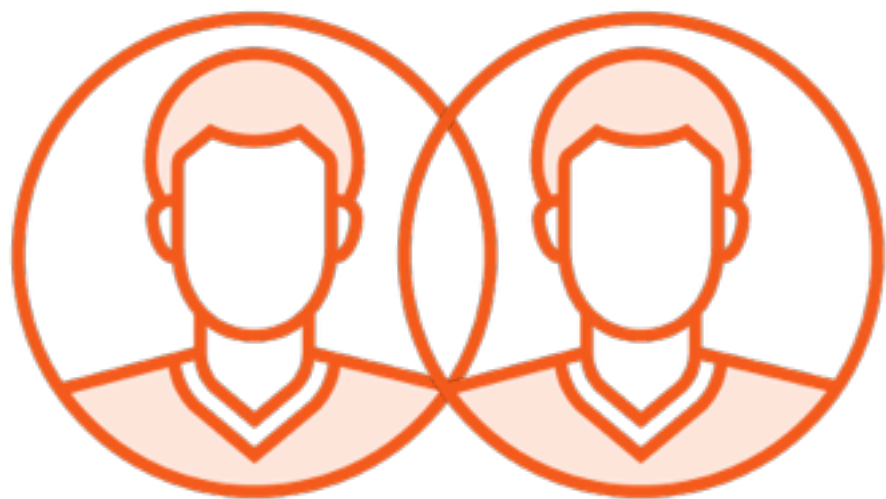
	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

Nearest Neighbors Model

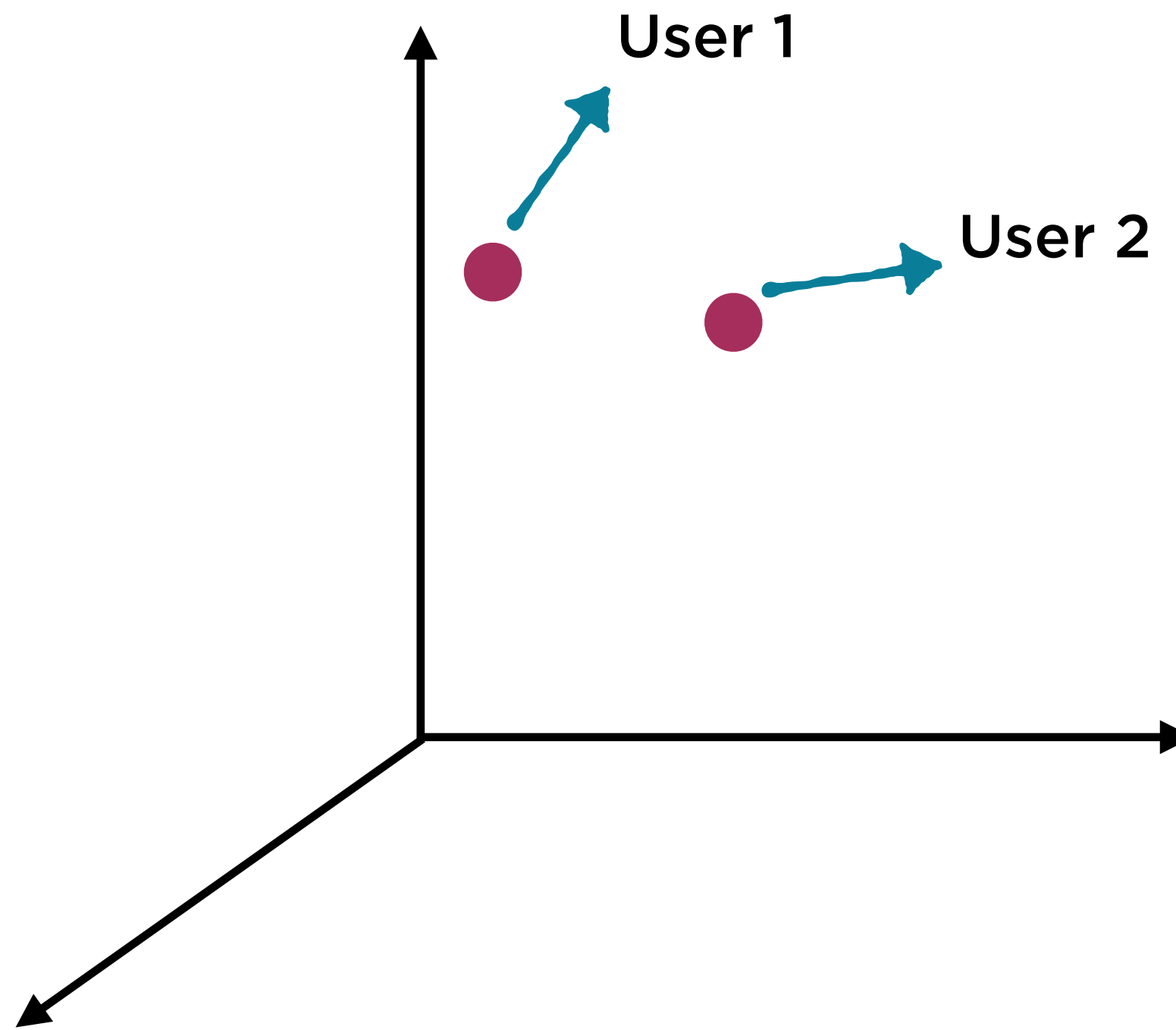


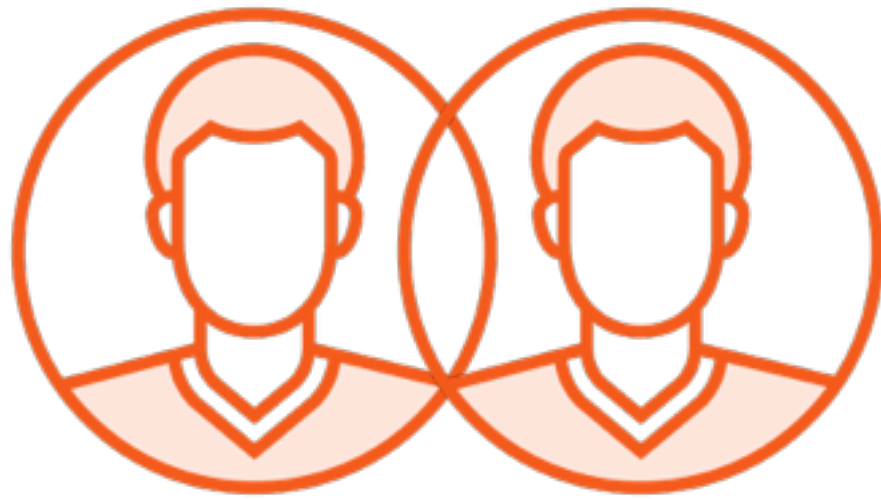
	P₁	P₂	P₃	P₄	P₅
U₁	3	4	-	-	4
U₂	3	5	3	4	5
U₃	4	2	-	5	4
U₄	3	-	4	5	2
U₅	1	-	4	2	1
U₆	3	4	-	2	5

Each user is
represented
as a set of
numbers

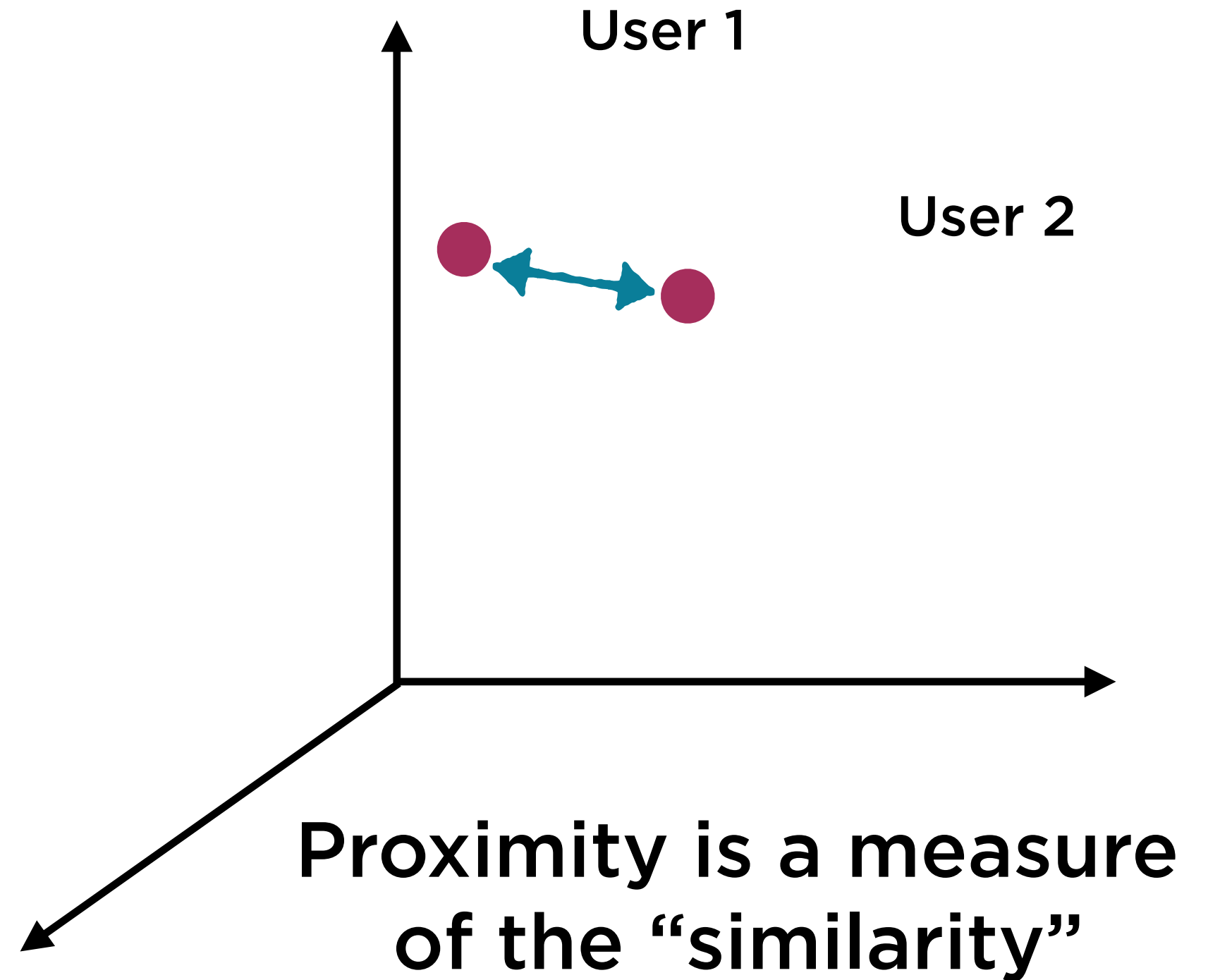


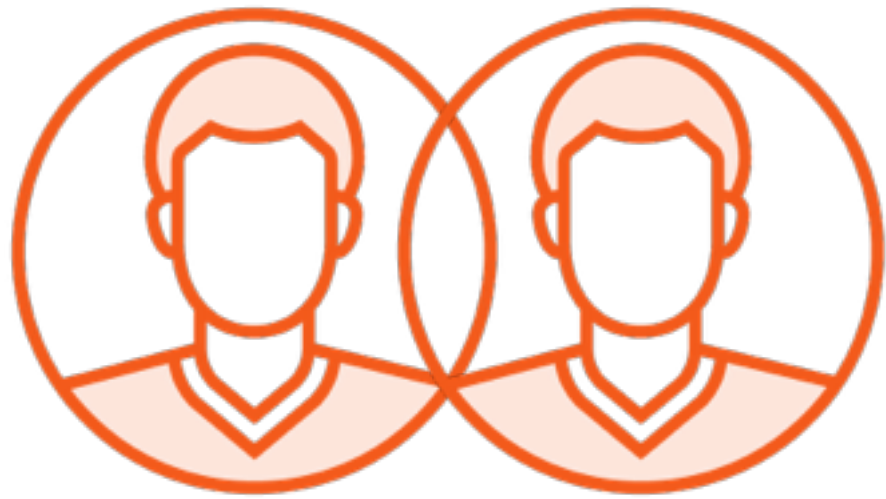
How do we measure
“similarity”?





How do we measure
“similarity”?

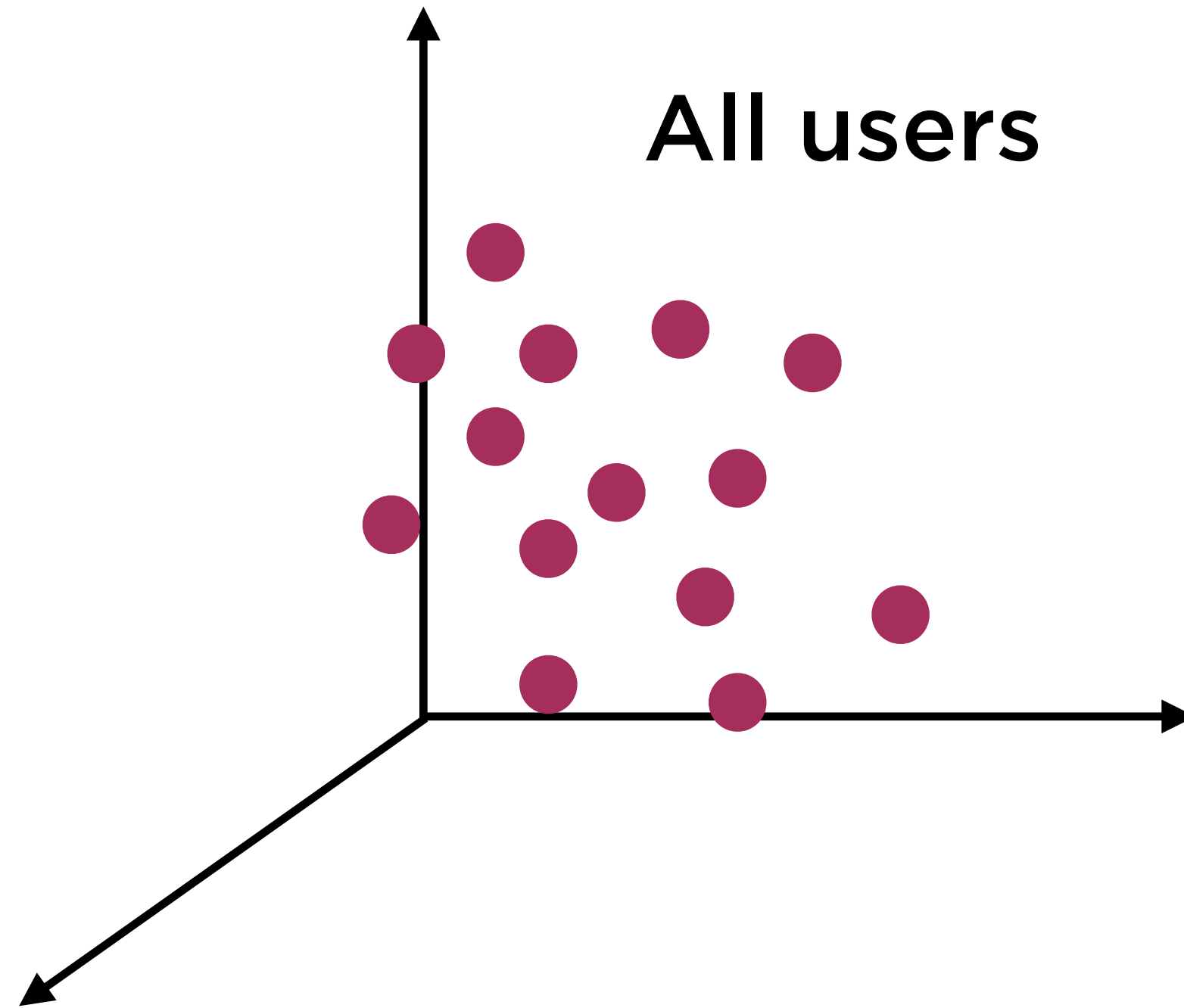




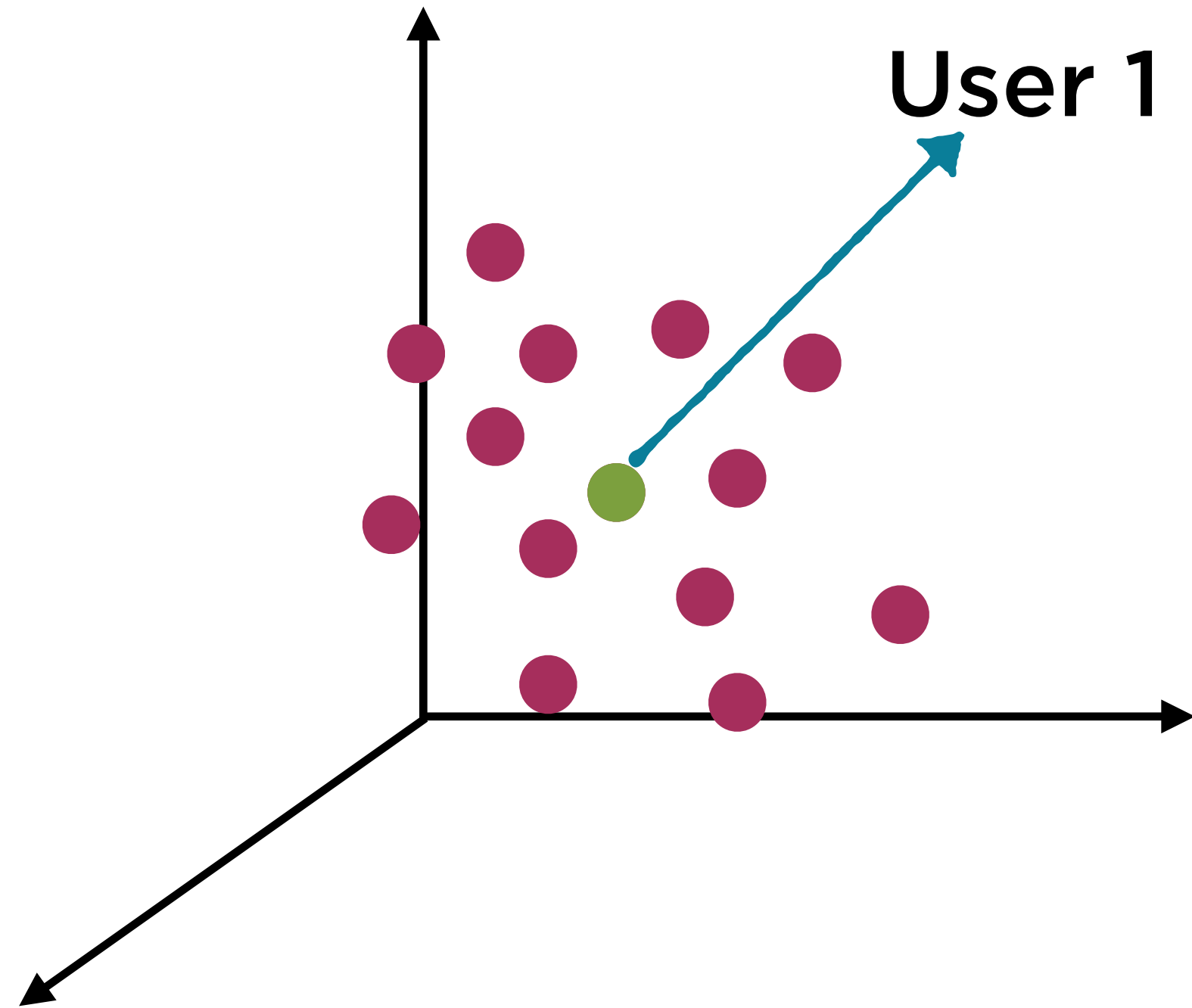
How do we measure
“similarity”?

**“Similarity” is measured
using distance metrics**

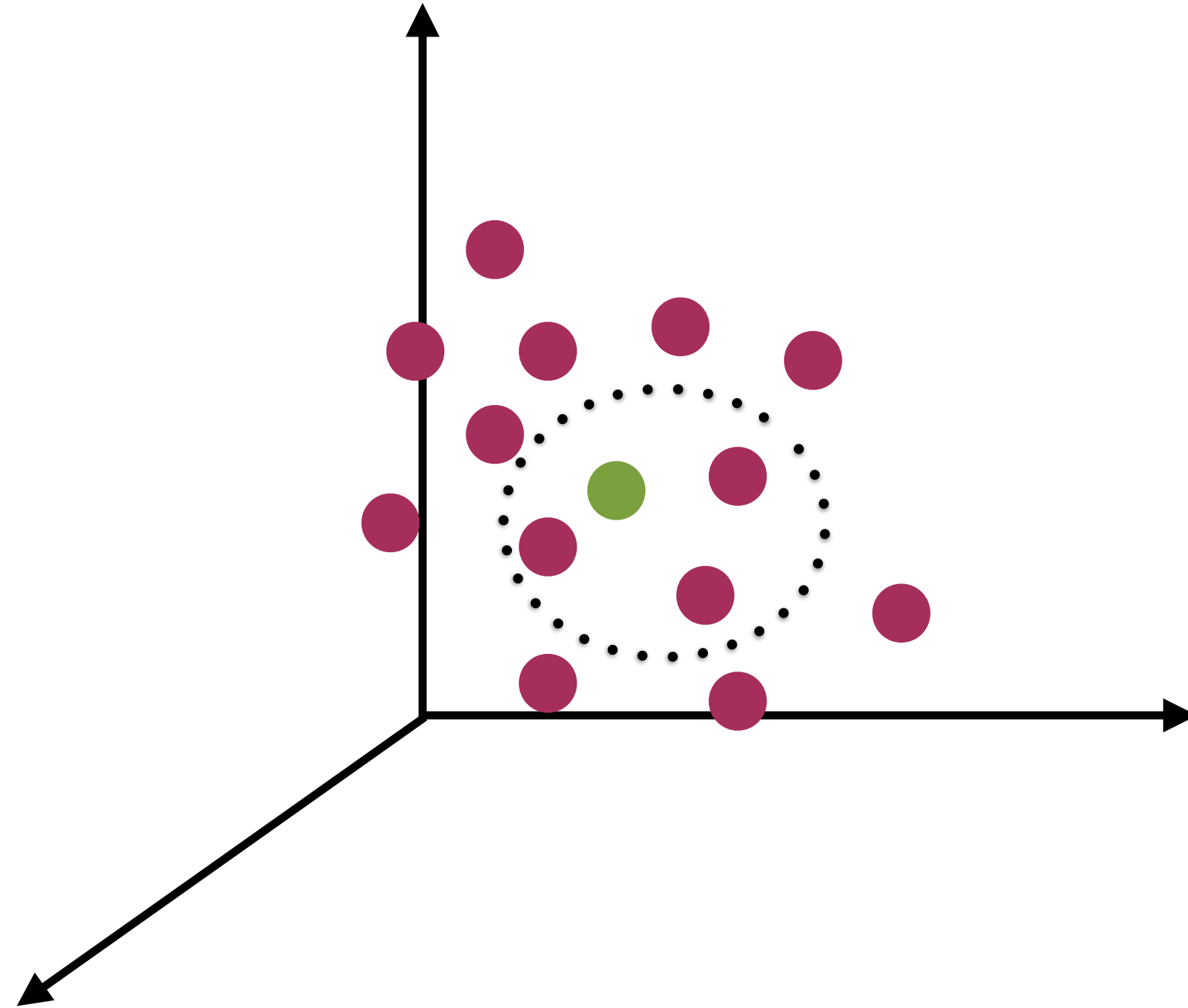
Nearest Neighbors Model



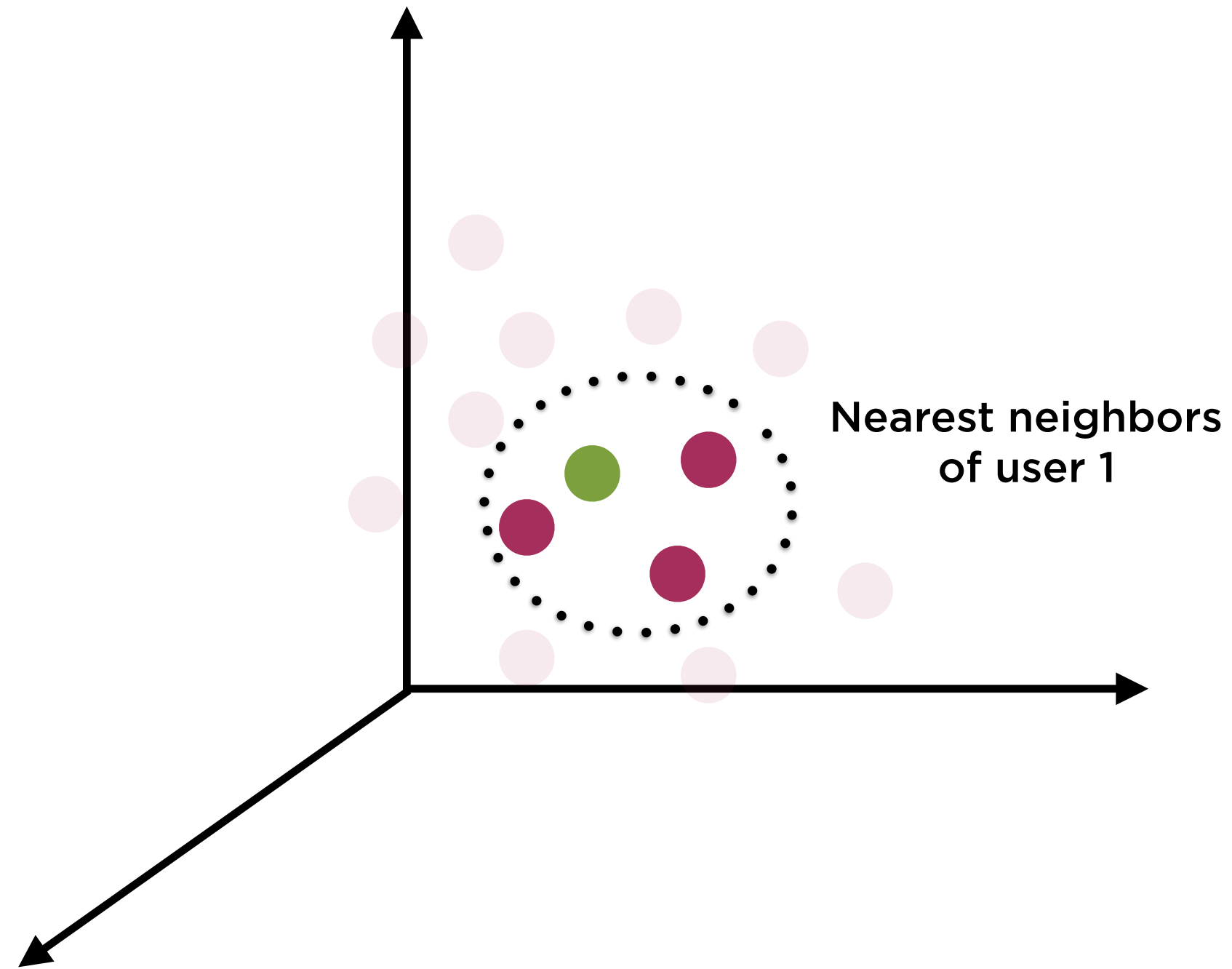
Nearest Neighbors Model



Nearest Neighbors Model



Nearest Neighbors Model



Finding Top 10 Book Recommendations

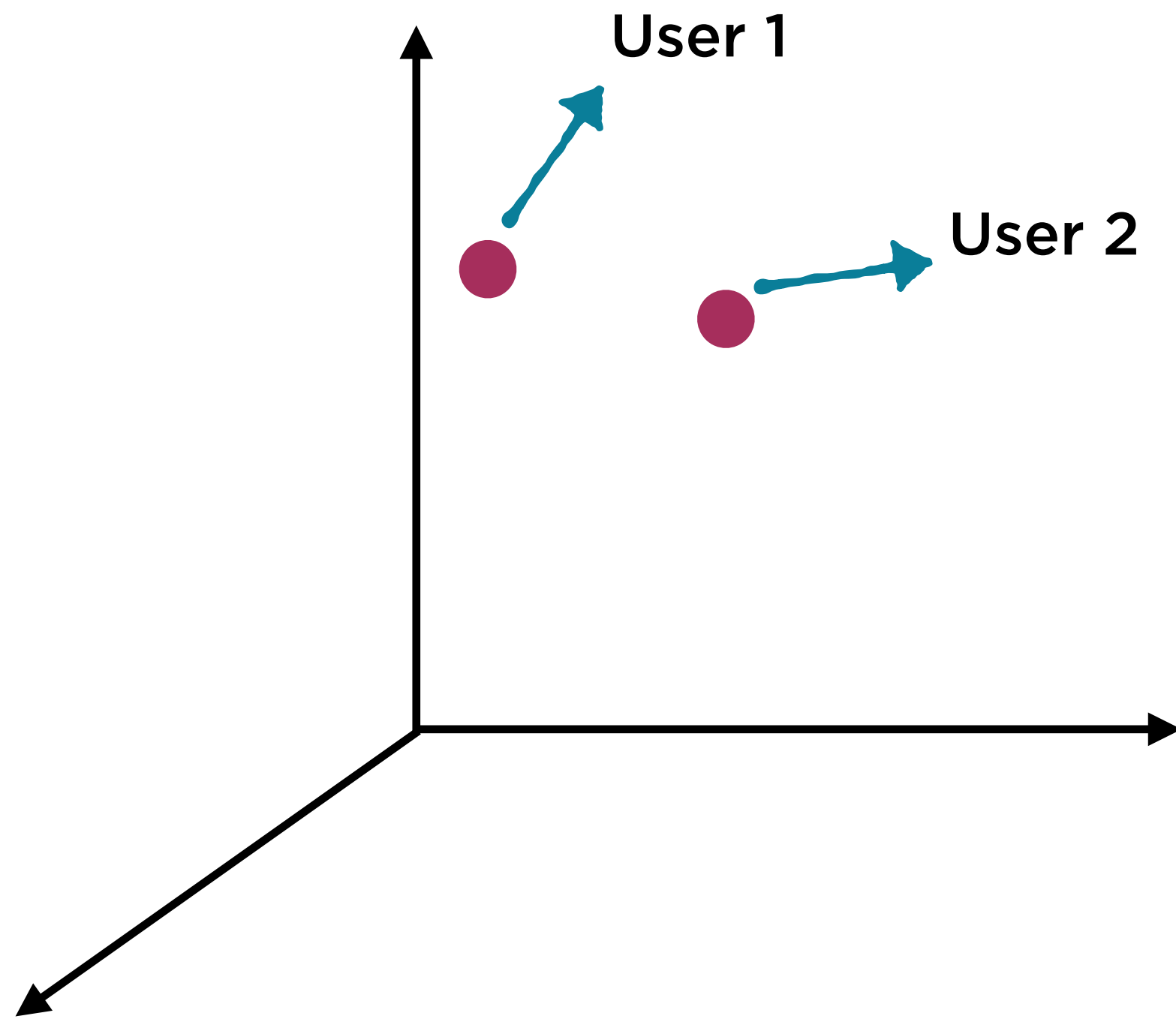
Nearest neighbors
of user 1



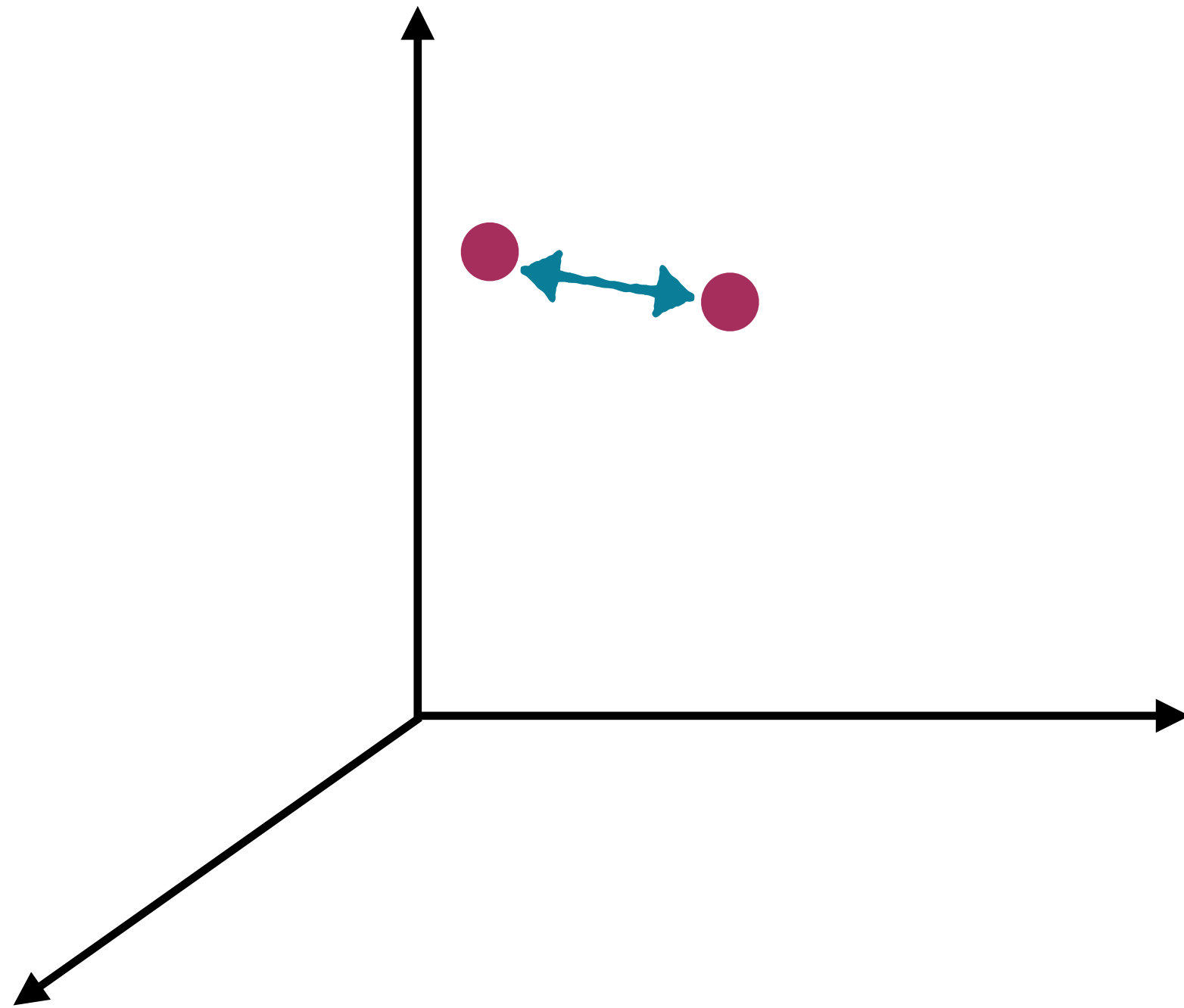
1. Find the K nearest neighbors of a user
2. Average the ratings of nearest neighbors for unrated books
3. Sort in descending order
4. Pick the top 10

Measuring Distance Between Users

Measure the distance
between users



Measure the distance
between users



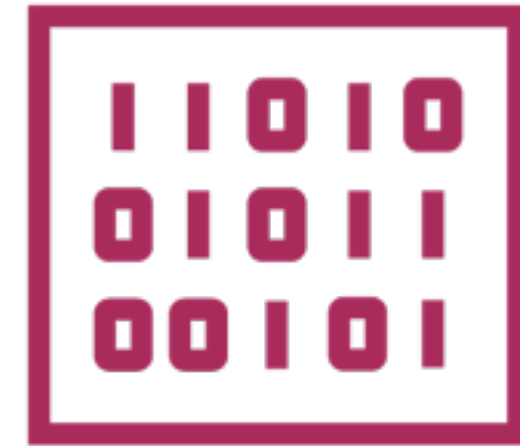
Distance Metrics



**Euclidean
Distance**



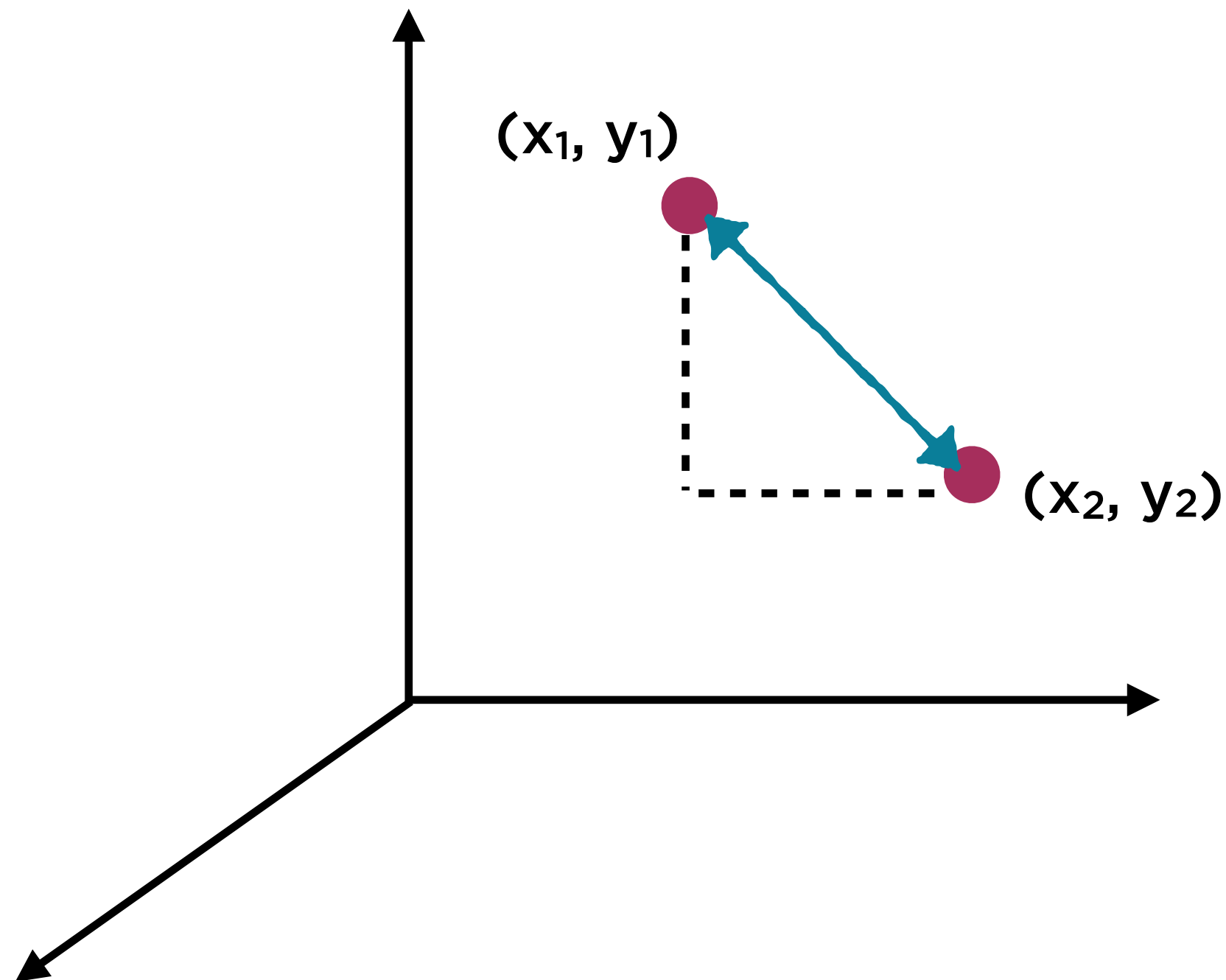
**Correlation
Distance**



**Hamming
Distance**

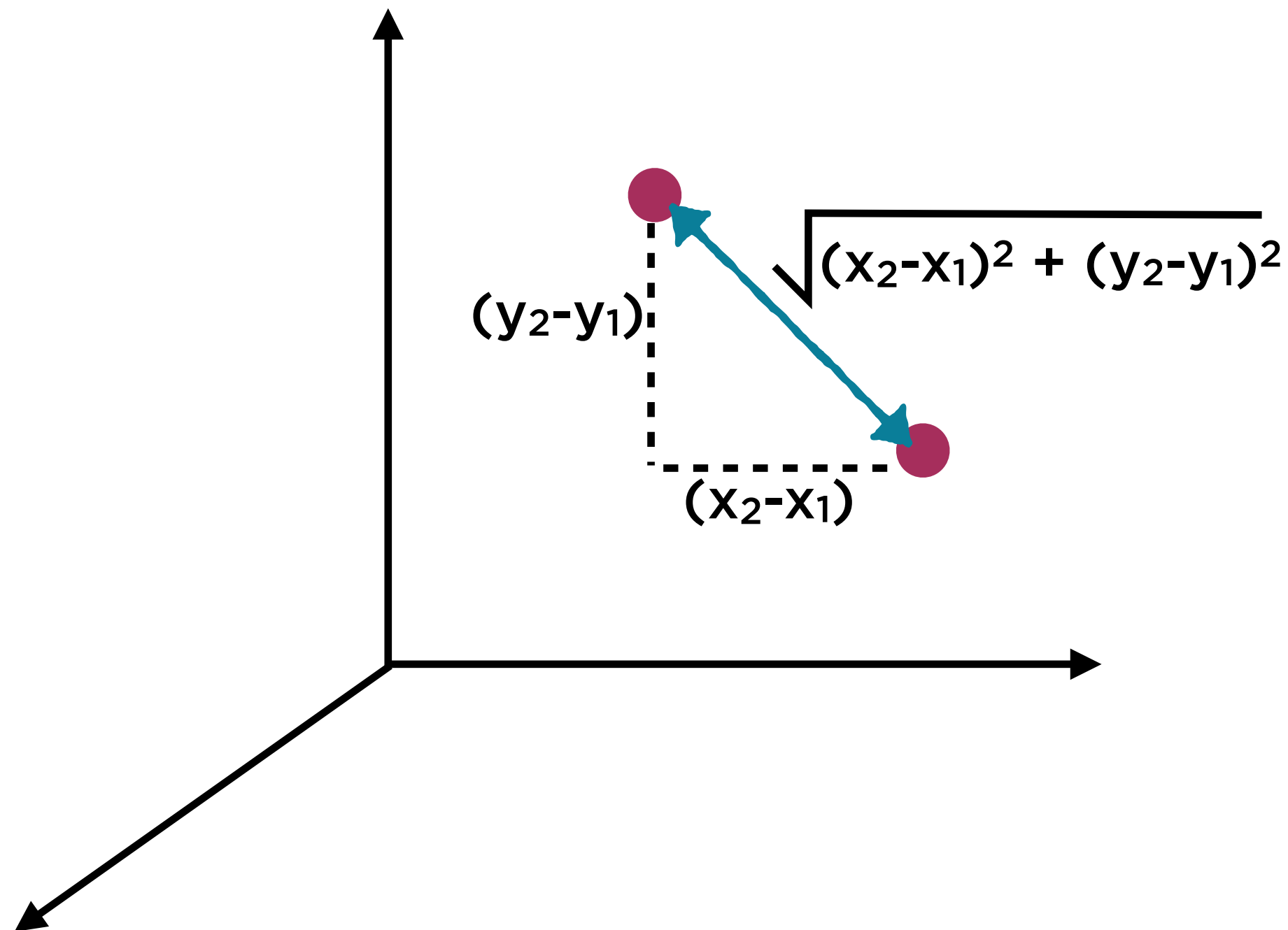


Euclidean
Distance





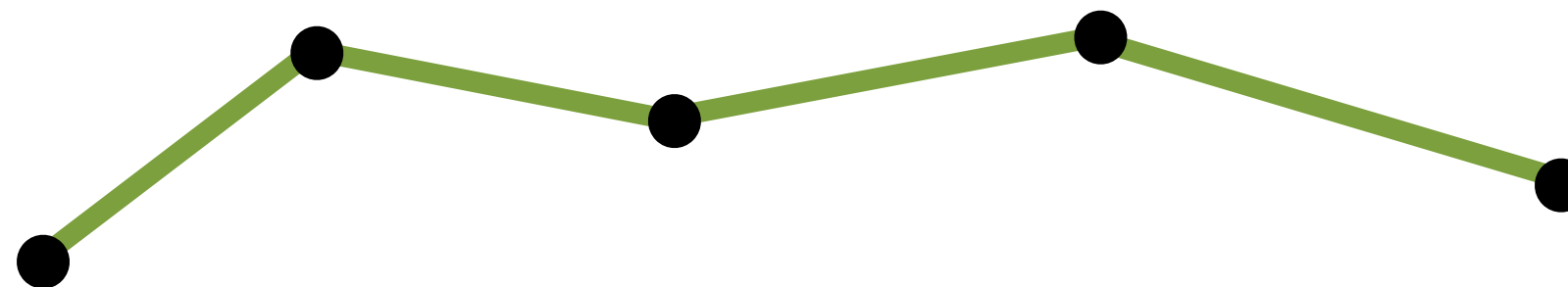
Euclidean
Distance



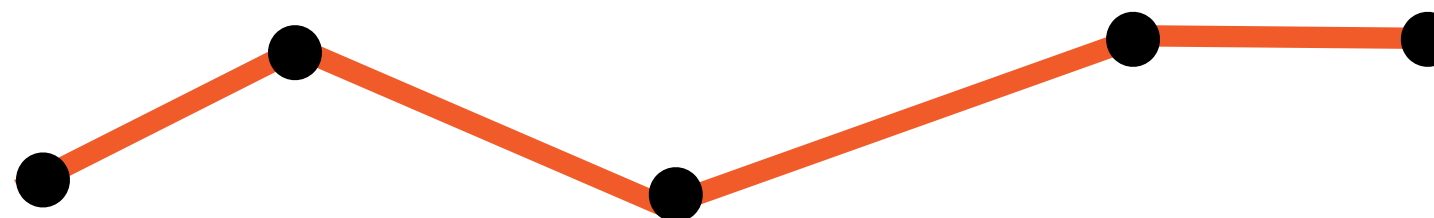


Correlation
Distance

U₁	1	4	3	4	2
----------------------	---	---	---	---	---

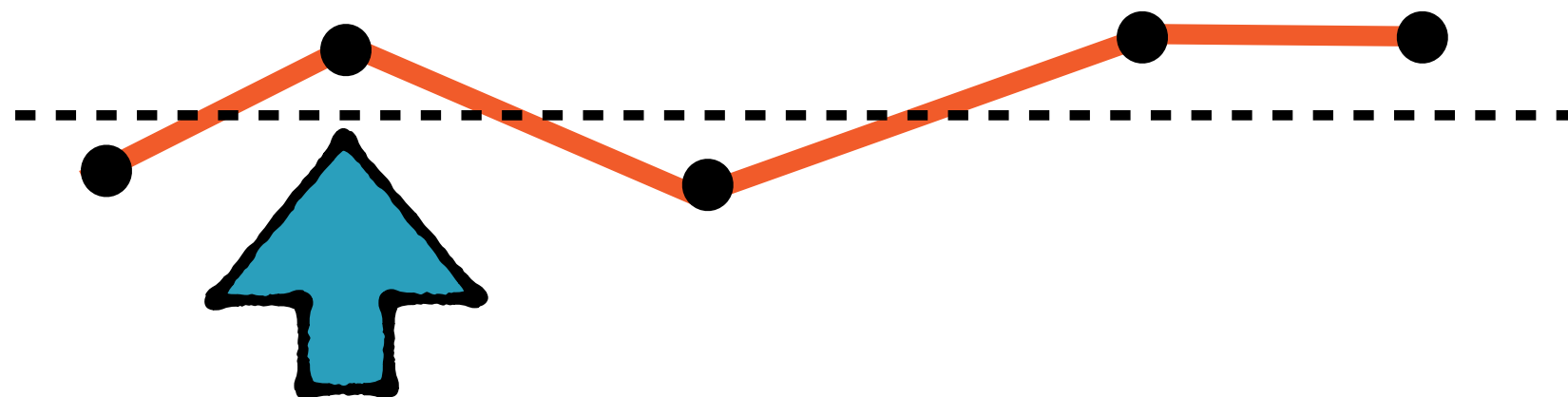
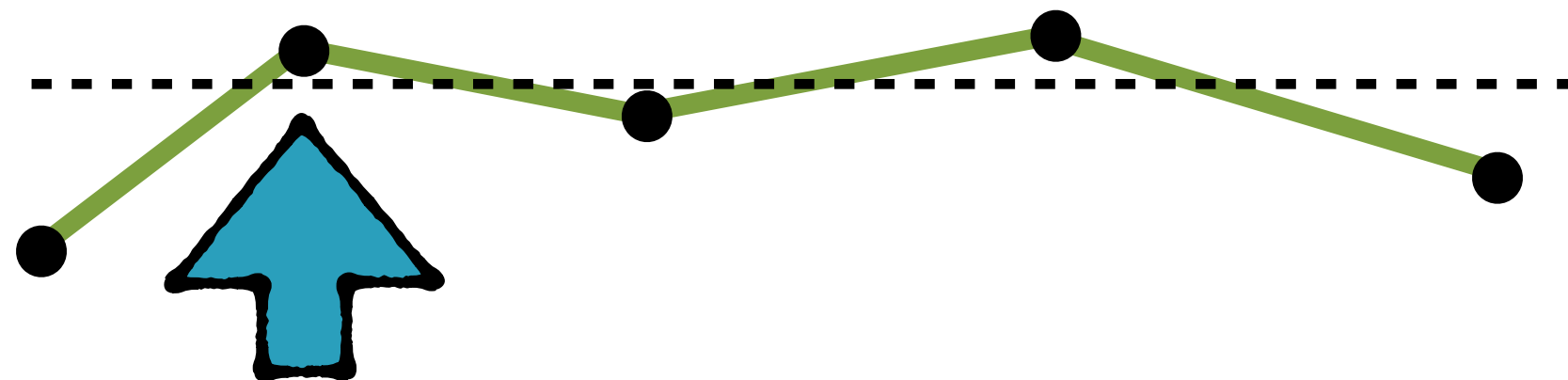


U₂	3	4	2	4	4
----------------------	---	---	---	---	---



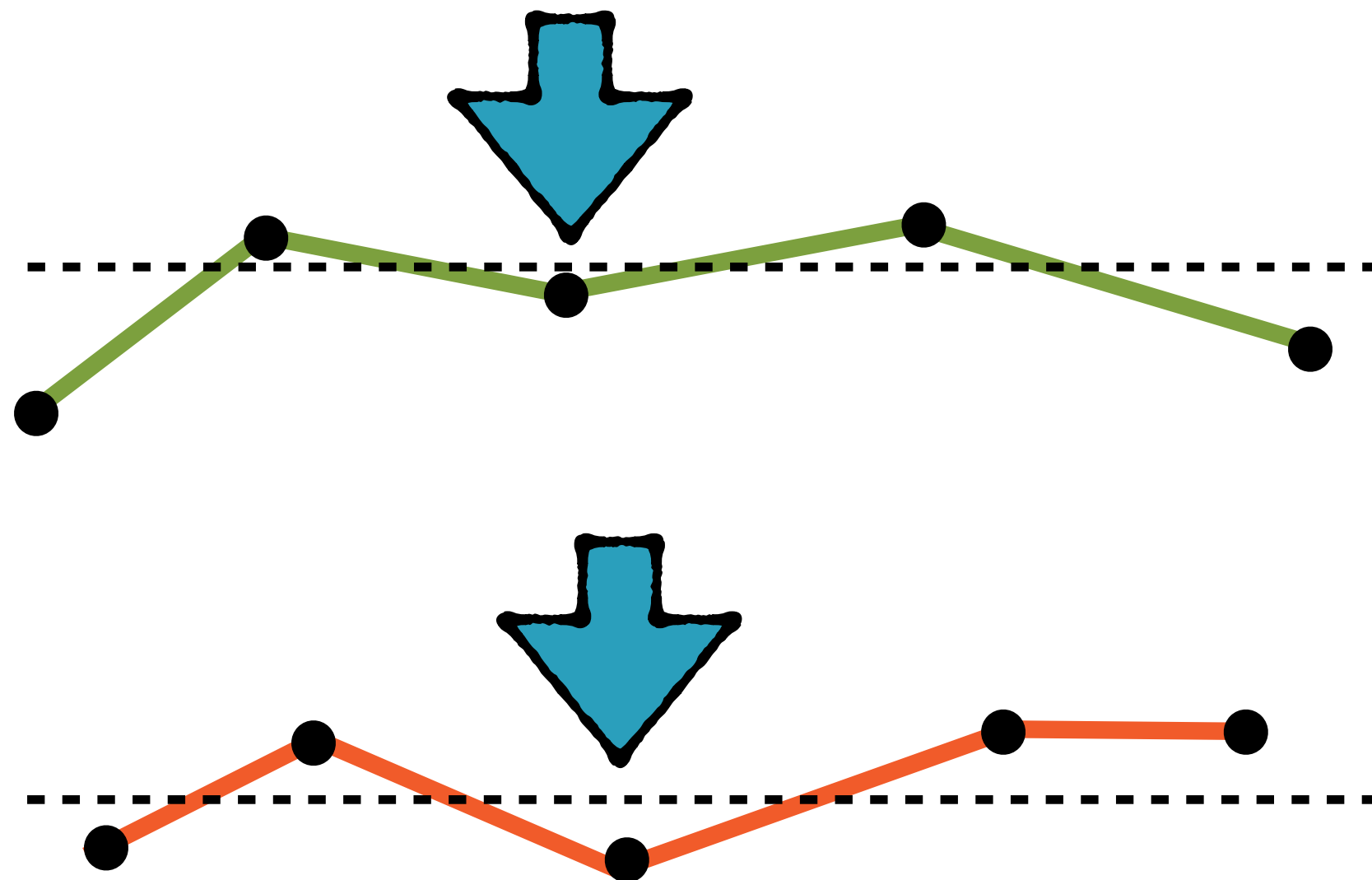


Correlation
Distance





Correlation
Distance





Correlation
Distance

U₁	x₁	x₂
----------------------	----------------------	----------------------	-----------	-----------	-----------

U₂	y₁	y₂
----------------------	----------------------	----------------------	-----------	-----------	-----------

Mean

\bar{x}

\bar{y}

$$Corr(x, y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$



**Correlation
Distance**

Correlation is a measure of similarity

Lies in the range of $[-1, 1]$

Correlation distance = $1 - \text{Correlation}$



Hamming
Distance

Hamming Distance = % Disagreement

1	1	0	1	0
0	1	0	1	1
0	0	1	0	1

Hamming
Distance

U₁	3	3	5	2	1
----------------------	---	---	---	---	---

U₂	4	5	2	2	1
----------------------	---	---	---	---	---

1	1	0	1	0
0	1	0	1	1
0	0	1	0	1

Hamming
Distance

U₁	3	3	5	2	1
----------------------	---	---	---	---	---

U₂	4	5	2	2	1
----------------------	---	---	---	---	---

Disagreement = 0.6

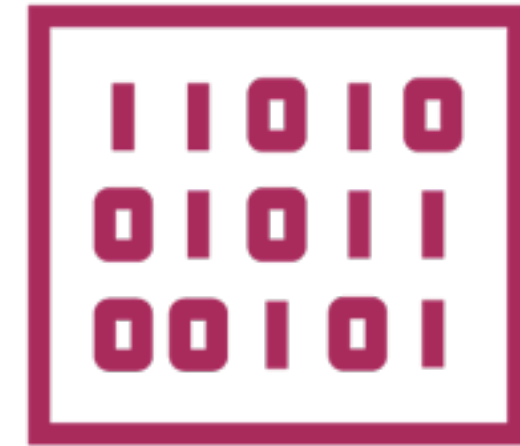
Distance Metrics



**Euclidean
Distance**



**Correlation
Distance**



**Hamming
Distance**

Implementing the Nearest Neighbors Model

Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Find the top N recommendations

Use the average ratings of
the K nearest neighbors

Finding Top N Recommendations

Set up the data

Functions to access
relevant information



Load 2 files

Ratings

User	ISBN	Rating

Book Metadata

ISBN	Title	Author

Finding Top N Recommendations

Set up the data

Functions to access
relevant information



**A function to lookup metadata
for an ISBN**

**A function to find the favorite
books for a user**

Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Find the top N recommendations

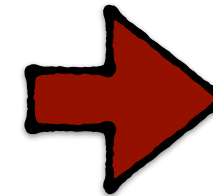
Use the average ratings of
the K nearest neighbors

Finding Top N Recommendations

Construct a rating matrix

The representation needed
for collaborative filtering

User	ISBN	Rating



	P ₁	P ₂	P ₃	P ₄	P ₅
U ₁	3	4	-	-	4
U ₂	3	5	3	4	5
U ₃	4	2	-	5	4
U ₄	3	-	4	5	2
U ₅	1	-	4	2	1
U ₆	3	4	-	2	5

`pandas.pivot_table`

Finding Top N Recommendations



Construct a rating matrix

The representation needed
for collaborative filtering

The rating matrix is sparse

**Restrict the size of the matrix
for better computational
performance**

Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Find the top N recommendations

Use the average ratings of
the K nearest neighbors

Finding Top N Recommendations

Find the K Nearest
Neighbors

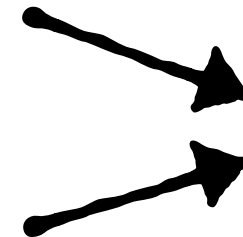
User 1

User 2

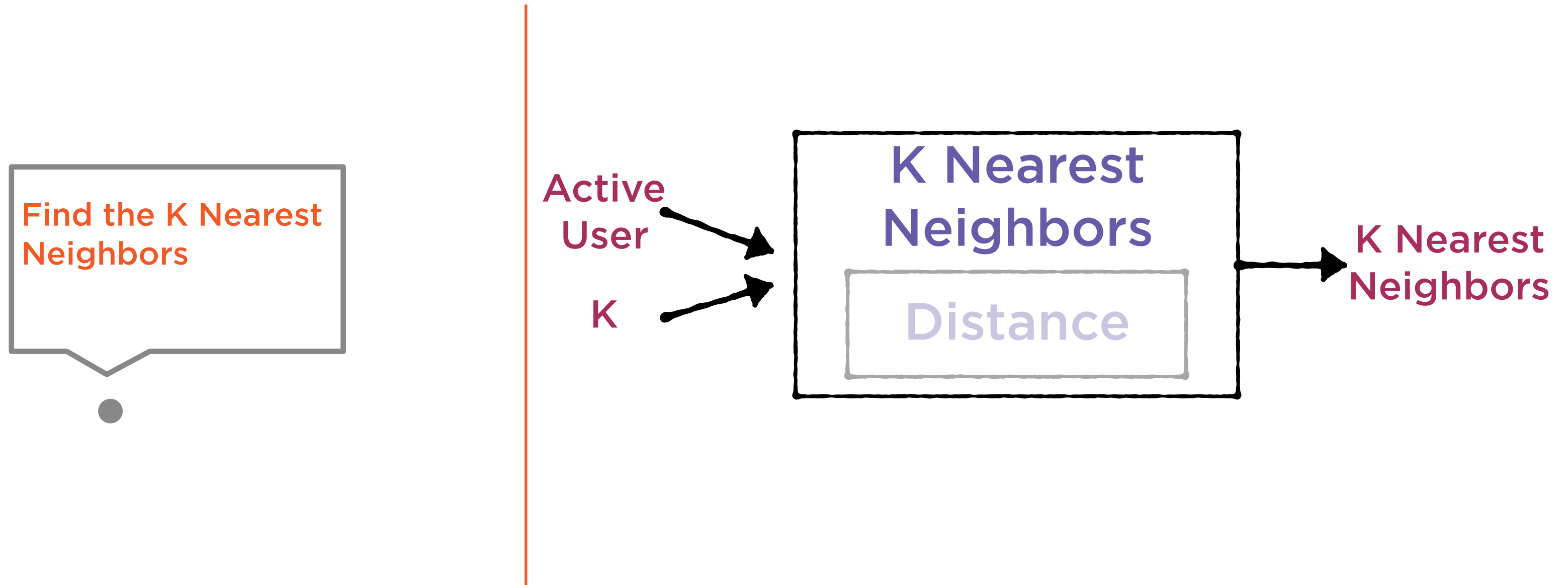
Distance

scipy

Distance
between
the users



Finding Top N Recommendations



Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Find the top N recommendations

Use the average ratings of
the K nearest neighbors

Finding Top N Recommendations



**Find the top N
recommendations**

Use the average ratings of
the K nearest neighbors

1. **Average the ratings of nearest neighbors for unrated books**
2. **Sort in descending order**
3. **Pick the top N**

Finding Top N Recommendations

Set up the data

Functions to access
relevant information



Demo

Download the Book Crossing ratings dataset

Set up a function to find book metadata

Set up a function to find the favorite books for a user

User Id	Location	Age

ISBN	Title	Author	Year of Publication	Publisher

User	ISBN	Rating

Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Construct a rating matrix

The representation needed
for collaborative filtering

Demo

**Construct a rating matrix
using book ratings in Python**

Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Finding Top N Recommendations

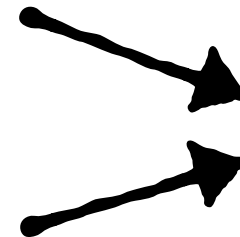
Find the K Nearest
Neighbors

User 1

User 2

Distance

Distance
between
the users



Demo

**Compute the distance
between a pair of users**

Finding Top N Recommendations

Set up the data

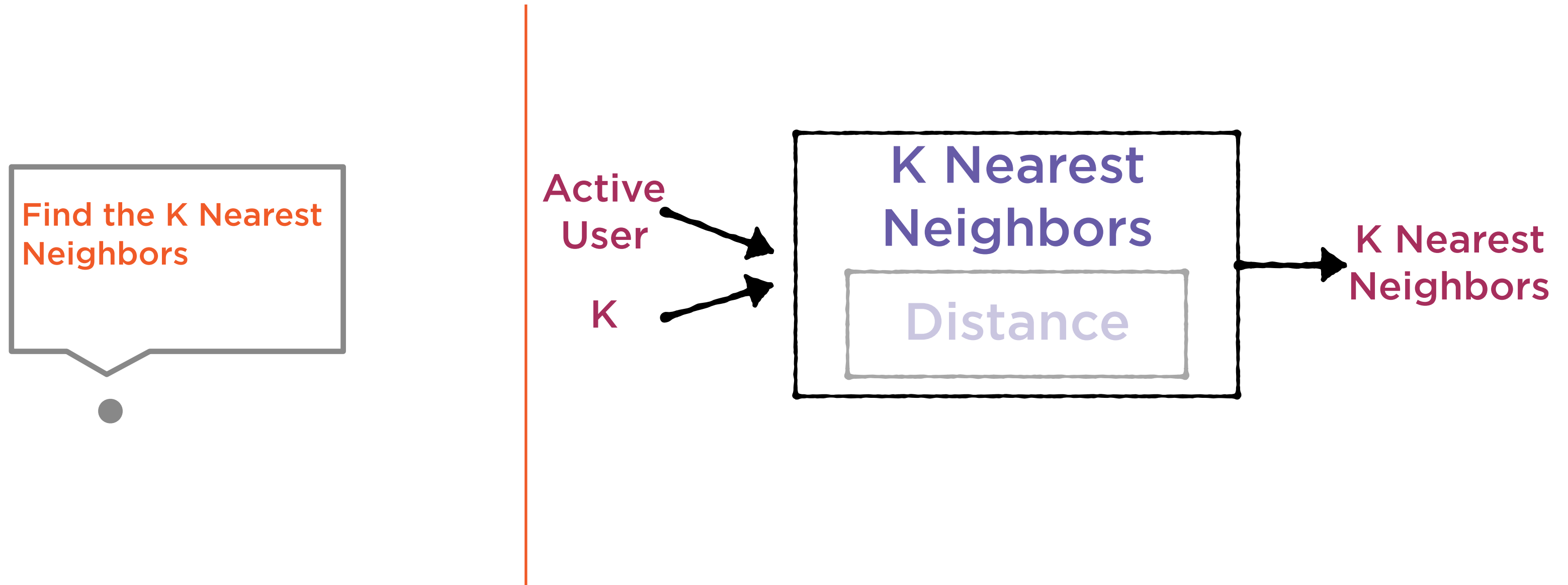
Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Finding Top N Recommendations



Demo

Find the nearest neighbors for a user

Finding Top N Recommendations

Set up the data

Functions to access
relevant information

Find the K Nearest Neighbors

Construct a rating matrix

The representation needed
for collaborative filtering

Find the top N recommendations

Use the average ratings of
the K nearest neighbors

Demo

Average the ratings of nearest neighbors for unrated books

Sort in descending order

Pick the top N

Summary

Understand the nearest neighbors model for collaborative filtering

Measure similarity of users using distance metrics

Find the top 10 book recommendations for a user