


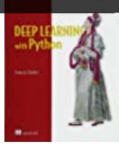
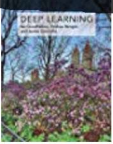
K-Nearest Neighbor (KNN)

Recommendation Systems

Examples - Amazon

amazon.com/Deep-Learning-Adaptive-Computation-Machine/dp/0262035618/ref=sr_1_18?keywords=machine+learning&qid=1572584650&sr=...

Frequently bought together



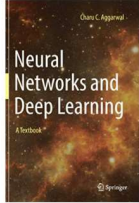
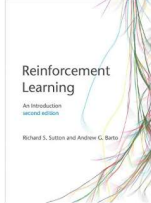
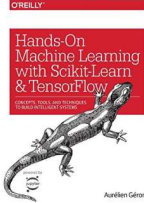
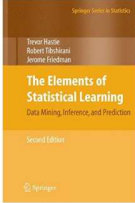
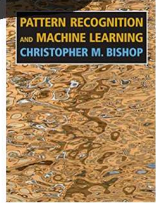
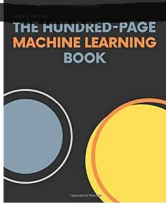
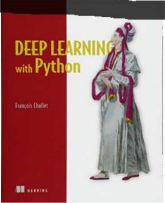
Total price: **\$139.08**

[Add all three to Cart](#)

[Add all three to List](#)

- ✓ **This item:** Deep Learning (Adaptive Computation and Machine Learning series) by Ian Goodfellow Hardcover **\$59.53**
- ✓ Deep Learning with Python by Francois Chollet Paperback **\$22.99**
- ✓ Pattern Recognition and Machine Learning (Information Science and Statistics) by Christopher M. Bishop Hardcover **\$56.56**

Customers who viewed this item also viewed



Book Title	Author	Format	Price	Prime
Deep Learning with Python	Francois Chollet	Paperback	\$22.99	✓
The Hundred-Page Machine Learning Book	Andriy Burkov	Paperback	\$30.94	✓
Pattern Recognition and Machine Learning (Information Science and Statistics)	Christopher M. Bishop	Hardcover	\$56.56	✓
The Elements of Statistical Learning: Data Mining, Inference, and Prediction	Trevor Hastie	Hardcover	\$63.05	
Hands-On Machine Learning with Scikit-Learn & TensorFlow	Aurélien Géron	Paperback	\$50.98	✓
Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning series)	Richard S. Sutton	Hardcover	\$49.98	
Neural Networks and Deep Learning: A Textbook	Charu C. Aggarwal	Hardcover	\$43.02	✓

Page 1 of 9

Examples - In e-commerce

Related hotels...



**Four Seasons Hotel
Chicago**
★★★★★ 948 Reviews
Chicago, IL

Show Prices

Produktempfehlungen (9)



privileg Waschmaschine »PAWF
3746 A« ...
★★★★★
399,00€



Whirlpool Toplader AWE
100...
~~319,00€~~ -6%
299,00€

You may also like



Taifun
~~£75.00~~
£70.00

Free delivery &
returns

Kunden interessierte auch



Read Commented Recommended



Germany Just Rejected The
Idea That The European Bailout
Fund Would Buy Spanish Debt

×



There Is Almost No Gold In The
Olympic Gold Medal

×

MOST POPULAR

RECOMMENDED

How to Break NRA's Grip on Politics: Michael R.
Bloomberg +

Growth in U.S. Slows as Consumers Restrain
Spending +

Examples - In Social Media

Who to follow · Refresh · View all

**Gnip, Inc.**  @gnip
Promoted · Follow

**Twitter**  @twitter
Followed by Michael Ekstrand and...
Follow

**Yong Zheng** @irecsys
Followed by sbourke
Follow

Personen, die du vielleicht kennst [Alle anzeigen](#)


4 gemeinsame Freunde
[FreundIn hinzufügen](#)



1 gemeinsame/r FreundIn
[FreundIn hinzufügen](#)


Jobs you may be interested in Beta [Email Alerts](#) [See More »](#)


**Technical Sales Manager - Europe** ×
Thermal Transfer Products - Home office

**Senior Program Manager (f/m)** ×
Johnson Controls - Germany-NW-Burscheid

Groups You May Like [More »](#)

**Advances in Preference Handling**
[Join](#)

**FP7 Information and Communication Technologies (ICT)**
[Join](#)

**The Blakemore Foundation**
[Join](#)





Picasa™ -Webalben [Startseite](#) [Meine Fotos](#) [Erkunden](#) [Hochladen](#)

Empfohlene Fotos [Alle anzeigen](#)







Examples - Mobile Apps

App Empfehlungen

			
Trailerpop: Trailers & trailerpop, inc.	IMDb Movies & TV IMDb	ABBY TextGrabber ABBY Software House	The Voice: On Stage StarMaker Interactive
★★★★★ FREE	★★★★★ FREE	★★★★★ 5,49 €	★★★★★ FREE

Musik-Apps

			
Speaker Booster Wait What	Google Play Music Google Inc.	Equalizer + mp3 Player DJIT	doubleTwist Music doubleTwist™
★★★★★ FREE	★★★★★ FREE	★★★★★ FREE	★★★★★ FREE

Definition - Problem domain

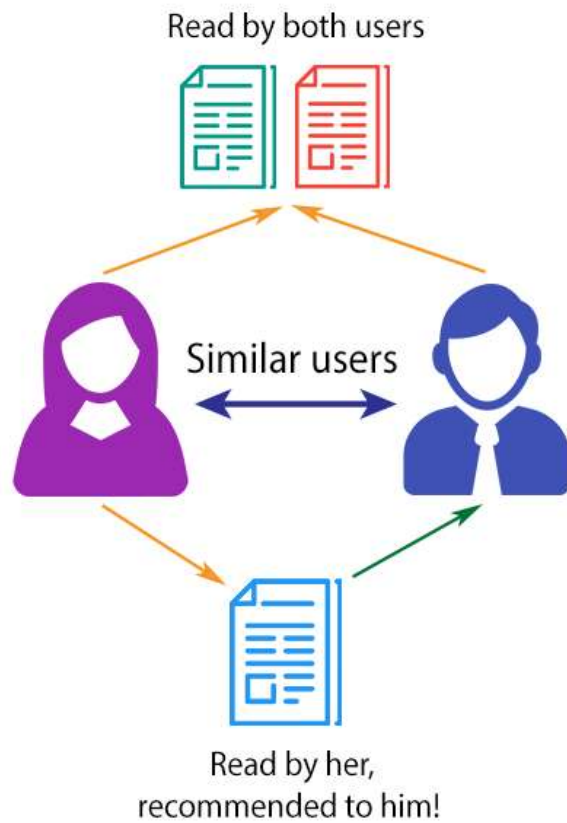
- Recommendation systems (RS) help to match **users with items**

Definition - Problem domain

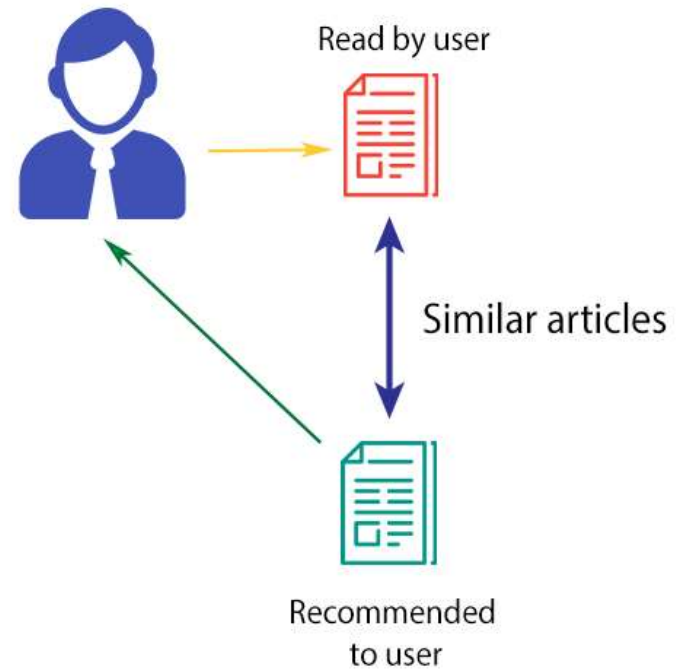
- RS are one of the **most successful and widespread applications** of machine learning technologies in business.

Two types of systems

COLLABORATIVE FILTERING



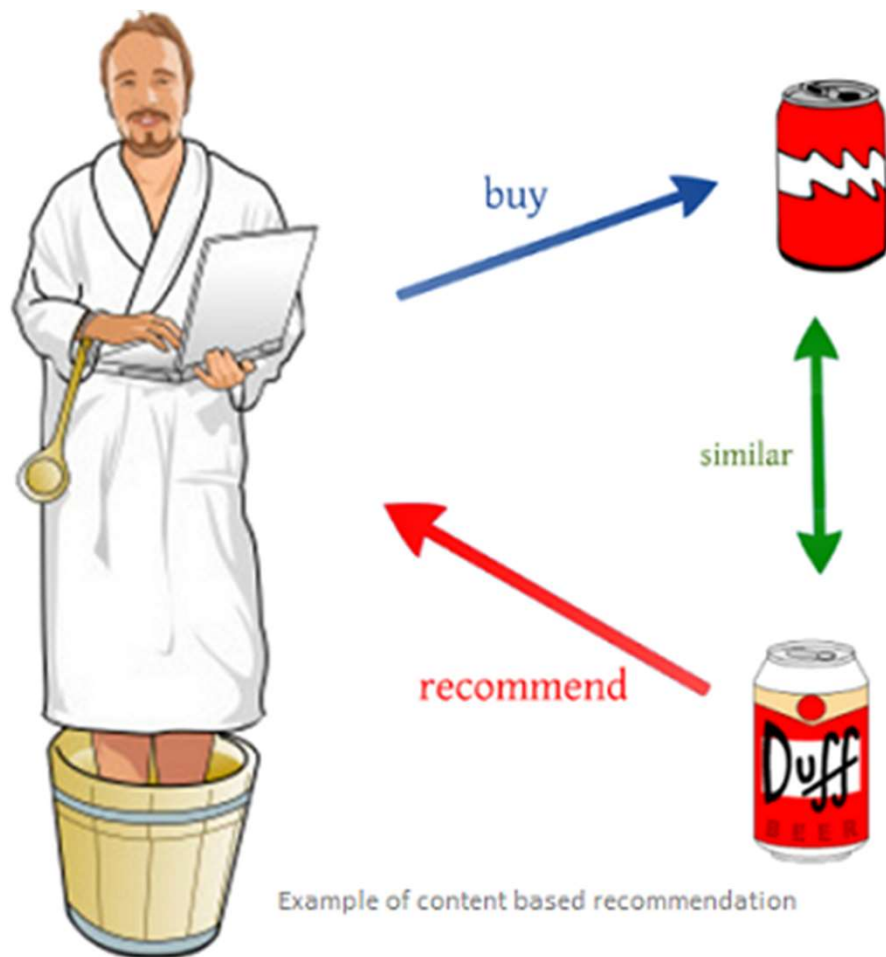
CONTENT-BASED FILTERING



Two types of systems

- **Content- Based Filtering:** Recommending to user A based on his/her existing profiles.
- **Collaborative Filtering:** Recommending to user A based on his/her community's profiles.

Content- Based Filtering

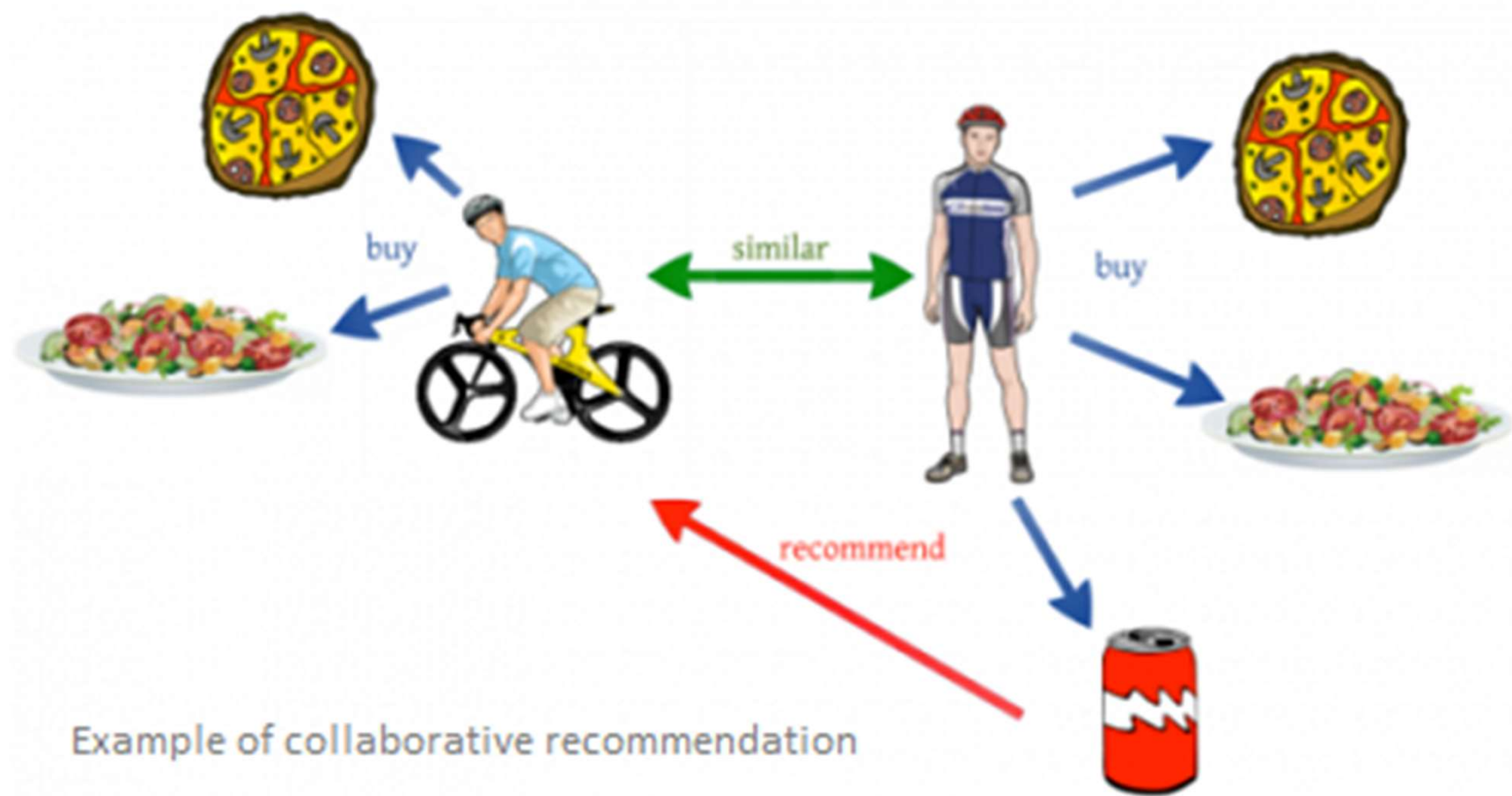


Example of content based recommendation

Content- Based Filtering

- Assume there are four categories of news A) Politics B) Sports C) Entertainment D) Technology
- User A who has read 10 articles related to Technology
- Recommend a new article in Technology for him to read.

Collaborative Filtering

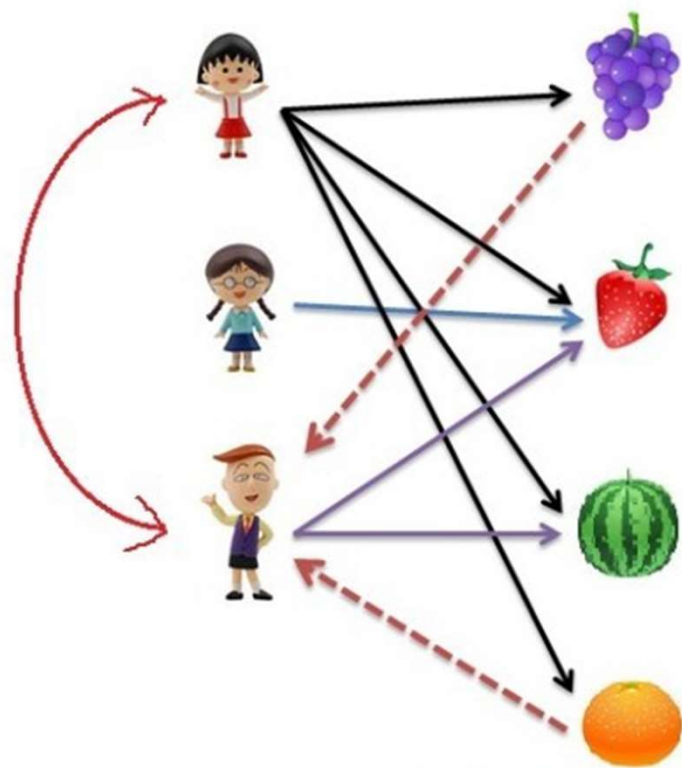


Example of collaborative recommendation

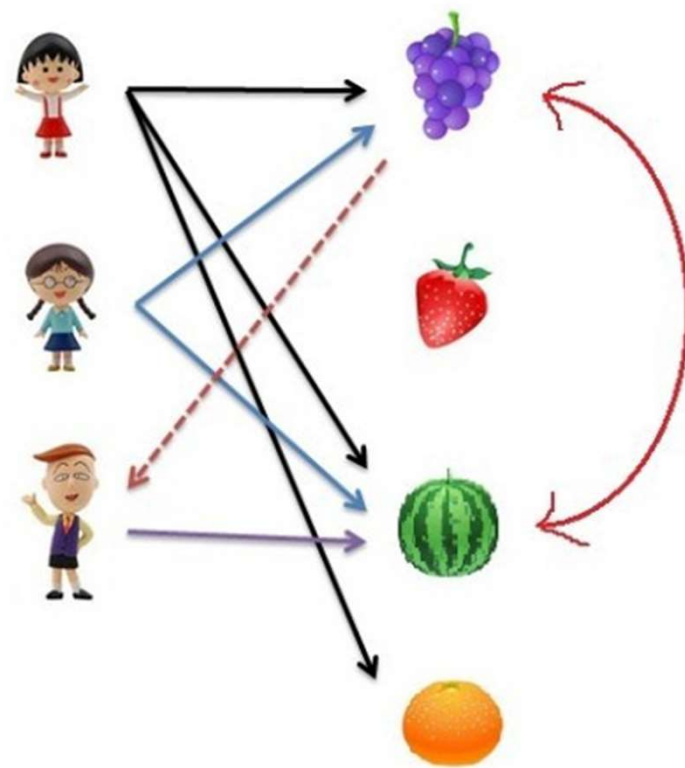
Collaborative Filtering

- Assume there are four categories of news A) Politics B) Sports C) Entertainment D) Technology
- User A who has read 10 articles related to Technology
- User B who has read **the same** 10 articles related to Technology and an X article in Sports.
- Recommend the article X to user A.

Collaborative Filtering: Two approaches



User-based filtering



Item-based filtering

Utility Matrix

- Utility Matrix contains ratings of users on items

	Item 1	Item 2	Item 3	Item 4	Item 5
Alice	5	3	4	4	???
User 1	3	1	2	3	3
User 2	4	3	4	3	5
User 3	3	3	1	4	4
User 4	1	5	5	2	1

	Item 1	Item 2	Item 3	Item 4	Item 5
Alice	5	3	4	4	???
User 1	3	1	2	3	3
User 2	4	3	4	3	5
User 3	3	3	1	4	4
User 4	1	5	5	2	1

x_1 x_2 x_3 x_4 γ

	Item 1	Item 2	Item 3	Item 4	Item 5
Alice	5	3	4	4	???
User 1	3	1	2	3	3
User 2	4	3	4	3	5
User 3	3	3	1	4	4
User 4	1	5	5	2	1

$$\Rightarrow \gamma = 2.1 \cdot x_1 + 3 \cdot x_2 - 4x_3 - 5.1x_4 + 3$$

	Item 1	Item 2	Item 3	Item 4	Item 5
γ Alice	5	3	4	4	???
x_1 User 1	3	1	2	3	3
x_2 User 2	4	3	4	3	5
x_3 User 3	3	3	1	4	4
x_4 User 4	1	5	5	2	1

$$\gamma = 3.1x_1 - 2.6x_2$$

$$- 3.1x_3 + 4x_4 + 2$$

Nearest-neighbors (kNN)

- A “pure” CF approach and traditional baseline
- Using the utility as inputs
- Returns a ranked list of items based on rating predictions

Nearest-neighbors (kNN)

– Assumptions

- If users had similar tastes in the past they will have similar tastes in the future
- User preferences remain stable and consistent over time

User-based KNN

	Item 1	Item 2	Item 3	Item 4	Item 5
Alice	5	3	4	4	???
User 1	3	1	2	3	3
User 2	4	3	4	3	5
User 3	3	3	1	4	4
User 4	1	5	5	2	1

- Find find k nearest neighbors of Alice.
- Use the average rating of the nearest neighbors on Item 5 as a prediction of Alice on Item 5.

User-based KNN

Let $A1$ is the distance from Alice to User 1 and so on. We have:

$$A1 = 3.60$$

$$A2 = 1.41$$

$$A3 = 3.60$$

$$A4 = 5$$

- For 3NN, the predicted rating of Alice for item 5 is the average of ratings on item 5 of her 3 nearest neighbors, User 1, 2 and 3.
- Predicted rating of Alice on item 5 is: $(3+5+4)/3 =$

Item-based KNN

	Item 1	Item 2	Item 3	Item 4	Item 5
Alice	5	3	4	4	???
User 1	3	1	2	3	3
User 2	4	3	4	3	5
User 3	3	3	1	4	4
User 4	1	5	5	2	1

- Find the k nearest neighbors of **Item 5**.
- The predicted rating of Alice on item 5 is the average rating of Alice on the nearest neighbors.

Item-based KNN

Let d_{54} be the distance of item 5 to item 4 and so on. We have

$$d_{54} = 2.23$$

$$d_{53} = 5.19$$

$$d_{52} = 5$$

$$d_{51} = 1.41$$

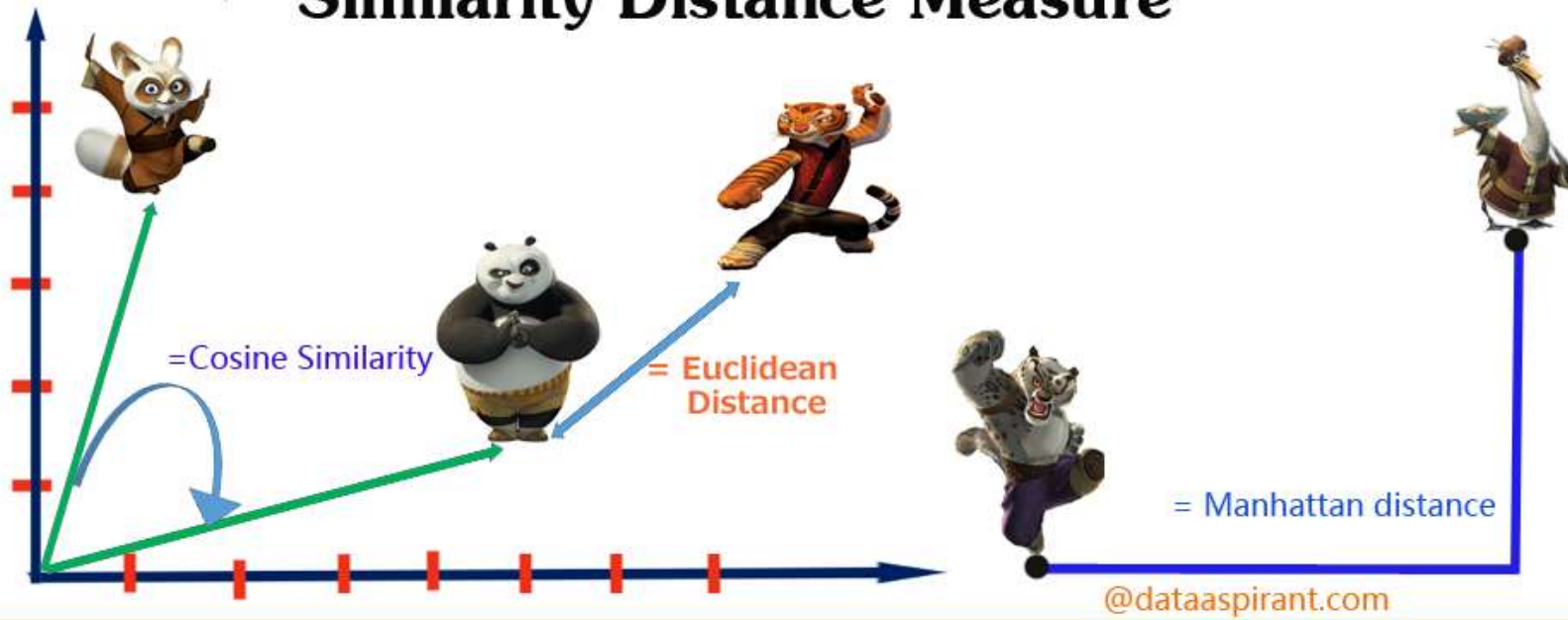
- For 3NN, the two nearest neighbors of Item 5 are Item 1, 4 and Item 2.
- Predicted rating of Alice on Item 5 is the average of her ratings on Item 1, 4 and 2, which is

Similarity Measure

- Neighborhood can be decided by **similarity** measures
- Similarity can be measured as the inverse of the Distance
- The possible similarity values are between 0 and 1, where values near to 1 indicate a strong similarity.
- There are many distance measure
- There are many similarity measure

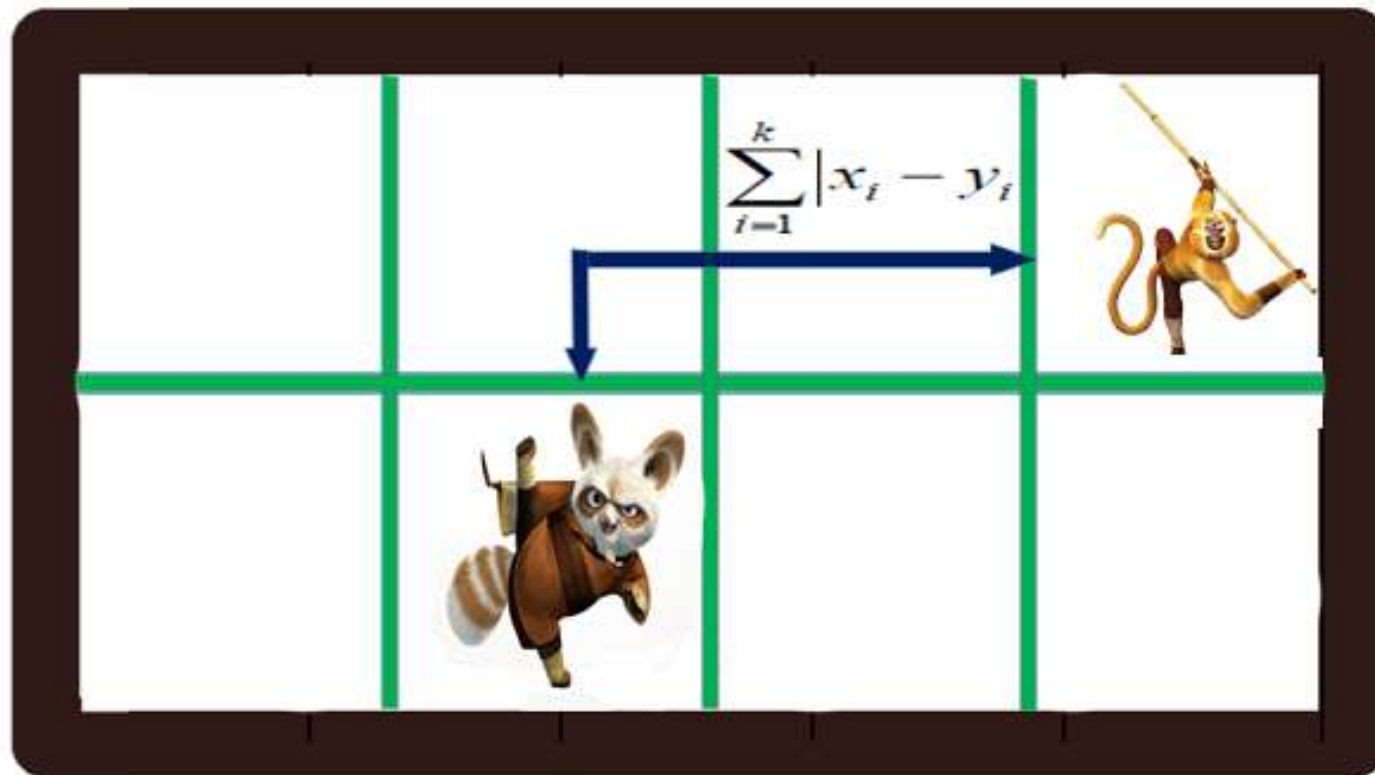
Similarity Measure

Similarity Distance Measure



Manhattan Distance

Manhattan Distance



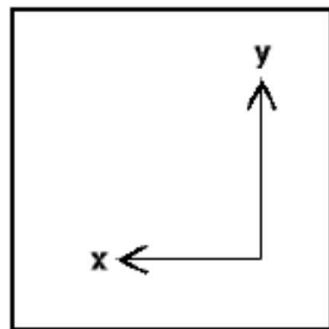
Manhattan Distance

- ManhattanDistance between Alice and User 1 (*A1*).

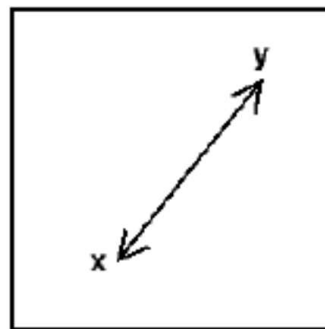
	Item 1	Item 2	Item 3	Item 4
Alice	5	3	4	4
User 1	3	1	2	3

$$A1 = |5 - 3| + |3 - 1| + |4 - 2| + |4 - 3| = 7$$

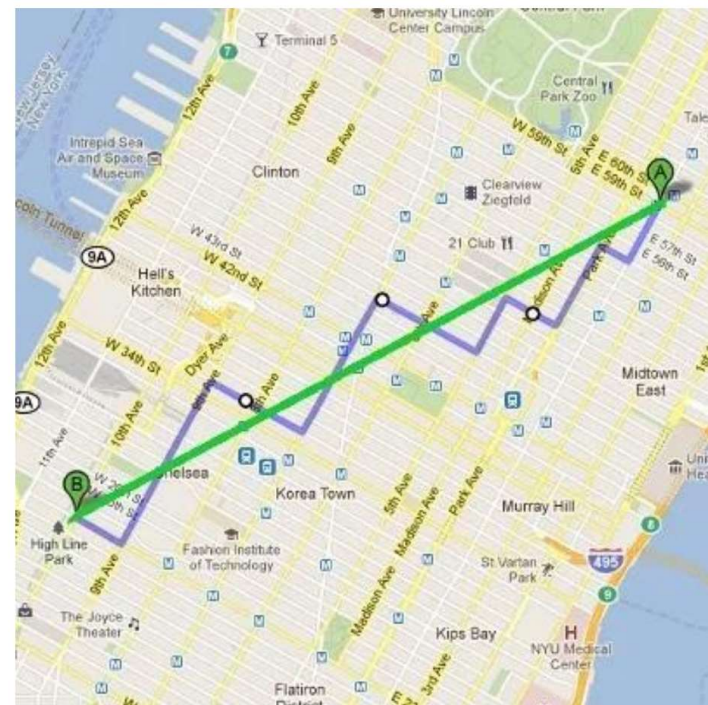
Manhattan vs. Euclidean



Manhattan

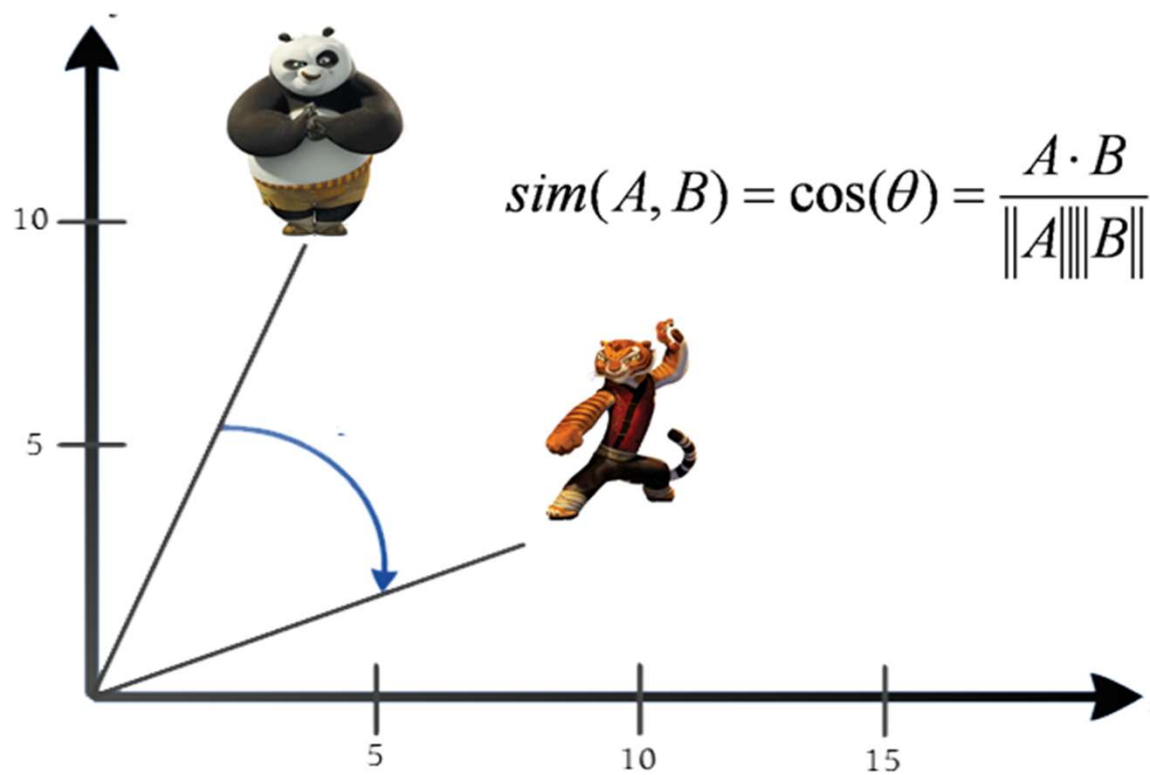


Euclidean



Cosine Similarity

Cosine Similarity



- **Cosine similarity** is established as the standard in Recommendation System.

Cosine Similarity Measure

- Cosine similarity between Alice and User 1 (S_1).

	Item 1	Item 2	Item 3	Item 4
Alice	5	3	4	4
User 1	3	1	2	3

$$\begin{aligned}
 & S1 \\
 = & \frac{5 \cdot 3 + 3 \cdot 1 + 4 \cdot 2 + 4 \cdot 3}{\sqrt{5^2 + 3^2 + 4^2 + 4^2} \cdot \sqrt{3^2 + 1^2 + 2^2 + 3^2}} \\
 & = 0.975
 \end{aligned}$$

The Netflix Challenge

[Link](#)