An Introduction to Recommendation Systems

- https://github.com/topspinj/recommender-tutorial/blob/master/part-1-building-from-scratch.ipynb

Agenda

- What is a Recommendation System?

- Recommendation Techniques

  1. Collaborative filtering

  2. Content-based filtering

- Tutorial using the MovieLens database

- Evaluation of Recommendation Systems

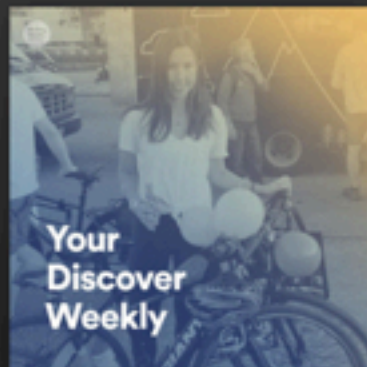- Implicit vs. Explicit Feedback

What is a recommendation system?

- An application of machine learning

- Predicts user preference, filters relevant content

- Clustering similar items and users together

Spotify

"Discover Weekly"

MADE FOR JILL

Discover Weekly

Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts chosen just for you. Updated every Monday, so save your...

Made for **Jill Cts** by **Spotify** • 30 songs, 1 hr 47 min

PLAY **FOLLOWING** ...

Your Discover Weekly

Q Filter   Download ⬤

TITLE	ARTIST	🗓
＋ The Weekend - Funk Wav Remix	SZA, Calvin H...	3 days ago
＋ You Say	Ehrling	3 days ago
＋ Grow Up	Bolier	3 days ago

amazon

Amazon

"Frequently bought together"

Customers who bought this item also bought

An Introduction to Statistical Learning: with Applications in R
Gareth James
★★★★★ 13
Hardcover
CDN$ 77.35 ✓prime

Deep Learning
Ian Goodfellow
★★★★½ 25
Hardcover
CDN$ 81.36

Applied Predictive Modeling
Max Kuhn
★★★★½ 8
Hardcover
CDN$ 97.10 ✓prime

Pattern Recognition and Machine Learning
Christopher M. Bishop
★★★½ 8
Hardcover
CDN$ 86.14 ✓prime

Machine Learning: A Probabilistic Perspective
Kevin P. Murphy
★★★★☆ 6
Hardcover
CDN$ 120.49 ✓prime

Hands-On Machine Learning with Scikit-Learn and TensorFlow:...
Aurélien Géron
★★★★½ 27
Paperback
CDN$ 45.06 ✓prime
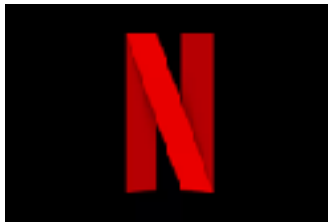
Netflix

"Because you watched this TV Show…"

OkCupid

"Finding your best match"

Twitter

"Who to follow"

Who to follow

Hillary Clinton ✓
@HillaryClinton
Wife, mom, grandma, women+kids advocate, FLOTUS, Senator, SecState, hair icon, pantsuit aficionado, 2016 presidential candidate.

[Follow]

Barack Obama ✓
@BarackObama
Dad, husband, President, citizen.
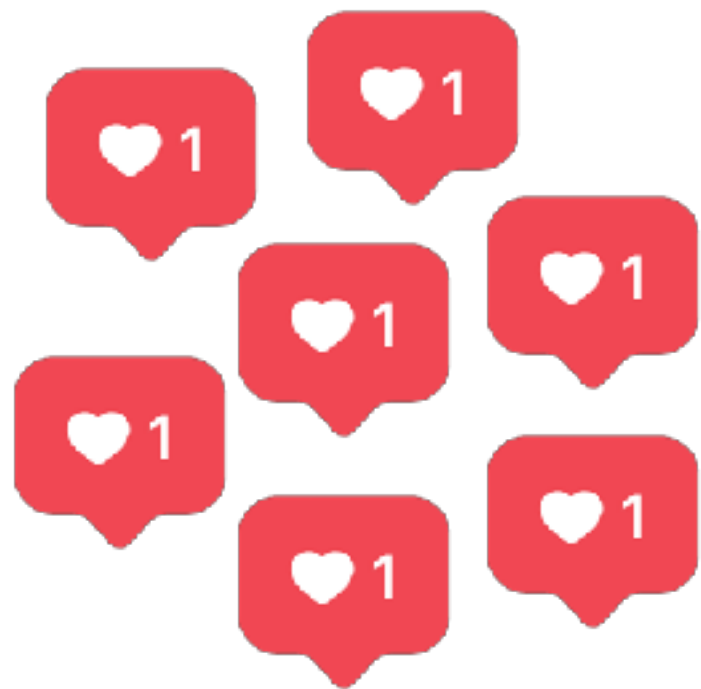
[Follow]

Ellen DeGeneres ✓
@TheEllenShow
Comedian, talk show host and ice road trucker. My tweets are real, and they're spectacular.

[Follow]

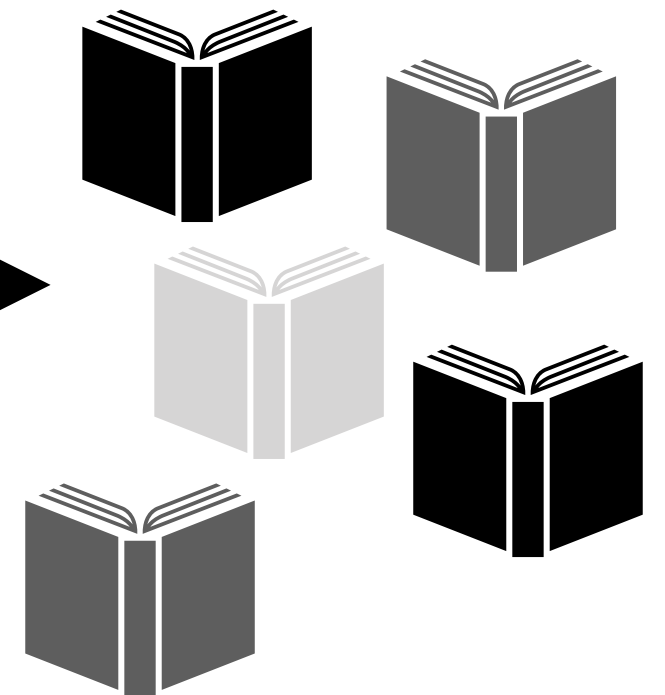# Recommendation Systems = Machine Learning

Historical data

Suggestions

model

- **direct preferences:** "liking", rating, "swiping right"
- **indirect behaviour:** watching, listening, shopping, viewing

- **predicting future behaviour**
- which book will I rate highly
- which song will I put on repeat?

Recommendation Systems = Machine Learning

Two main techniques:

1. Collaborative Filtering

2. Content-based Filtering

Unsupervised Learning

(Un)supervised Learning

Supervised Learning

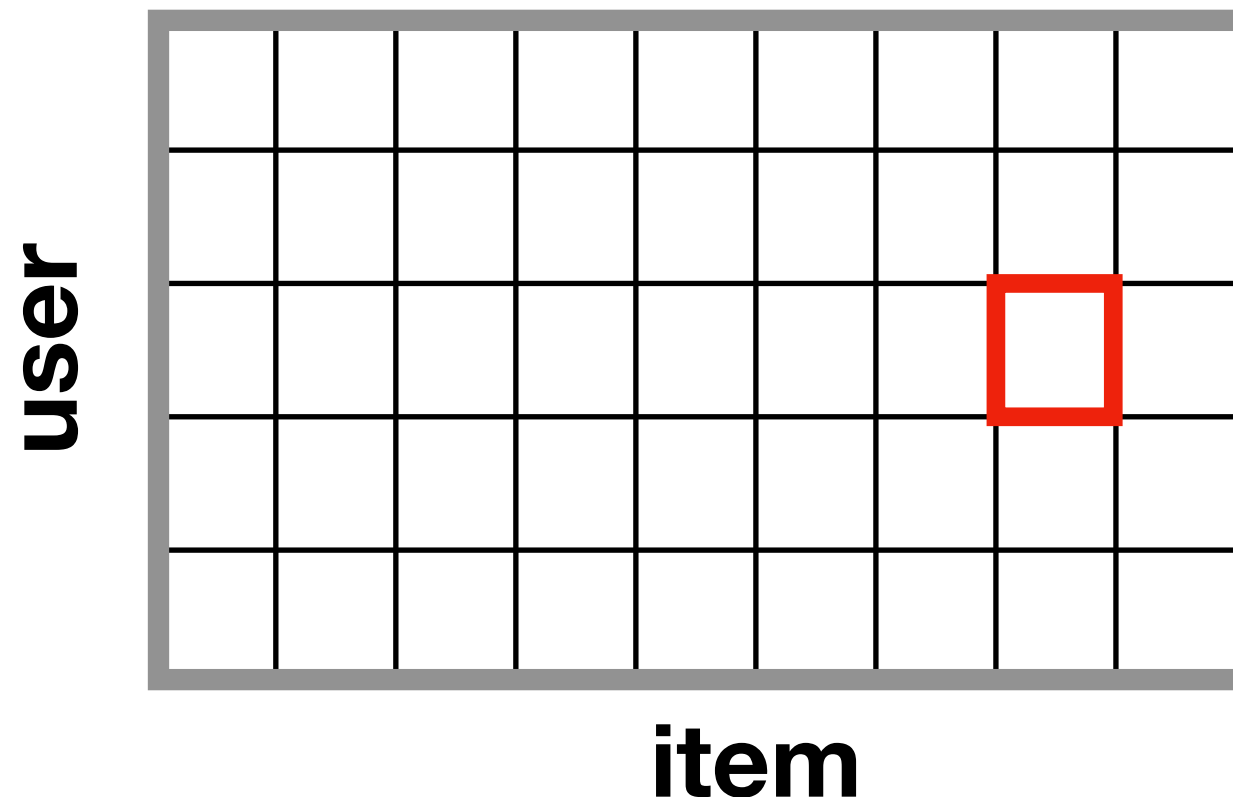| id	age	gender	profession	ticket price	cabin	survived?
1	24	M	S	10	middle	N
2	60	F	D	80	upper	Y
3	33	M	R	5	lower	N
4	7	F	NA	1	upper	Y
…	…	…	…	…	…	
X	27	F	S	25	middle	?

Unsupervised Learning

clustering things together
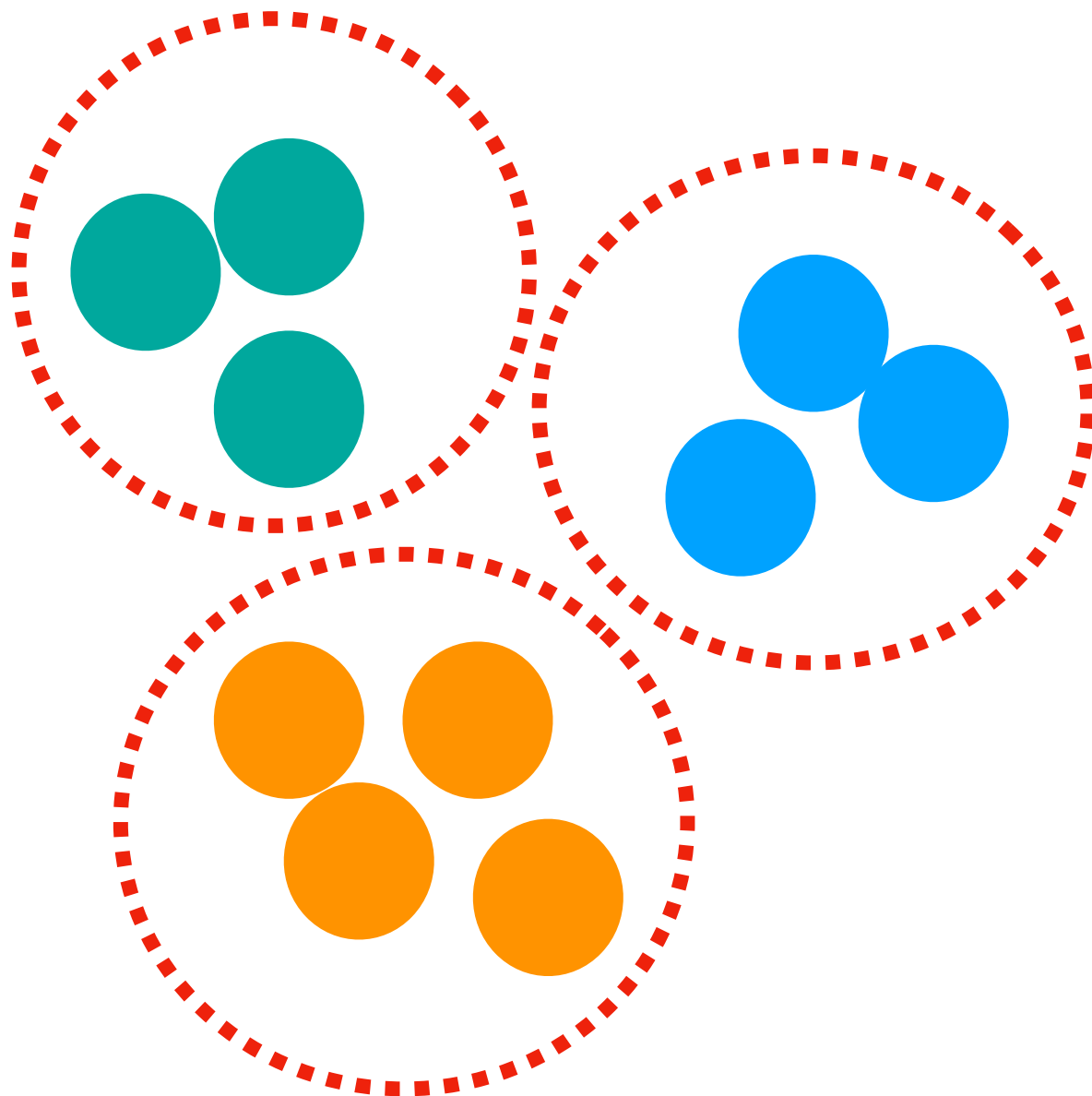
t-SNE plot

Collaborative Filtering

- A type of unsupervised learning

- Assumption: "similar users like similar items"

- Uses a utility (user-item) matrix

user

item

	movie 1	movie 2	movie 3	movie 4	movie 5
user 1	1	5	1		
user 2	1	5		4	2
user 3	5	4		2	
user 4	2		5		
user 5	1.5	4.5	1	5	3
user 6		2		4.5	5
user 7	2		4		5

	movie 1	movie 2	movie 3	movie 4	movie 5
user 1	1	5	1		
user 2	1	5	?	4	2
user 3	5	4		2	
user 4	2		5		
user 5	1.5	4.5	1	5	3
user 6		2		4.5	5
user 7	2		4		5

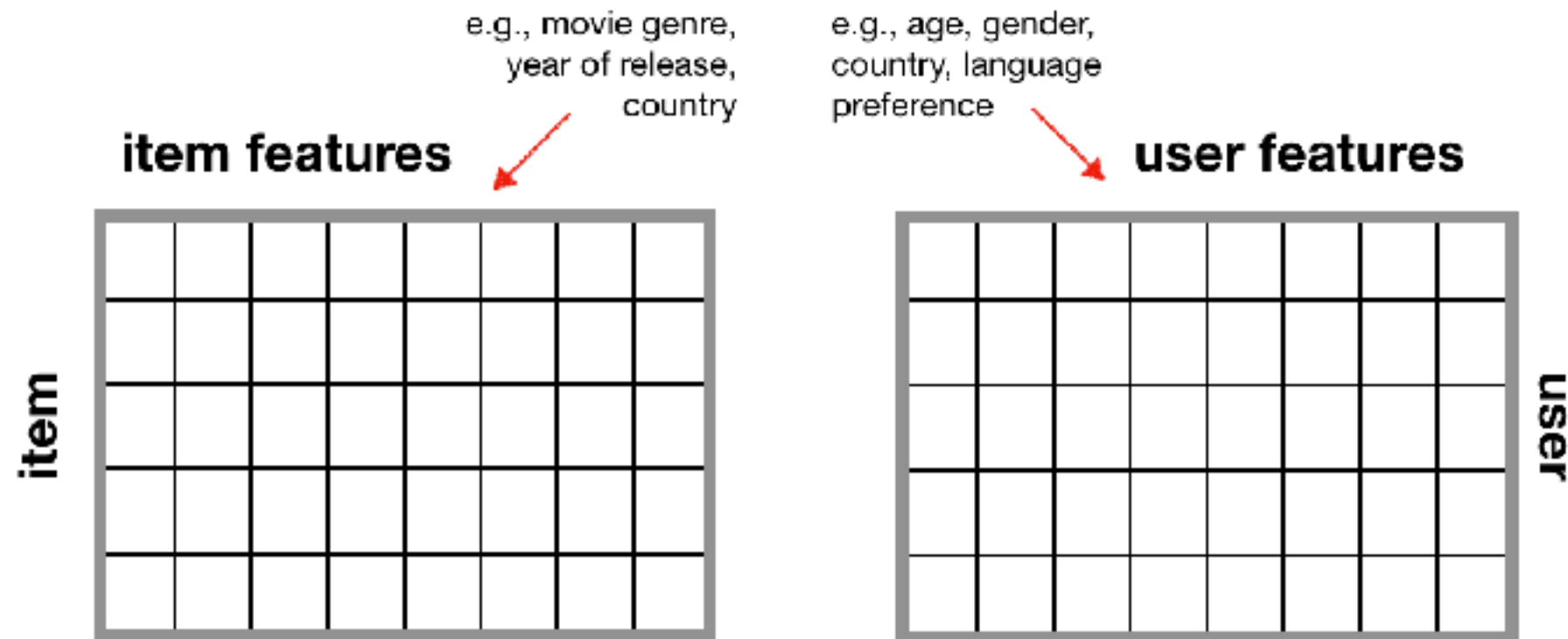|	movie 1	movie 2	movie 3	movie 4	movie 5
user 1	1	5	1		
user 2	1	5	1	4	2
user 3	5	4		2	
user 4	2		5		
user 5	1.5	4.5	1	5	3
user 6		2		4.5	5
user 7	2		4		5

Collaborative Filtering

- k-Nearest Neighbours (kNN)

  - Computes similarity between users and items in
    a vector space

  - "Users who liked this item also liked"

- Matrix factorization

$$sim(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$
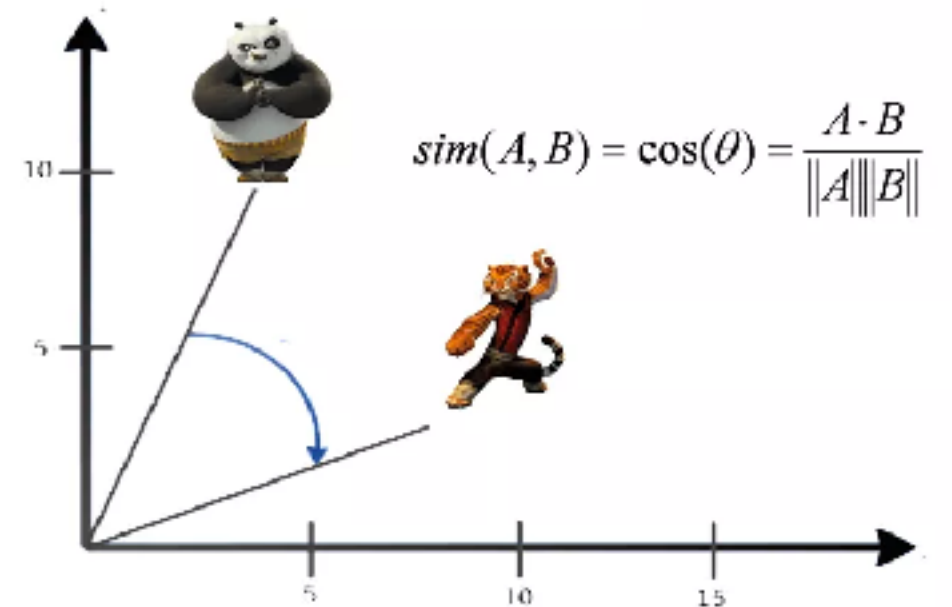
Content-Based Filtering

- Can be unsupervised or supervised learning

- Looks at item and user features to make predictions

e.g., movie genre, year of release, country

e.g., age, gender, country, language preference

item features

user features

item

user

Content-Based Filtering

| movie	year	country	genre	rating
Titanic	1997	USA	drama	3.5
Shawshank	1994	USA	drama	4
Lamerica	1994	Italy	drama	5
Boss Baby	2017	USA	comedy	2
…	…	…	…	
X	**2010**	**Canada**	**comedy**	**?**

Tutorial

- **MovieLens Database**

  - created in 1997 by a research group at the University of Minnesota

  - recommendation system for movies

  - open-source data, very active in the machine learning community

  - 26,000,000 ratings and 750,000 tag applications applied to 45,000 movies by 270,000 users
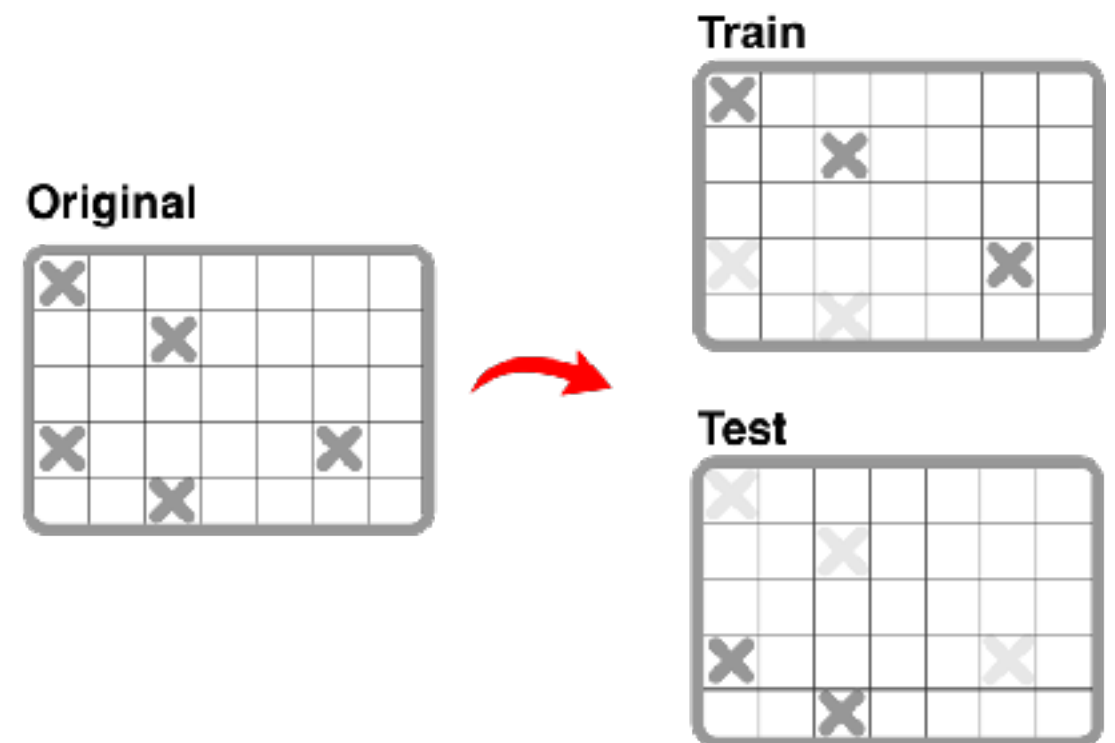
How do we evaluate recommendations?

- Gold standard: A/B testing

- Accuracy (content-based filtering)

- Precision and Recall

How do we evaluate recommendations?

Traditional ML
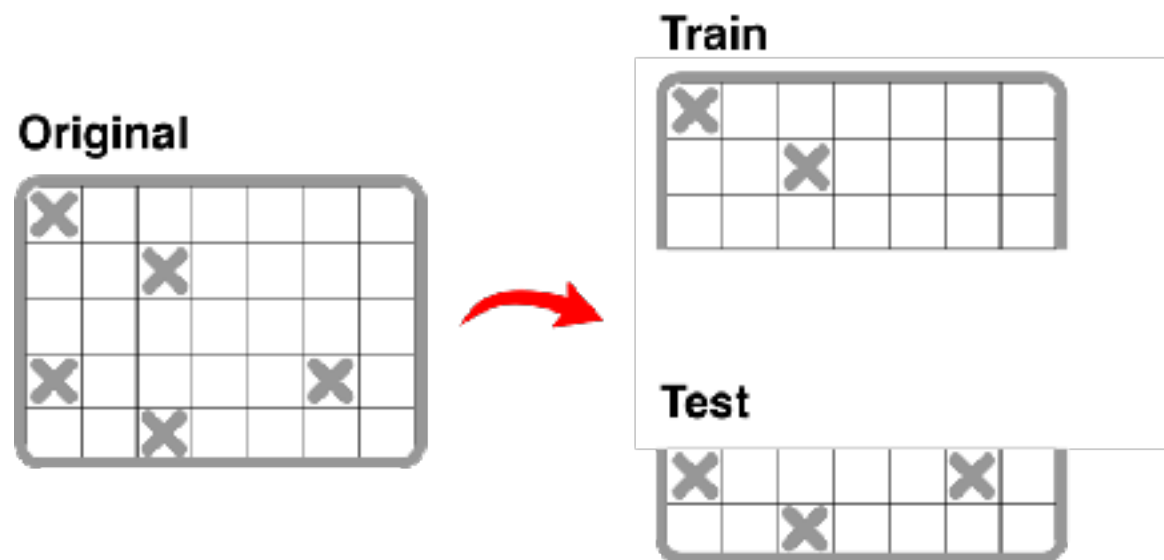
Recommendation Systems

Precision@K and Recall@K

Precision@K
Of the top k recommendations, what proportion are actually "relevant"?

Recall@K
Proportion of items that were found in the top k recommendations.

Reality

	liked	did not like
Predicted liked	True positive	False positive
did not like	False negative	True negative

$$\text{precision} = \frac{TP}{TP + FP} \qquad \text{recall} = \frac{TP}{TP + FN}$$

Precision@K and Recall@K

Reality

Precision@K
Of the top k recommendations, what proportion are actually "relevant"?

Recall@K
Proportion of items that were found in the top k recommendations.

	liked	did not like
liked	True positive	False positive
did not like	False negative	True negative

Predicted

$$\text{precision} = \frac{TP}{TP + FP} \qquad \text{recall} = \frac{TP}{TP + FN}$$

Implicit vs. Explicit Feedback

- Explicit feedback

  - user directly rates their preference towards an item

- Implicit feedback

  - user behaviour towards an item

  - e.g., listening to song multiple times = positive feedback

The Netflix Challenge

- Open competition in 2009

- Challenge: improve current recommendation system by 10%

- Dataset included **100,480,507 ratings** that ~500K users gave to ~18K movies

  ➡ (user_id, movie_id, rating_ts, rating)

- BellKor's Pragmatic Chaos improved score by 10.05%