Recommendation Systems - 1

[ML-2]

→ Introduction

→ A priori Algorithm

TT - 5th Nov

Can you give examples of rec. sys?

→ E-commerce

↳ Customers who bought this also bought ----

↳ Similar products

→ Social media

↳ You may also know....

→ Feed:

↳ Similar posts / blogs / news videos / songs

Instagram, netflix, spotify, youtube

> inventory plan

> search algo

Poll: Can a search algo like "google search" be interpreted as a rec sys??

a) YES

b) NO

1 min rec-sys

In facebook, you get suggestions such as:

You may also know:

→ Gyan Arora

→ Patrick peterson

→ ⋮

How can you design this system?

Why are they useful?

→ increase in revenue !!

→ better user experience !!

E-commerce

↓

people may buy more products

↓

more money $$

Social media

↓

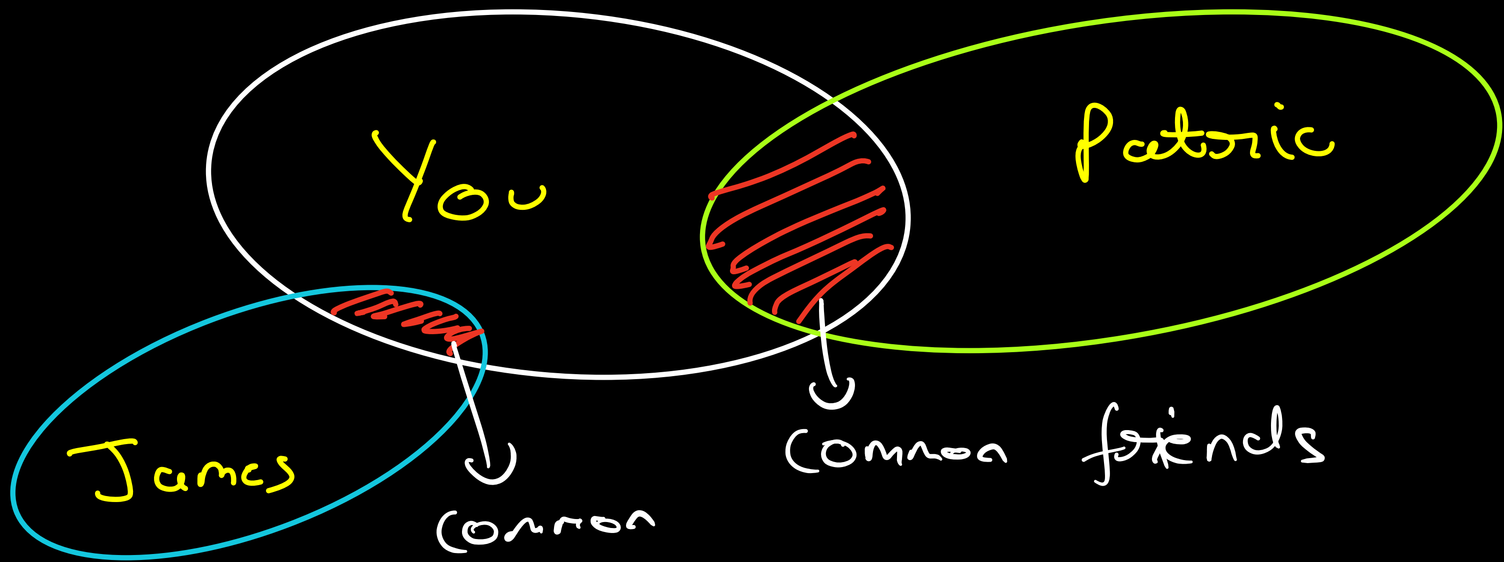
• People will stay "engaged longer".

↳ I show add after 5 songs

• More users will join

→ more ads / more users

→ more money $




```
for u in all_users :
    if len(u.friends.intersection(my.friends))
            > threshold
        recommend = True
```

variation

convert to %

based

E-commerce

↓

Similar products

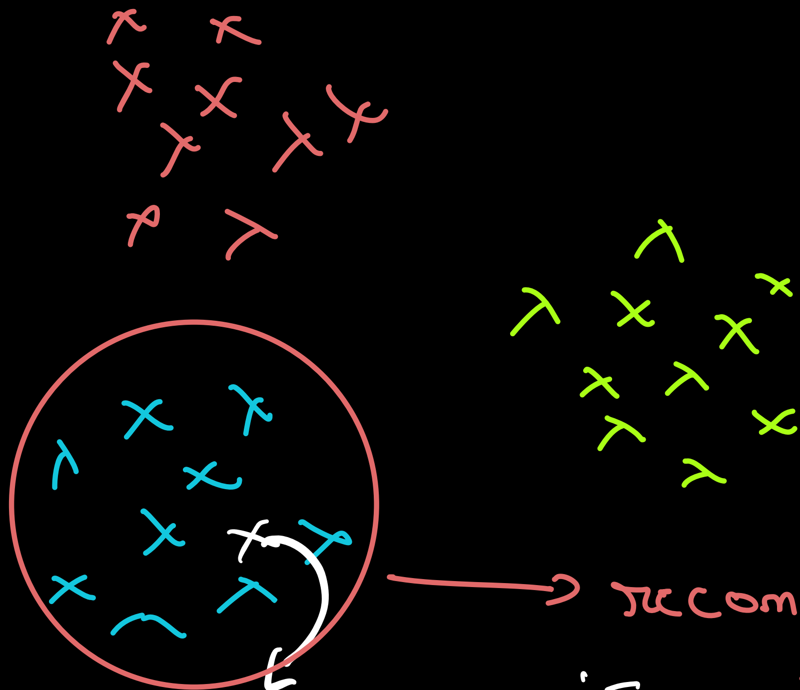
RAM	ROM	Processor	Price	OS ...

User is looking at : Dell XPS 8 GB - i5

Recomendation — ?

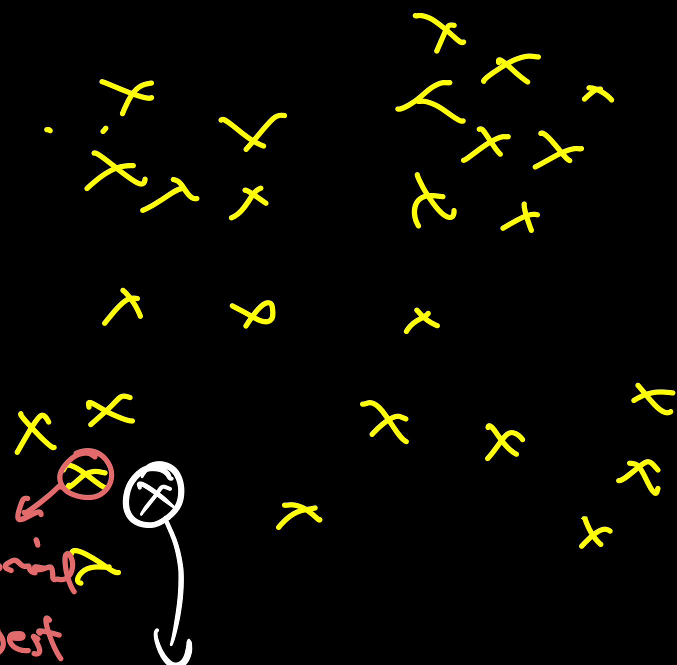
List of suggestions ?

↓ ?



dell XPS 8-i5

recommend from this cluster.



recommend closest products

dell XPS - 8GB i5

Distance can be

→ cosine

→ euclidean ets

People who bought this also bought ...

- iphone : case
  - airpods
  - iwatch
  - warranty
- bread : milk
  - butter
  - Jam
  - eggs

Data based :

↓

Apriori

How many times did it happen in the past ?

- flight tickets to Kashmir
  - hotels in Kashmir
  - sweaters.

Youtube

Sports car video

news of cars
→ recent

My video history

Features
→ Text : description
→ video : images
→ tabular : likes / com / shares / tags

Similarity w.r.t other vids → Shortest dist
↳ more popular

Broad Ideas

Simple logic based

Similarity based

group based

Rec Sys

co-occurrence freq based

Regression based

Other / misc / combinations

popularity / recency

Collaborative filtering

* could be many more types of ideas

Feedback Signals

Q: How can you tell if a user would appreciate your recomendat'n on any item in general ?

→ Buy ? How many ?

→ Wishlist,

→ View page # times, # mins

→ Watch video ?

↳ % watched

→ like / share / com

→ Add to playlist / download / etc . . . .

Apriori Algorithm

Frequency based !

People who bought this also bought ...

CASE STUDY

E-commerce sales data.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/10 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	01/12/10 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/10 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/10 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01/12/10 8:26	3.39	17850.0	United Kingdom

Idea:

Conditional probability using past data:

→ if customer is buying 'A'.

↳ in the past which products are bought with A ?

Eg: {milk, eggs} = 300

{milk, bread} = 600

{milk, jam} = 20

⋮

if A = milk

B = ?

Representing Data

Let's denote a transaction (invoice) by T.

$$T_1 = \{0, 13, 4, 7\}$$

↳ product codes.

$$T_2 = \{1, 2, 0\}$$

$$\vdots$$

Let's say total # products $n = 100$

Q: Do I need to record how many

units of each product was bought?

a) Yes

b) No

From all transactions, we can seperate out the subsets

{1} {2} ⋮ — user only bough 1 product.

→ Does not help much

{1, 3}

{1, 5}

⋮

{1, ⋮ 5, 6 ⋮} — mutiple products.

Q: How many subsets will there be?

a) $^nC_1 + {^nC_2} + {^nC_3}$ b) $^nC_1 + {^nC_n}$ c) 2^n

We can't really compute for so many subsets, hence we can remove products which sell rarely.

Support : $\dfrac{\#\text{ times subset } S \text{ occurred}}{\#\text{ transactions}}$

Hyperparam 1 -> min_support

$$\text{Support (Itemset)} = \frac{\text{Frequency of Itemset (Support Count)}}{\text{Total Number of Transactions}}$$

TID	Items
1	{Bread, Milk}
2	{Bread, Diapers, Beer, Eggs}
3	{Milk, Diapers, Beer, Cola}
4	{Bread, Milk, Diapers, Beer}
5	{Bread, Milk, Diapers, Cola}

Itemset	**Support**	**%**
[Beer]	3/5	0.6
[Bread]	4/5	0.8
[Cola]	2/5	0.4
[Diapers]	4/5	0.8
[Milk]	4/5	0.8
[Eggs]	1/5	0.2

→ Flaw → if 'A' is very popular, many products will have high support w.r.t A

Association Rules

How to make recommendations?

Q: If I rec B after A, can I rec. A after B?

a) Yes

b) No

query product = A

↳ search for all subsets with 'A'

$$\text{Confidence}(B \mid A) = \frac{\text{Freq}(A \text{ with } B)}{\text{Freq}(A)}$$

↳ This is called associatⁿ rule mining or 'market basket analysis'

→ Recom. products with highest

$$Con(x_i / A)$$

Flaw: If B is very popular, it will have high confidence w.r.t many items.

eg: Bread, Milk

low freq — $\boxed{\text{toy, Milk}}$ → high freq (Milk) → high conf

eggs, milk

Lift

→ Another metric for recomendations.

$$\text{Lift}(B \mid A) = \frac{\text{confidence}(B \mid A)}{P(B)}$$

↳ Overall Prob of B selling

what is the confidence of 'milk' given 'toy' while adjusting for the fact that 'milk' itself is very popular.