Recommender Systems || Lecture 2

Summer 2022

Dr. Karthik Mohan

Univ. of Washington, Seattle

June 24 2022

Motivation for Recommender Systems

1. Where you have a web based product sales, you need recommendations
2. Top companies deploy simple to sophisticated recommendation systems depending on their needs
3. Recommendations drives sales, revenue and customer base (e.g. Amazon)
4. Example: Amazon, Walmart, Facebook, YouTube, Twitter and so many more!
5. Scalability issues are rampant and bring in interesting solutions to Recommender systems
6. The course will discuss real case-studies and help students get hands on in thinking about building scalable recommender systems
7. The course will be focused on concepts and practical aspects of recommender systems. Hence all assessments will be through programming assignments and mini-projects hosted on Kaggle.

Week by Week Break Down (Tentative)

Week	Lecture Material	Assignment
1	Intro to Recommender Systems	Sambazon case study
2	Recommender System Baselines	Shopify case study
3	Matrix Factorization methods	Twitter case study
4	Matrix Factorization methods	Twitter case study
5	Deep Learning based recommendations	Walmart case study
6	ML Pipeline for Recommender systems	Amazon case study
7	Real-time Recommendations	Amazon Fresh case study
8	Diversity and Relevance	Final Project
9	Scaling Recommender systems	Final Project
10	Special Topics	Final Project

Summer!

Logistics

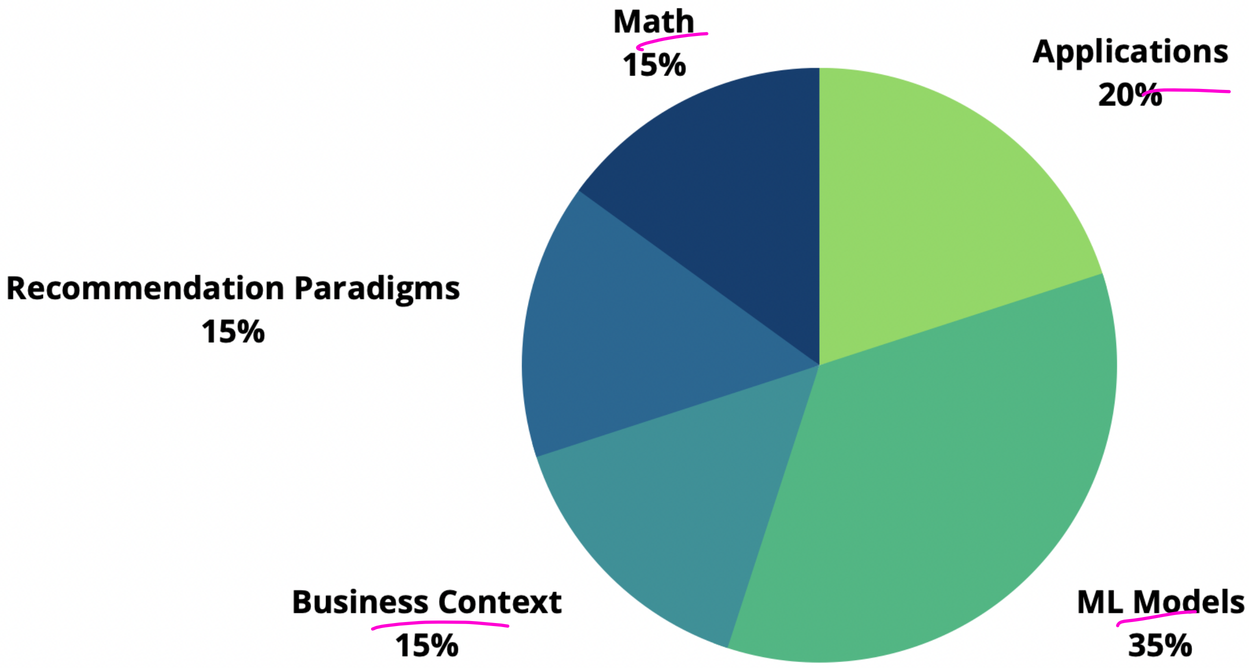
6PM (in-person)

Zoom

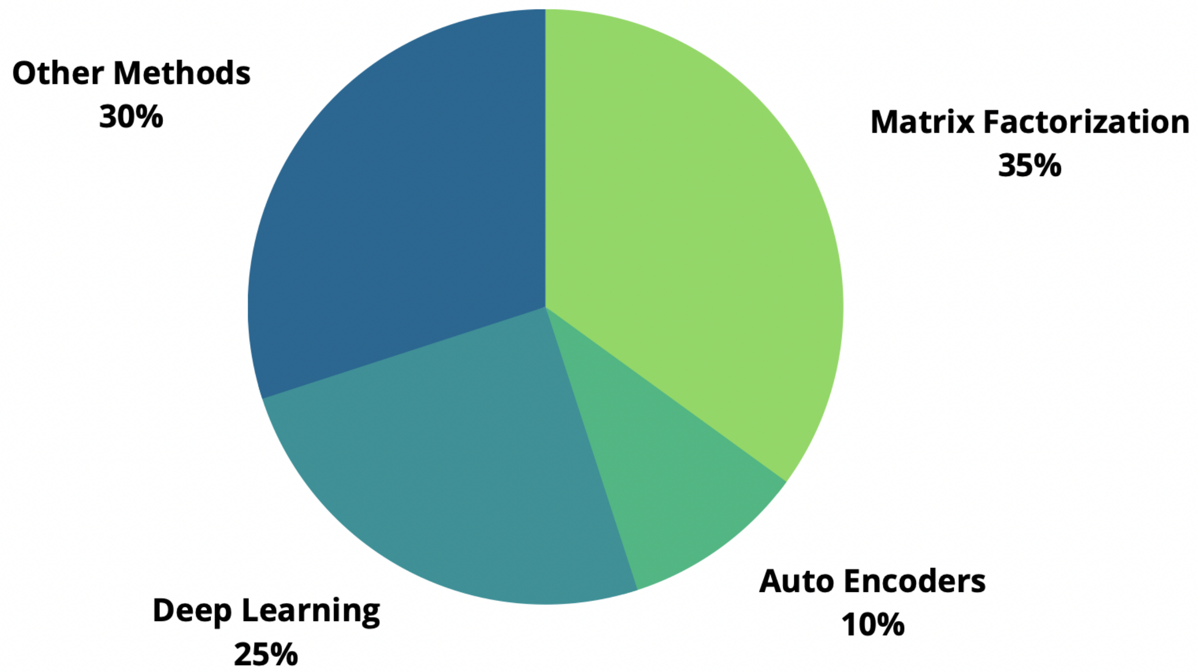
- Lectures on Monday and Wednesday 4 pm pst
- Monday lecture is in-person and Wednesday is online.
- **TA Quiz Section:** F: 4-5PM (Zoom) → online
- **TA Office Hours:** Tu: 4-5PM (EE 433 / ~~Zoom~~) → in-person
- **Karthik Office Hours:** Monday 5 pm, EEB M258 / (unless indicated)

Zoom

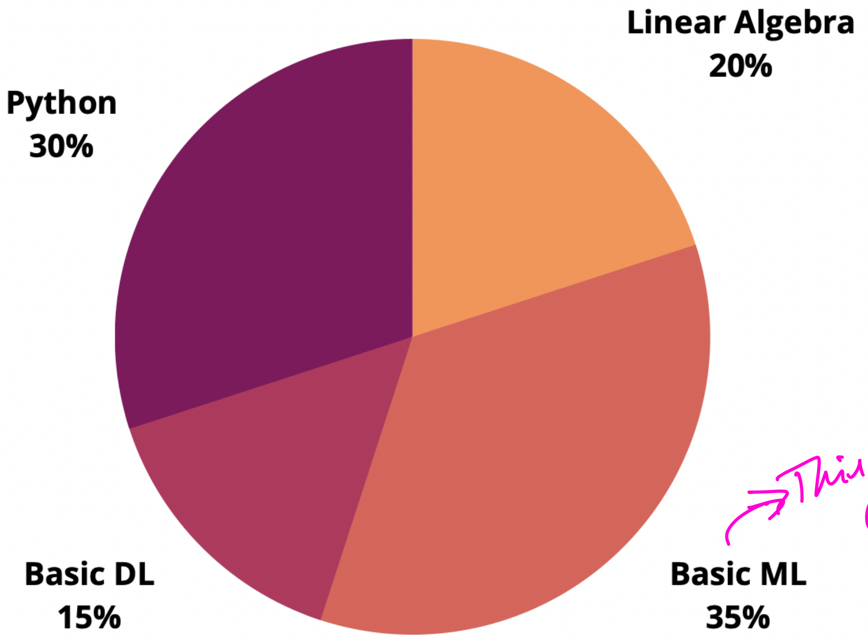
Content


Math
15%
Applications
20%
Recommendation Paradigms
15%
Business Context
15%
ML Models
35%

ML Methods

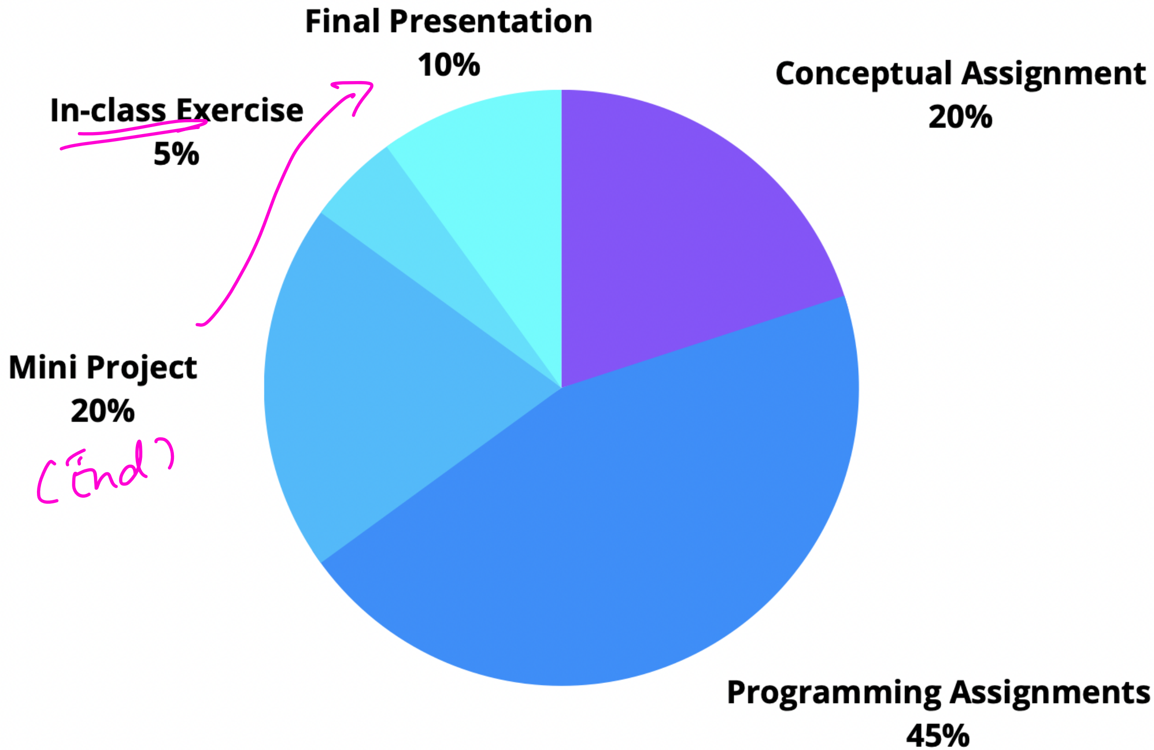

Other Methods
30%
Matrix Factorization
35%
Auto Encoders
10%
Deep Learning
25%

Pre-requisites



This is a requirement! (catch up on your own)

Assessments



Linear and Matrix Algebra Recap

Scalars & Vectors & Matrices & Tensors

$(\mathbb{R})$ $\quad (\mathbb{R}^n)$ $\quad (\mathbb{R}^m \times \mathbb{R}^n)$ $\quad (\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^a)$

Complexity & dimensionality

vectors: p_1, p_2

Add:- $p_1 + p_2$

multiply vectors:- 1) $p_1 \cdot p_2 = p_1^T p_2$ (more frequently used)

$[\quad\quad] \begin{bmatrix} \\ \\ \\ \end{bmatrix} = \text{scalar!}$

2) $p_1 \circ p_2 = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$ (Less used)

Hadamard product

$$\|p_1\|_2^2 = p_1 \cdot p_1 = \begin{bmatrix} & p_1^T & \end{bmatrix} \begin{bmatrix} \\ p_1 \\ \\ \end{bmatrix}$$

$\|p_1\|_2$ = magnitude of p_1

$$= \sqrt{p_1 \cdot p_1}$$

$$\|X\|_F^2 = \langle X, X \rangle = X \cdot X$$

F ↳ Frobenius

$$= Tr(X^T X)$$

$$X \cdot Y = Tr(X^T Y)$$

$\|X\|_F$ = magnitude of matrix X

$$= \sqrt{Tr(X^T X)}$$

$Tr(Z) = ?$ = Sum of diagonals of Z

$$= Z_{11} + Z_{22} + Z_{33} + \dots + Z_{nn}$$

Example:-

$$Z = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 7 & 8 \\ 9 & 3 & -2 \end{bmatrix}$$

$Tr(Z) = ?$

$$= 1 + 7 - 2 = 6$$

$$\|Z\|_F = \sqrt{Tr(Z^T Z)}$$

$Z^T Z = ?$

$$Z^T = \begin{bmatrix} 1 & 2 & 9 \\ 4 & 7 & 3 \\ 5 & 8 & -2 \end{bmatrix}$$

$$= \underset{Z^T}{\begin{bmatrix} 1 & 2 & 9 \\ 4 & 7 & 3 \\ 5 & 8 & -2 \end{bmatrix}} \underset{Z}{\begin{bmatrix} 1 & 4 & 5 \\ 2 & 7 & 8 \\ 9 & 3 & -2 \end{bmatrix}}$$

$$= \begin{bmatrix} 86 & & \\ & 74 & \\ & & 93 \end{bmatrix}$$

$$Tr(Z^T Z) = 253 = (86 + 74 + 93)$$

$$\|Z\|_F = \sqrt{253} \rightarrow \text{Frobenius norm of } Z!!$$

Euclidean norm of a vector $\Longleftrightarrow$ Frobenius norm of a matrix.

Application of matrix

Dot product??

→ Image Search → $I_1 : I_2$

→ Image Similarity → $I_1 : I_2$

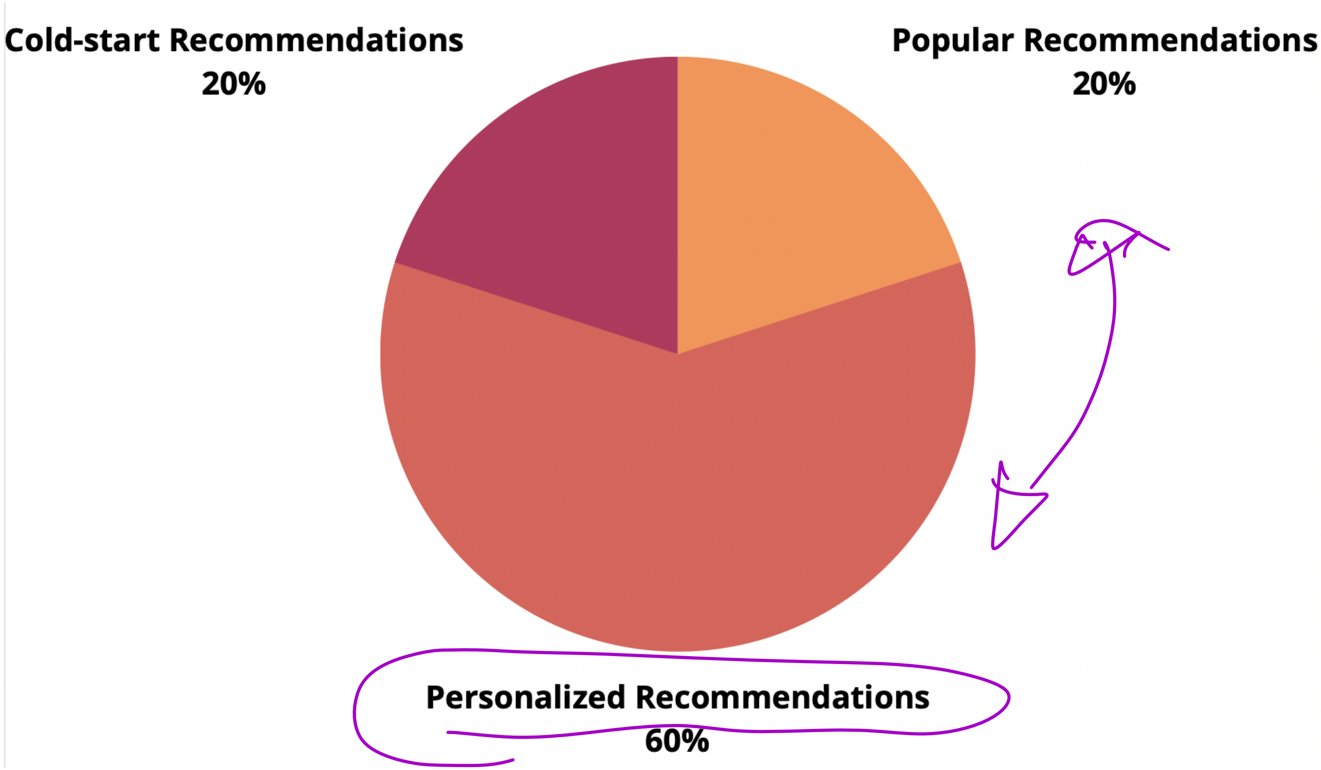
→ Image Distance

$$\|I_1 - I_2\|_F$$

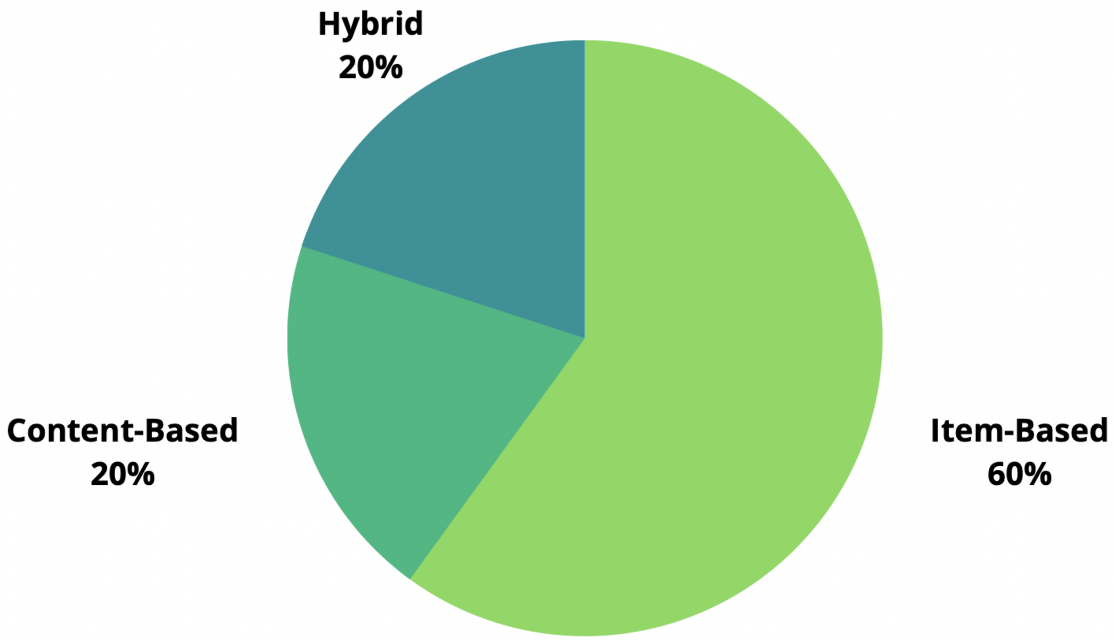
↓

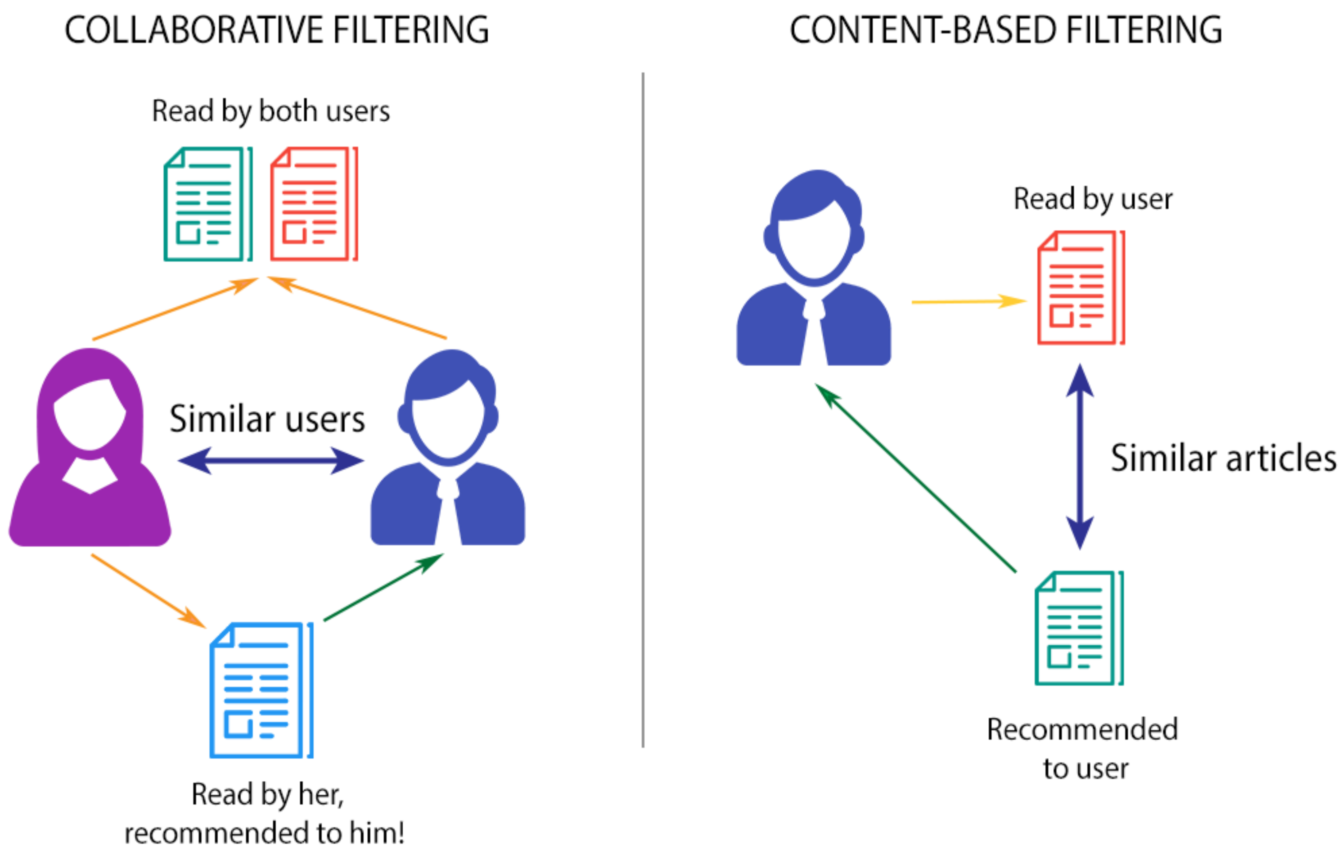
Distance between I_1 & I_2!

Recommender Types

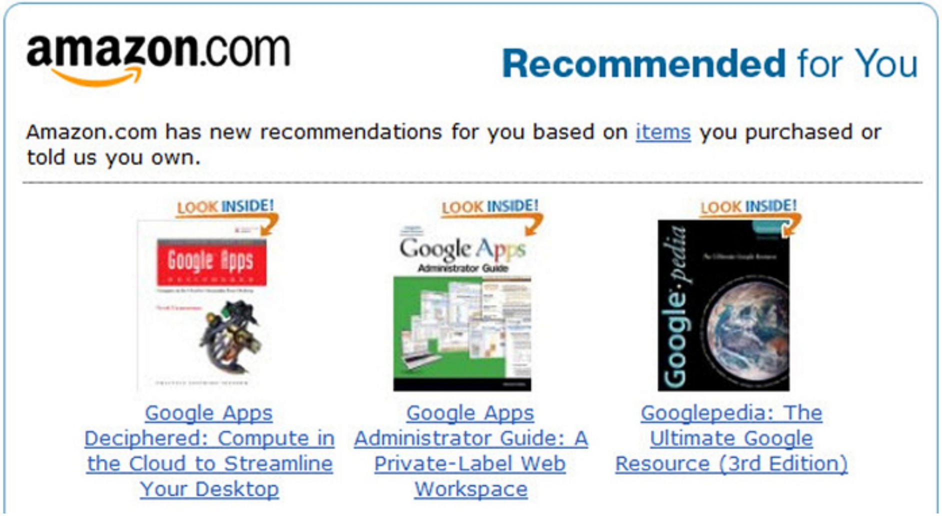


Recommenders

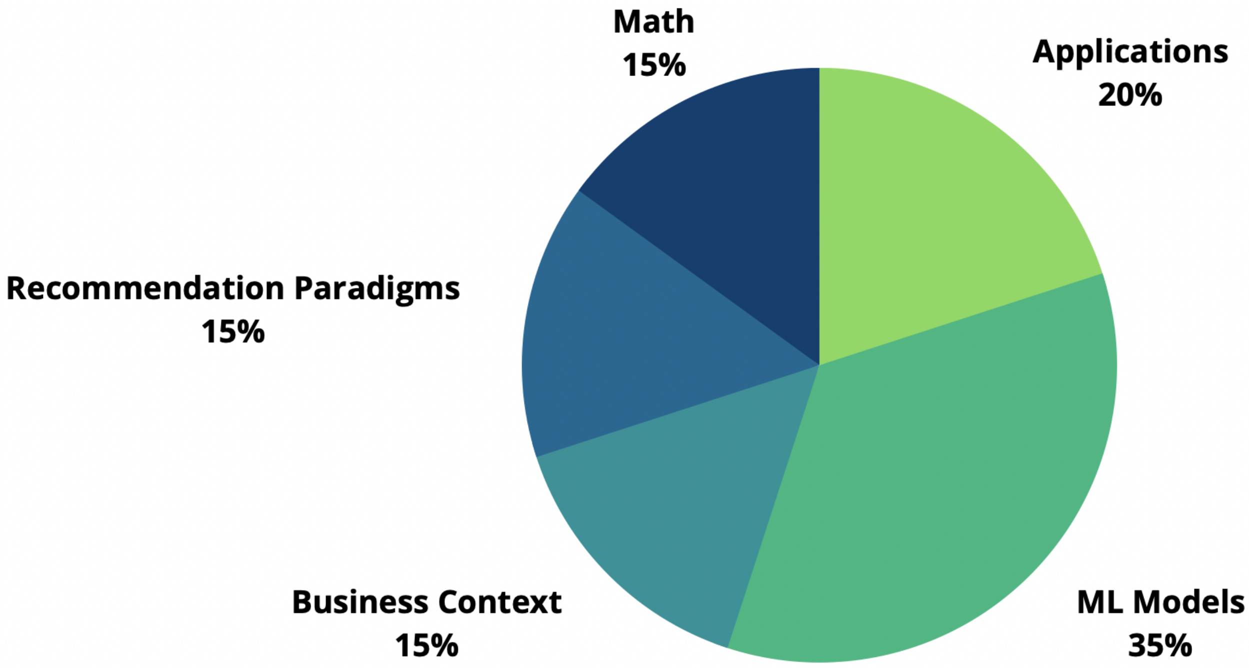



COLLABORATIVE FILTERING
Read by both users
Similar users
Read by her,
recommended to him!
CONTENT-BASED FILTERING
Read by user
Similar articles
Recommended
to user

Item based recommendations



Content


Math
15%
Applications
20%
Recommendation Paradigms
15%
Business Context
15%
ML Models
35%

Amazon Recommendations

Explore the Amazon Webpage

Amazon home page

Identifying the types of recommendations

(widgets)

1. Past incomplete purchase recs

Identifying the types of recommendations

1. Past incomplete purchase recs
2. Buy it again recs

Identifying the types of recommendations

1. Past incomplete purchase recs
2. Buy it again recs
3. Frequently bought recs

Identifying the types of recommendations

1. Past incomplete purchase recs
2. Buy it again recs
3. Frequently bought recs
4. Customers who bought this also bought

Identifying the types of recommendations

1. Past incomplete purchase recs
2. Buy it again recs
3. Frequently bought recs
4. Customers who bought this also bought
5. Similar items to items viewed recs

Identifying the types of recommendations

1. Past incomplete purchase recs
2. Buy it again recs
3. Frequently bought recs
4. Customers who bought this also bought
5. Similar items to items viewed recs
6. Similar items to items purchased recs

Recommendation systems modeling

1. Goal may not be building one 'ultimate' recommender widget

Recommendation systems modeling

1. Goal may not be building one ‘ultimate’ recommender widget
2. Variety of widgets might get better bang for the buck as we saw in earlier example

Breakout discussion #1

(Statistical) 2. (A/B testing) → online / Live Testing

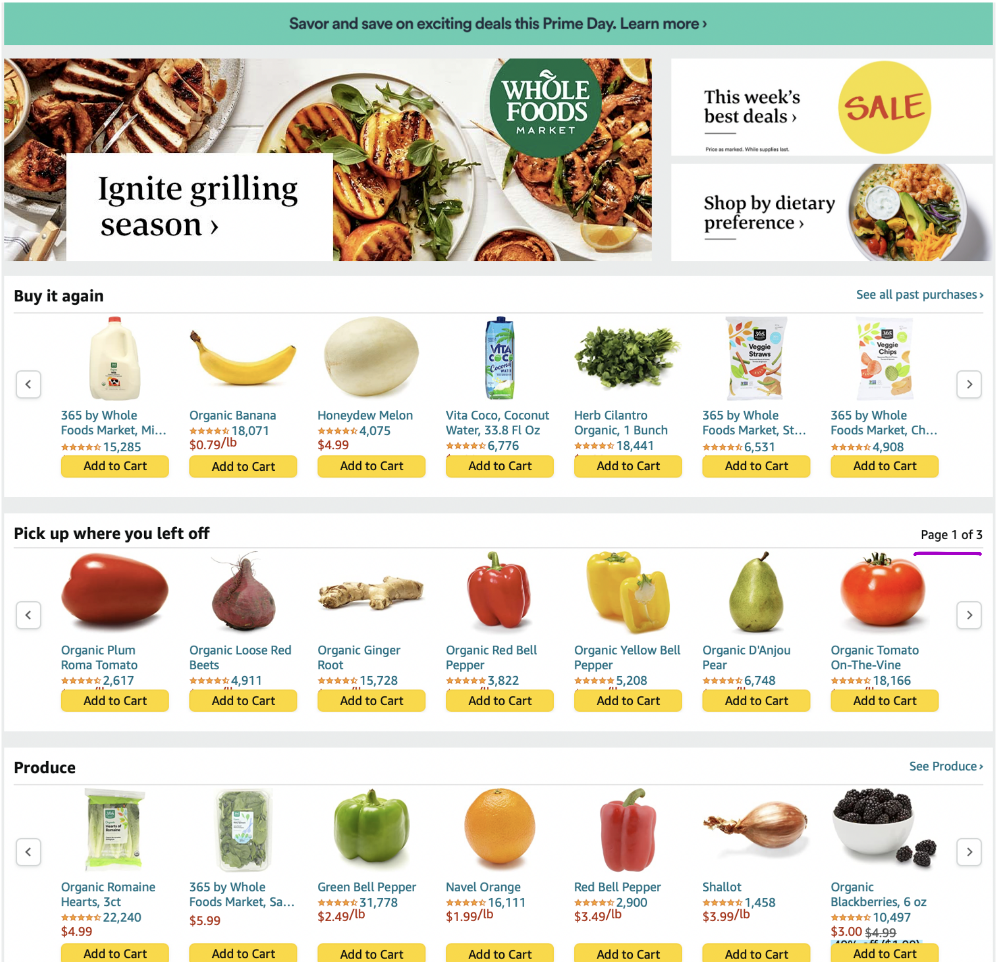
1. Offline Testing → Past Data

Evaluating widgets

You are working in a team of engineers, data scientists and UX designers on a new home page for fashion products. As an ML engineer, you want to build out recommendation widgets on your fashion homepage to help customers find fashion products faster. You decide that you want to have 10 different widgets for different channels of recommendations on the home page to help customers and also maximize sales. The UX designer you work with says that would be not fit with their design specs and asks you to limit it to 3 widgets. Your manager asks you and the UX designer to come to consensus so the team can proceed with executation. What steps/data driven arguments would you take and/or make to resolve this situation or come to a consensus that best meets the common goals that your team has for the home page?

(metrics)

Widgets Example



widgets

carousel

Simple Baseline Recommender

Cosine Similarity

Let's say you want to recommend products from some categories (e.g. groceries, music and fitness to customers) in a way that is personalized and so they find it relevant. Here's a simple baseline model:

1. **Representation:** Find a way to represent your product by a fixed length vector of some dimension d. d here usually captures the number of relevant "features".

or a Matrix (Tensor)

Cosine Similarity

Let's say you want to recommend products from some categories (e.g. groceries, music and fitness to customers) in a way that is personalized and so they find it relevant. Here's a simple baseline model:

1. **Representation:** Find a way to represent your product by a fixed length vector of some dimension d. d here usually captures the number of relevant "features".

   Purchased

2. **Cosine Similarity:** Let's say you have a reference product p and candidate products to recommend $p_1, p_2, \ldots$. Evaluate the cosine similarity between product p and all the other products.

Rep → Earphone (p)
Rep → Another " (p_1)
Rep → watch (p_2)

Cosine Similarity

Let's say you want to recommend products from some categories (e.g. groceries, music and fitness to customers) in a way that is personalized and so they find it relevant. Here's a simple baseline model:

1. **Representation:** Find a way to represent your product by a fixed length vector of some dimension d. d here usually captures the number of relevant "features".
2. **Cosine Similarity:** Let's say you have a reference product p and candidate products to recommend $p_1, p_2, \ldots$. Evaluate the cosine similarity between product p and all the other products.
3. **Rank:** Rank the candidates for recommendation based on cosine simialrity scores (highest to lowest).

Simple Baseline Recommender

Cosine Similarity

Let's say you want to recommend products from some categories (e.g. groceries, music and fitness to customers) in a way that is personalized and so they find it relevant. Here's a simple baseline model:

1. **Representation:** Find a way to represent your product by a fixed length vector of some dimension d. d here usually captures the number of relevant "features".
2. **Cosine Similarity:** Let's say you have a reference product p and candidate products to recommend $p_1, p_2, \ldots$. Evaluate the cosine similarity between product p and all the other products.
3. **Rank:** Rank the candidates for recommendation based on cosine simialrity scores (highest to lowest).
4. **Truncate:** Pick top k from the ranked list to recommend. Where k is usually decided by specifications of your recommender widget.

Cosine Similarity

Definition

Let's say we have two vectors p_1, p_2. The cosine similarity between the two vectors is given by: $\underbrace{\frac{p_1 \cdot p_2}{\|p_1\|_2 \|p_2\|_2}}_{= \cos(\theta)}$

$= Cos(\theta)$

-1 & 1

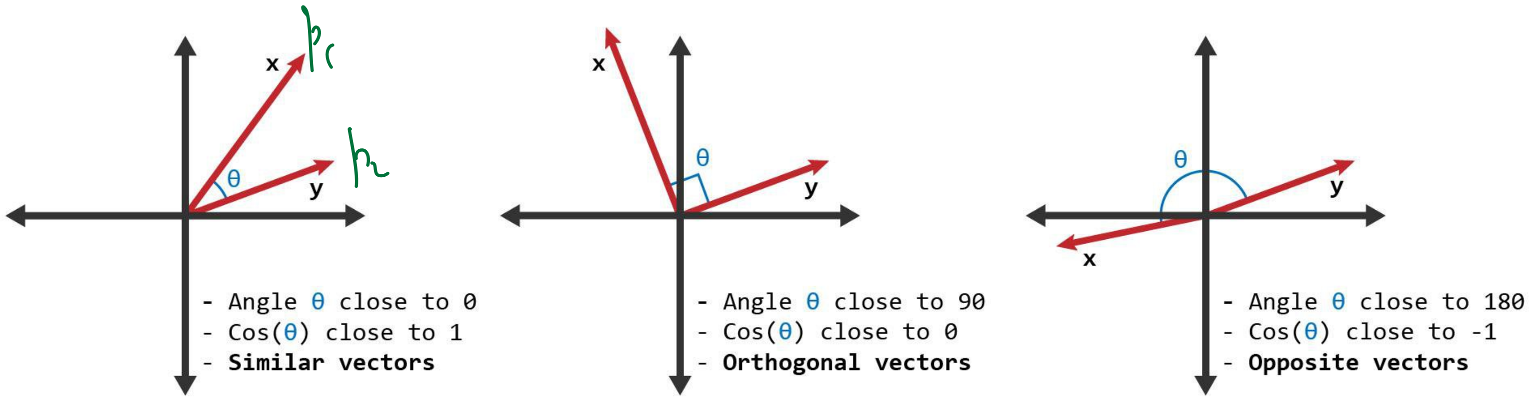


Image reference

ICE #1

Cosine Similarity

Poll

Let's say we have 3 products with attributes as fitness (0 or 1), groceries (0 or 1) and music (0 or 1). Each product can now be represented by a 3 dimensional binary vector. Let $P1 = [0, 1, 1], P2 = [1, 1, 0], P3 = [1, 0, 0]$. If a customer has bought $P1$, you want your baseline recommender to pick between $P2$ and $P3$ as your next recommendation for the customer. You do this by evaluating cosine similarities between the appropriate products. The resulting similarities ~~that~~ will be used ~~compared for~~ for the purpose of making a recommendation are as follows:

1. $[0, 0.5]$
2. $[0.5, 0.71]$
3. $[0.71, 0]$
4. None of the above

dot product

$\cos(P_1, P_2)$ $\qquad$ $\cos(P_1, P_3)$

$$= \frac{[0,1,1]\cdot[1,1,0]}{\|1\|_2 \; \|P_2\|_2}$$

$$= \frac{[0\ 1\ 1]\cdot[1\ 0\ 0]}{\|\cdot\|_2 \; \|\cdot\|_2} = 0$$

$$= \frac{1}{\sqrt{2}\,\sqrt{2}} = \frac{1}{2} = 0.5$$

Connecting Similarity with Distance

Euclidean Distance

Euclidean distance between two vectors p_1 and p_2 is given by

$$\|p_1 - p_2\|_2^2 = \|p_1\|_2^2 + \|p_2\|_2^2 - 2\cancel{p_1, p_2} \quad p_1 \cdot p_2$$

Connecting Similarity with Distance

Euclidean Distance

Euclidean distance between two vectors p_1 and p_2 is given by
$\|p_1 - p_2\|_2^2 = \|p_1\|_2^2 + \|p_2\|_2^2 - 2p_1, p_2$

Euclidean Distance with normalization

Assume that representation of two products p_1 and p_2 are normalized so that $\|p_1\|_2$ and $\|p_2\|_2$ are equal to 1. Then,

$$\|p_1 - p_2\|_2^2 = 1 + 1 - 2\underbrace{p_1 \cdot p_2}_{\|p_1\|_2 \|p_2\|_2}$$

$=$ 2 - 2×Cosine Similarity
$\sim 1 -$ Cosine Similarity

$\boxed{\text{dist}^2 \propto 1 - \text{Cosine Similarity}}$

Connecting Similarity with Distance

Euclidean Distance

Euclidean distance between two vectors p_1 and p_2 is given by
$\|p_1 - p_2\|_2^2 = \|p_1\|_2^2 + \|p_2\|_2^2 - 2p_1, p_2$

Euclidean Distance with normalization

Assume that representation of two products p_1 and p_2 are normalized so that $\|p_1\|_2$ and $\|p_2\|_2$ are equal to 1. Then,

$$\|p_1 - p_2\|_2^2 = 1 + 1 - 2p_1 \cdot p_2$$

$=$ 2 - 2$\times$Cosine Similarity
$\sim 1 -$ Cosine Similarity

So with normalization of vectors, distance between products is proportional to 1 minus the similarity between products

Connecting Similarity with Distance

Euclidean Distance

Euclidean distance between two vectors p_1 and p_2 is given by $\|p_1 - p_2\|_2^2 = \|p_1\|_2^2 + \|p_2\|_2^2 - 2p_1, p_2$

Euclidean Distance with normalization

Assume that representation of two products p_1 and p_2 are normalized so that $\|p_1\|_2$ and $\|p_2\|_2$ are equal to 1. Then,

$$\|p_1 - p_2\|_2^2 = 1 + 1 - 2p_1 \cdot p_2$$

$=$ 2 - 2$\times$Cosine Similarity
$\sim 1 -$ Cosine Similarity

So with normalization of vectors, distance between products is proportional to 1 minus the similarity between products
Higher the similarity, lower the distance and vice versa.

ICE #2

$\text{dist}^2 = \|p_1 - p_2\|_2^2 = 2 - 2 \times \text{cosine Similarity}$

Bounds on distance

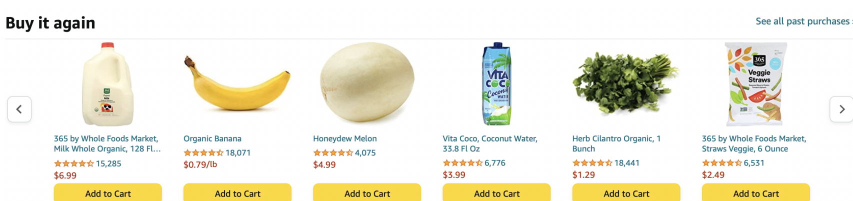
Let's say we want to find the distance between two products p_1 and p_2. Assume that the vector representation of these products is normalized, i.e. $\|p_1\|_2 = \|p_2\|_2 = 1$. What can you say about the bounds on distance between p_1, p_2, i.e. $\|p_1 - p_2\|_2$?

1. Between 0 and 1
2. Between -1 and 1
3. Between 0 and 2
4. Between -1 and $\sqrt{2}$

Breakout discussion #2 (Come back Later)

Repeat purchase recommendations

Assume you are working in a team at Sambazon that is building a repeat purchase recommender system. You have access to anonymized data on customers past purchase history over the past year for popular categories of products. What would be a simple baseline model for repeat purchase recommendations? What factors would you consider when you design your baseline model (your factors have to tie in with the bottom line/top line metrics that your team/project cares about) How would you evaluate the goodness of your model? What specific quantiative metrics would you use to measure the goodness? What are the pros/cons of your baseline and how would you improve it?



Behavioral vs Content Based Recommendations

1. **Behavioral:** Recommendations on understanding customer behaviors based on data (purchases, views and other data). **Collaborative Filtering** is a modeling approach in recommendations that is solely dependent on behavioral data.
   Customers who bought this also bought this.. is an example of a behavioral recommender.
2. **Content:** Recommenders that use content information on items or products are called content based recommenders.
   Items similar to what you purhcased ... is an example of this.
3. **Hybrid:** Often times, a mixture of behavioral and content based data leads to better overall recommender widgets.

Main Signal:- Behavior
Secondary Signal:- Content

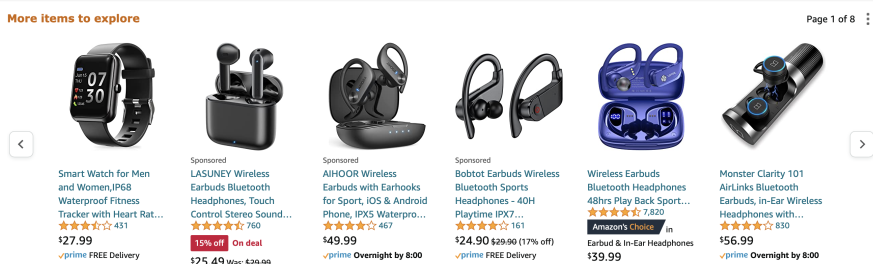
Breakout discussion #3

Customers who bought this also bought ...

Assume that your team would like you to build a baseline recommender for a newly launched fashion product line of the type, “Customers who bought this also bought ...” Your UX designer mentions that you can only surface 5 products at a time on the widget you are building. Discuss and outline step by step how you would arrive at the top 5 products to surface given a single fashion product as input. What if now the input is 3 past fashion purchases of the customer instead of one? How do you convince your team that your widget is doing well enough to be considered for a live launch? How do you evaluate your widget performance when you go live and launch it?

Back Test

A/B



Content based recommendations

1. Use features (or content) from the products to make recommendations.

Content based recommendations

1. Use features (or content) from the products to make recommendations.
2. This includes product categories and product brand

Content based recommendations

1. Use features (or content) from the products to make recommendations.
2. This includes product categories and product brand
3. Anything else? | hash tags, Keywords, description

Breakout discussion #4

New product on the block

Your team is launching new products and product categories to the massive product line of Sambazon. You are the new data scientist hired to surface these new products to customers so your team can drive the sales for the launched products. As of now, there is no easy way for a customer to see the new products, as it isn't yet popular and also because the new products have very little sales to be surfaced by other widgets that exist on the Sambazon recommendations pipeline. How would you leverage what you know about the new product line including any product descriptions and meta data to drive up the sales for the products. Assume the new products are in a product category that a significant percentage of Sambazon customers have purchased from in the past. Come up with a clear step-by-step model/algorithm/approach and also discuss how you will evaluate your recommender model?

Next Lecture

- SVD based recommendations
- Other baseline methods for recommendations
- Introduction to Matrix Factorization