Embeddings:
Vector & Semantic Search

Applications | Concepts | Examples

January 23 2025 | Lecture at University of Washington, Seattle by Dr. Karthik Mohan

Today's Talk

1. Introduction and Motivation

Recommender Systems

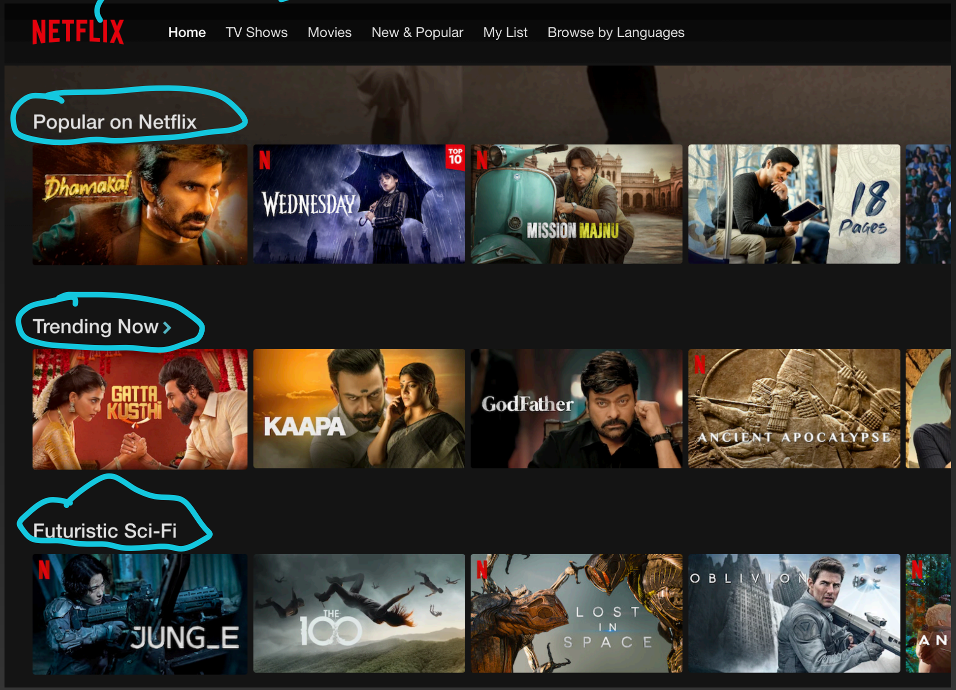
2. Semantic Search

MP1 → Embeddings
→ Semantic Search
→ streamlit Demo

Netflix Recommendations

→ Rec Sys

MULTIPLE CAROUSELS



Personalized

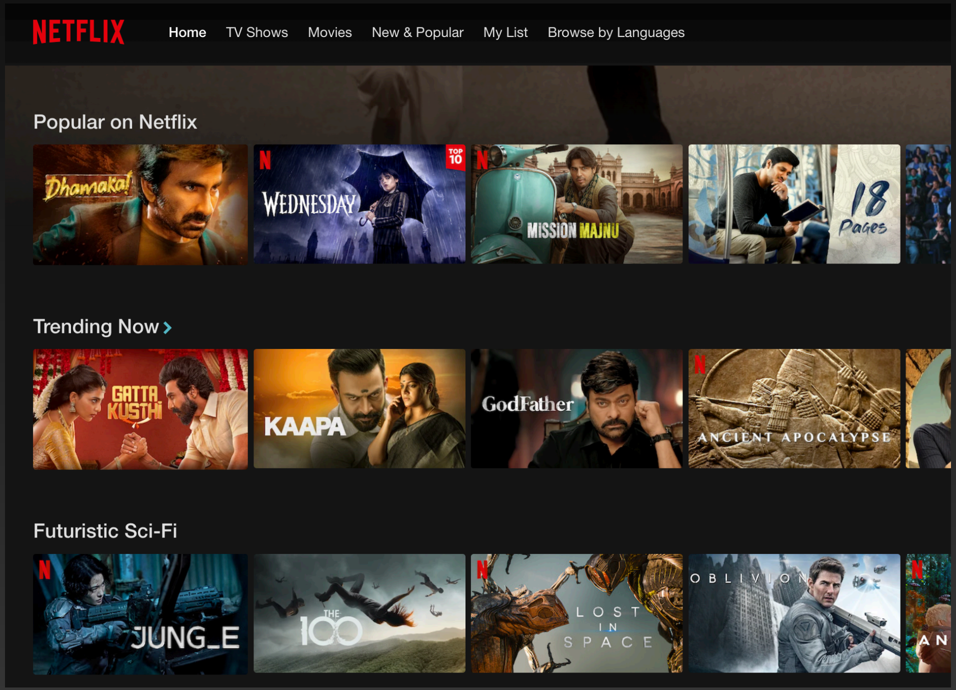
J Row

Netflix Recommendations

POPULAR yet PERSONALIZED

PERSONALIZED TRENDS

PERSONALIZED GENRE

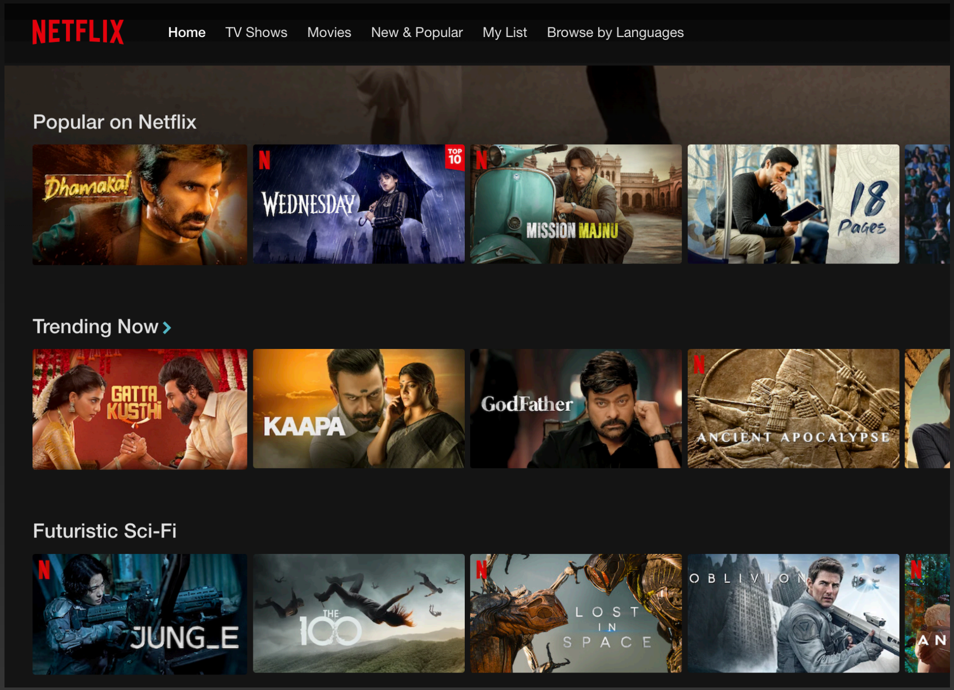


Netflix Recommendations

POPULAR yet PERSONALIZED

PERSONALIZED TRENDS

PERSONALIZED GENRE



Netflix Recommendations

POPULAR yet PERSONALIZED

PERSONALIZED TRENDS

PERSONALIZED GENRE

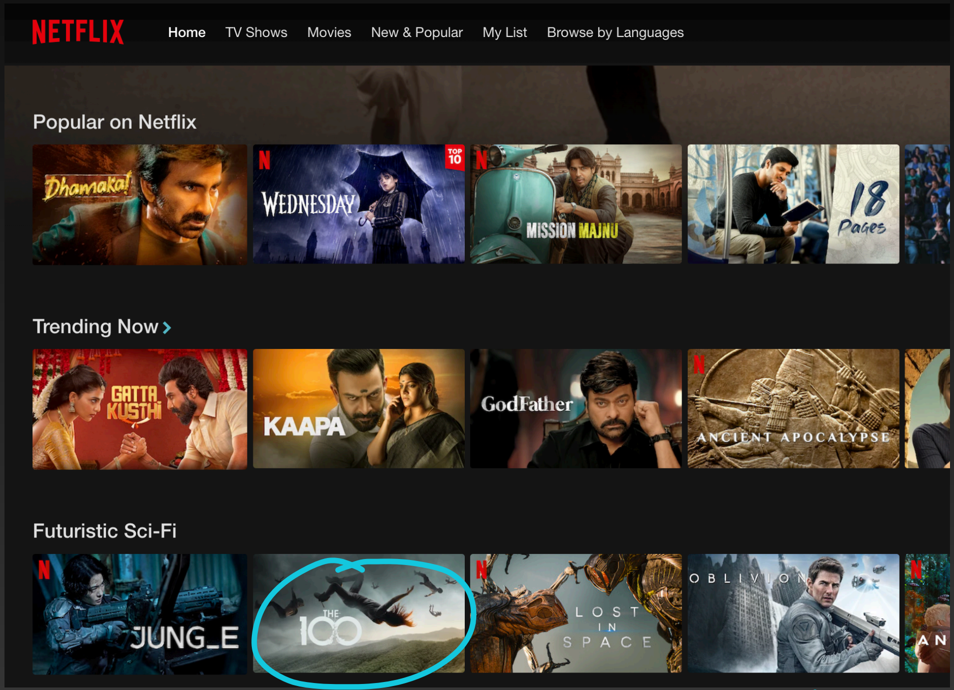


Image is also personalized

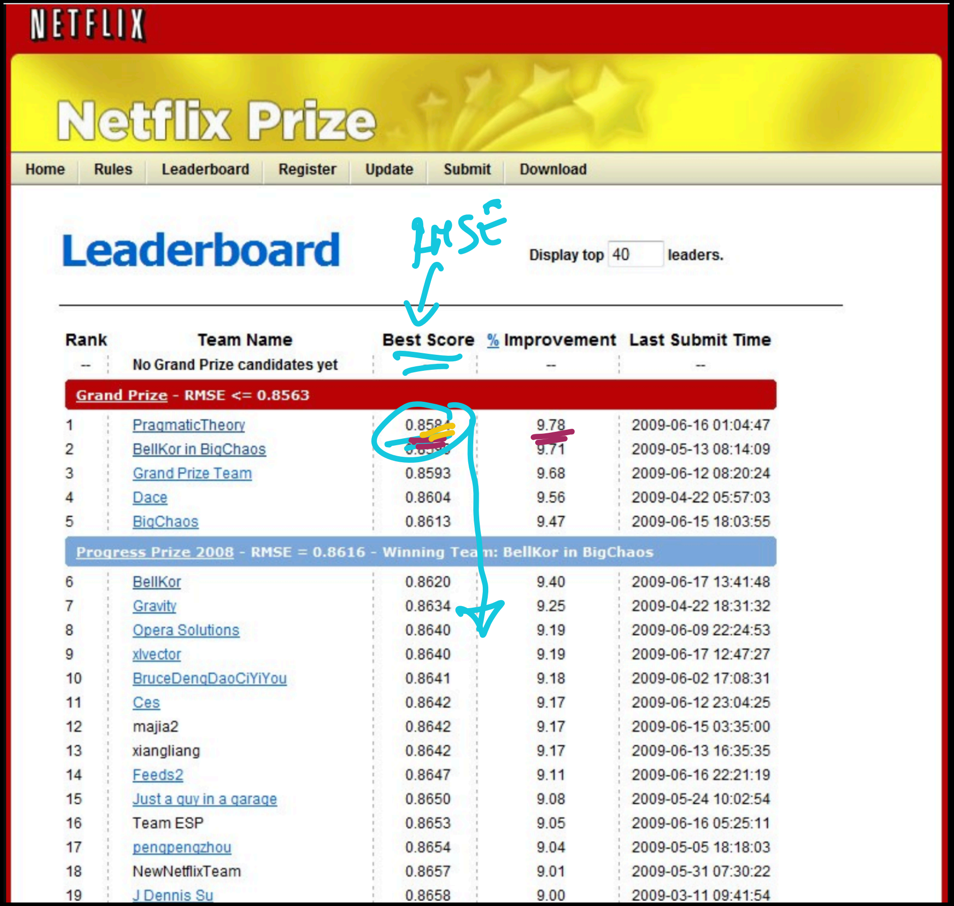
Netflix Million Dollar Prize!

Movie → Good

Bad 1 2 3 4

← Algos. that can improve accuracy of movie recs

Netflix Million Dollar Prize!



NETFLIX

Netflix Prize

Home | Rules | Leaderboard | Register | Update | Submit | Download

Leaderboard

RMSE

Display top 40 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
--	No Grand Prize candidates yet	--	--	--
Grand Prize - RMSE <= 0.8563				
1	PragmaticTheory	0.858[illegible]	9.78	2009-06-16 01:04:47
2	BellKor in BigChaos	[illegible]	9.71	2009-05-13 08:14:09
3	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
4	Dace	0.8604	9.56	2009-04-22 05:57:03
5	BigChaos	0.8613	9.47	2009-06-15 18:03:55
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
6	BellKor	0.8620	9.40	2009-06-17 13:41:48
7	Gravity	0.8634	9.25	2009-04-22 18:31:32
8	Opera Solutions	0.8640	9.19	2009-06-09 22:24:53
9	xlvector	0.8640	9.19	2009-06-17 12:47:27
10	BruceDengDaoCiYiYou	0.8641	9.18	2009-06-02 17:08:31
11	Ces	0.8642	9.17	2009-06-12 23:04:25
12	majia2	0.8642	9.17	2009-06-15 03:35:00
13	xiangliang	0.8642	9.17	2009-06-13 16:35:35
14	Feeds2	0.8647	9.11	2009-06-16 22:21:19
15	Just a guy in a garage	0.8650	9.08	2009-05-24 10:02:54
16	Team ESP	0.8653	9.05	2009-06-16 05:25:11
17	pengpengzhou	0.8654	9.04	2009-05-05 18:18:03
18	NewNetflixTeam	0.8657	9.01	2009-05-31 07:30:22
19	J Dennis Su	0.8658	9.00	2009-03-11 09:41:54

Kaggle

$RMSE = 0.86$

$\Rightarrow$ on an avg the diff in true rating & predicted rating ≈ 1

$$RMSE = \sqrt{\frac{\sum_{i,u} (\hat{r}_{iu} - r_{iu})^2}{N}}$$

Avg RMSE

Predict	True	Err
1	3	2
4	1	3
5	4	1

Collaborative Filtering

Users

Rishi

Michael

Karthik

Roshin

Amy

Collaborative Filtering

Avatar | Arrival | When Harry | Before Sunrise | Minions

5 movies

Rishi

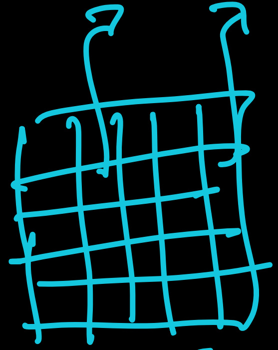
Michael

Karthik

Roshin

Amy

r_{22} r_{25}



5x5 matrix

r_{25}

– Rating of movie 5 by user 2

Collaborative Filtering

					
	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	?	👎		👍
Roshin	?	👎	👍	👍	
Amy	?	👍	?	👍	👎

Rating :-
1 to 5

Simplify Ratings
Like | Dislike
1 | 0

$\hat{y} \rightarrow 1$ or 0

1. Explicit feedback
1-5 vs Like Imp?
2. Implicit feedback
I watched TV show for 30 mins

Collaborative Filtering


	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	?	👎	?	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	?	👎	?	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

Collaboration

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍 ✓	👍	👎 ✓	👎	👍 ✓
Michael	👎	?	👍	👍	👍
Karthik	👍 ✓	? Like	👎 ✓	? Dislike	👍 ✓
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	?	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	?	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	?	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	?	👎	👍	👍	?
Amy	?	👍	?	👍	👎

Collaborative Filtering

Collaborative Filtering

					
	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	? ✗	👍 ✓	👍 ✓	👍
Karthik	👍	👍	👎	👎	👍
Roshin	?	👎	👍 ✓	👍 ✓	? ✓
Amy	?	👍	👍	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	?	👎	👍	👍	👍
Amy	?	👍	👍	👍	👎

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	?	👎	👍	👍	👍
Amy	?	👍	👍	👍	👎

Collaborative Filtering

Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	?	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	👎	👎	👍	👍	👍
Amy	?	👍	👍	👍	👎

Collaborative Filtering

					
	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	👎	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	👎	👎	👍	👍	👍
Amy	?	👍	👍	👍	👎

Collaborative Filtering

					
	Avatar	Arrival	When Harry	Before Sunrise	Minions
Rishi	👍	👍	👎	👎	👍
Michael	👎	👎	👍	👍	👍
Karthik	👍	👍	👎	👎	👍
Roshin	👎	👎	👍	👍	👍
Amy	?	👍	👍	👍	👎

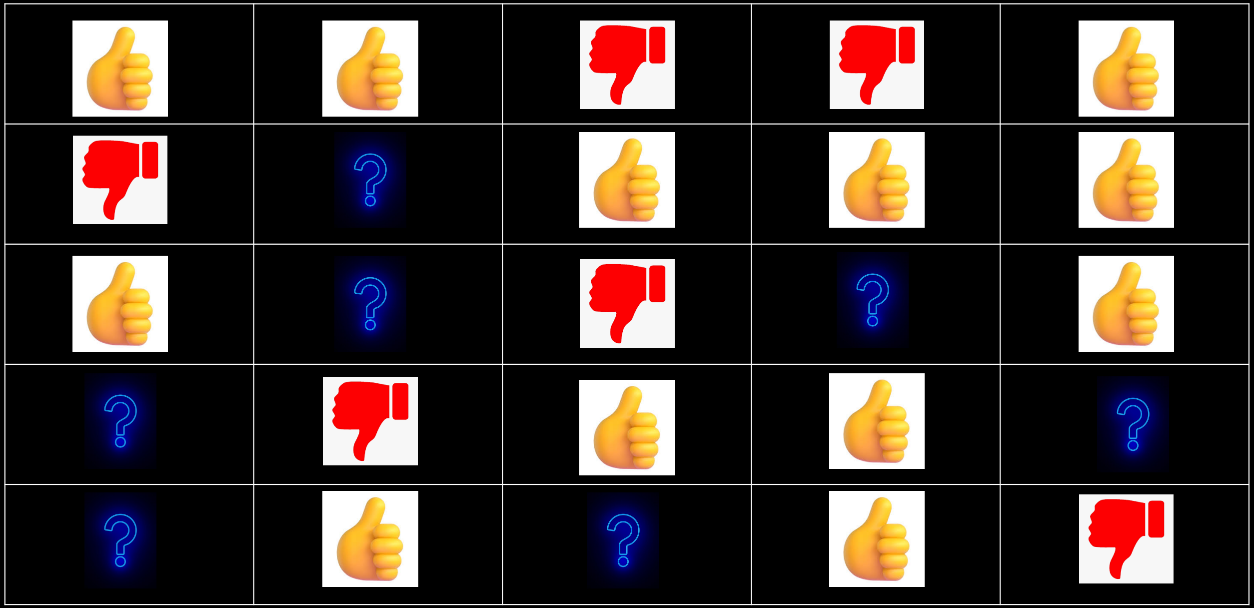
Collaborative Filtering

using user behaviour / user psychology to inform prediction

Filled in the missing ratings based on Collaborative Filtering

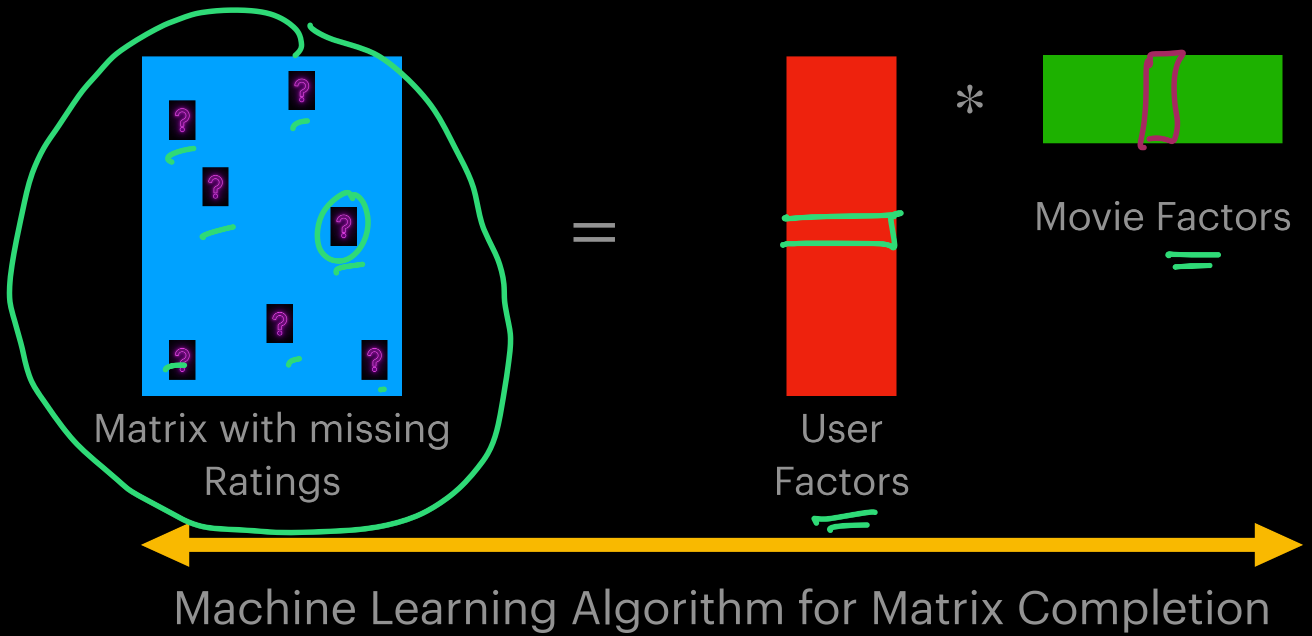
Collaborative Filtering

Through Matrix Completion!



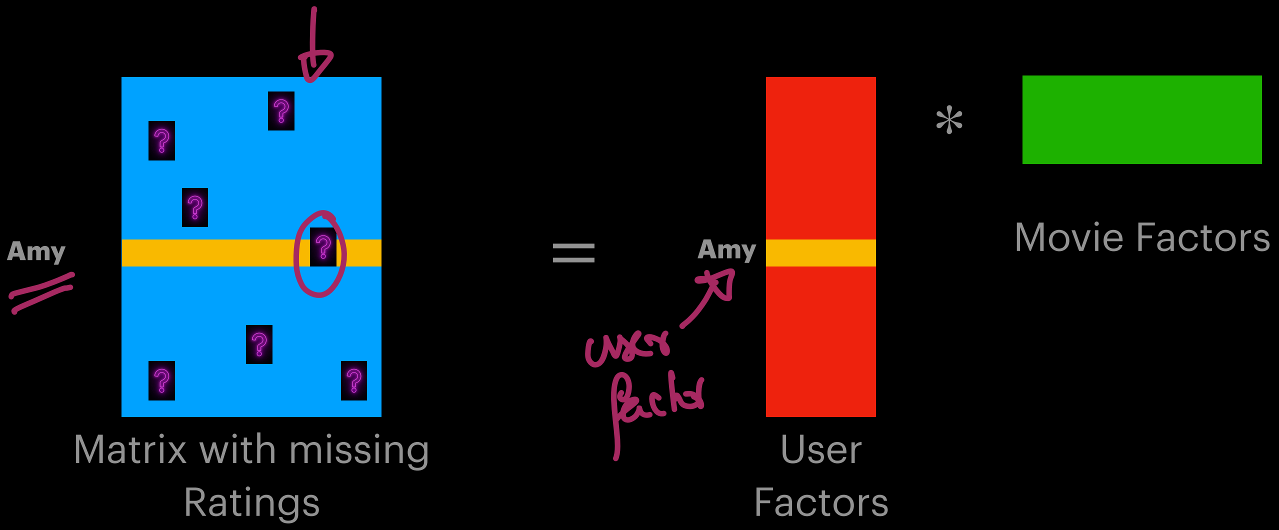
Collaborative Filtering

Through Matrix Completion!



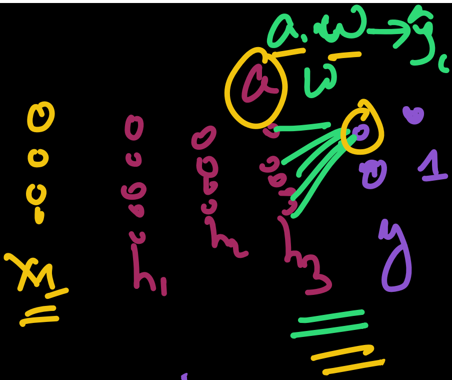
Collaborative Filtering

Through Matrix Completion!



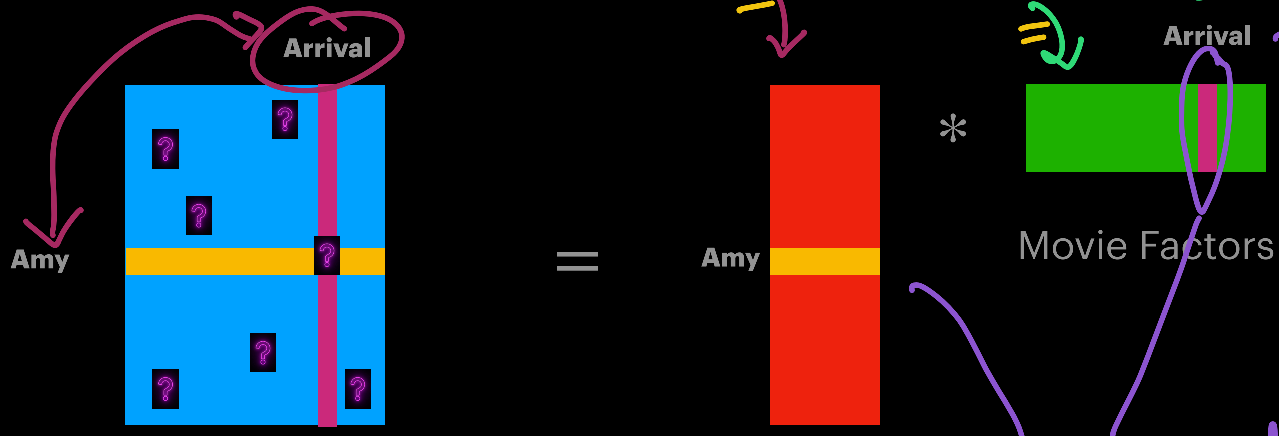
Collaborative Filtering

Through Matrix Completion!



W^U → user weights

W^M → movie weights



Matrix with missing Ratings

User Factors

Movie Factors

→ movie factor for Arrival

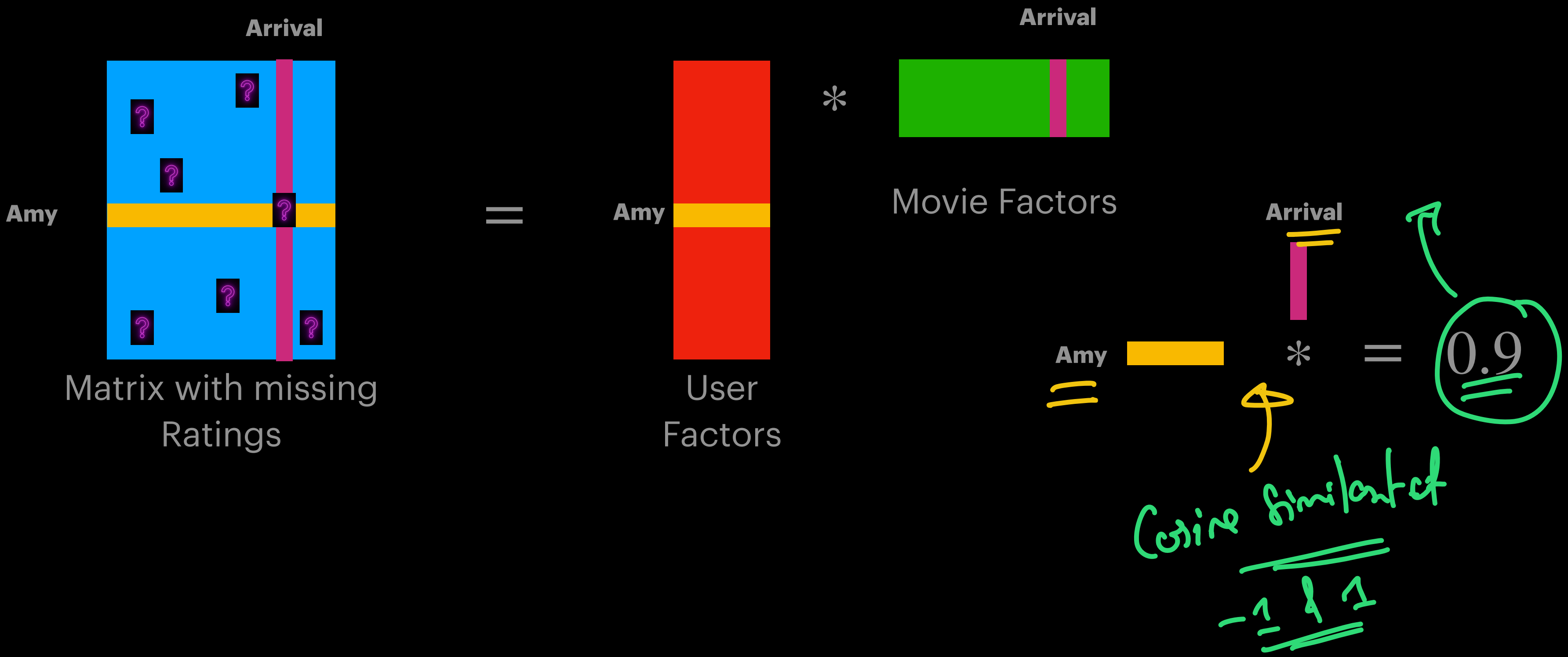
W^U & W^M — weights are learned from data thru training

Latent (Hidden) Representations

Train Data

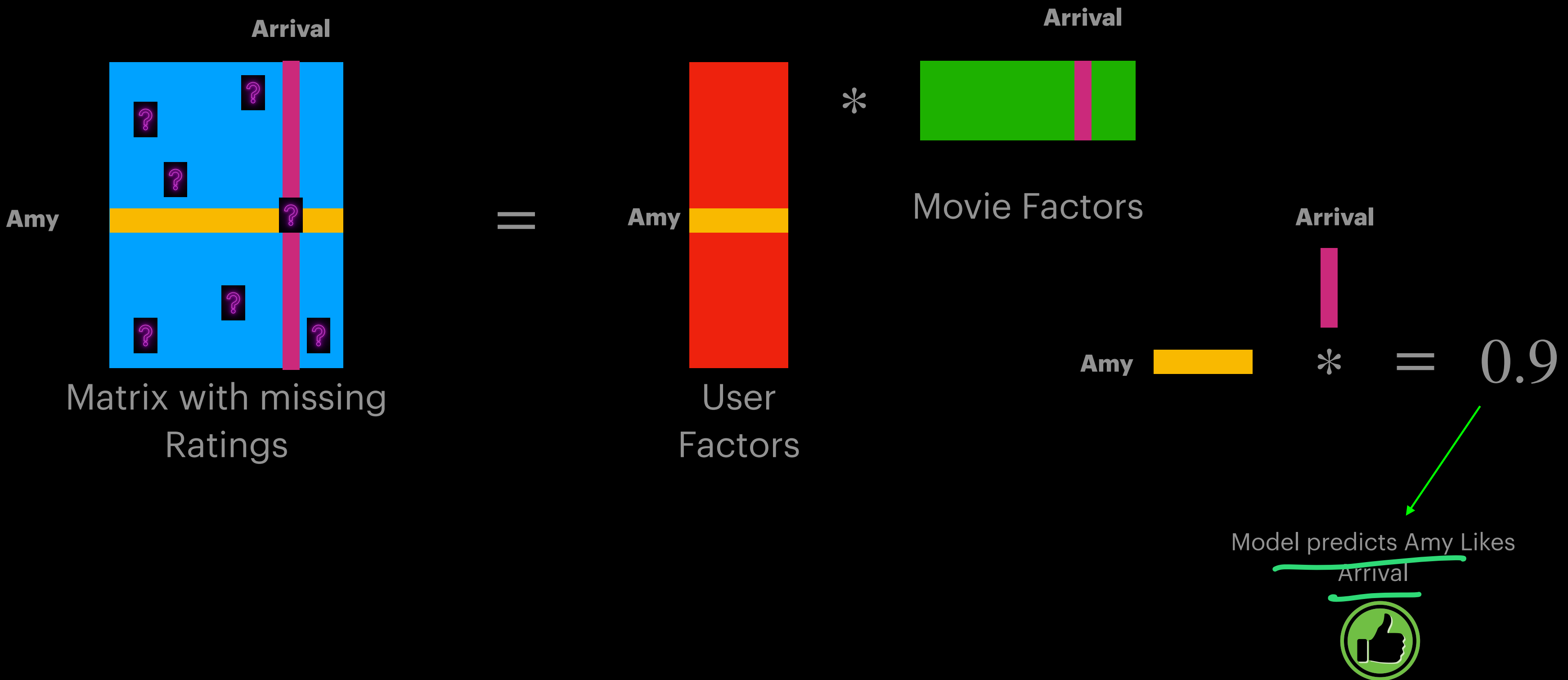
1000's of such data { u_1, m_{10} → Like; u_{200}, m_{1000} → Dislike } → learn W^U, W^M

Collaborative Filtering
Through Matrix Completion!



Collaborative Filtering

Through Matrix Completion!



Collaborative Filtering

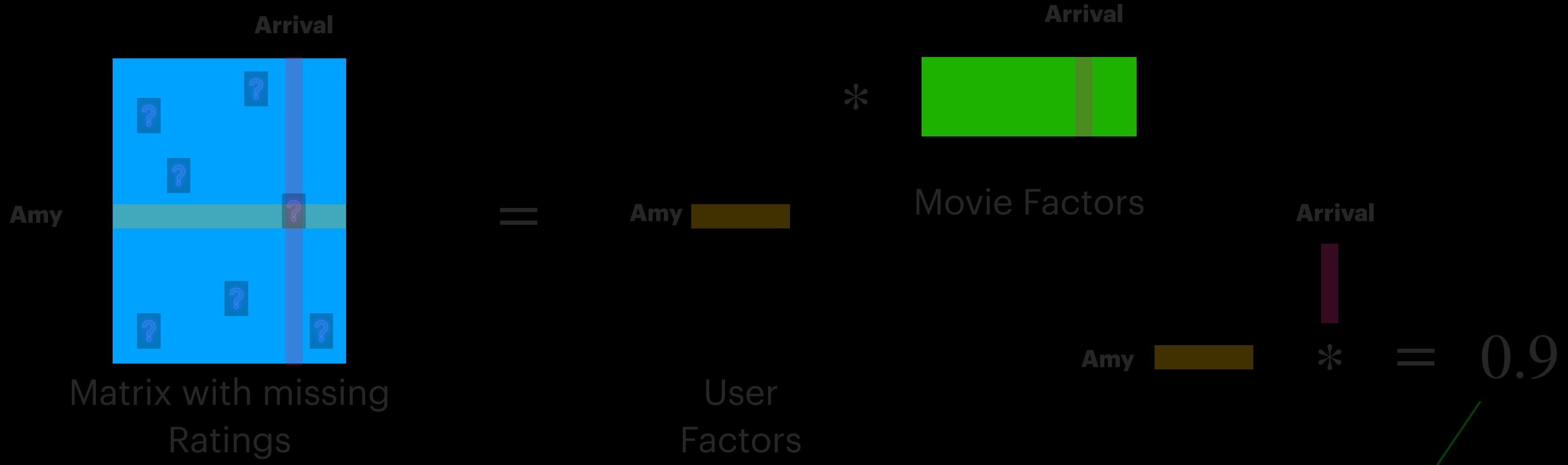
Through Matrix Completion!

Linear Algebra in Action

Q: What is the closest Linear Algebra method that looks similar to the above factorization?

Model predicts Amy Likes Arrival

Collaborative Filtering
Through Matrix Completion!



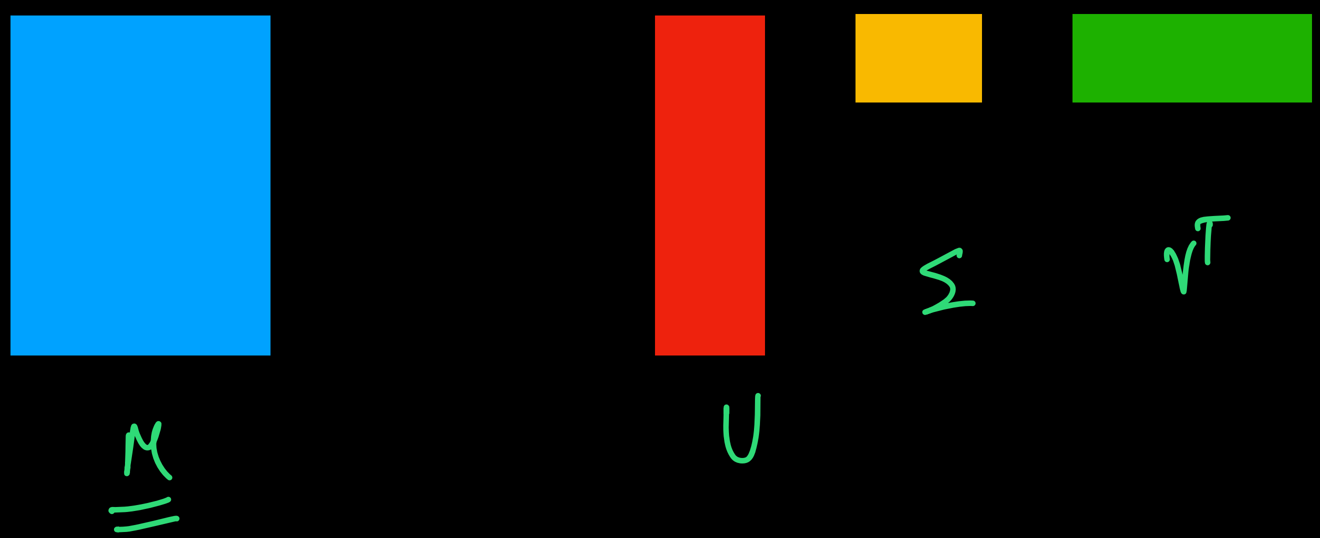
Model predicts Amy Likes Arrival

Linear Algebra in Action

Q: What is the closest Linear Algebra method that looks similar to the above factorization?

A: SVD = Singular Value Decomposition

SVD of a matrix



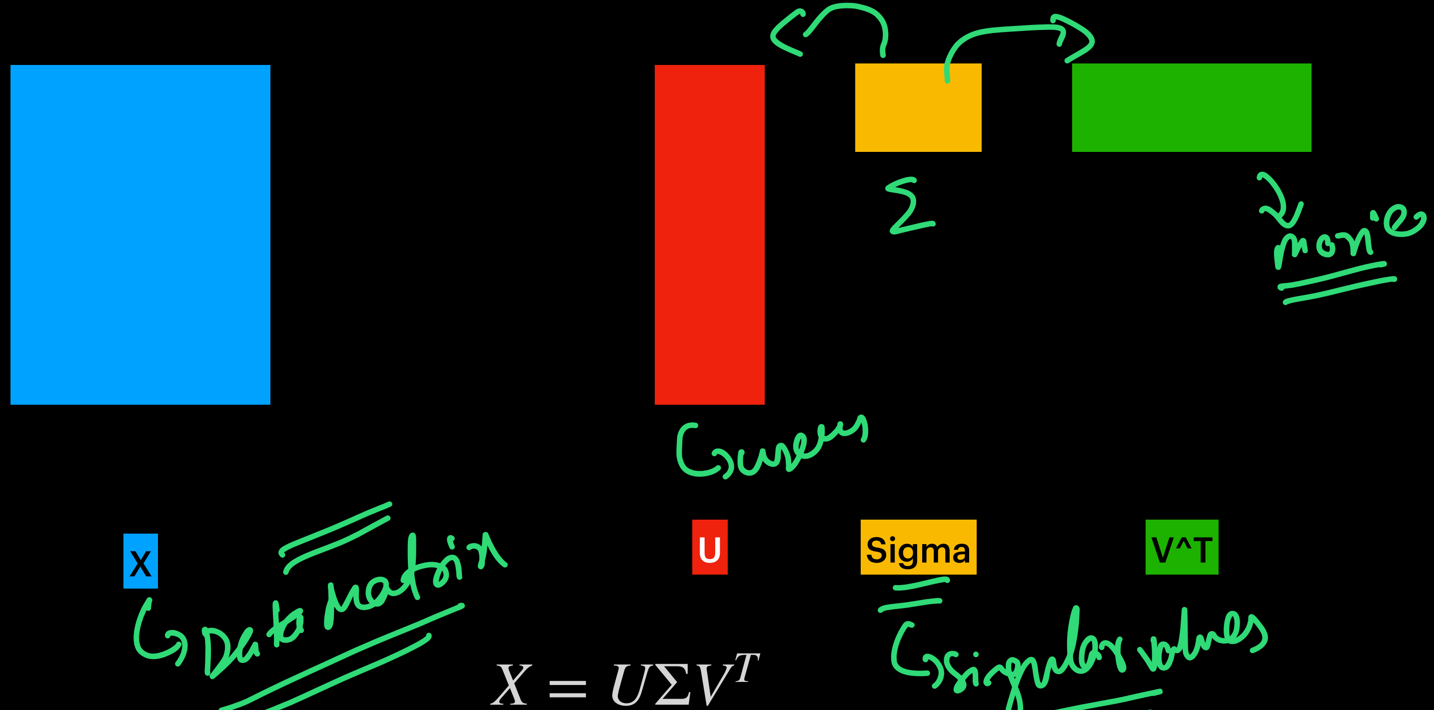
SVD of a matrix

SVD of a matrix

$$X = U\Sigma V^T$$

SVD of a matrix

Every matrix has an SVD!



$$X = U\Sigma V^T$$

SVD of a matrix

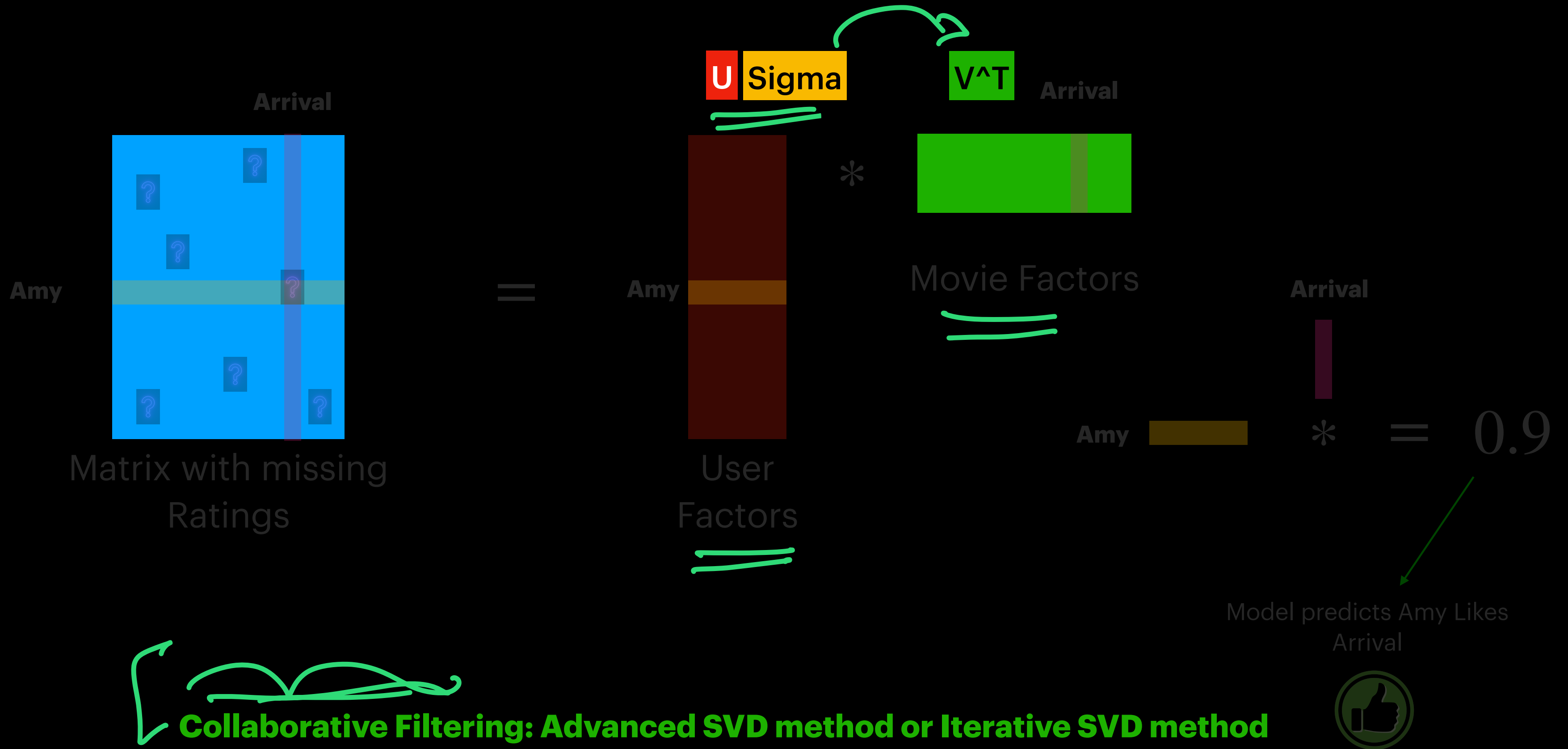
Hence: Data Matrix also has an SVD!

$$X = U\Sigma V^T$$

importance

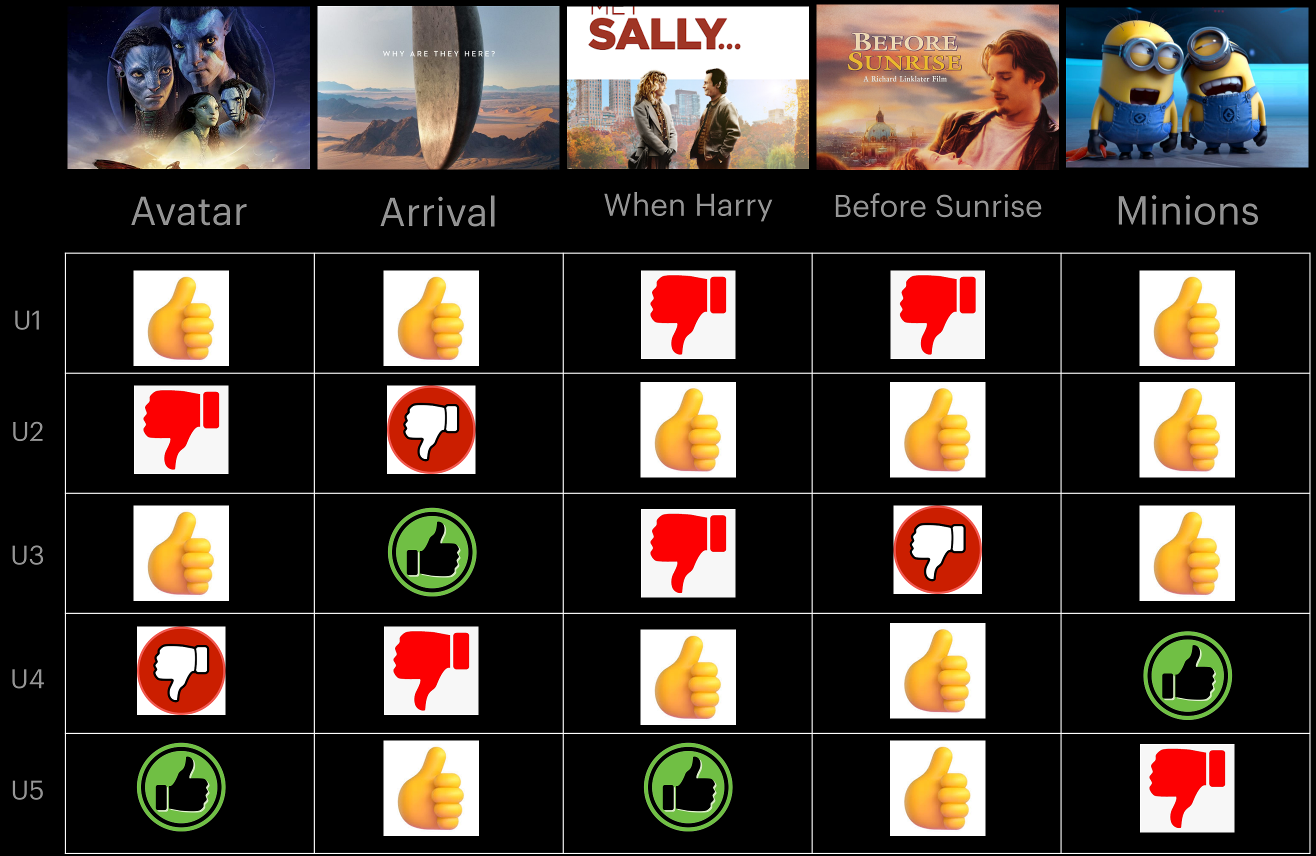
Collaborative Filtering

Through Matrix Completion!



Collaborative Filtering: Advanced SVD method or Iterative SVD method

Collaborative Filtering

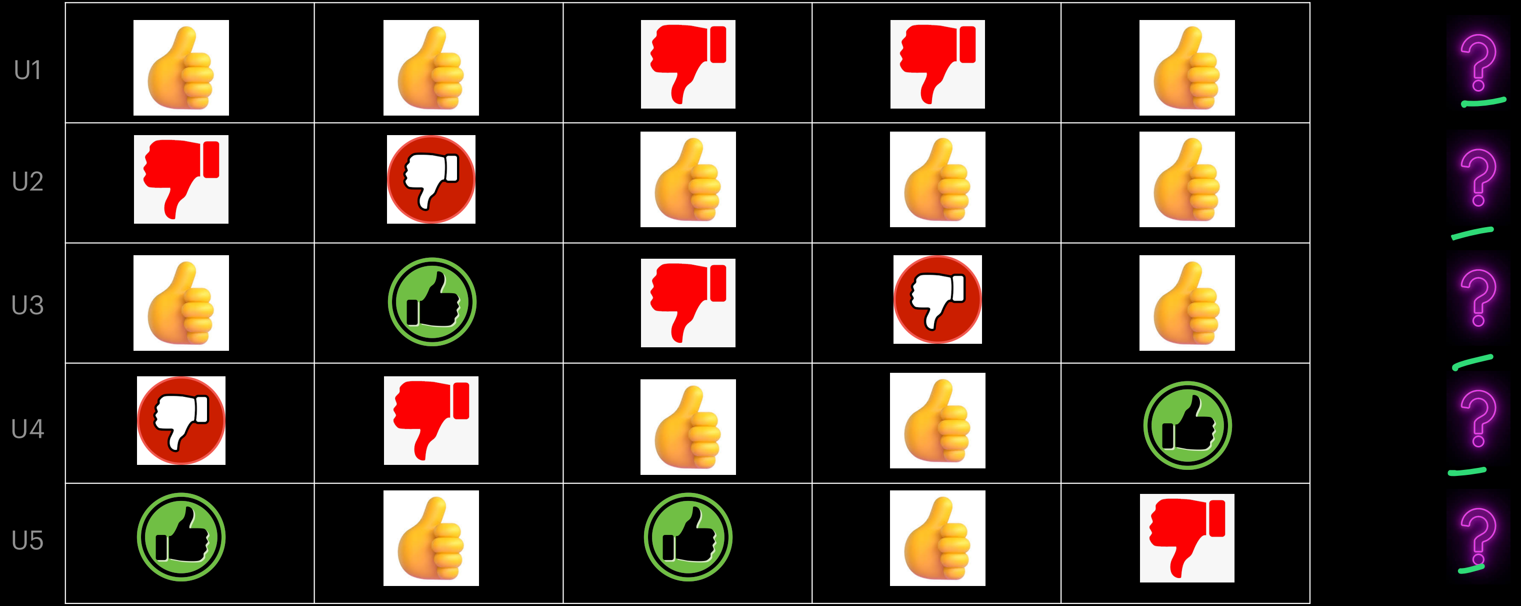


Men in Black

Collaborative Filtering

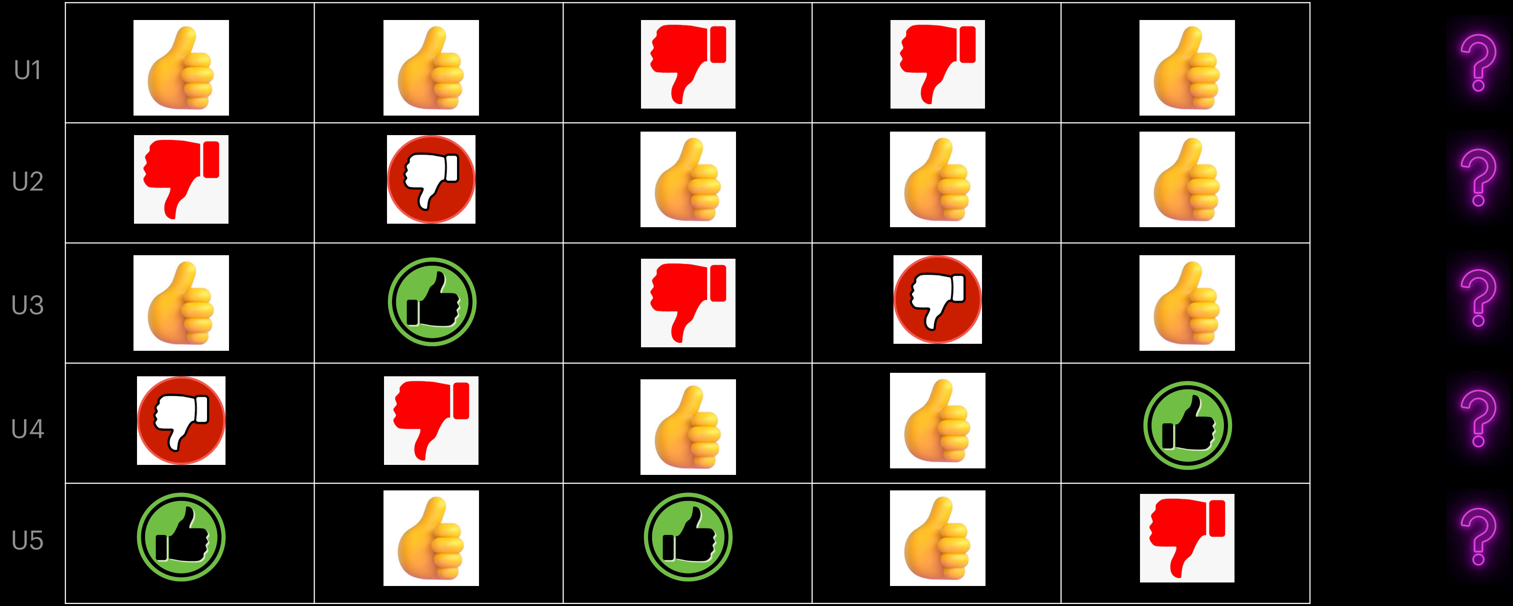
New Movie

Avatar | Arrival | When Harry | Before Sunrise | Minions | Men in Black II/III



Collaborative Filtering

Avatar | Arrival | When Harry | Before Sunrise | Minions | Men in Black



user Behavior Standpoint

Cold Start Problem

Movie is new

⇒ Cold Start!

Content Based Filtering

vs Collaborative Filtering

Avatar

Arrival

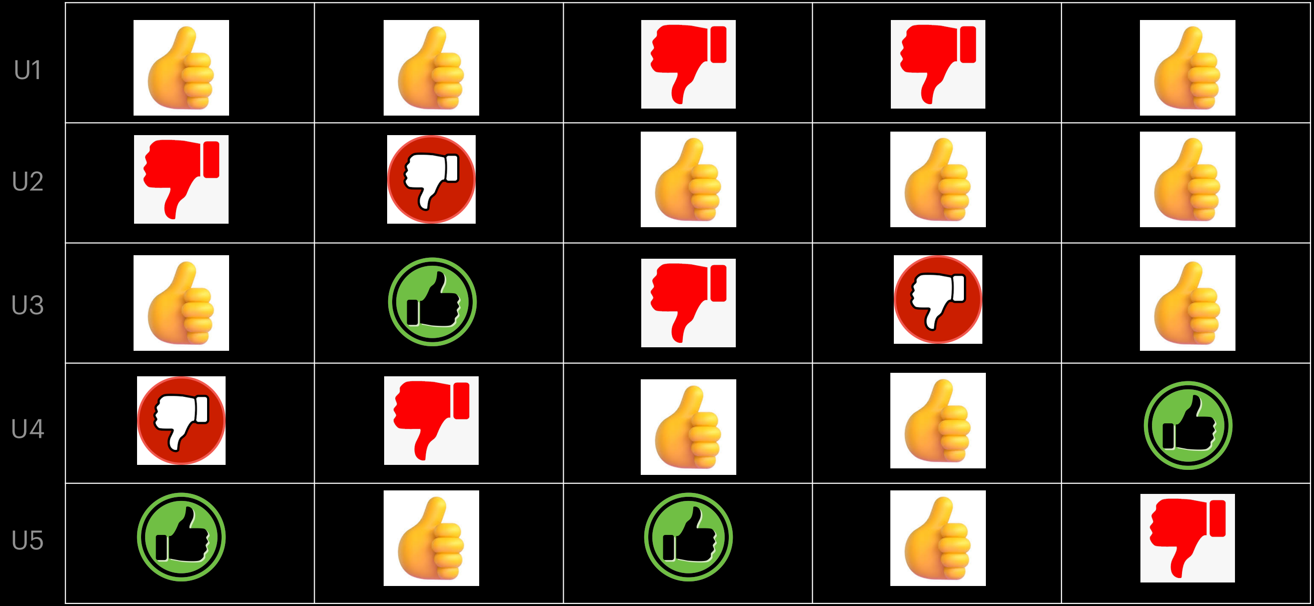
When Harry

Before Sunrise

Minions

Men in Black

Karthik

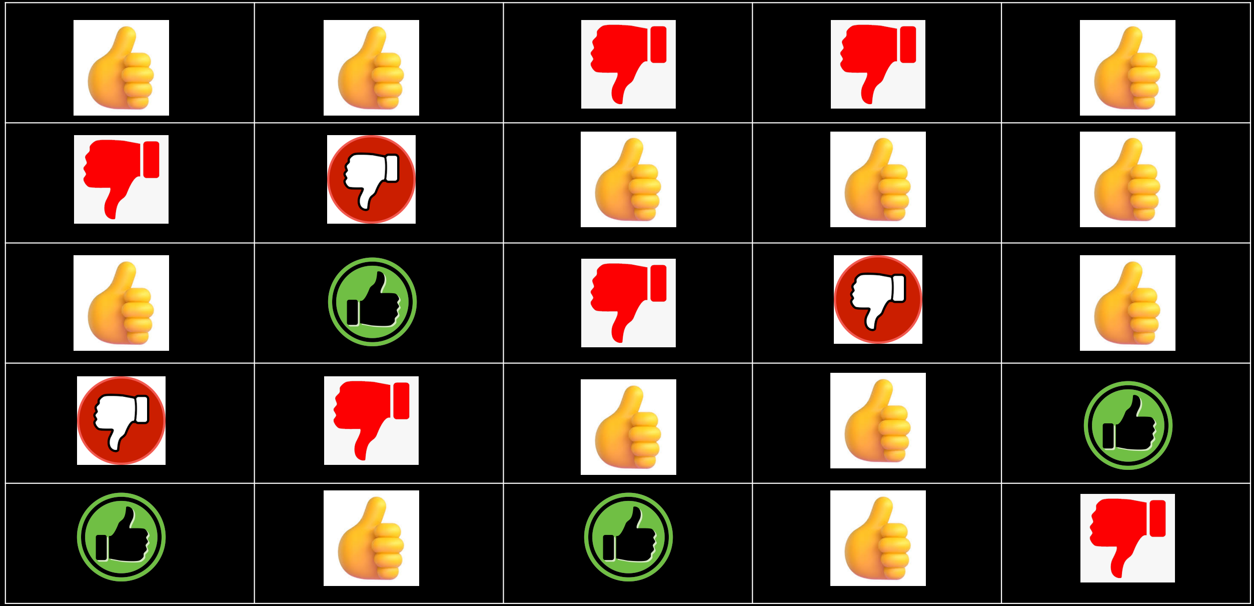


Content Based Filtering

vs Collaborative Filtering

Avatar | Arrival | When Harry | Before Sunrise | Minions

U1
U2
U3
U4
U5

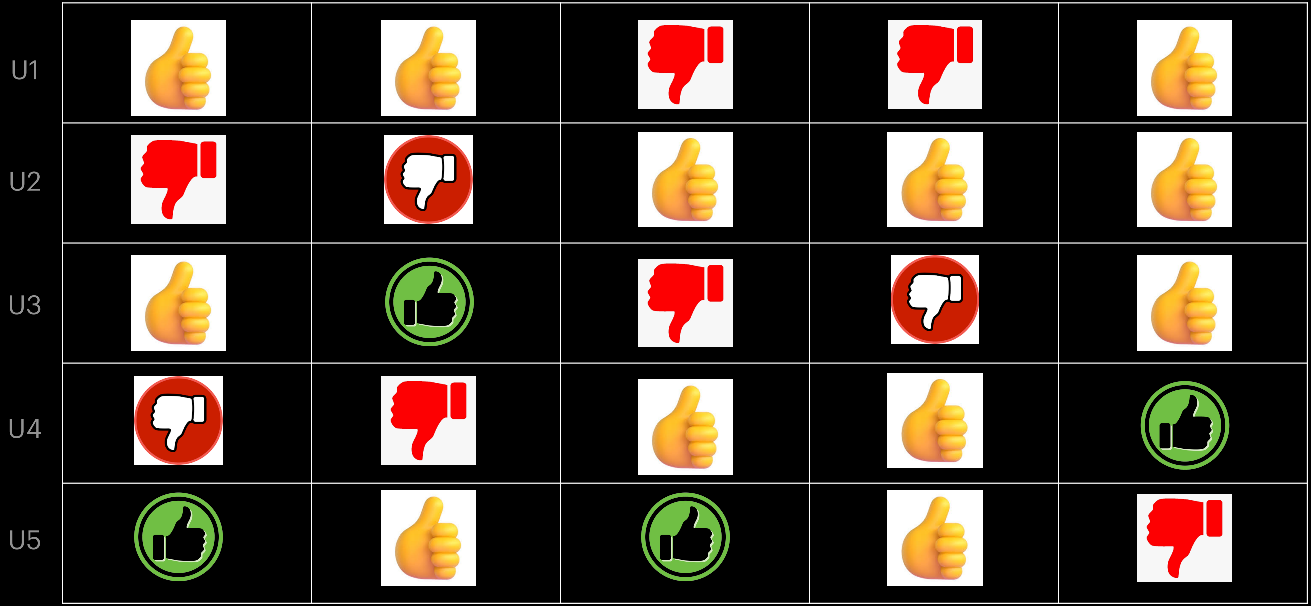


Karthik

Content Based Filtering

vs Collaborative Filtering

Avatar | Arrival | When Harry | Before Sunrise | Minions



Karthik

Watched

Sci-fi

Arrival

Content Based Filtering
vs Collaborative Filtering

Avatar

Arrival

When Harry

Before Sunrise

Minions

	Avatar	Arrival	When Harry	Before Sunrise	Minions
U1	👍	👍	👎	👎	👍
U2	👎	👎	👍	👍	👍
U3	👍	👍	👎	👎	👍
U4	👎	👎	👍	👍	👍
U5	👍	👍	👍	👍	👎

Karthik

Watched

Aliens
Sci-fi

Arrival

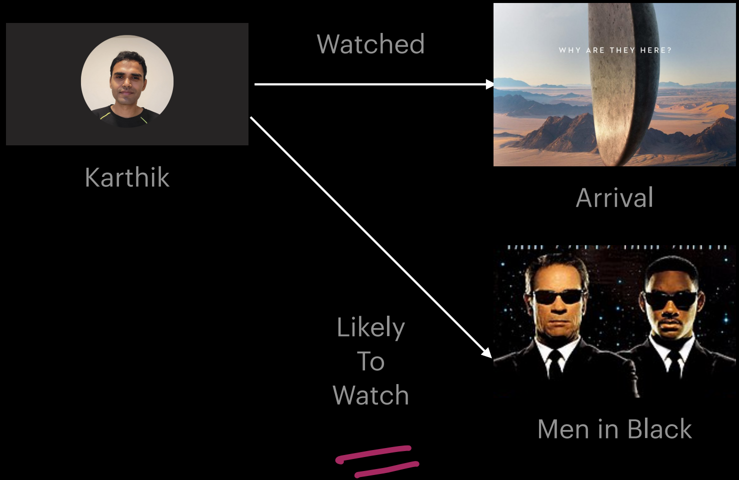
Men in Black

Sci-fi
Aliens

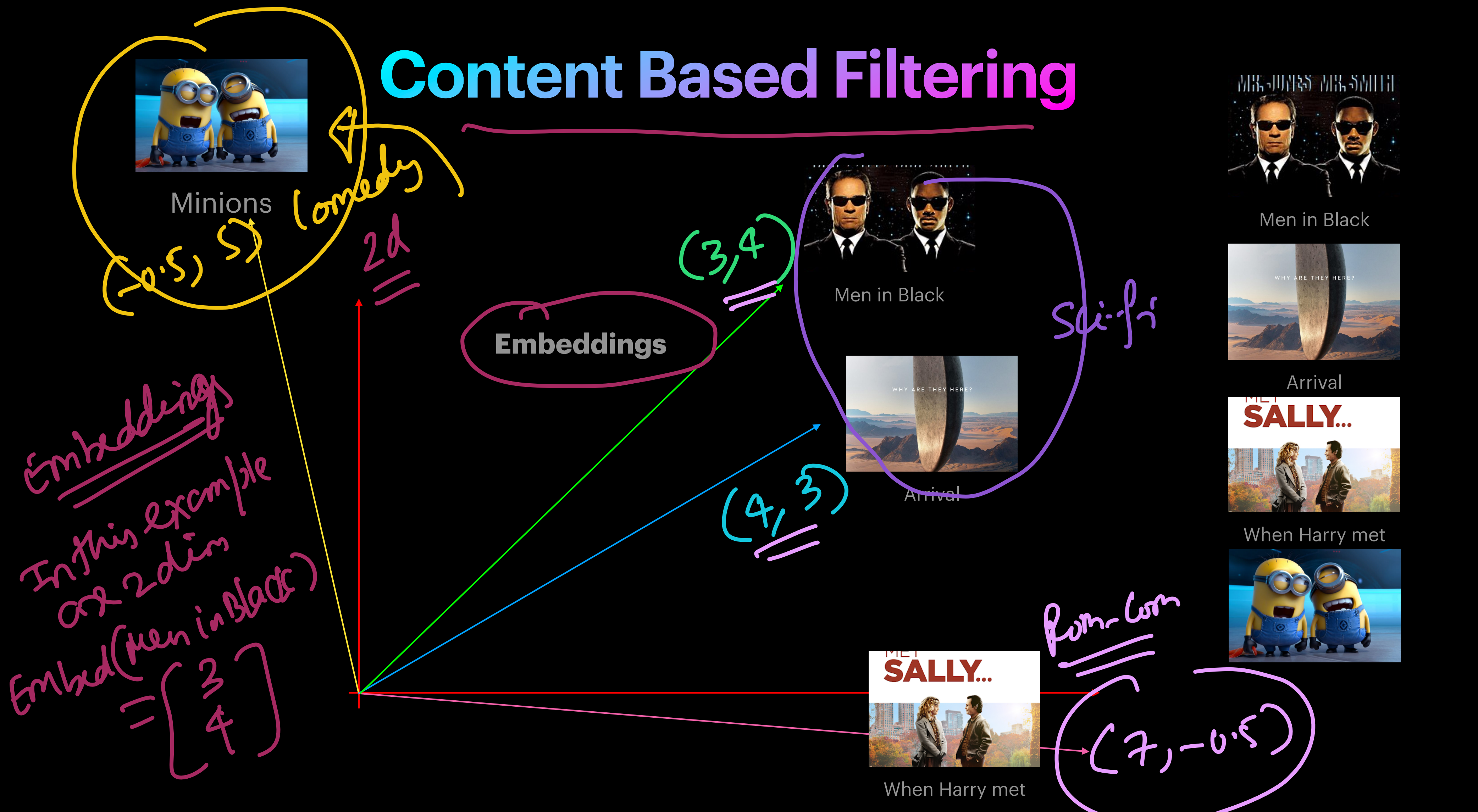
Content Based Filtering

vs Collaborative Filtering

	Avatar	Arrival	When Harry	Before Sunrise	Minions
U1	👍	👍	👎	👎	👍
U2	👎	👎	👍	👍	👍
U3	👍	👍	👎	👎	👍
U4	👎	👎	👍	👍	👍
U5	👍	👍	👍	👍	👎



Content Based Filtering



Content Based Filtering

Embeddings

Men in Black

Arrival

When Harry met

Typically 128 or 256 latent dimensions

(words)

Glove: -

[ ] 50

[ ] 100

MP1

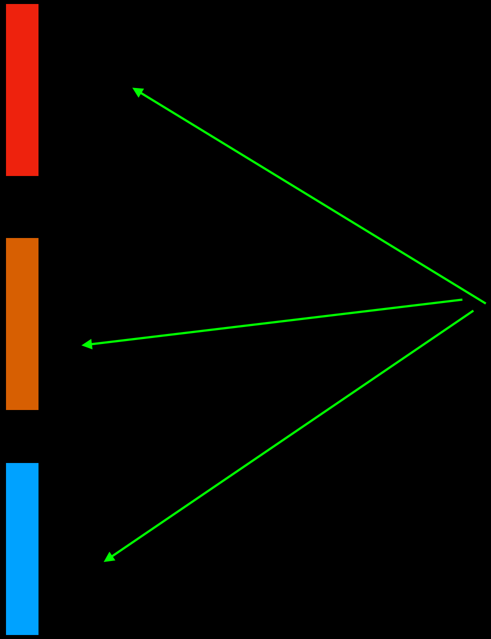
Content Based Filtering

Embeddings

Men in Black

Arrival

When Harry met



How do we
Obtain these embeddings?

Content Based Filtering

Embeddings

Men in Black

Arrival

When Harry met

How do we Obtain these embeddings?

A: Through a DL model!
Maybe last but one hidden layer activations

Take any
LLM → Embedding for a sentence
↳ Movie

Embeddings Interpretation

Embeddings | **Latent Dimensions** → **Interpretation**

Men in Black

Arrival

When Harry met

? → Comedy + Romance
? → Thriller + Adventure
? → Comedy + Thriller
? → Adventure
⋮

Embeddings | Vector Representations

Embeddings

Latent Dimensions → Interpretation

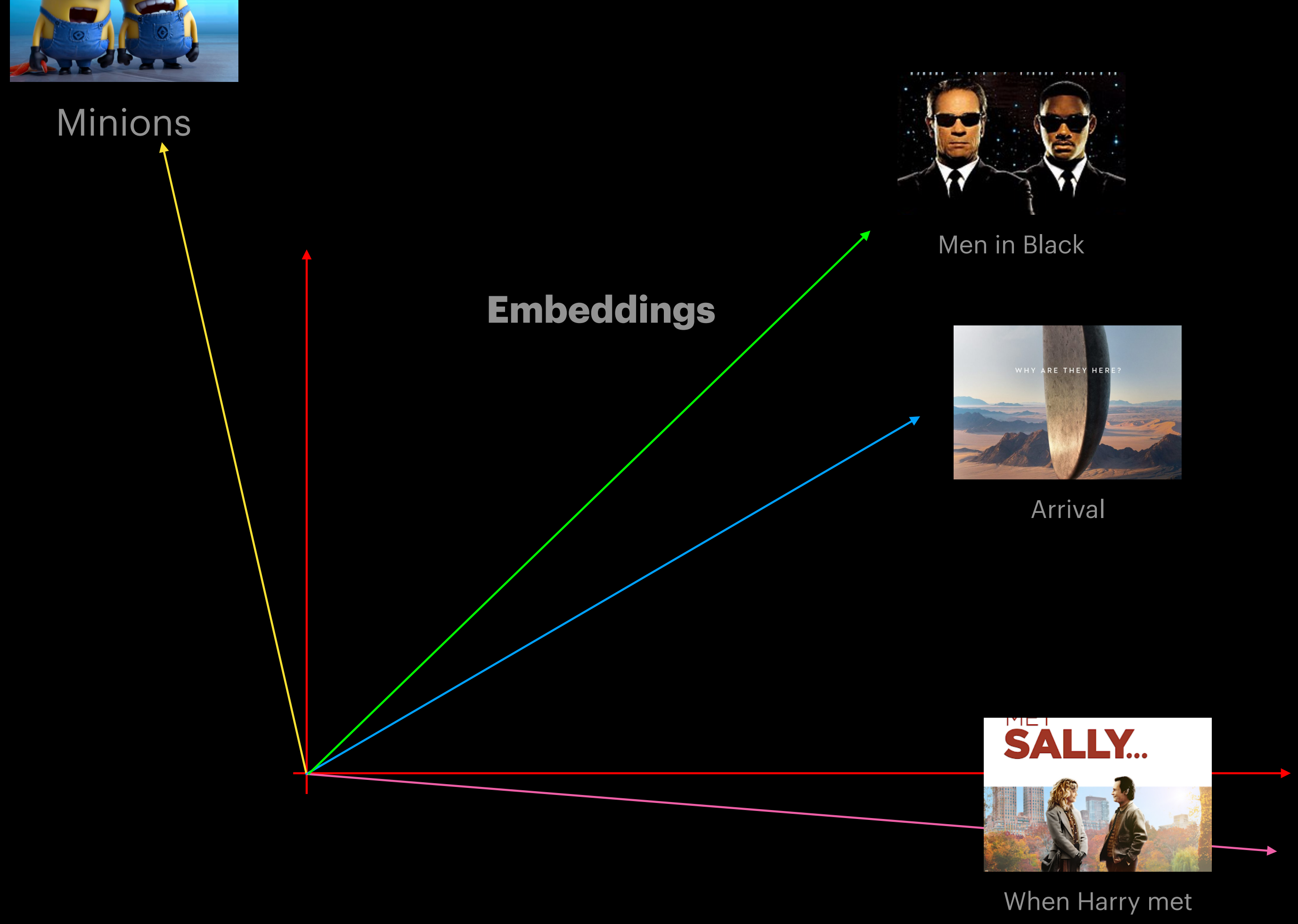
Men in Black

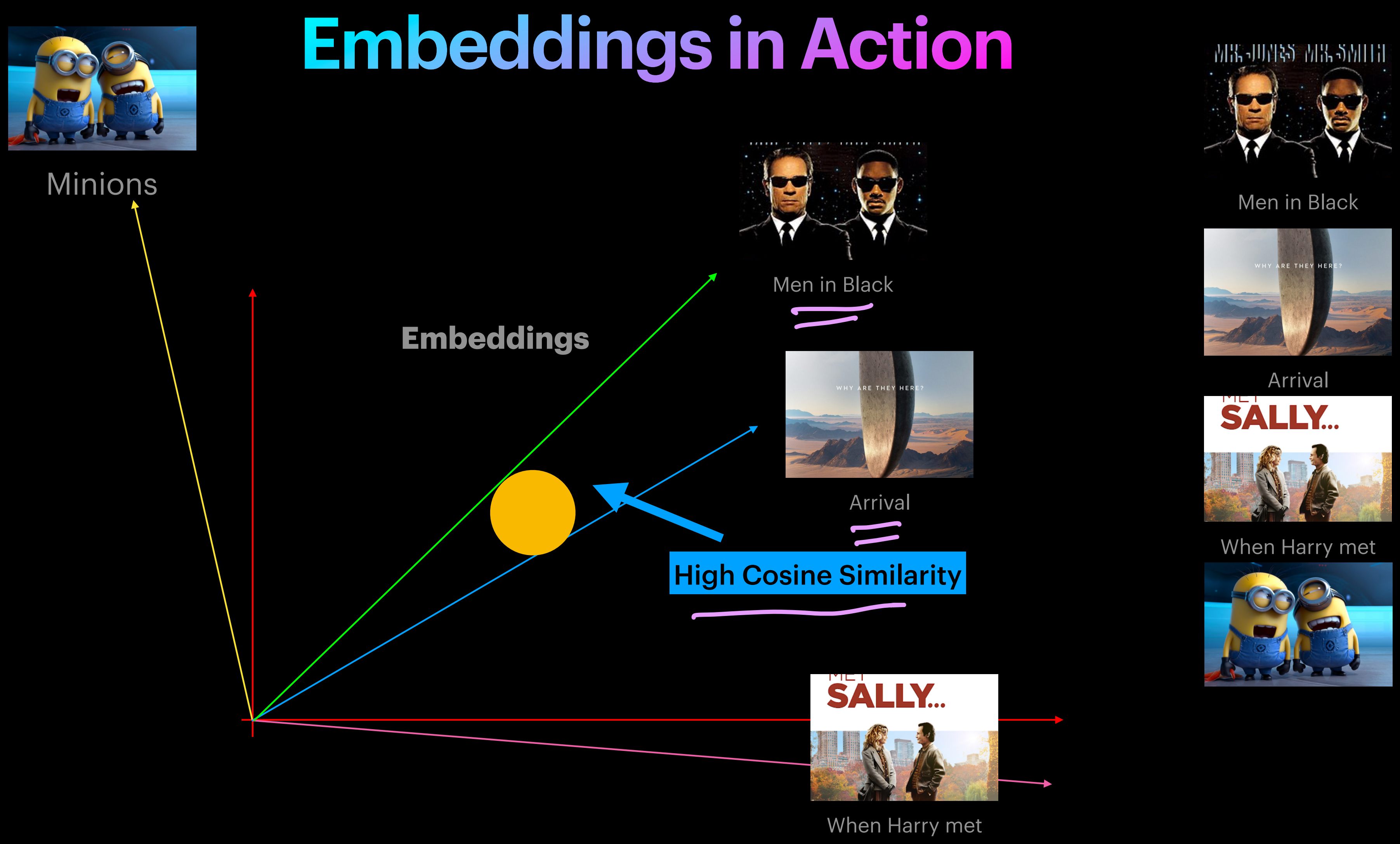
Arrival

When Harry met

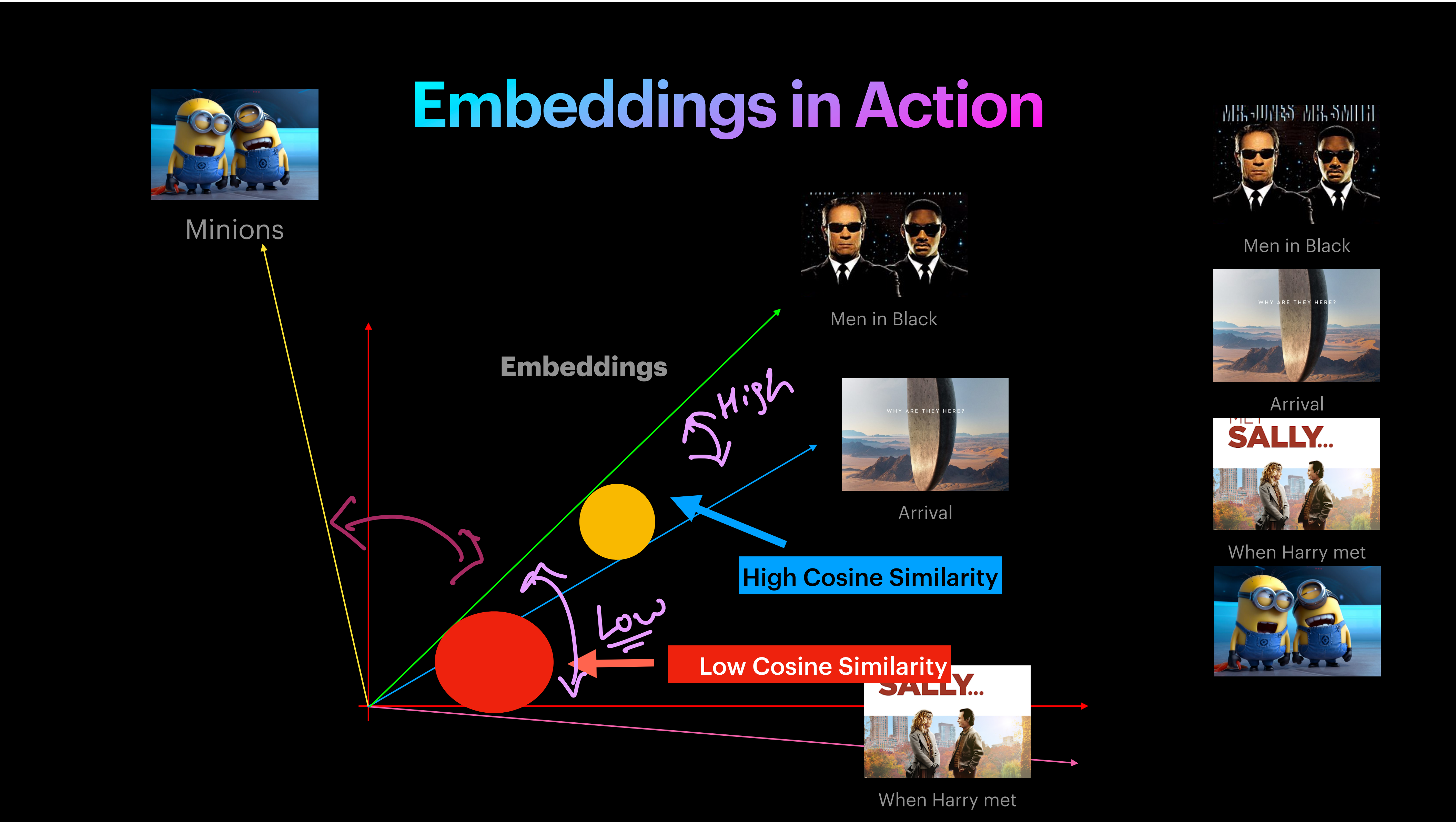
? → Comedy + Romance
? → Thriller + Adventure
? → Comedy + Thriller
? → Adventure
⋮
⋮

Embeddings in Action



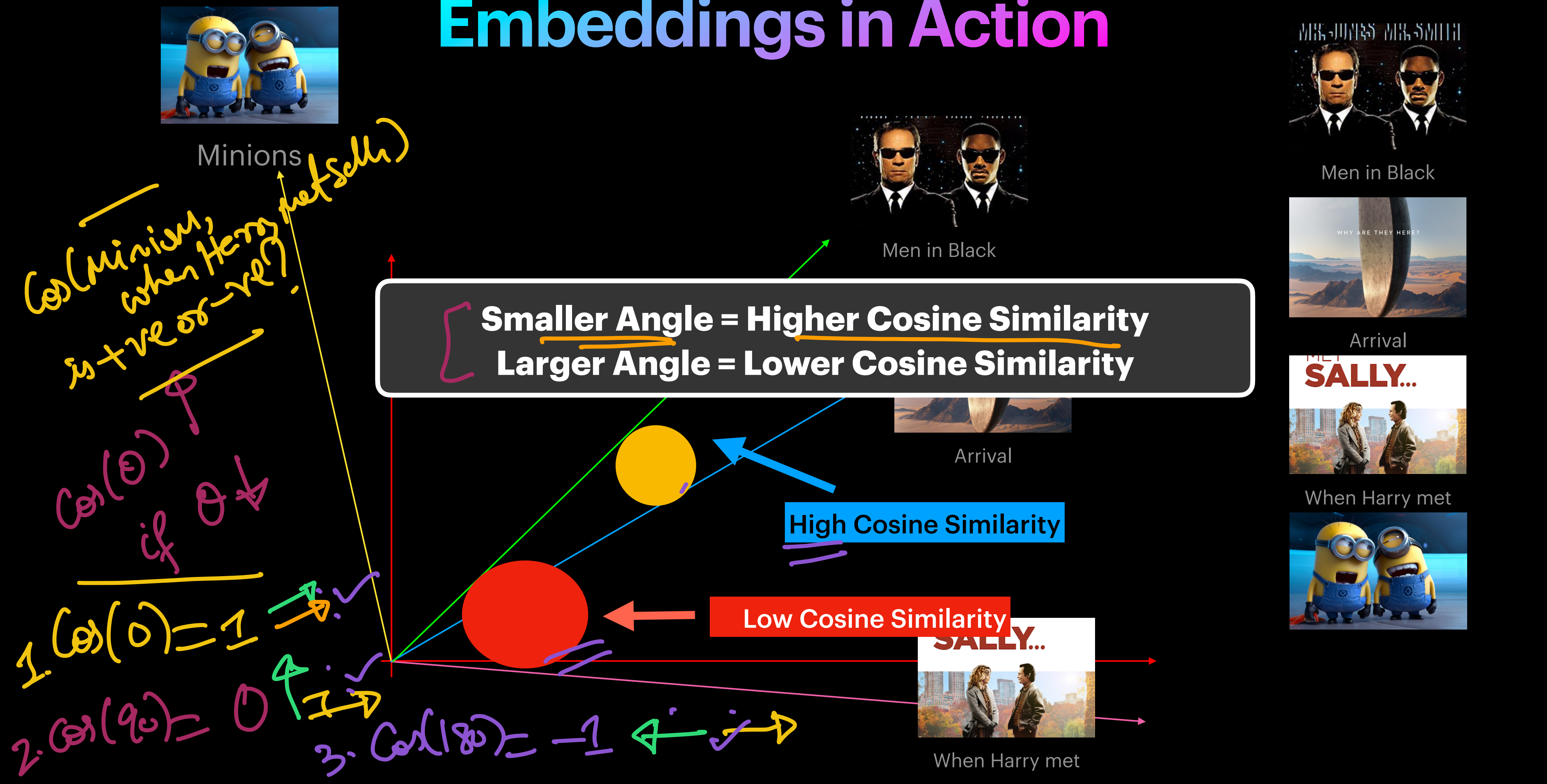


Embeddings in Action
Minions
Embeddings
Men in Black
WHY ARE THEY HERE?
Arrival
High Cosine Similarity
SALLY...
When Harry met
Men in Black
WHY ARE THEY HERE?
Arrival
SALLY...
When Harry met



Embeddings in Action
Minions
Embeddings
Men in Black
High
Arrival
High Cosine Similarity
Low
Low Cosine Similarity
SALLY...
When Harry met
MR. JONES MR. SMITH
Men in Black
WHY ARE THEY HERE?
Arrival
SALLY...
When Harry met

Embeddings in Action



Embeddings in Action

Minions

Men in Black

Men in Black

Arrival

When Harry met

Embeddings

Smaller Angle = Higher Cosine Similarity
Larger Angle = Lower Cosine Similarity

High Cosine Similarity

Low Cosine Similarity

$$\mathrm{Emb}(\mathrm{MIB}) = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\mathrm{Emb}(\mathrm{Arrival}) = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$\mathrm{Cos}(\mathrm{MIB}, \mathrm{Arrival}) = \frac{\begin{bmatrix} 3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 3 \end{bmatrix}}{\|\cdot\|_2 \, \|\cdot\|_2}$$

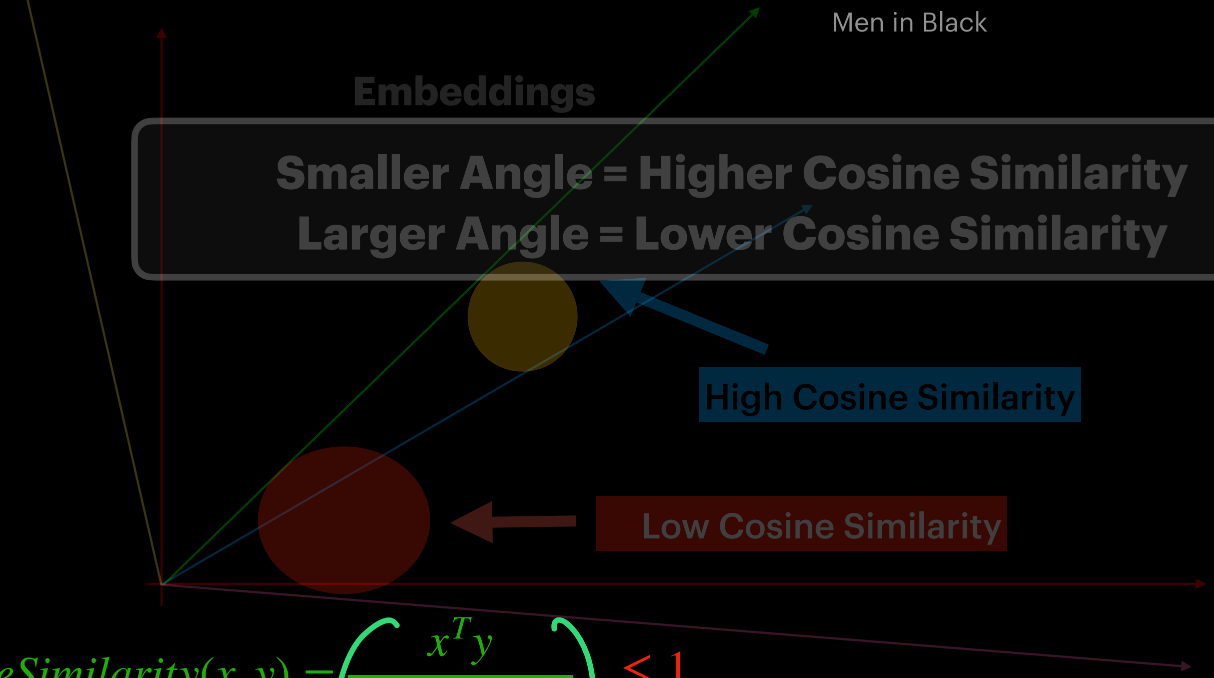
$$= \frac{24 \; (3\cdot4 + 4\cdot3)}{\sqrt{3^2+4^2}\,\sqrt{4^2+3^2}} = \frac{24}{5\cdot5} = \frac{24}{25} = 0.9$$

$$? \leq \quad CosineSimilarity(x,y) = \frac{x^T y}{||x||\,||y||} \quad \leq ?$$

Embeddings in Action

Minions

Men in Black



Men in Black

Arrival

When Harry met

$$-1 \leq CosineSimilarity(x, y) = \left(\frac{x^T y}{||x||\,||y||}\right) \leq 1$$

Cauchy-Schwarz Inequality!

Most Dissimilar / opposite

$$-1 \leq CosineSimilarity(x, y) = \frac{x^T y}{||x||\,||y||} \leq 1$$

Most similar

0 → orthogonal (Apples & oranges)

Cauchy-Schwarz Inequality!

$$-1 \leq \; CosineSimilarity(x, y) = \frac{x^T y}{||x||\,||y||} \leq 1$$

$$|x^T y| <= ||x||\,||y||$$

Cauchy-Schwarz Inequality!

Semantic search

- search for similar objects based on semantics/content

$$-1 \leq CosineSimilarity(x, y) = \frac{x^T y}{||x||\,||y||} \leq 1$$

VIMP metric to use embeddings to find similar movies/objects

$$|x^T y| <= ||x||\,||y||$$

$$||x||_2$$

Euclidean Norm of x

Semantic search

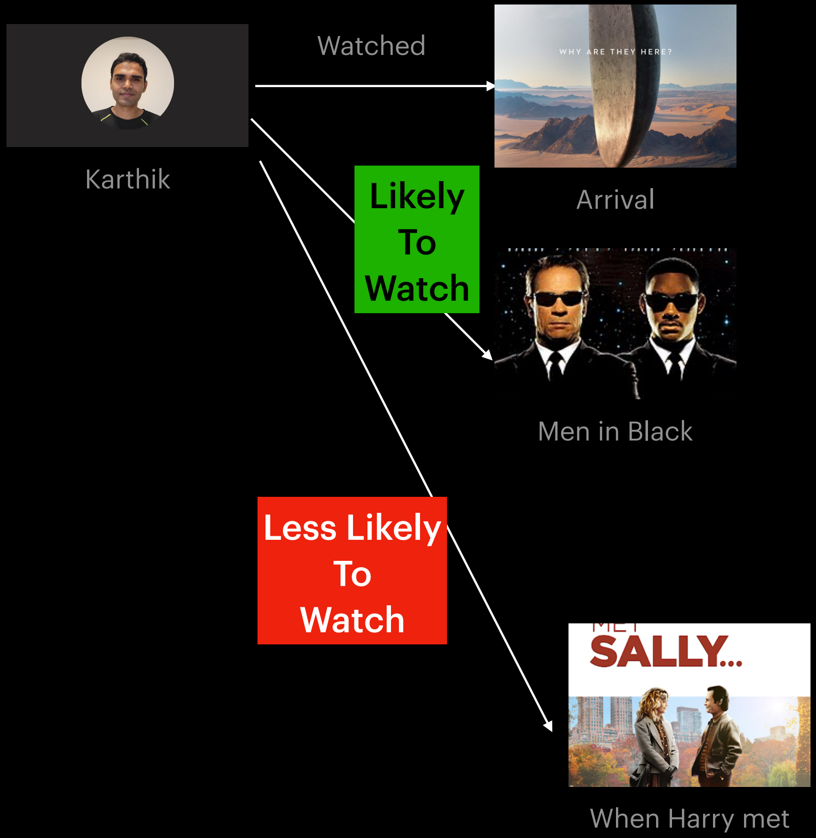
Ref word:- Mountain []

Cosine Sim

Candidate words:- valley, Bike, Road, class

[]

Embeddings in Action



Ref. movie

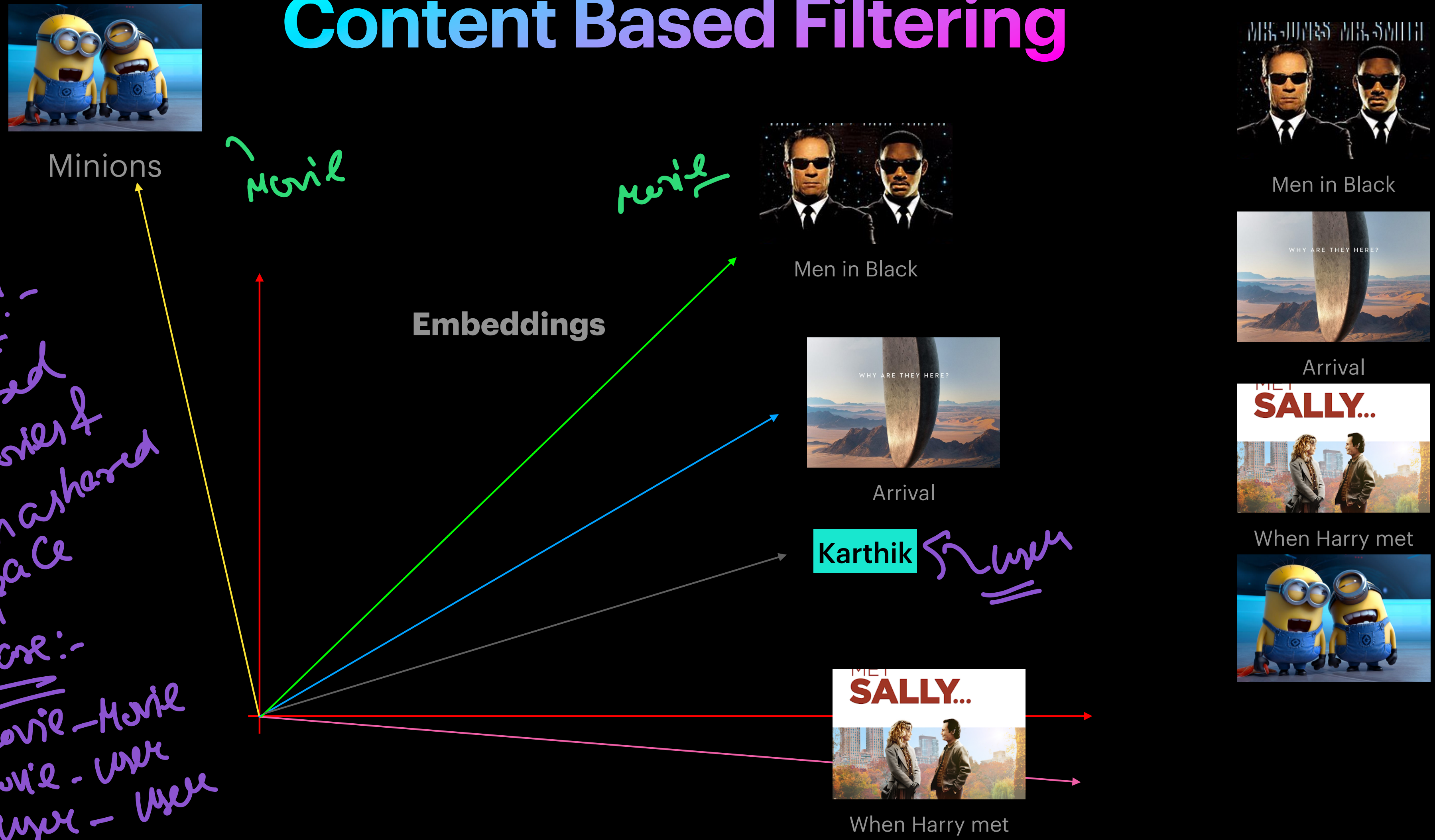
Arrival [ ]

Candidates

[ ] MIB, minions, ...

What if I like both sci-fi and romance?

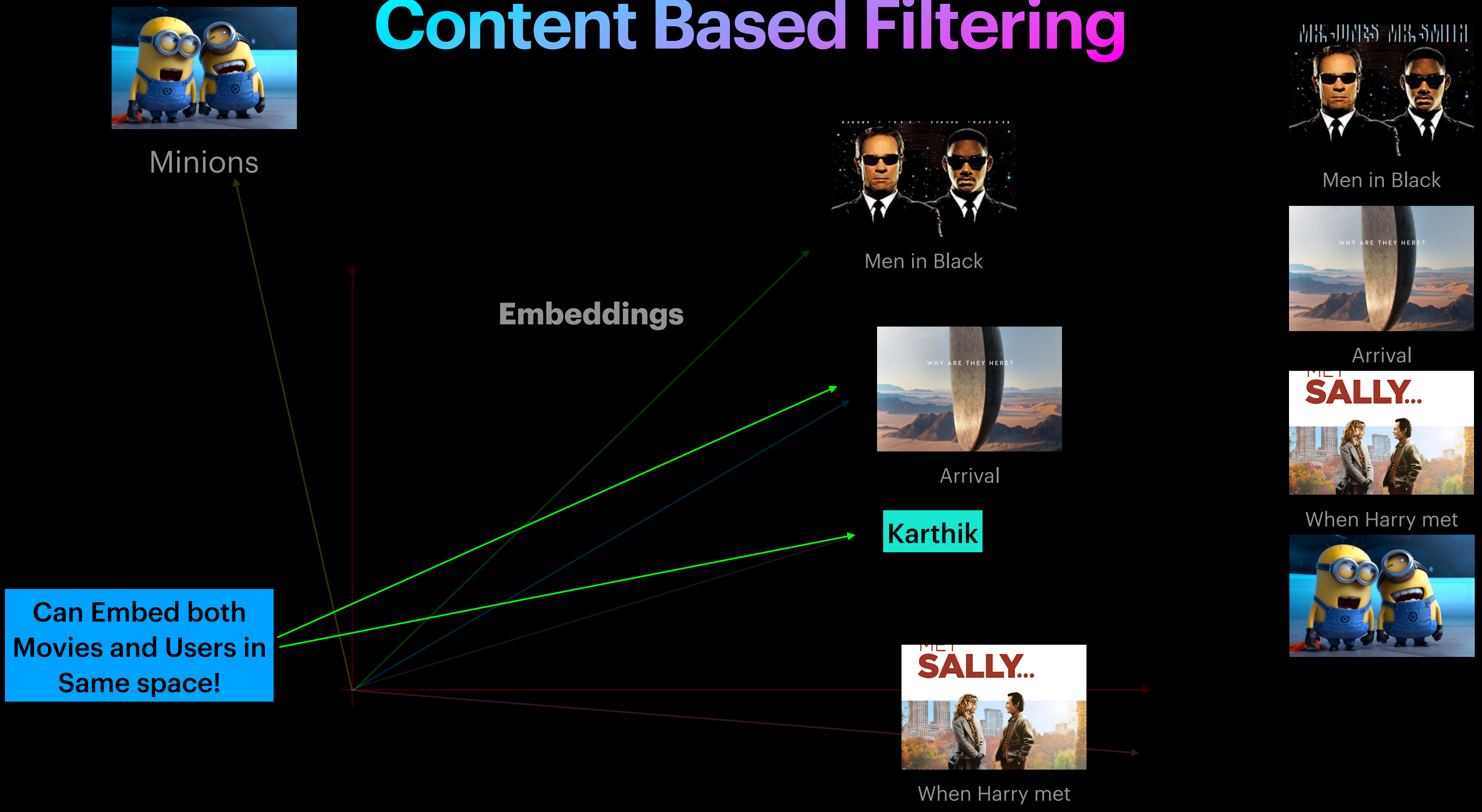
Content Based Filtering



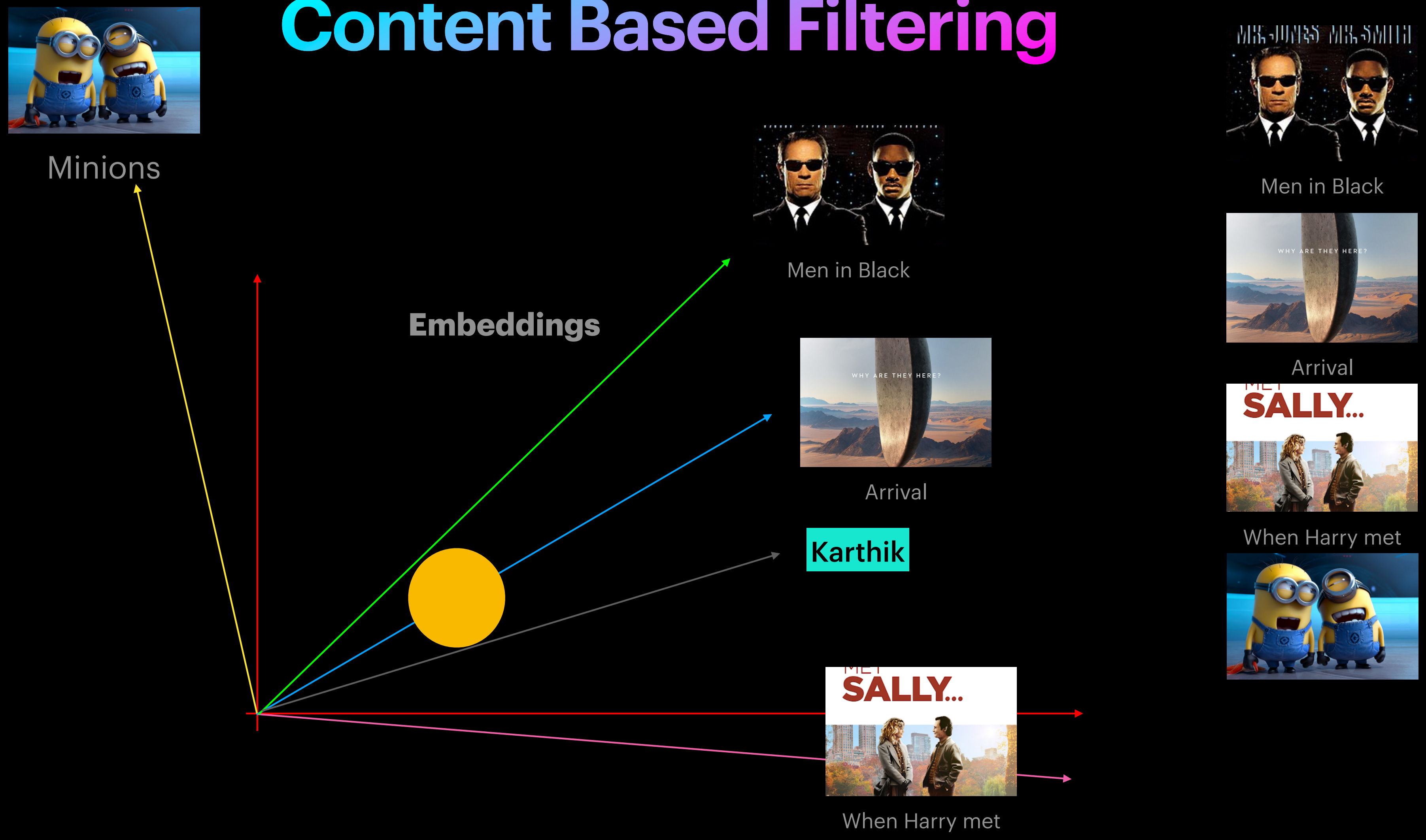
Idea:-
Can embed both movies & users in a shared space

Compare:-
Movie–Movie
Movie - User
User – User

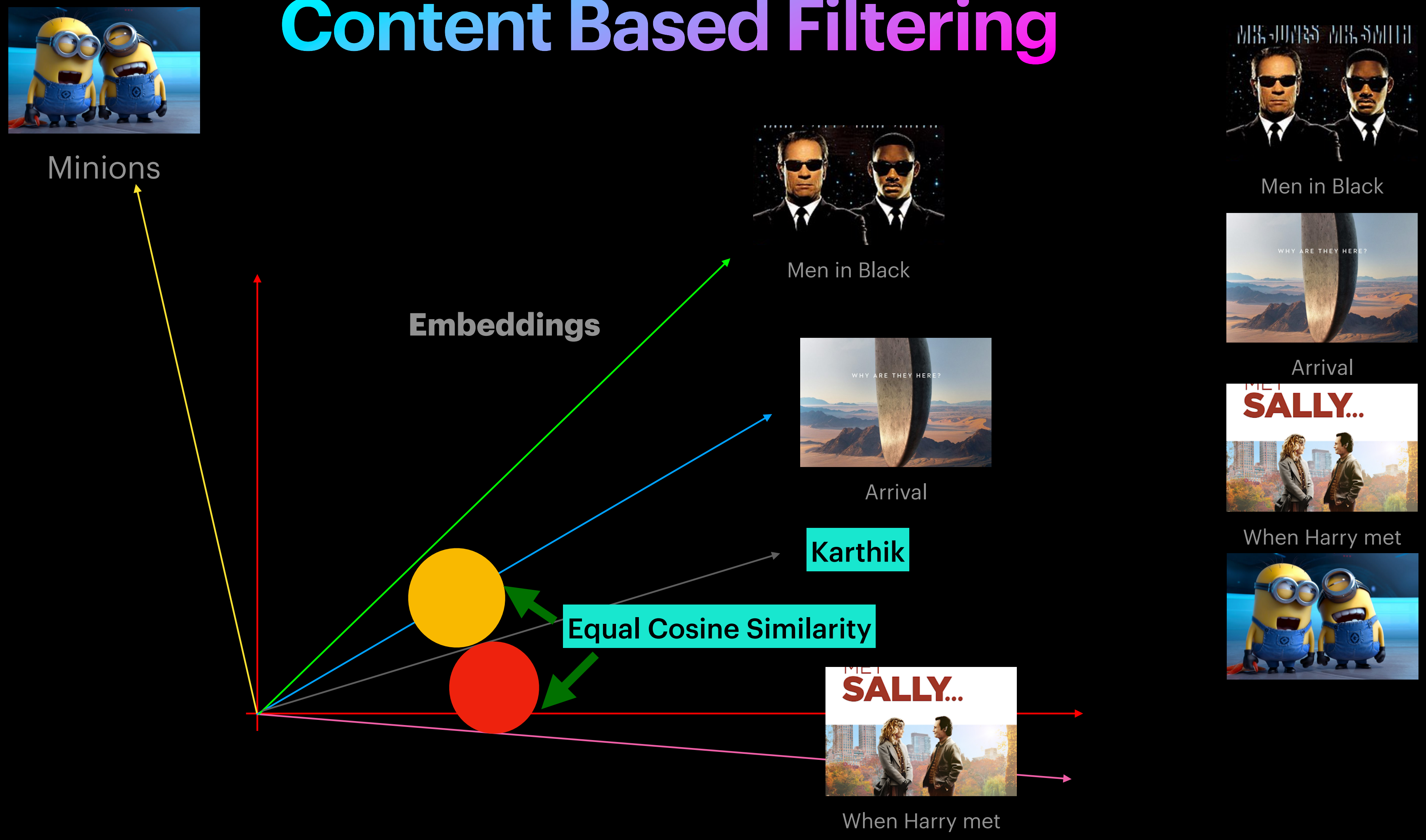
Content Based Filtering



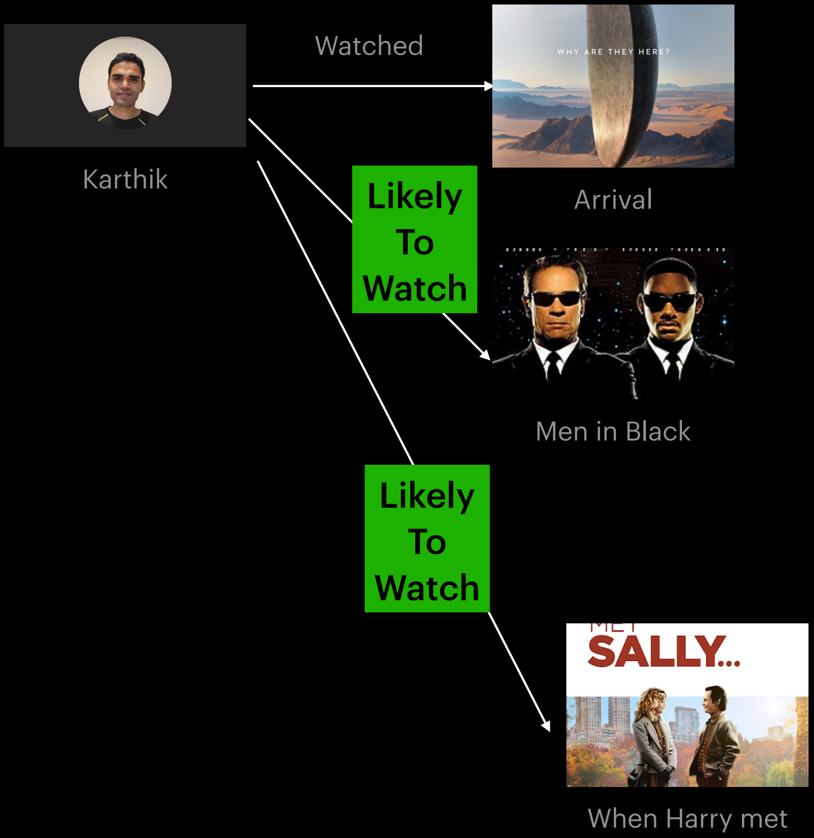
Content Based Filtering



Content Based Filtering



Like both Sci-fi and Romance



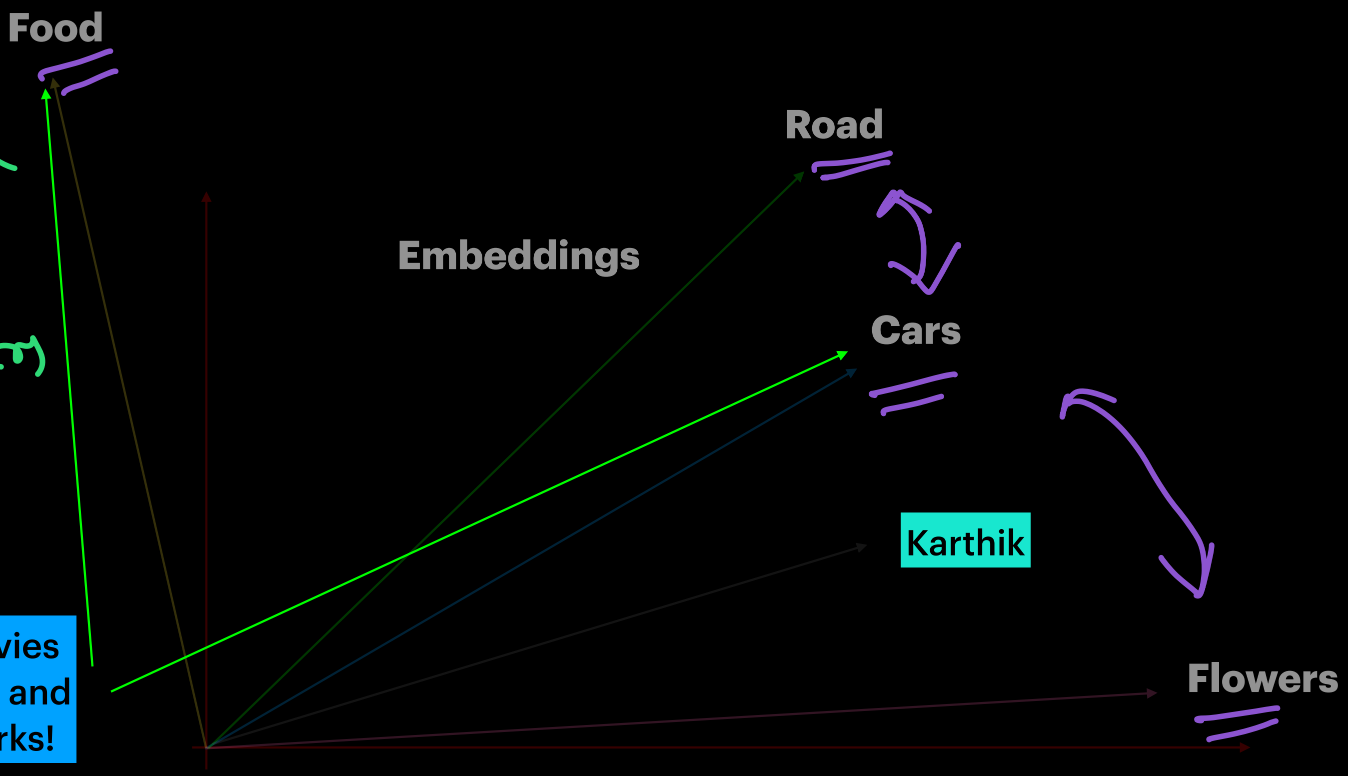
Word Embeddings

MP1

Glove Embeddings

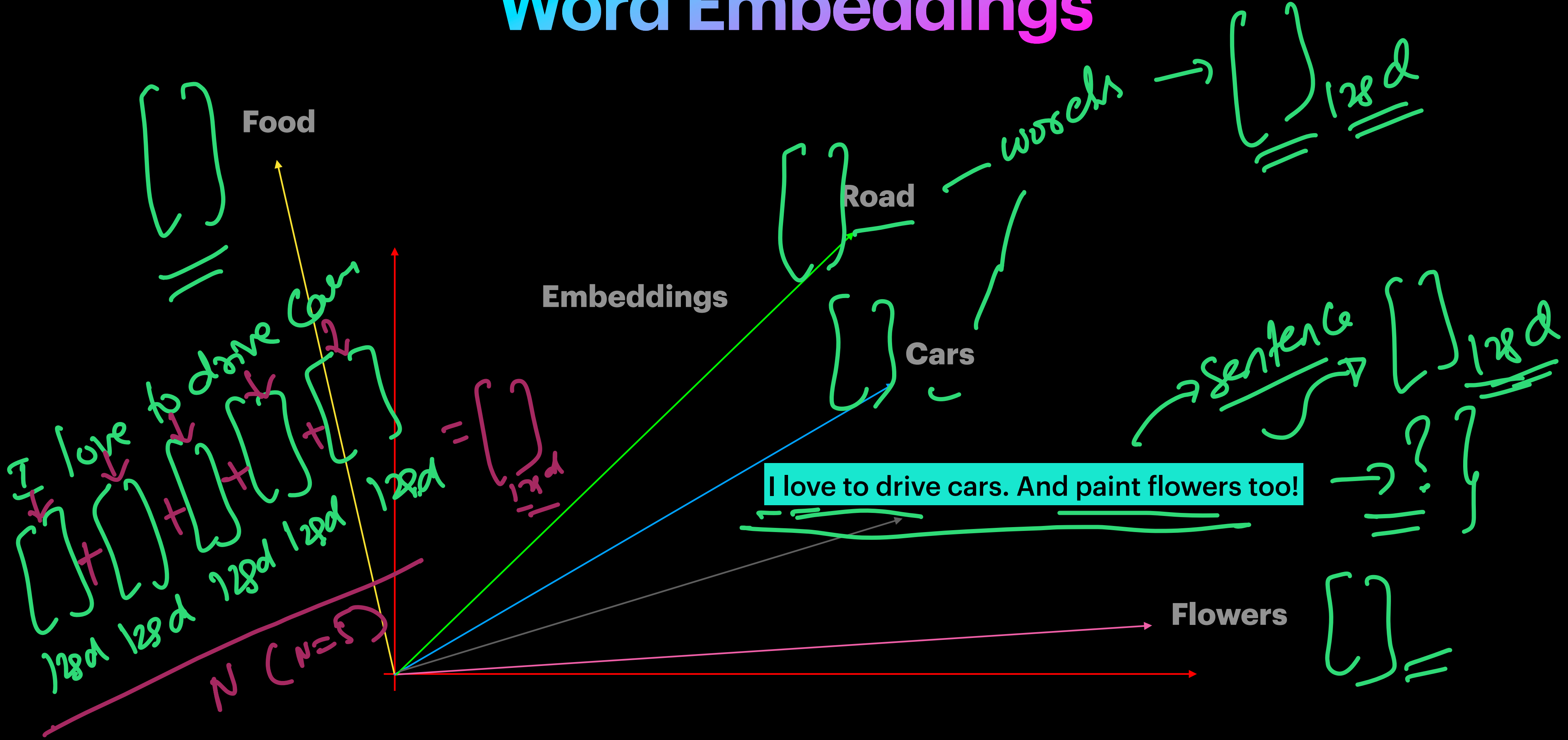
Embed for every word in English dictionary

Movies → Words



Replace movies with "words" and This still works!

Word Embeddings



Example of need for positional context

Ex 1

[ ] + [ ]
Milk chocolate → chocolate

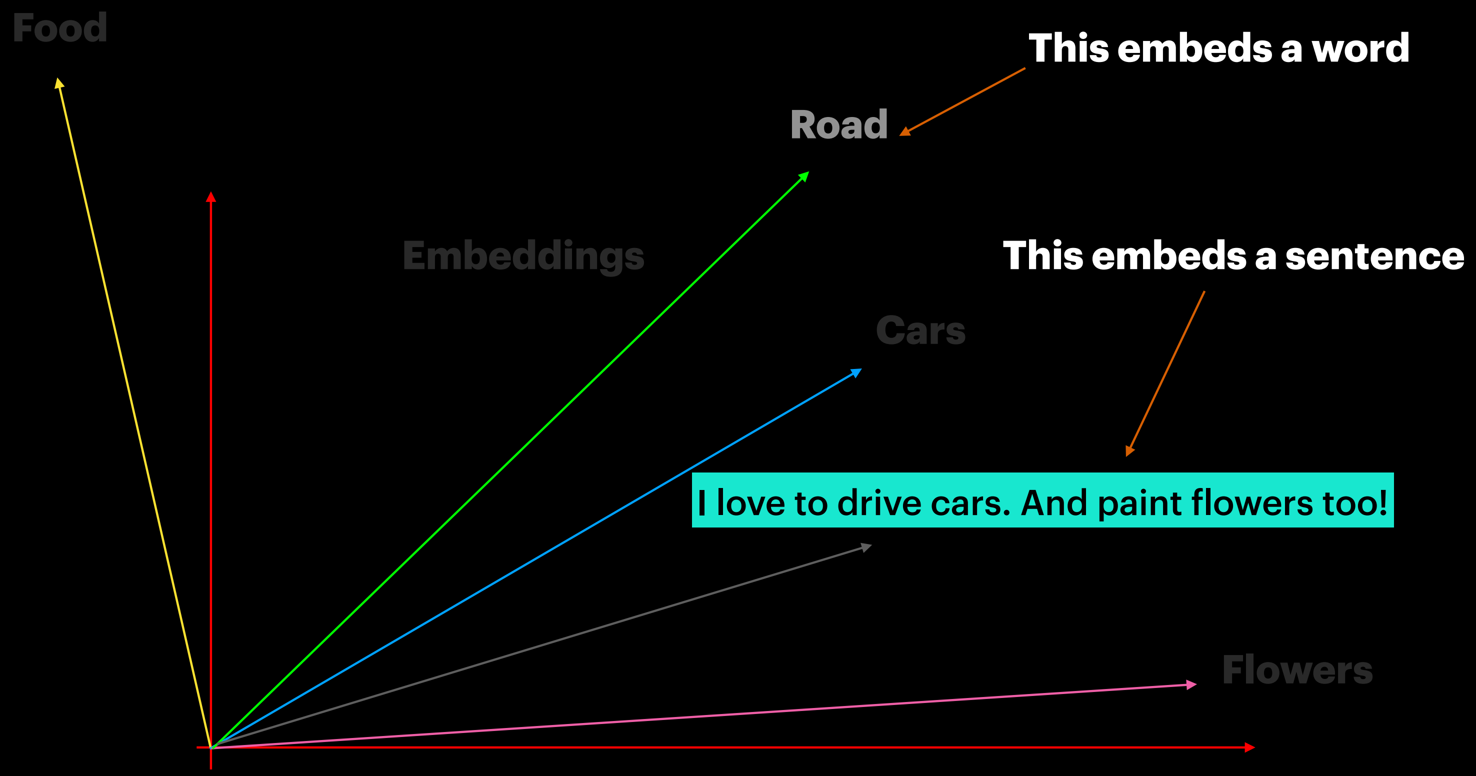
Chocolate milk → milk

[ ] + [ ]

Summary:- Average of word embedding given the same embedding for — milk chocolate & chocolate milk

× Drawback of simple avg!

Word and Sentence Embeddings



Word and Sentence Embeddings

How do we
Obtain these embeddings?

A: Through a DL model!
Maybe last but one hidden layer activations

Word and Sentence Embeddings

How do we
Obtain these embeddings?

A: Through a DL model!
Maybe last but one hidden layer activations

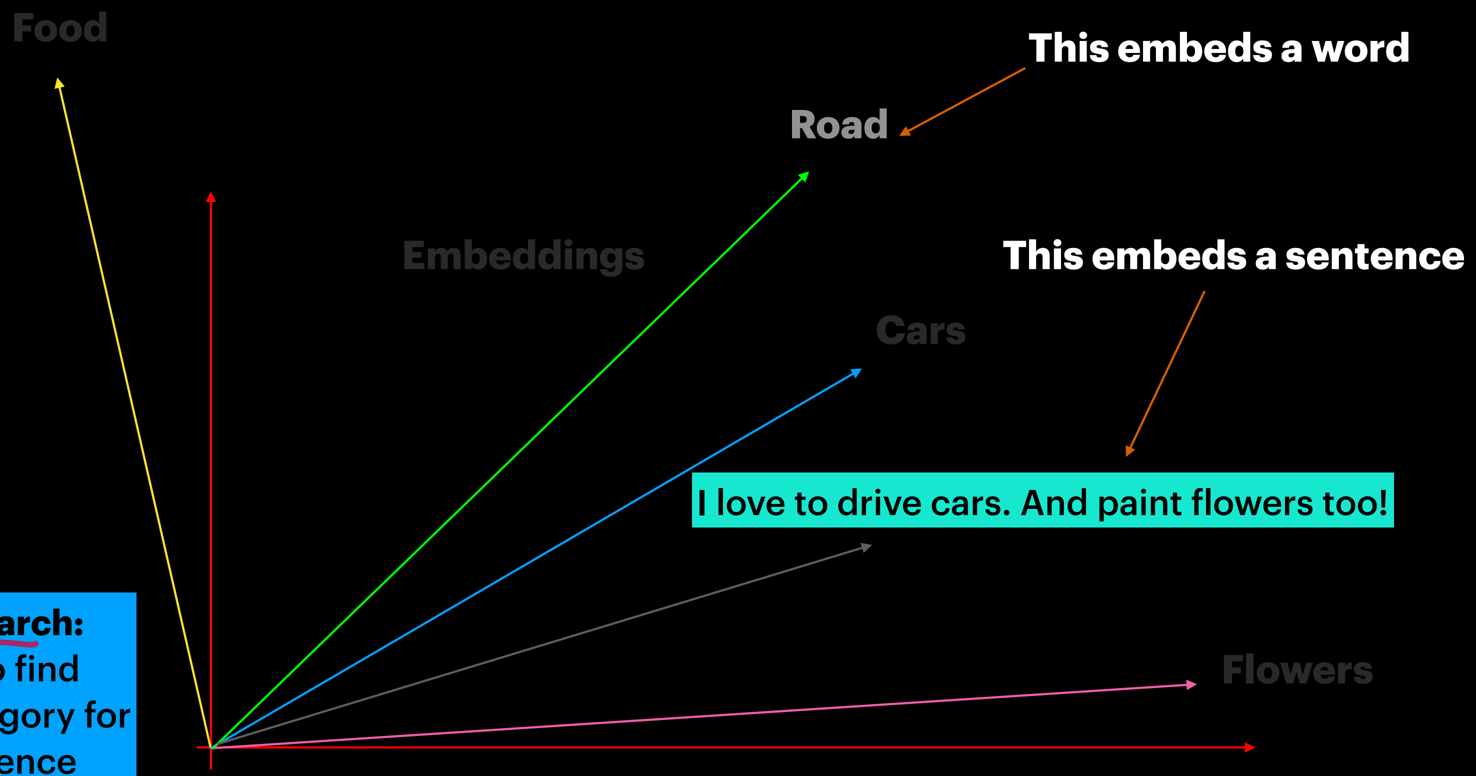
MP1

Glove Embeddings

Sentence
BERT

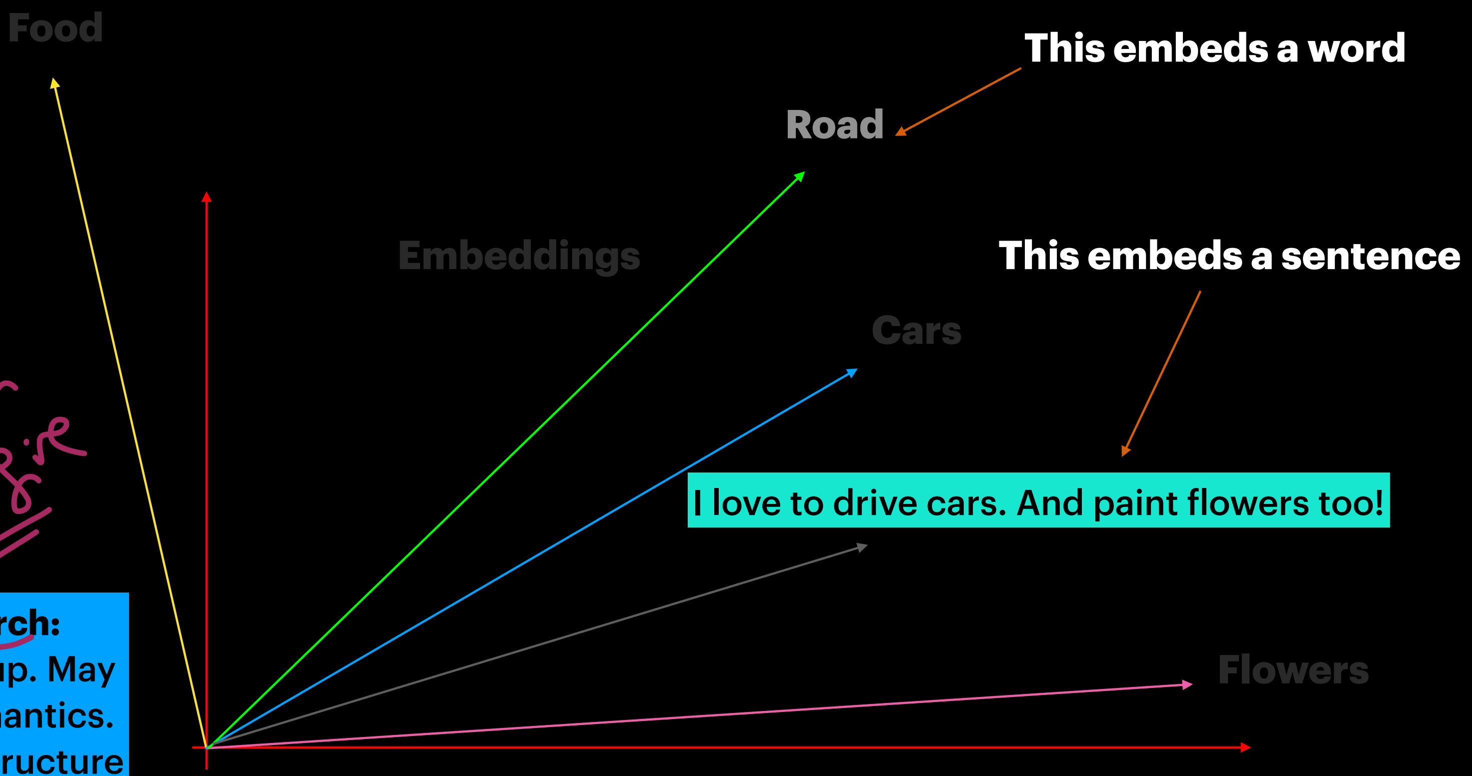
We will cover Sentence BERT when we
Get to Transformers!

Semantic Search | Vector Search



Semantic Search:
Enables us to find
the closest category for
A given sentence

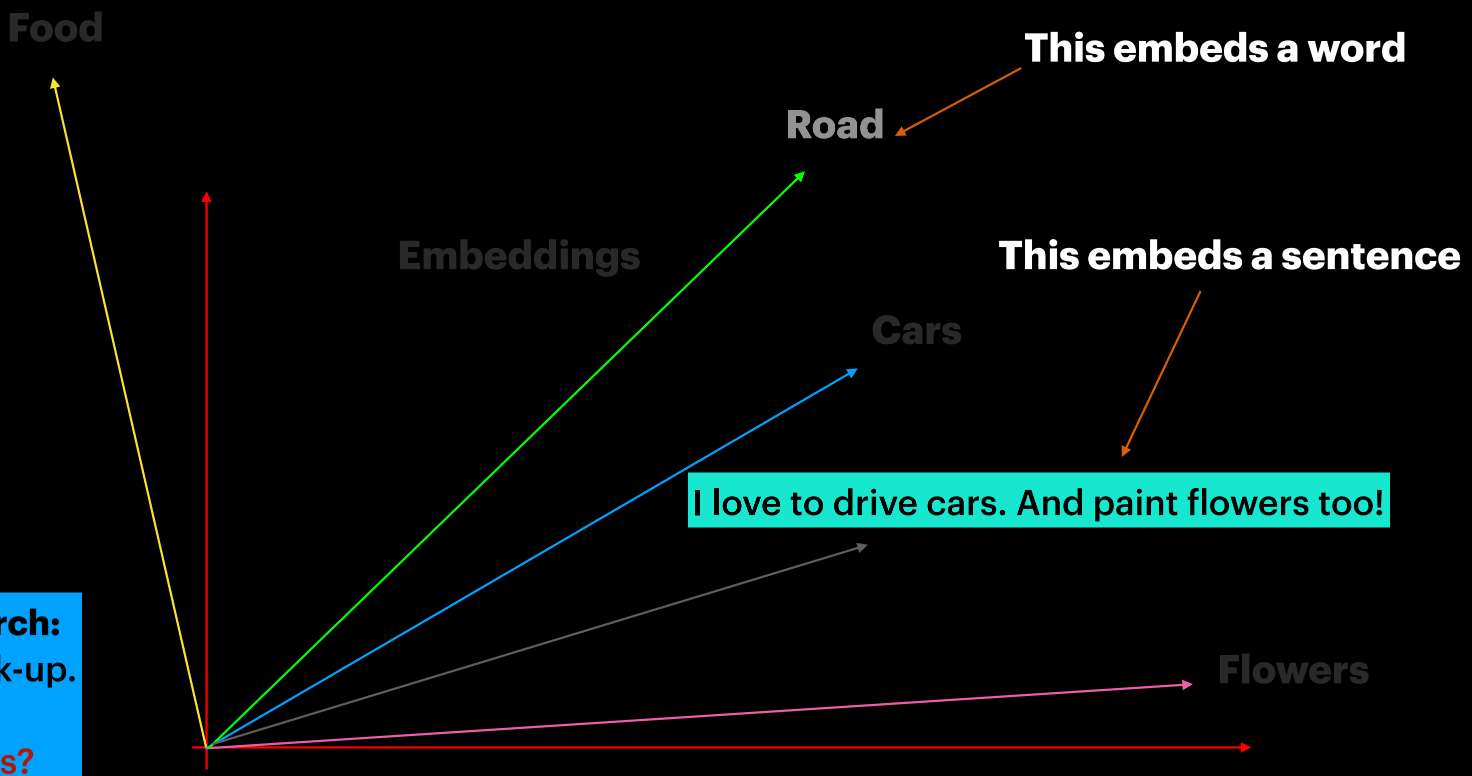
Semantic Search | Vector Search



search engine

Typical Search:
Based on look-up. May
Not handle semantics.
Uses Trie Datastructure

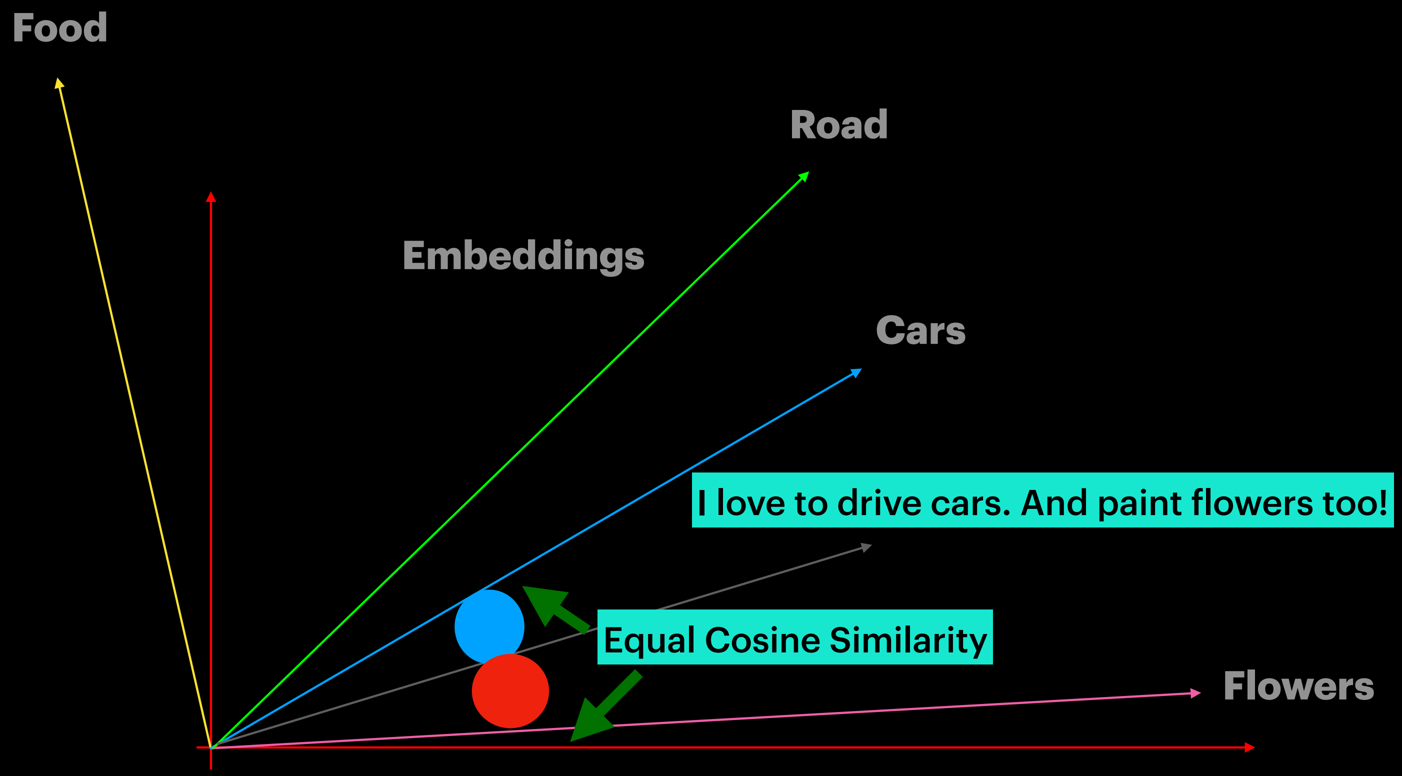
Semantic Search | Vector Search



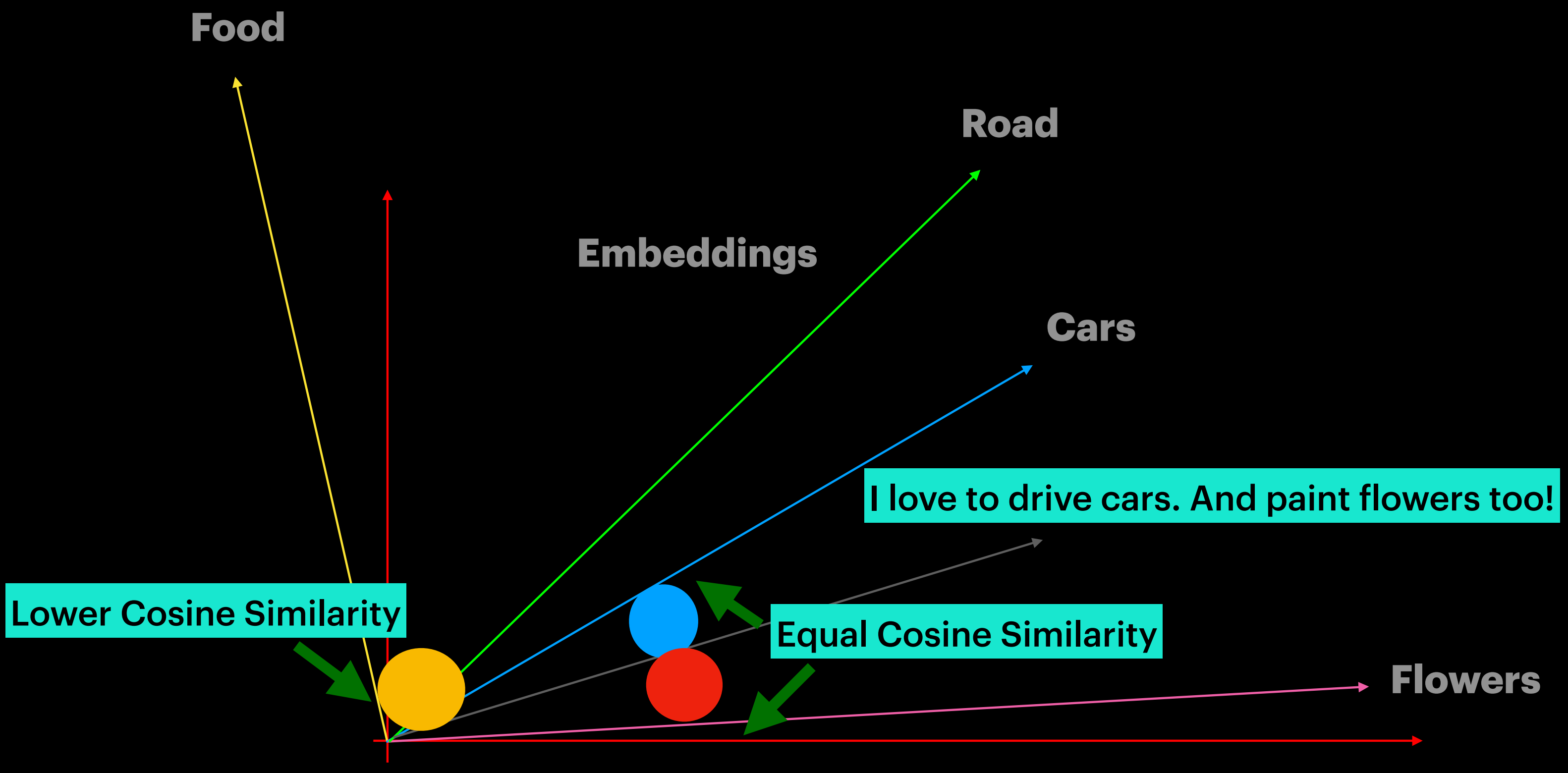
Typical Search:
Based on look-up.

Down-sides?

Semantic Search | Vector Search

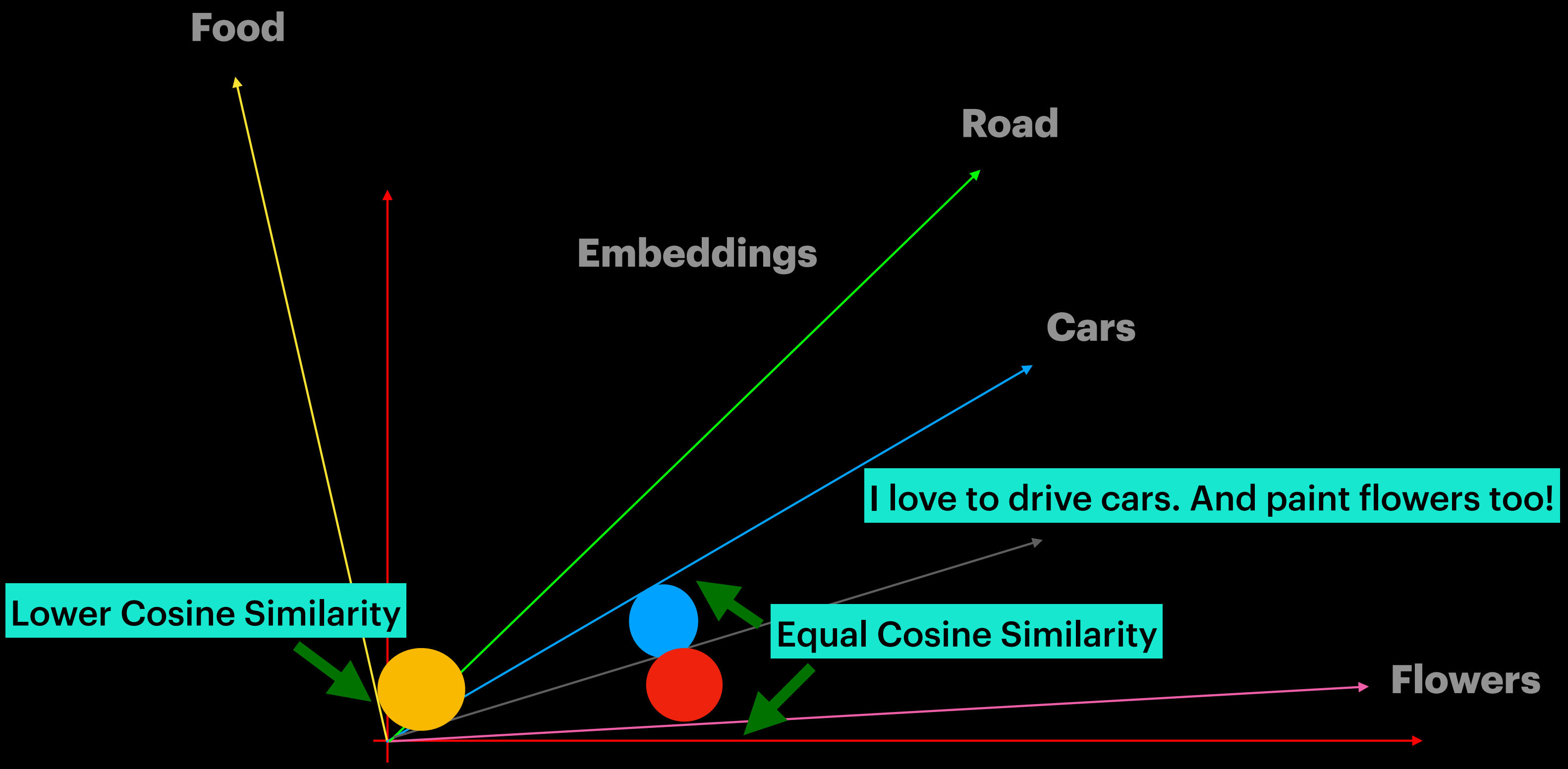


Semantic Search | Vector Search



What’s the closest category
for the following sentence? “I love to drive cars. And paint flowers too!"

Semantic Search | Vector Search



What's the closest category
for the following sentence? "I love to drive cars. And paint flowers too!"

Highest cosine similarity based on vector search: Flowers and Cars

Vector Arithmetic!

What is King - Man + Woman?

$$[\ ] - [\ ] + [\ ] = [\ ]$$

128d 128d 128d

result

= Queen

semantic see

man [ ]

King [ ]

Queen [ ]

Girl [ ]

Demo on Semantic Search