

A NEAREST NEIGHBOR APPROACH TO MAKING RECOMMENDATIONS

Lecture 8

by Marina Barsky

A nearest neighbor approach to making recommendations

Knowing that lots of people liked something is not enough. *Who* liked it is important.

[Underworld: Awakening 3D](#)

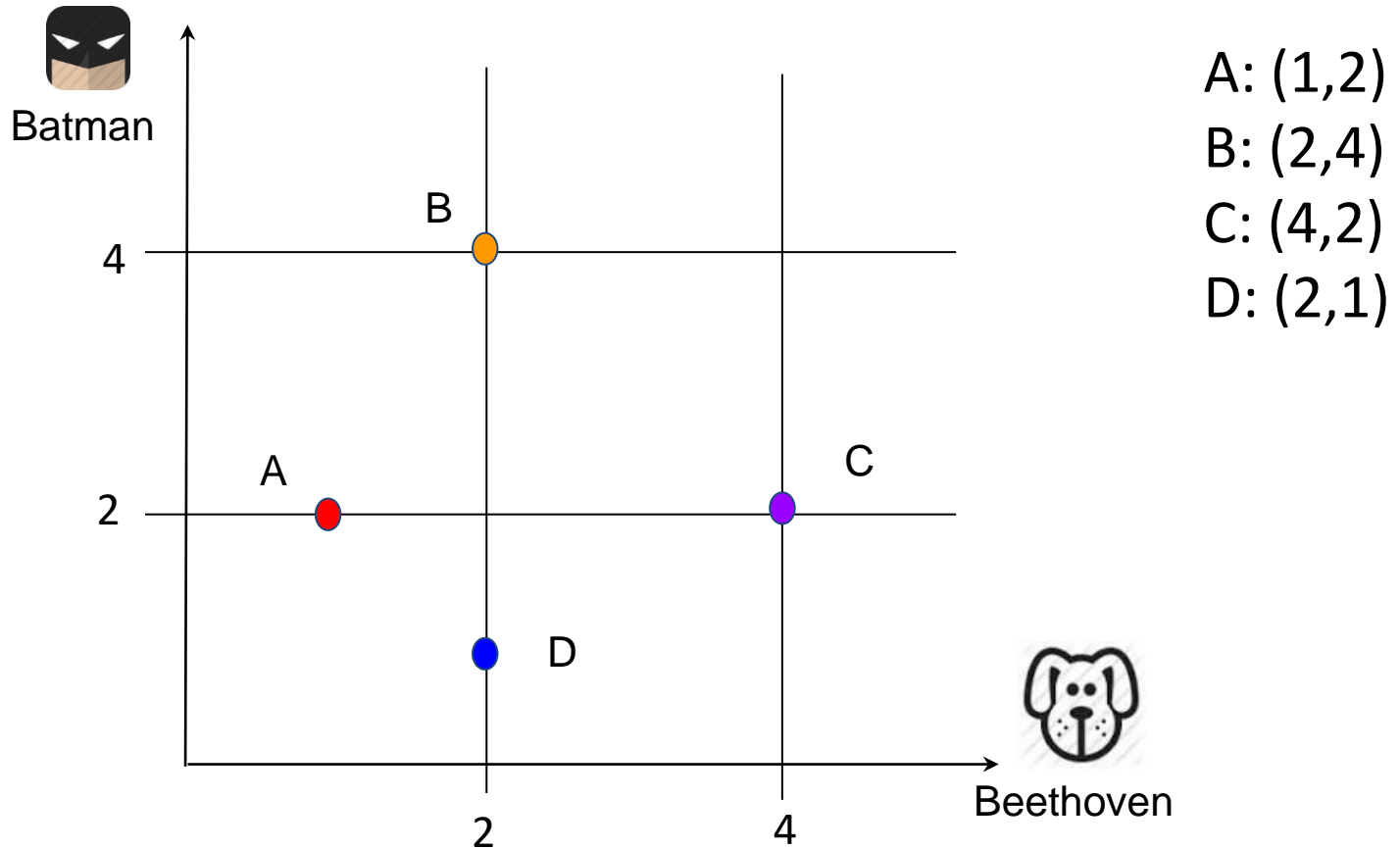


[Reviews](#) - [Trailer](#) - [IMDb](#)

1hr 28min - Rated 18-A - Action/Adventure/Scifi/Fantasy/Horror - English
Director: Mans Marlind - Cast: Kate Beckinsale, Scott Speedman, India Eisley, Charles Dance, Michael Ealy - :

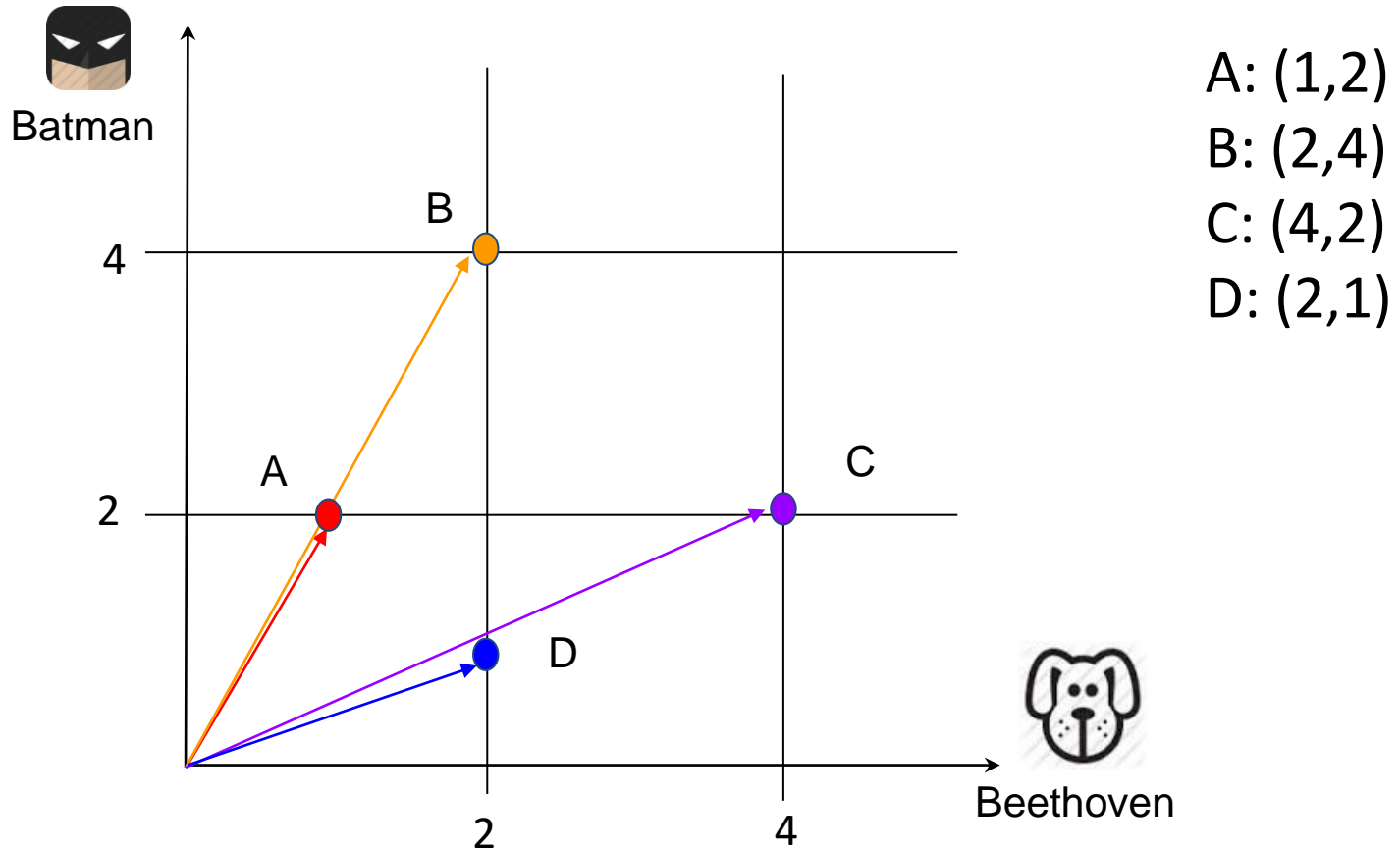
Selene escapes imprisonment to find herself in a world where humans have discovered the existence of both Vampire and Lycan clans, and are conducting an all-out war to eradicate both immortal species.

Coordinate-based similarity: Euclidean distance



Movie rankings

Vector-based similarity: cosine of an angle between vectors



Movie rankings

K-NN prediction

Item for which prediction is sought

	i_1	i_2	...	i_j	i_m
u_1							
u_2				5			
...							
u_a				?			
...							
...				5			
u_n							

Active user

K-NN prediction

Item for which prediction is sought

	i_1	i_2	...	i_j	i_m
u_1							
u_2				5			
...							
u_a				?			
...							
...				5			
u_n							

Active user

K-NN



predicted rank (i_j, u_a)

K-NN recommender (collaborative filtering)

Recommended items

	i_1	i_2	...	i_j	i_m
u_1							
u_2		4			5		
...							
u_a		-			-		
...							
...		5			3		
u_n							

Active user

K-NN recommender

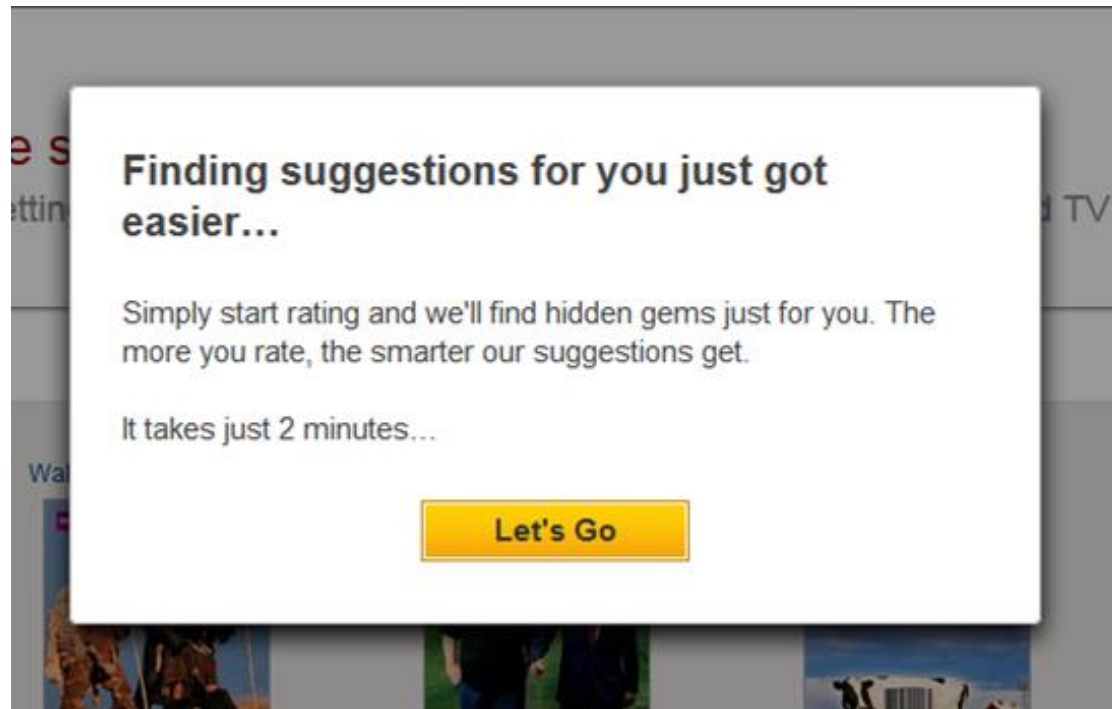


List of top ranked items for u_a

Automated recommender system (collaborative filtering)

- I. Build profile for active user
- II. Compare active profile with the profiles of other customers, locate similar “neighbors”
- III. Find and combine ratings of a peer group selected by similar tastes for the items that active user did not rank yet
- IV. Rank predictions and output top-scored ones

Creating customer profile



Netflix.ca

Example: music recommender

You:

{Lady Gaga, Katy Perry, Justin Bieber, Maroon 5}

II. Finding users with similar tastes

1. April:
{The Beatles, Lady Antebellum}
2. Ben:
{Lady Gaga, Adele, Kelly Clarkson, The Dixie Chicks, Lady Antebellum}
3. Cory:
{Kelly Clarkson, Lady Gaga, Katy Perry, Justin Bieber, Lady Antebellum}
4. Dave:
{The Beatles, Maroon 5, Lady Antebellum}
5. Edgar:
{Adele, Maroon 5, Katy Perry, Bruno Mars}

II. Finding users with similar tastes

	Adele	The Beatles	Justin Bieber	The Dixie Chicks	Kelly Clarkson	Lady Gaga	Lady Antebellum	Maroon 5	Bruno Mars	Katy Perry
April		Yes					Yes			
Ben	Yes			Yes	Yes	Yes	Yes			
Cory			Yes		Yes	Yes	Yes			Yes
Dave		Yes					Yes	Yes		
Edgar	Yes							Yes	Yes	Yes

You

Similarity measures for asymmetric binary data

Simple matching coefficient:

number of matches / total attributes

→ Jaccard index:

number of matches / total not-both-null attributes

II. Finding users with similar tastes (Jaccard index)

	Adele	The Beatles	Justin Bieber	The Dixie Chicks	Kelly Clarkson	Lady Gaga	Lady Antebellum	Maroon 5	Bruno Mars	Katy Perry
April		Yes					Yes			
Ben	Yes			Yes	Yes	Yes	Yes			
Cory			Yes		Yes	Yes	Yes			Yes
Dave		Yes					Yes	Yes		
Edgar	Yes							Yes	Yes	Yes

You

		Yes			Yes		Yes		Yes
--	--	-----	--	--	-----	--	-----	--	-----

You vs April: $0/6$

You vs Ben: $1/8$

You vs. Cory: $3/6$

You vs. Dave: $1/6$

You vs. Edgar: $2/6$



Your peer group:

Cory: similarity 0.50

Dave: similarity 0.17

Edgar: similarity 0.33

III. Combine ratings for new items (weighted voting)

	Adele	The Beatles	Justin Bieber	The Dixie Chicks	Kelly Clarkson	Lady Gaga	Lady Antebellum	Maroon 5	Bruno Mars	Katy Perry
April		Yes					Yes			
Ben	Yes			Yes	Yes	Yes	Yes			
Cory			Yes		Yes	Yes	Yes			Yes
Dave		Yes					Yes	Yes		
Edgar	Yes							Yes	Yes	Yes

You

		Yes			Yes		Yes		Yes	
--	--	-----	--	--	-----	--	-----	--	-----	--

Your peer group:

Cory: similarity 0.50

Dave: similarity 0.17

Edgar: similarity 0.33

Predicted likes for new items:

Adele: 1 like * 0.33 = 0.33

The Beatles: 1 like * 0.17 = 0.17

Kelly Clarkson: 1 like * 0.50 = 0.50

Lady Antebellum: 1 like * 0.50 + 1 like * 0.17 = 0.67

Bruno Mars: 1 like * 0.33 = 0.33

IV. Output top-ranked

	Adele	The Beatles	Justin Bieber	The Dixie Chicks	Kelly Clarkson	Lady Gaga	Lady Antebellum	Maroon 5	Bruno Mars	Katy Perry
April		Yes					Yes			
Ben	Yes			Yes	Yes	Yes	Yes			
Cory			Yes		Yes	Yes	Yes			Yes
Dave		Yes						Yes		
Edgar	Yes							Yes	Yes	Yes

You Yes Yes Yes Yes

Your peer group:

Cory: similarity 0.50

Dave: similarity 0.17

Edgar: similarity 0.33

Top ranked:

Lady Antebellum: 0.67

Kelly Clarkson: 0.50

These are your recommendations !

To improve recommender

- Choice of distance/similarity
- Weighted voting
- Optimal number of neighbors k
- Domain knowledge ...