



# **CIS 635 Knowledge Discovery & Data Mining**

Association Learning



# Association Learning

- Market basket analysis
- Collaborative filtering

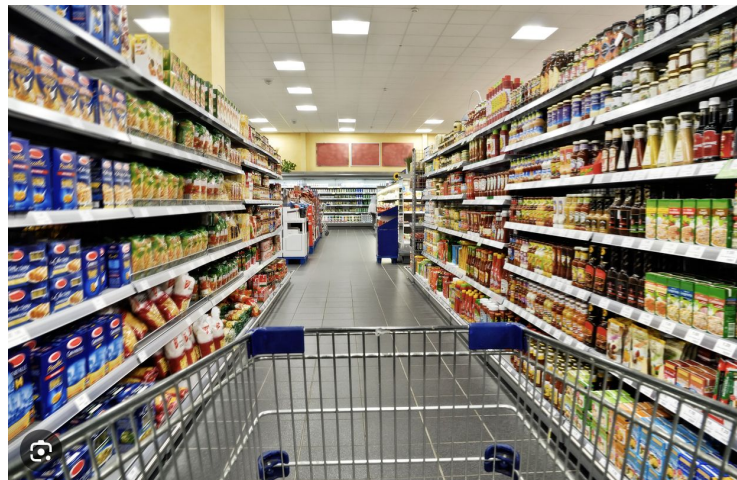


# Association Learning

- Market basket analysis

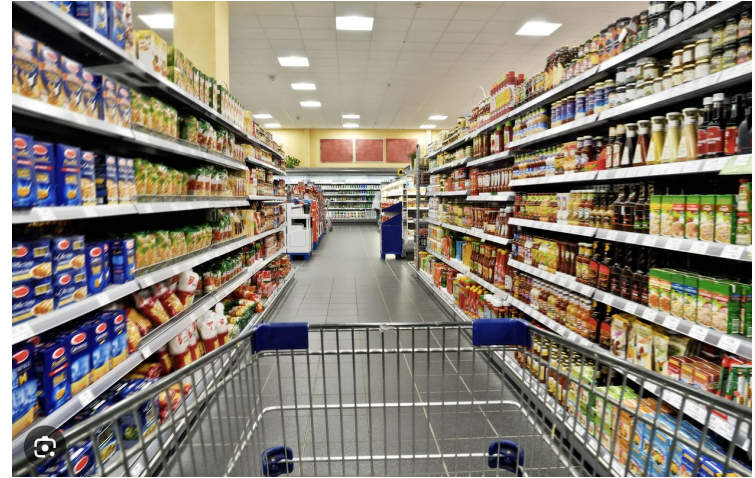
# Introduction

- Supermarket items are displayed not random, they follow some patterns



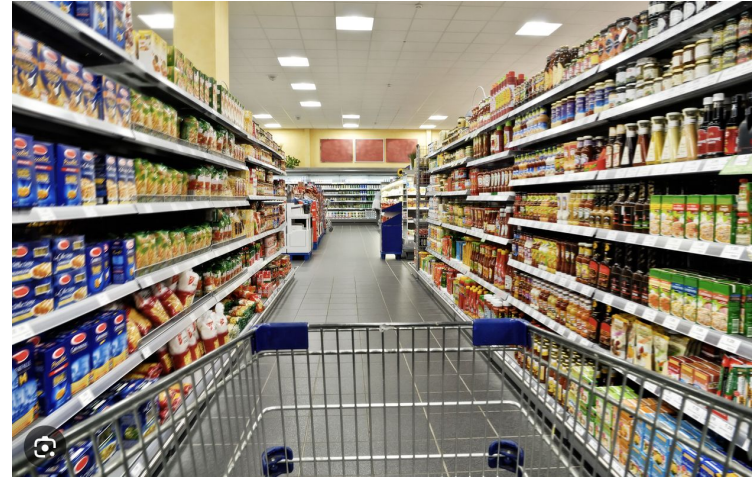
# Introduction

- Supermarket items are displayed not random, they follow some patterns
  - The objective is, clients can access/find items easily



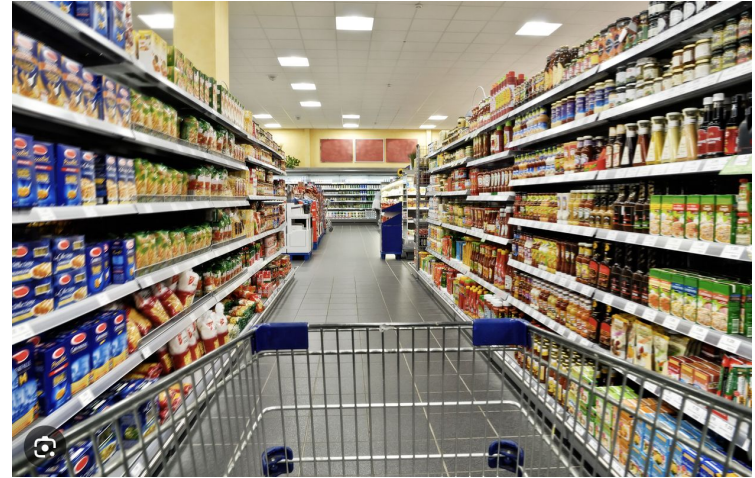
# Introduction

- Supermarket items are displayed not random, they follow some patterns
  - The objective is, clients can access/find items easily
  - **More importantly, the business goal is to increase sales; the more the sales the higher the profit**



# Introduction

- Supermarket items are displayed not random, they follow some patterns
  - The objective is, clients can access/find items easily
  - More importantly, the business goal is to increase sales; the more the sales the higher the profit
- **This is a known concept for years**



# Beer Diaper Story

## Beer and Diapers – the Perfect Couple

👤 Jim Borden   📁 family, humor, technology   ⌚ December 7, 2018   📖 3 Minutes

The topic for my class today we discussed was data mini



The **story goes something like this:**

“One Midwest grocery chain used the data mining capacity of Oracle software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer

2002

[Ref: Article published in 2018](#)



# Beer Diaper Story

## Beer and Diapers – the Perfect Couple

👤 Jim Borden   📁 family, humor, technology   ⌚ December 7, 2018   📖 3 Minutes

The topic for my class today we discussed was data mini



The **story goes something like this:**

“One Midwest grocery chain used the data mining capacity of Oracle software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer

2002

[Ref: Article published in 2018](#)

# Beer Diaper Story

## Beer and Diapers – the Perfect Couple

👤 Jim Borden   📁 family, humor, technology   ⌚ December 7, 2018   📖 3 Minutes

The topic for my class today we discussed was data mini



The **story goes something like this:**

“One Midwest grocery chain used the data mining capacity of Oracle software to analyze local buying patterns. They discovered that when men bought diapers on Thursdays and Saturdays, they also tended to buy beer. Further analysis showed that these shoppers typically did their weekly grocery shopping on Saturdays. On Thursdays, however, they only bought a few items. The retailer concluded that they purchased the beer to have it available for the upcoming weekend. The grocery chain could use this newly discovered information in various ways to increase revenue. For example, they could move the beer

2002

[Ref: Article published in 2018](#)

---

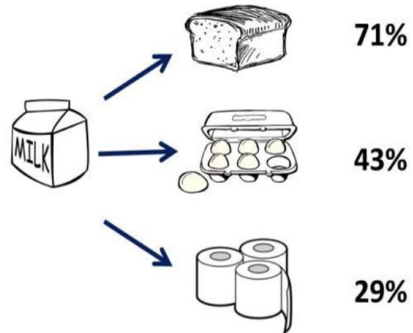
# Market Basket Analysis

- The purchase patterns of clients/users when they buy items together.



# Market Basket Analysis

- The purchase patterns of clients/users when they buy items together.



# Association rule learning

To create the association rules, we use some evaluation metrics:

**Support:** Percentage of that sequence of items appearing in orders

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

market  
basket  
transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

$$\frac{\{\text{Diapers, Beer}\}}{\text{count transactions}} = \frac{3}{5} = 60\%$$



# Association rule learning

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

market  
basket  
transactions

**{Diapers, Beer}** Example of a frequent itemset

**{Diapers} → {Beer}** Example of an association rule

To create the association rules, we use some evaluation metrics:

**Support:** Percentage of that sequence of items appearing in orders

$$\frac{\{\text{Diapers, Beer}\}}{\text{count transactions}} = \frac{3}{5} = 60\%$$

**Confidence:** It's the percentage of seeing the consequent in a transaction, given that it also contains the antecedent.

$$\frac{\text{Support}\{\text{Diapers, Beer}\}}{\text{Support}\{\text{Diapers}\}} = \frac{(3 \div 5)}{(4 \div 5)} = 75\%$$



# Association rule learning

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

market  
basket  
transactions

**{Diapers, Beer}** Example of a frequent itemset

**{Diapers} → {Beer}** Example of an association rule

To create the association rules, we use some evaluation metrics:

**Support:** Percentage of that sequence of items appearing in orders

$$\frac{\{\text{Diapers, Beer}\}}{\text{count transactions}} = \frac{3}{5} = 60\%$$

**Confidence:** It's the percentage of seeing the consequent in a transaction, given that it also contains the antecedent.

$$\frac{\text{Support}\{\text{Diapers, Beer}\}}{\text{Support}\{\text{Diapers}\}} = \frac{(3 \div 5)}{(4 \div 5)} = 75\%$$

**Lift:** This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is.

$$\frac{\text{Confidence}\{\text{Diapers, Beer}\}}{\text{Support}\{\text{Beer}\}} = \frac{\frac{(3 \div 5)}{(4 \div 5)}}{3 \div 5} = 1.25$$

# Association rule learning

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

market  
basket  
transactions

{Diapers, Beer} Example of a frequent itemset

{Diapers} → {Beer} Example of an association rule

To create the association rules, we use some evaluation metrics:

**Support:** Percentage of that sequence of items appearing in orders

$$\frac{\{\text{Diapers, Beer}\}}{\text{count transactions}} = \frac{3}{5} = 60\%$$

**Confidence:** It's the percentage of seeing the consequent in a transaction, given that it also contains the antecedent.

$$\frac{\text{Support}\{\text{Diapers, Beer}\}}{\text{Support}\{\text{Diapers}\}} = \frac{(3 \div 5)}{(4 \div 5)} = 75\%$$

**Lift:** This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is.

$$\frac{\text{Confidence}\{\text{Diapers, Beer}\}}{\text{Support}\{\text{Beer}\}} = \frac{\frac{(3 \div 5)}{(4 \div 5)}}{3 \div 5} = 1.25$$

**Lift Value:**

*value* = 1 implies no association between items.

*value* > 1 means that item Y is likely to be bought if the X is bought.

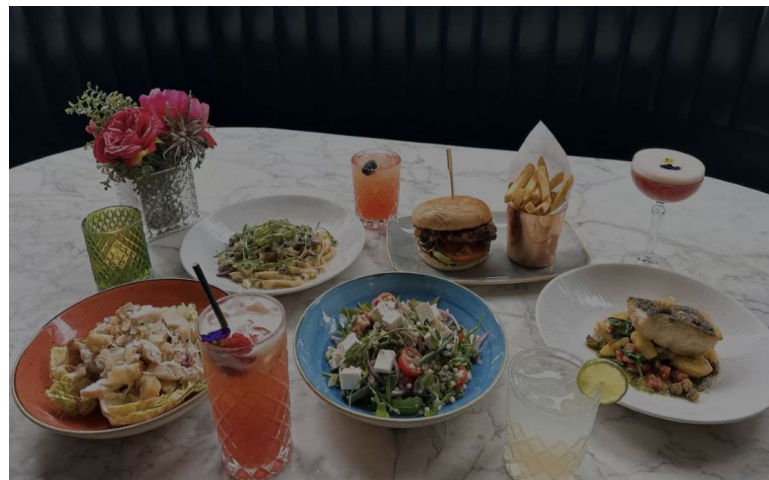
*value* < 1 means that item Y is unlikely to be bought if item X is bought.

Ref: Notebook presentation



## Similar cases

- The purchase patterns of clients/users when they buy items together
- **Restaurant orders.**



# Notebook presentation!

- Market basket analysis of two restaurants menus (what items people ordered together)
- Comparison of their business in terms of their menus offering, etc!



**YOUR LOGO**

**★ Sandwiches ★**

Signature Sandwich Platter	\$55.99
Classic Sandwich Platter	\$49.99
Classic Sandwich Boxed Lunch	\$7.99
Signature Sandwich Boxed Lunch	\$8.99

**★ Salads & Sides ★**

Thai chicken, summer greens, shiitake mushrooms, cucumbers, red peppers, grape tomatoes, red onions, basil, sesame noodles, to main noodles, sesame seeds, toasted almonds, and your choice of dressing.

Signature Summer Salad	\$45.99
Garden Salad	\$34.99
Southwest Salad	\$45.99
Greek Salad	\$45.99
Salad Boxed Lunch	\$9.99
Caesar Salad	\$34.99
Thai Salad	\$45.99
Chicken Caesar Salad Platter	\$39.99
Caribbean Chicken Salad	\$45.99
Seasonal Fruit Salad	\$39.99
Chef Salad	\$45.99
Quart Chicken Pecan Salad	\$11.99
Quart Crispers Tuna Salad	\$11.99
Quart Potato Salad	\$10.49

**SPECIAL OFFER: 25 DEC - 1 JAN**

**Gallon Unsweet Tea**

**Gallon Sweet Tea**

**★ Soups ★**

Quart Broccoli Cheddar Soup	\$11.99
Quart Tomato Bisque Soup	\$11.99
Vegetarian Vegetable Soup	\$11.99
Quart Loaded Potato Soup	\$11.99
Quart Butternut Squash Soup	\$11.99
Quart Chicken Noodle Soup	\$11.99
Quart Kickin' Crab Chowder Soup	\$12.99

**Desserts**

Assorted Large Cookies	\$7.59
Nibbler Cookies & Brownies	\$34.99

**NEW Sandwich BOXED LUNCHEON**

Each box includes your choices of sandwich, side, and cookie.



# Association Learning

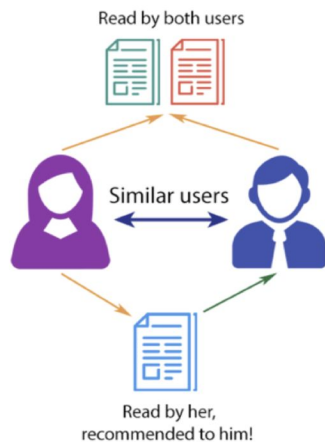
- Market basket analysis
- Collaborative filtering



# Recommender Systems

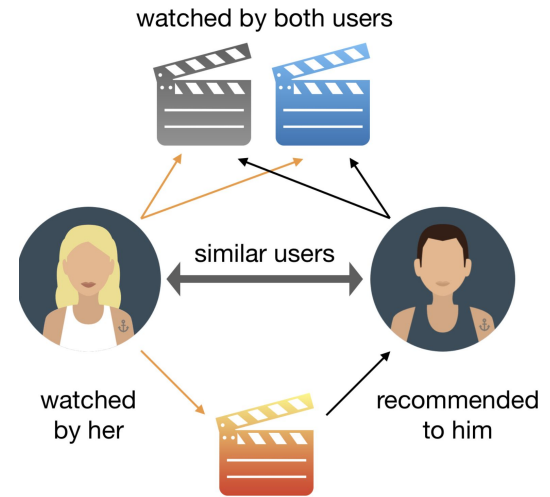
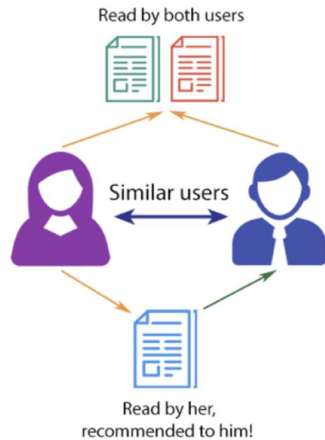
# Recommender Systems

## COLLABORATIVE FILTERING



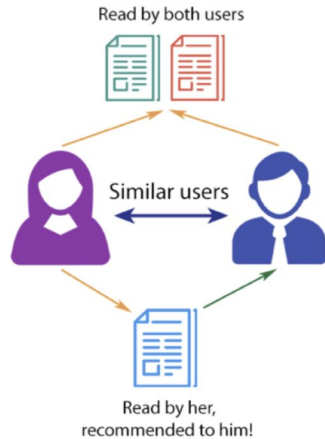
# Recommender Systems

## COLLABORATIVE FILTERING

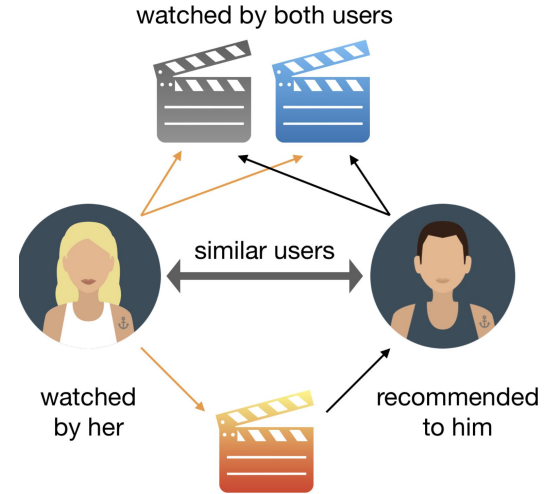


# Recommender Systems

## COLLABORATIVE FILTERING

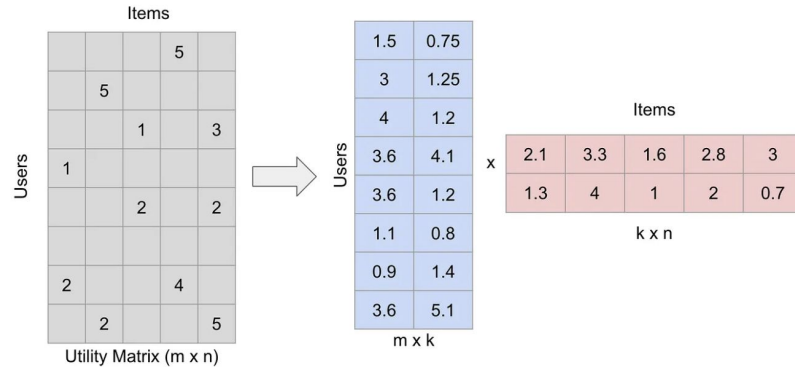
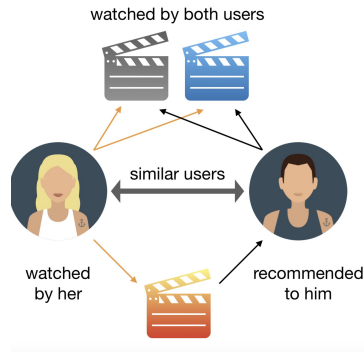


## Collaborative filtering



# Recommender Systems

## Collaborative filtering

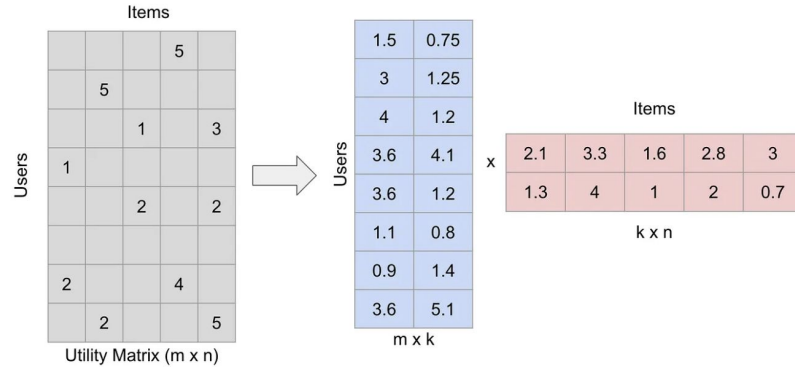
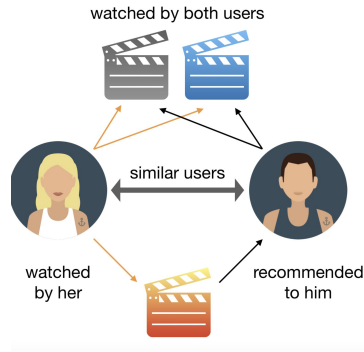


Matrix factorization ( numbers are random for the sake of illustration) (Image by author)



# Recommender Systems

## Collaborative filtering

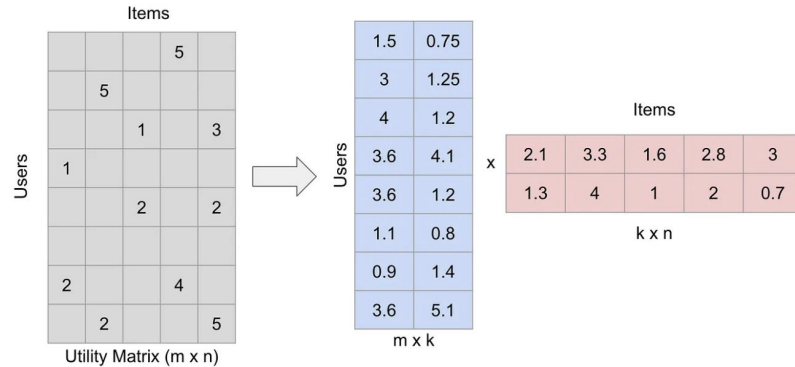
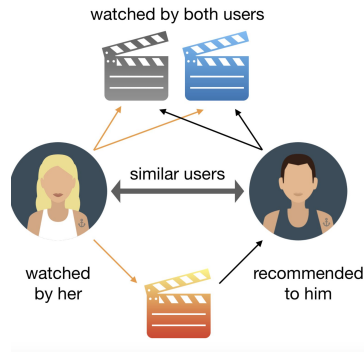


Matrix factorization ( numbers are random for the sake of illustration) (Image by author)

$$M \times N = M \times K \times K \times N$$

# Recommender Systems

## Collaborative filtering

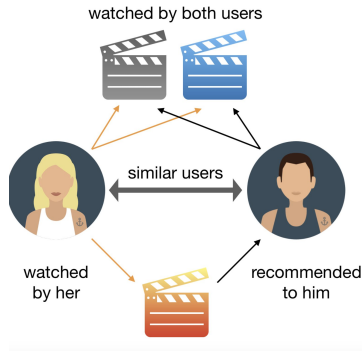


Matrix factorization ( numbers are random for the sake of illustration) (Image by author)

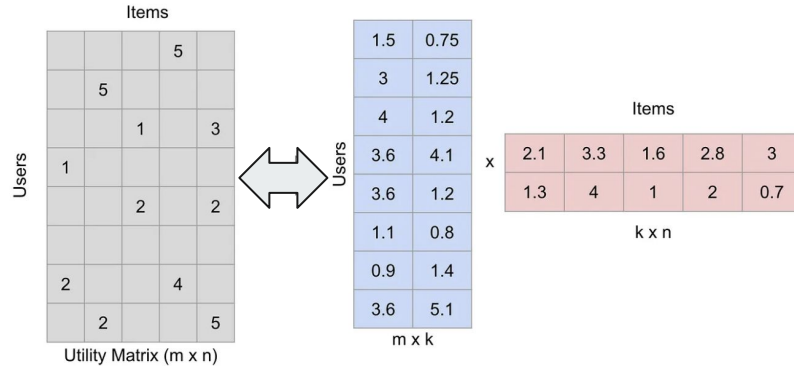
$$8 \times 4 = 8 \times 2 \times 2 \times 4$$

# Recommender Systems

## Collaborative filtering



## optimization

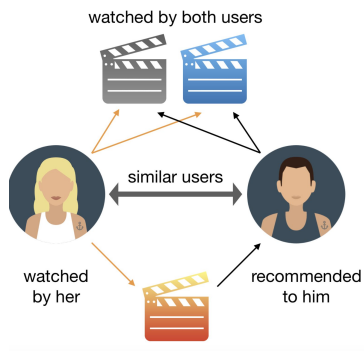


Matrix factorization ( numbers are random for the sake of illustration) (Image by author)

$$8 \times 4 = 8 \times 2 \times 2 \times 4$$

# Recommender Systems

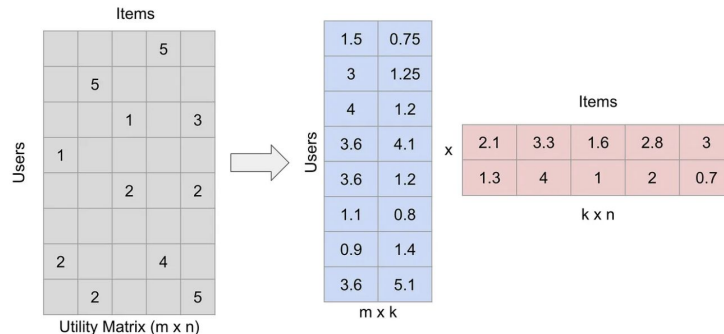
## Collaborative filtering



where  $U$  is  $m \times k$  and  $V$  is  $n \times k$ .  $U$  is a representation of users in some low dimensional space, and  $V$  is a representation of items. For a user  $i$ ,  $u_i$  gives the representation of that user, and for an item  $e$ ,  $v_e$  gives the representation of that item. The rating prediction for a user-item pair is simply:

$$\hat{y}_{ij} = u_i \cdot v_j$$

Rating prediction for a user-item pair is simply the dot product of the user and item representations



- Optimize for new ratings of an item
  - Given by 2 users
  - And for previously learned users and items
  - What item values  $\{=(?, ?)\}$  can explain the two new ratings given that users are fixed (learned).

- Given by 2 users
- And for previously learned users and items
- What item values  $\{=(?, ?)\}$  can explain the two new ratings given that users are fixed (learned).

	Items					
				5		4
		5				?
			1		3	?
Users	1					?
			2		2	?
						?
	2			4		2
		2			5	?

Utility Matrix (m x n)

Users	1.5	0.75
	3	1.25
	4	1.2
	3.6	4.1
	3.6	1.2
	1.1	0.8
	0.9	1.4
	3.6	5.1
m x k		

3

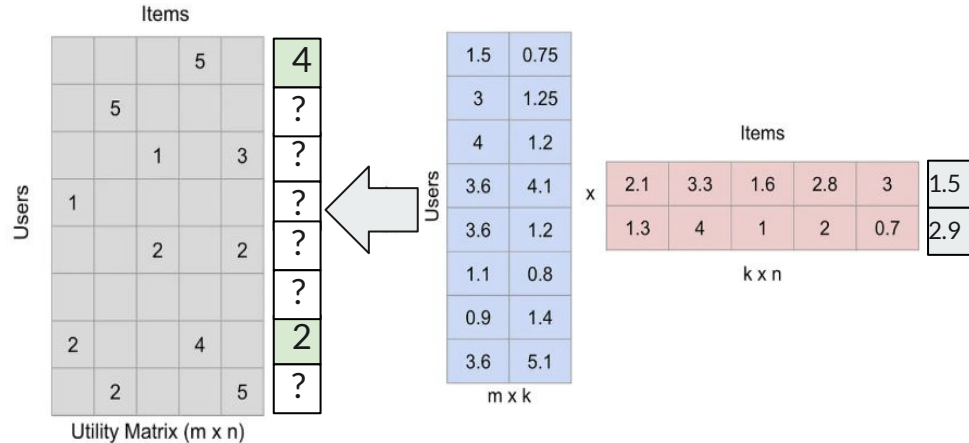
2.1	3.3	1.6	2.8	3	?
1.3	4	1	2	0.7	?

 $k \times n$

## A new movie, lets rated by 2 users

Optimize for new ratings of an item

- Given by 2 users
- And for previously learned users and items
- What item values  $\{=(?, ?)\}$  can explain the two new ratings given that users are fixed (learned).
- Then fill the ?? is in the utility matrix.





**QA**

