

# Recommending Movies from User Ratings

Sue Lim  
September 11, 2023

# I. PROJECT BACKGROUND

# Goal: Take Advantage of Existing User Ratings

- Our client is a **movie recommendation website**
- It has **millions of user ratings** to take advantage of
- It wants make **personalized recommendations** to its customers

# Recommendation

- Use **model-based** collaborative filtering (vs memory-based)
  - It gives the **best performance**
  - It **scales well** once trained
- Use known predictors as much as possible

# MovieLens Dataset

- **100,000 ratings** between 1996 and 2018
  - **600 users**
  - **9,000 movies**

- Example

userId	movieId	rating	timestamp
1	1	4	1225734739
1	110	4	1225865086
1	158	4	1225733503
1	260	4.5	1225735204
1	356	5	1225735119

## II. MAIN MODEL

# Training: Different Models

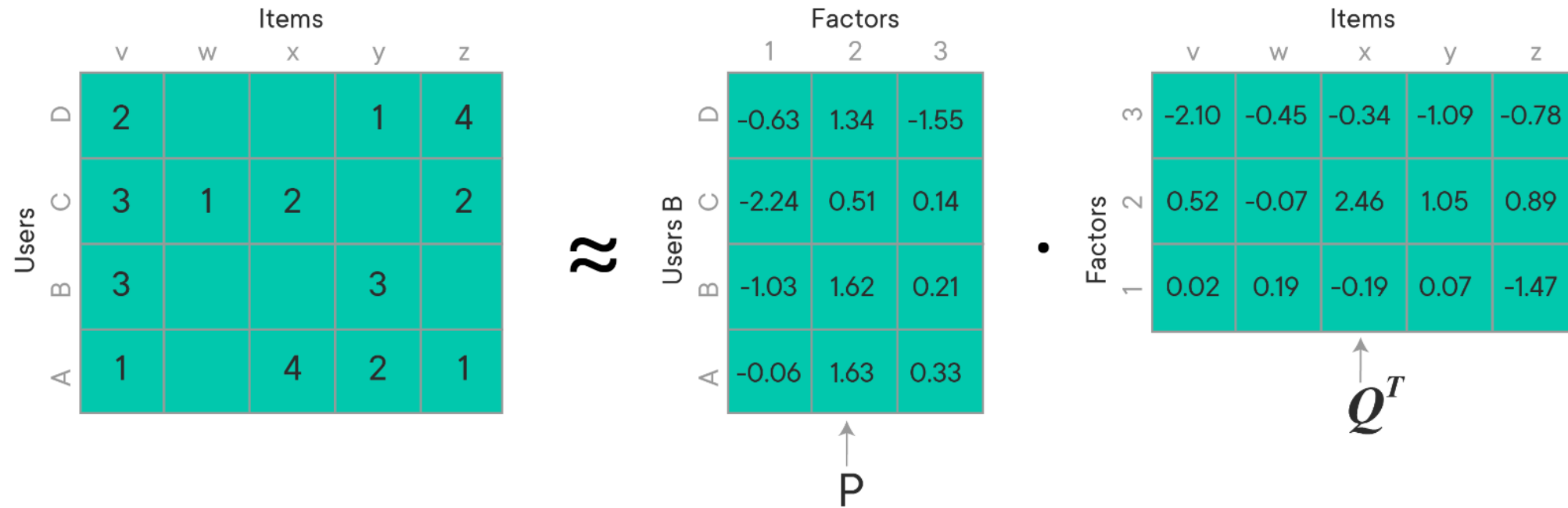
- **Two Model-Based Models**

- Singular Value Decomposition (“SVD”)
- Alternating Least Squares (“ALS”)

- **Three Memory-Based Models**

- KNNBasic
- KNN-w-Means
- KNN-w-Baseline

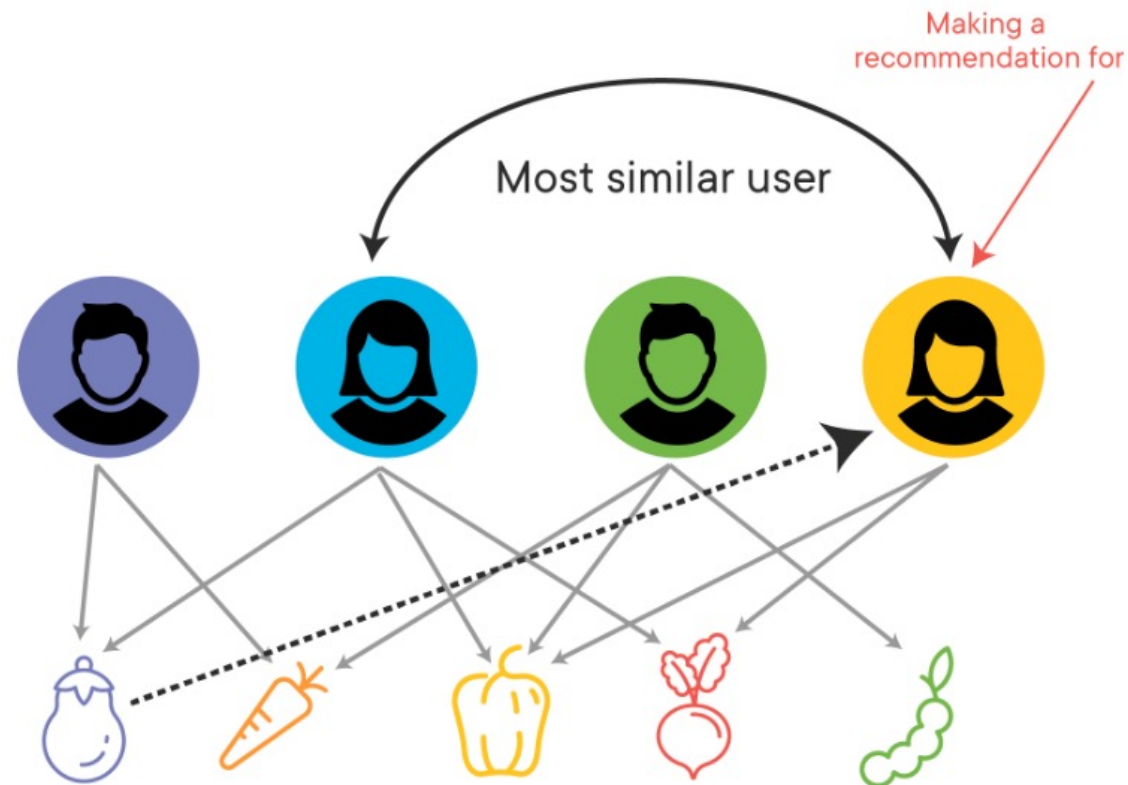
# Model-Based Collaborative Filtering



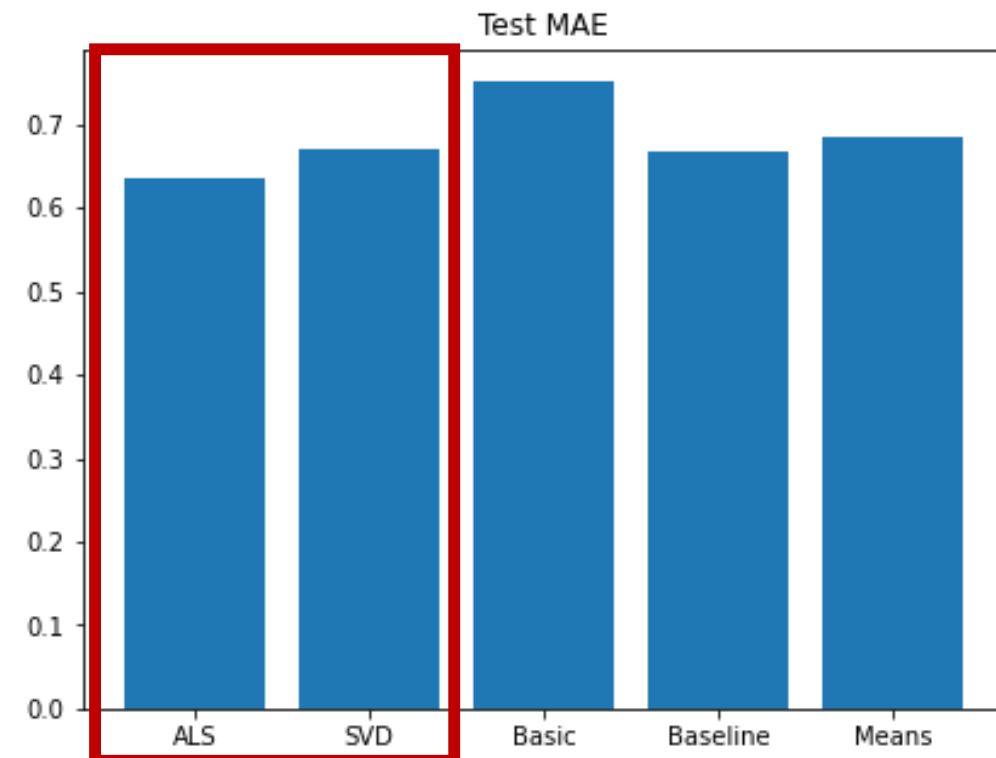
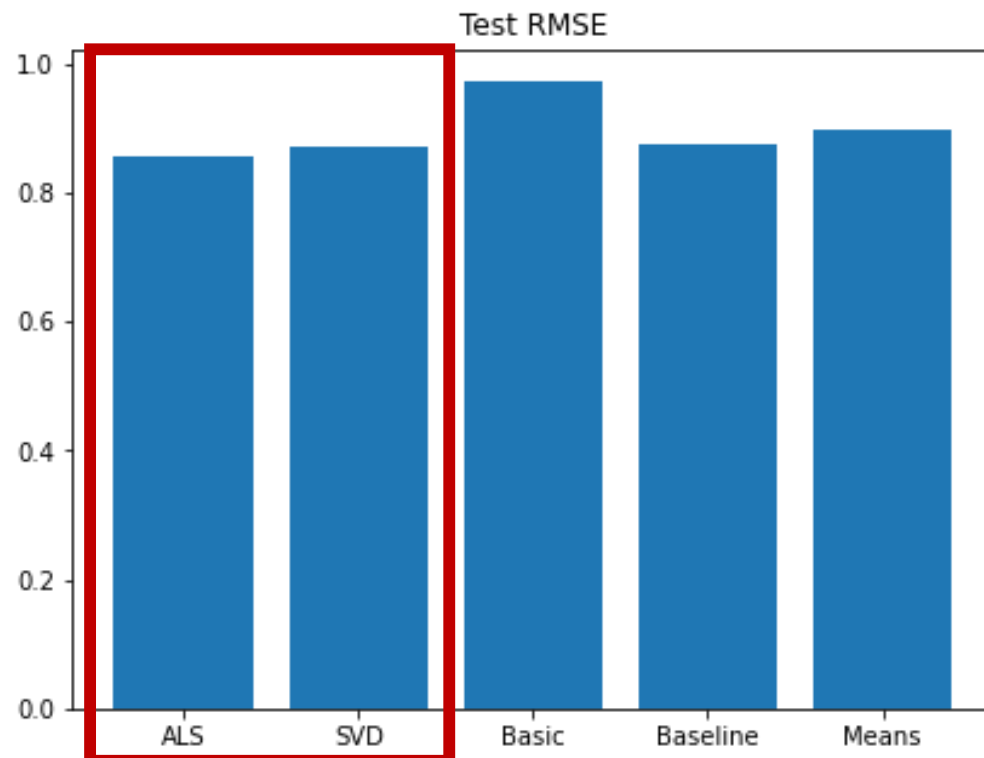


# Memory-Based Collaborative Filtering

## User-Based Collaborative Filtering



# Result: RMSE and MAE Comparison



# Result: Model-Based vs Memory-Based Methods

Memory-Based	Model-Based
complete input data is required	abstraction (model) that represents input data
does not scale well	scales well
pre-computation not possible	pre-computation possible
relies on similarity metrics between users and items	relies on matrix factorization

# ALS: Pre-Defined Predictors

- Movie Genres
- Movie Year
- Rating Year

		Items				
		v	w	x	y	z
Users	D	2			1	4
	C	3	1	2		2
	B	3			3	
	A	1		4	2	1

 $\approx$ 

		Factors		
		1	2	3
Users	D	-0.63	1.34	-1.55
	C	-2.24	0.51	0.14
	B	-1.03	1.62	0.21
	A	-0.06	1.63	0.33

$\uparrow$   
 $P$

		Items				
		v	w	x	y	z
Factors	3	-2.10	-0.45	-0.34	-1.09	-0.78
	2	0.52	-0.07	2.46	1.05	0.89
	1	0.02	0.19	-0.19	0.07	-1.47

$\uparrow$   
 $Q^T$

# Result: ALS Example

Title	Genres	Rating
Out of Sight (1998)	Comedy Crime Drama Romance Thriller	5.0
Like Water for Chocolate (Como agua para choco...	Drama Fantasy Romance	5.0
Iron Giant, The (1999)	Adventure Animation Children Drama Sci-Fi	5.0
Usual Suspects, The (1995)	Crime Mystery Thriller	5.0
Young Frankenstein (1974)	Comedy Fantasy	5.0
Brazil (1985)	Fantasy Sci-Fi	5.0
Enemy of the State (1998)	Action Thriller	4.0
Jungle Book, The (1967)	Animation Children Comedy Musical	4.0
Metropolitan (1990)	Comedy	4.0
Elizabeth (1998)	Drama	4.0
Independence Day (a.k.a. ID4) (1996)	Action Adventure Sci-Fi Thriller	4.0
Sex, Lies, and Videotape (1989)	Drama	4.0
Crying Game, The (1992)	Drama Romance Thriller	4.0
Star Wars: Episode I - The Phantom Menace (1999)	Action Adventure Sci-Fi	4.0
Midnight in the Garden of Good and Evil (1997)	Crime Drama Mystery	3.0
My Best Friend's Wedding (1997)	Comedy Romance	2.0

# Result: ALS Example

Factor	Coefficient
user_factor_1	0.020771
user_factor_2	-0.010140
user_factor_3	0.000579
user_factor_4	0.003673
user_factor_5	0.027567
user_factor_6	-0.005106
user_factor_7	0.002362
user_factor_8	0.001284
user_factor_9	-0.001211
user_factor_10	-0.010920
user_factor_11	0.000000
user_factor_12	-0.009925
user_factor_13	0.039773
user_factor_14	-0.051627
user_factor_15	0.024520

**Comedy** →

user_factor_16	-0.021288
user_factor_17	0.013482
user_factor_18	0.060813
user_factor_19	0.038530
user_factor_20	0.005831
user_factor_21	0.005831
user_factor_22	0.000000
user_factor_23	0.000000
user_factor_24	0.049703
user_factor_25	0.067923
user_factor_26	0.000000
user_factor_27	0.000000
user_factor_28	-0.014799
user_factor_29	-0.008877
user_factor_30	0.000000
user_factor_intercept	0.396125

← **Thriller**

← **Fantasy**

## II. CONCLUSION

# Recommendation & Key Takeaways

- Use **model-based** collaborative filtering
  - It gives the **best performance**
  - It **scales well** once trained
- Use known predictors

Memory-Based	Model-Based
complete input data is required	abstraction (model) that represents input data
does not scale well	scales well
pre-computation not possible	pre-computation possible
relies on similarity metrics between users and items	relies on matrix factorization



# Next Steps

- Look for ways to gather meaningful predictors for users and movies
  - E.g., gender, race, etc. for users and features for movies

# Thank You

## Sue Lim

- Email: [sue.suyeonlim@gmail.com](mailto:sue.suyeonlim@gmail.com)
- GitHub: <https://github.com/suesuyeonlim>
- LinkedIn: <https://www.linkedin.com/in/suelim/>