Matrix Methods in Information Technology

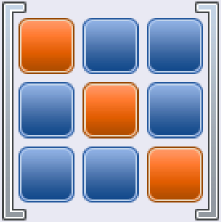
The Mathematics of Recommendation Systems

Yuriy Golovaty

Ivan Franko National University of Lviv

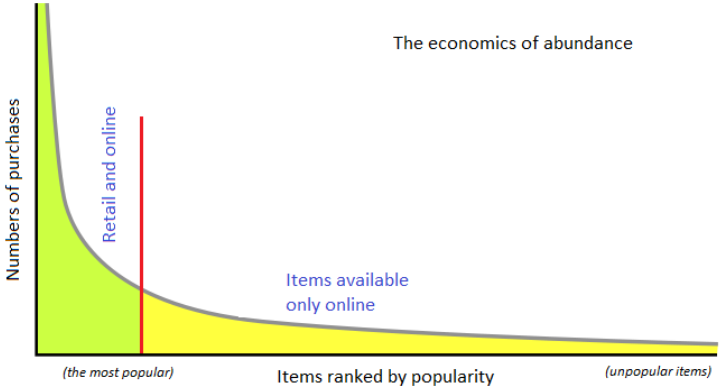
Data Engineering and Security 2020
Lviv, 22 January, 2020

Part 1
Introduction
to Recommendation Systems



A *recommendation system*, or a *recommender system*, is a subclass of information filtering system that seeks to predict the "preference" or "rating" a user would give to an item.

Long tail phenomenon



Recommendation systems

Formal model

- $\mathcal{C}$ is a set of Customers
- $\mathcal{I}$ is a set of Items
- Utility function $U : \mathcal{C} \times \mathcal{I} \to \mathcal{R}$, where $\mathcal{R}$ is a totally ordered set called a set of ratings

Utility matrix

	Item1	Item2	Item3	Item4
Customer1	2	5	1	4
Customer2	3	0	5	3
Customer3	1	1	1	0

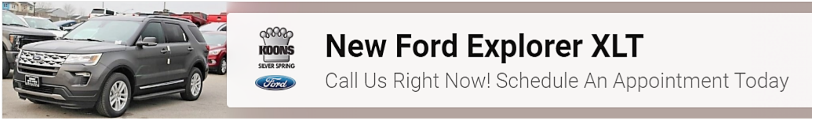
Key problem of recommendation systems

Utility matrix is sparse!

- Most people have not rated most items
- Cold start problem:
  - *New items have no ratings*
  - *New users have no history*

	Item1	Item2	Item3	Item4	NewItem
Customer1		5			
Customer2	3		4		
NewCustomer					
Customer3		1		0	

Extrapolating Utilities



Different approaches to recommendation systems

- Content-based filtering: methods are based on a description of the item and a profile of the user's preferences.
- Collaborative filtering: methods are based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past. The system generates recommendations using only information about rating profiles for different users or items.
- Latent factor based methods is a class of collaborative filtering algorithms based on a matrix factorization.

No advertising!

Item and User profiles

Features

- Movies: genre, author, director, country, actor1, actor2,. . .
- People: set of friends, location, gender, age, profession, . . .
- Notebooks: weight, screen resolution, SSD, RAM, . . .

Item profile

Matrix	1	1	0	1	0	1	0	0	1	0	1	. . .	0

HP Probook 450	6	2.1	0	1	1920	1	8	. . .	23

User profile

Oliver	0	1	0	1	0	0	0	0	1	1	1	. . .	0

Making predictions

Which item would the user prefer?

User	1	0	1	1	0	0	1	1	1	0

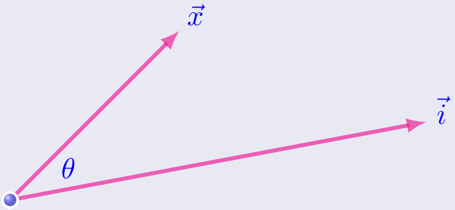
Item1	0	1	0	0	1	1	0	0	1	1
Item2	1	1	1	1	0	0	1	0	1	0
Item3	0	1	0	1	0	0	1	1	0	0
Item4	1	0	1	0	0	0	1	0	1	0
Item5	0	1	0	0	1	1	0	0	0	1

Measure of similarity

Cosine similarity

- $\vec{x}$ is the vector of user profile
- $\vec{i}$ is the vector of item profile

$$\text{similarity}(\vec{x},\vec{i}) = \cos\theta = \frac{\vec{x}\cdot\vec{i}}{|\vec{x}|\,|\vec{i}|}$$



Making predictions

Which item would the user prefer?

User	1	0	1	1	0	0	1	1	1	0

Item1	0	1	0	0	1	1	0	0	1	1
Item2	1	1	1	1	0	0	1	0	1	0
Item3	0	1	0	1	0	0	1	1	0	0
Item4	1	0	1	0	0	0	1	0	1	0
Item5	0	1	0	0	1	1	0	0	0	1

Predictions

	Item1	Item2	Item3	Item4	Item5
Cos	0.183	**0.834**	0.471	**0.730**	0

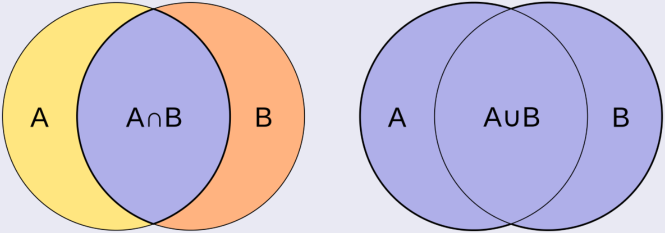
Measure of similarity

Jaccard similarity coefficient

- A and B are finite sets

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

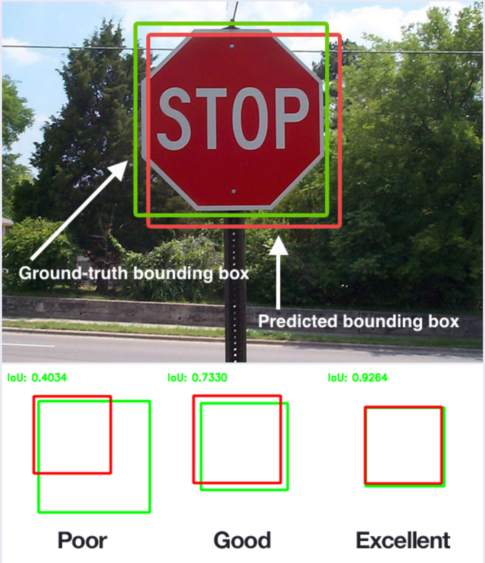
Intersection and union of sets



Measure of similarity

Jaccard similarity coefficient

$$|A \cap B| = 2, \quad |A \cup B| = 7 \quad \Longrightarrow \quad J(A,B) = \frac{2}{7} = 0.286$$


STOP
Ground-truth bounding box
Predicted bounding box
IoU: 0.4034
IoU: 0.7330
IoU: 0.9264
Poor
Good
Excellent

Making predictions

Which item would the user prefer?

User	1	0	1	1	0	0	1	1	1	0

Item1	0	1	0	0	1	1	0	0	1	1
Item2	1	1	1	1	0	0	1	0	1	0
Item3	0	1	0	1	0	0	1	1	0	0
Item4	1	0	1	0	0	0	1	0	1	0
Item5	0	1	0	0	1	1	0	0	0	1

Predictions

	Item1	Item2	Item3	Item4	Item5
Cos	0.183	**0.834**	0.471	**0.730**	0
Jaccard	0.100	**0.714**	0.286	**0.571**	0

Making predictions with missing data

Which item would the user prefer?

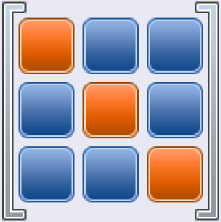
User	1	0			0	0			1	0

Item1	0	1	0	0	1	1	0	0	1	1
Item2	1	1	1	1	0	0	1	0	1	0
Item3	0	1	0	1	0	0	1	1	0	0
Item4	1	0	1	0	0	0	1	0	1	0
Item5	0	1	0	0	1	1	0	0	0	1

Predictions

	Item1	Item2	Item3	Item4	Item5
Jaccard	0.167	**0.334**	0	**0.400**	0

Part 2
Categories, Factorization and Data Storage



Utility matrices

	M1	M2	M3	M4	M5
U1	3	3	3	3	3
U2	3	3	3	3	3
U3	3	3	3	3	3
U4	3	3	3	3	3

	M1	M2	M3	M4	M5
U1	1	3	2	5	4
U2	2	1	1	1	5
U3	3	2	3	1	5
U4	2	4	1	5	2

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2					
U3	3	1	1	3	1
U4					

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2					
U3	3	1	1	3	1
U4					

Dependence

User1 = User3

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3			3	
U2	1			1	
U3	3			3	
U4	4			4	

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3			3	
U2	1			1	
U3	3			3	
U4	4			4	

$$\text{Movie1} = \text{Movie4}$$

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1					
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1					
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependence

$$\text{User4} = \text{User2} + \text{User3}$$

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1		1	1		1
U2		2	4		3
U3		1	1		1
U4		3	5		4

Dependent rows and columns

Utility matrix

	M1	M2	M3	M4	M5
U1		1	1		1
U2		2	4		3
U3		1	1		1
U4		3	5		4

Dependence

$$\text{Movie5} = \mathit{Average}(\text{Movie2}, \text{Movie3})$$

Rank of matrix

Matrix as a set of rows

$$\begin{pmatrix} 2 & 7 & 5 & 3 \\ 1 & 3 & 0 & 2 \\ 1 & 4 & 5 & 1 \end{pmatrix} \rightsquigarrow \left\{ \begin{pmatrix} 2 & 7 & 5 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 0 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 4 & 5 & 1 \end{pmatrix} \right\}$$

Matrix as a set of columns

$$\begin{pmatrix} 2 & 7 & 5 & 3 \\ 1 & 3 & 0 & 2 \\ 1 & 4 & 5 & 1 \end{pmatrix} \rightsquigarrow \left\{ \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} \right\}$$

Row rank

The row rank of a matrix A is the maximal number of linearly independent rows of A.

Column rank

The column rank of a matrix A is the maximal number of linearly independent columns of A.

Rank of matrix

Rank of matrix

$$\text{rank}\, A = \text{Row rank} A = \text{Column rank} A$$

Properties of rank

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{12} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{12} & \dots & a_{mn} \end{pmatrix}$$

- $0 \leq \text{rank}\, A \leq \min\{m, n\}$.
- Only a zero matrix has rank zero.
- $\text{rank}\, A = \text{rank}\, A^T = \text{rank}\, AA^T = \text{rank}\, A^T A$.

Utility matrices

	M1	M2	M3	M4	M5
U1	3	3	3	3	3
U2	3	3	3	3	3
U3	3	3	3	3	3
U4	3	3	3	3	3

$\operatorname{rank} U = 1$

	M1	M2	M3	M4	M5
U1	1	3	2	5	4
U2	2	1	1	1	5
U3	3	2	3	1	5
U4	2	4	1	5	2

$\operatorname{rank} U = 4$

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

$\operatorname{rank} U = 3$

Classification and categorical data

Classification and categorical data

Categories=Features

F1=Actions
F2=Comedies

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Classification and categorical data

Categories=Features

F1=Actions
F2=Comedies

	M1	M2	M3	M4	M5
F1	3	1	1	3	1
F2	1	2	4	1	3

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Classification and categorical data

Categories=Features

F1=Actions
F2=Comedies

	M1	M2	M3	M4	M5
F1	3	1	1	3	1
F2	1	2	4	1	3

	F1	F2
U1	1	0
U2	0	1
U3	1	0
U4	1	1

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

Classification and categorical data

Categories=Features

F1=Actions
F2=Comedies

	M1	M2	M3	M4	M5
F1	3	1	1	3	1
F2	1	2	4	1	3

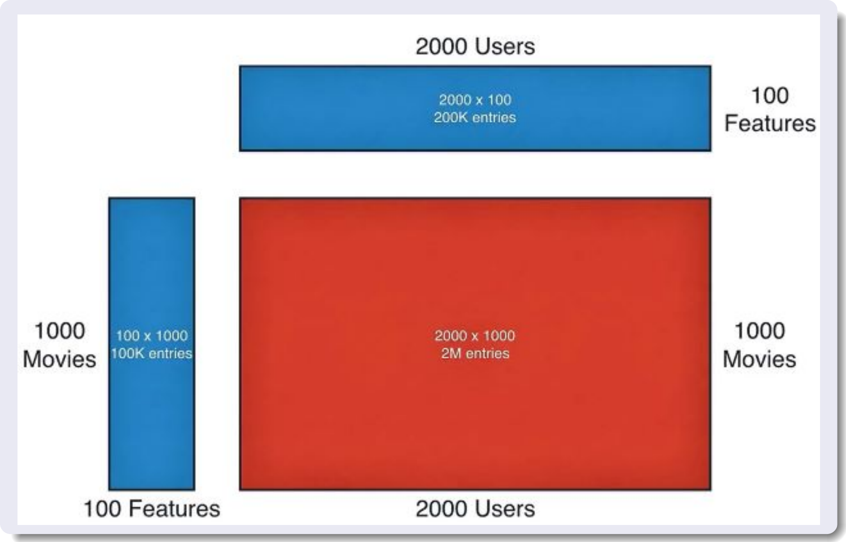
	F1	F2
U1	1	0
U2	0	1
U3	1	0
U4	1	1

	M1	M2	M3	M4	M5
U1	3	1	1	3	1
U2	1	2	4	1	3
U3	3	1	1	3	1
U4	4	3	5	4	4

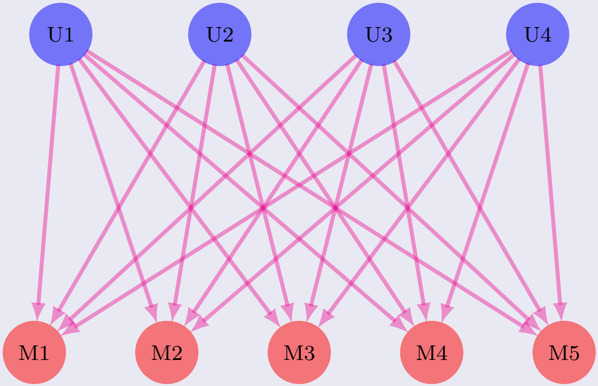
Matrix factorization

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 & 3 & 1 \\ 1 & 2 & 4 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 1 & 3 & 1 \\ 1 & 2 & 4 & 1 & 3 \\ 3 & 1 & 1 & 3 & 1 \\ 4 & 3 & 5 & 4 & 4 \end{pmatrix}$$

Data storage



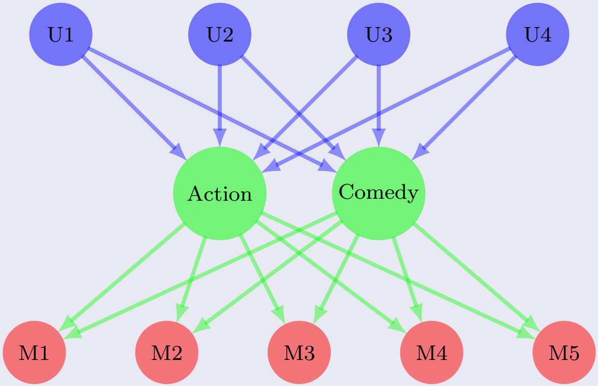
Data storage



$$\text{Users} \times \text{Movies} = \text{Links}$$

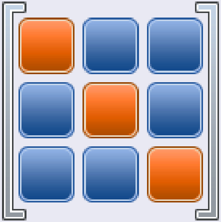
$$2\,000 \times 1\,000 \quad = 2\,000\,000$$

Data storage



$$
\begin{aligned}
(\text{Users} + \text{Movies}) \times \text{Features} &= \text{Links} \\
(2\,000 + 1\,000) \times 100 \qquad &= \color{magenta}{300\,000}
\end{aligned}
$$

Part 3
Singular Value Decomposition



Advantages of SVD

- the dimensionality reduction
- latent factors in a dataset
- latent concepts and their importance in a dataset
- the elimination of the less important information from a dataset

Square matrices

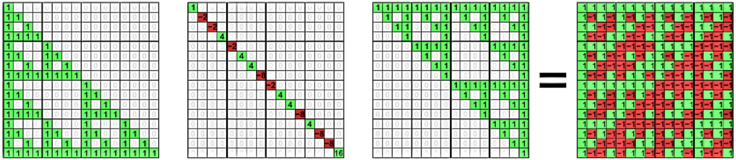
Types of square matrices

Diagonal matrix	$\begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$
Upper triangular matrix	$\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$
Lower triangular matrix	$\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$
Symmetric matrix $A^T = A$	$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$
Orthogonal matrix $A^T = A^{-1}$	$\begin{bmatrix} \vert & \vert & \vert \\ v_1 & v_2 & v_3 \\ \vert & \vert & \vert \end{bmatrix}$ $v_l^T v_k = \delta_{lk}$

Matrix factorization

Decomposition of square matrices

$A = LU$	L is lower triangular U is upper triangular	Gaussian Elimination
$A = LDU$	L is lower triangular D is diagonal U is upper triangular	Gaussian Elimination
$A = QR$	Q is orthogonal R is upper triangular	Gram-Schmidt process
$A = PJP^{-1}$	J is a Jordan matrix P is invertible	Jordan decomposition
$S = Q\Lambda Q^{-1}$	Λ is diagonal Q is orthogonal	Spectral decomposition of a symmetric matrix



Factorization of rectangle matrices

Let A be the $m \times n$ matrix.

- The eigenvalue problem is meaningless for rectangular matrices!
- AA^T is a $m \times m$ matrix
- $A^T A$ is a $n \times n$ matrix
- $\text{rank}\, AA^T = \text{rank}\, A^T A = \text{rank}\, A$

Factorization of rectangle matrices

Matrix A

$$A = \begin{bmatrix} 1 & -4 \\ 2 & -3 \\ 3 & 2 \\ 4 & 1 \end{bmatrix}$$

Matrix AA^T

$$AA^T = \begin{bmatrix} 1 & -4 \\ 2 & -3 \\ 3 & 2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ -4 & -3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 14 & -5 & 0 \\ 14 & 13 & 0 & 5 \\ -5 & 0 & 13 & 14 \\ 0 & 5 & 14 & 17 \end{bmatrix}$$

Matrix $A^T A$

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ -4 & -3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ 2 & -3 \\ 3 & 2 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 0 \\ 0 & 30 \end{bmatrix}$$

Factorization of rectangle matrices

Property 1

All eigenvalues of AA^T and A^TA are non-negative.

Property 2

- If λ_j is a positive eigenvalue of AA^T, then λ_j is a positive eigenvalue of A^TA and vice versa.
- The number of positive singular values σ_k is equal to $r = \operatorname{rank} A$.

$$\sigma(AA^T) \cap \sigma(A^TA) = \{\text{positive eigenvalues}\}$$
$$\sigma(AA^T) = \sigma(A^TA) \cup \{0, 0, \ldots, 0\}$$

Singular values

Numbers $\sigma_j = \sqrt{\lambda_j}$ are called the singular values of A.

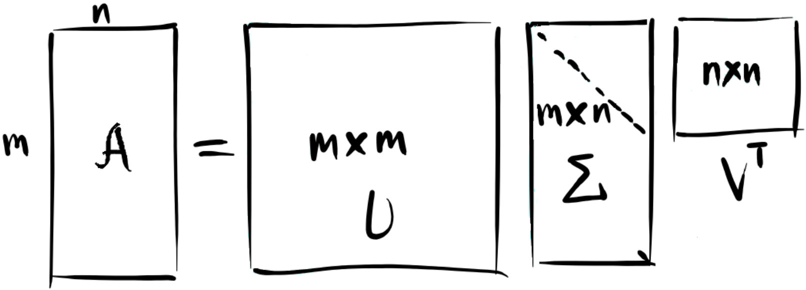
Singular value decomposition

SVD Theorem

Every $m \times n$ matrix A can be written as

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

where U and V are orthogonal and Σ is an $m \times n$ matrix with singular values of A on its main diagonal and zeros otherwise.



Singular value decomposition

Matrix Σ

$$\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The singular values are ordered in decreasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s \geq 0$.

Matrix U

$$\begin{bmatrix} u_{11} & u_{12} & \dots & u_{1m} \\ u_{21} & u_{22} & \dots & u_{2m} \\ \vdots & \vdots & & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mm} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{bmatrix}$$

The vectors u_k are the left singular vectors of A,

$$AA^T u_k = \sigma_k^2 u_k.$$

Singular value decomposition

Matrix Σ

$$\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

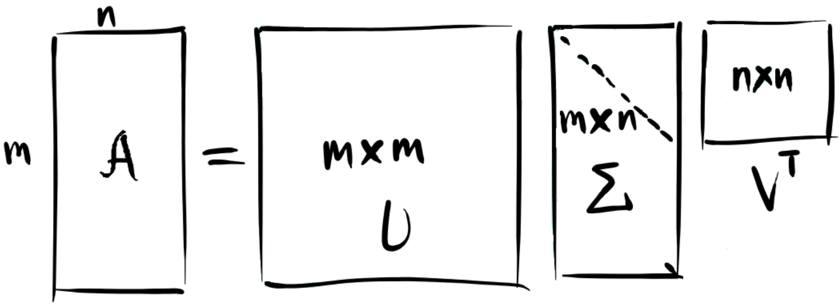
The singular values are ordered in decreasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s \geq 0$.

Matrix V^T

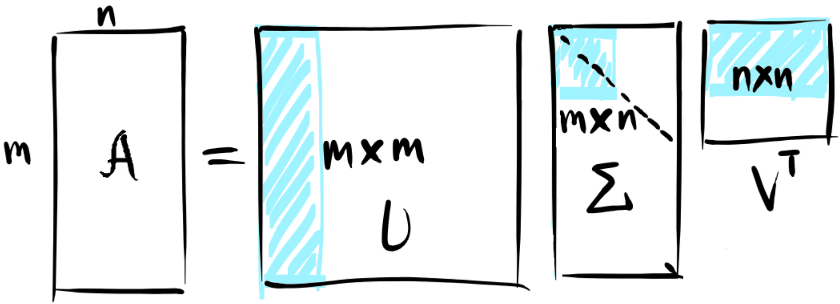
$$\begin{bmatrix} v_{11} & v_{21} & \dots & v_{m1} \\ v_{12} & v_{22} & \dots & v_{m2} \\ \vdots & \vdots & & \vdots \\ v_{1m} & v_{2m} & \dots & v_{mm} \end{bmatrix} = \begin{bmatrix} \rule{2em}{0.4pt} & v_1^T & \rule{2em}{0.4pt} \\ \rule{2em}{0.4pt} & v_2^T & \rule{2em}{0.4pt} \\ \dots & \dots & \dots \\ \rule{2em}{0.4pt} & v_n^T & \rule{2em}{0.4pt} \end{bmatrix}$$

The vectors v_k are the right singular vectors of A, i.e., $A^T A v_k = \sigma_k^2 v_k$, such that $A v_k = \sigma_k u_k$ or the same $A^T u_k = \sigma_k v_k$.

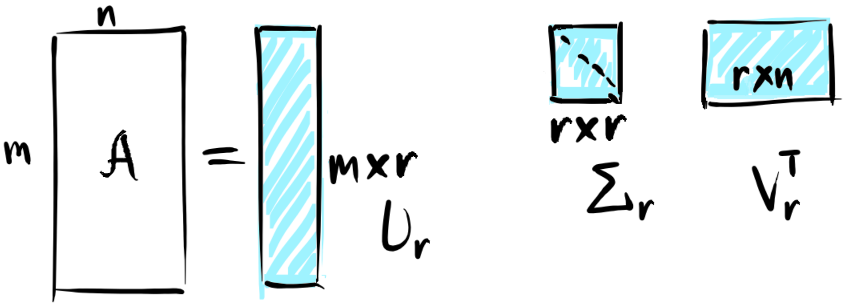
Reduced SVD



Reduced SVD

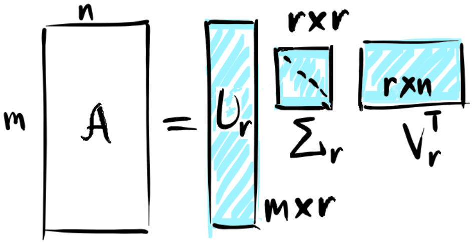


Reduced SVD



Reduced SVD

$$A = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_r \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \text{——} & v_1^T & \text{——} \\ \text{——} & v_2^T & \text{——} \\ \cdots & \cdots & \cdots \\ \text{——} & v_r^T & \text{——} \end{bmatrix}$$



A as the sum of rank-1 matrices

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

Dimensionality reduction

Rank-k approximation of A

$$A_1 = \sigma_1 u_1 v_1^T, \qquad A_2 = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T, \ldots$$

$$A_k = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T, \quad k = 1, 2, \ldots, r$$

$$A_k = U_r \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} V_r^T$$

The matrix A_k is the best rank-k approximation to A, i.e.,

$$\min_{B, \text{rank } B = k} \|A - B\| = \|A - A_k\|.$$

SVD and Image compression

Interpretation of SVD

Netflix utility matrix

	SF1	SF2	SF3	R1	R2
B1	1	1	1	0	1
B2	3	3	3	0	0
B3	4	4	4	0	1
B4	5	5	5	0	0
G1	0	2	0	4	4
G2	0	0	0	5	5
G3	0	1	0	2	2

SF = Science Fiction, R = Romance

Interpretation of SVD

Netflix utility matrix

	SF1	SF2	SF3	R1	R2
B1	1	1	1	0	1
B2	3	3	3	0	0
B3	4	4	4	0	1
B4	5	5	5	0	0
G1	0	2	0	4	4
G2	0	0	0	5	5
G3	0	1	0	2	2

SF = Science Fiction, R = Romance

Interpretation of SVD

	SF1	SF2	SF3	R1	R2
B1	1	1	1	0	1
B2	3	3	3	0	0
B3	4	4	4	0	1
B4	5	5	5	0	0
G1	0	2	0	4	4
G2	0	0	0	5	5
G3	0	1	0	2	2



Interpretation of SVD

Utility matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 1 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}, \quad \operatorname{rank} A = 4$$

SVD of A

$$A = U\Sigma V^T, \quad \Sigma = \begin{bmatrix} 12.574 & 0 & 0 & 0 \\ 0 & 9.449 & 0 & 0 \\ 0 & 0 & 1.362 & 0 \\ 0 & 0 & 0 & 0.857 \end{bmatrix}$$

Interpretation of SVD

Utility matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 1 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}, \quad \operatorname{rank} A = 4$$

SVD of A

$$A = U\Sigma V^T, \quad \Sigma = \begin{bmatrix} 12.574 & 0 & 0 & 0 \\ 0 & 9.449 & 0 & 0 \\ 0 & 0 & 1.362 & 0 \\ 0 & 0 & 0 & 0.857 \end{bmatrix}$$

Interpretation of SVD

Utility matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 1 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}, \quad \operatorname{rank} A = 4$$

SVD of A

$$A_2 = U\Sigma_2 V^T, \quad \Sigma_2 = \begin{bmatrix} 12.574 & 0 & 0 & 0 \\ 0 & 9.449 & 0 & 0 \\ 0 & 0 & 0.000 & 0 \\ 0 & 0 & 0 & 0.000 \end{bmatrix}$$

Interpretation of SVD

Reduced SVD

$$\begin{bmatrix} 0.149 & 0.033 \\ 0.403 & -0.119 \\ 0.552 & -0.086 \\ 0.672 & -0.198 \\ 0.190 & 0.575 \\ 0.120 & 0.725 \\ 0.095 & 0.287 \end{bmatrix} \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{bmatrix} 0.551 & 0.588 & 0.551 & 0.123 & 0.179 \\ -0.175 & -0.023 & -0.175 & 0.688 & 0.682 \end{bmatrix}$$

Rank-2 approximation of A

$$A_2 = \begin{bmatrix} 0.975 & 1.092 & 0.975 & 0.442 & 0.545 \\ 2.988 & 3.010 & 2.988 & -0.147 & 0.142 \\ 3.964 & 4.102 & 3.9634 & 0.296 & 0.687 \\ 4.981 & 5.015 & 4.981 & -0.244 & 0.237 \\ 0.361 & 1.277 & 0.361 & 4.029 & 4.132 \\ -0.369 & 0.730 & -0.369 & 4.895 & 4.941 \\ 0.181 & 0.639 & 0.181 & 2.015 & 2.066 \end{bmatrix}$$

Interpretation of SVD

Reduced SVD

$$\begin{bmatrix} 0.149 & 0.033 \\ 0.403 & -0.119 \\ 0.552 & -0.086 \\ 0.672 & -0.198 \\ 0.190 & 0.575 \\ 0.120 & 0.725 \\ 0.095 & 0.287 \end{bmatrix} \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{bmatrix} 0.551 & 0.588 & 0.551 & 0.123 & 0.179 \\ -0.175 & -0.023 & -0.175 & 0.688 & 0.682 \end{bmatrix}$$

Rank-2 approximation of A

$$A = \begin{bmatrix} 1.000 & 1.000 & 1.000 & 0.000 & 1.000 \\ 3.000 & 3.000 & 3.000 & 0.000 & 0.000 \\ 4.000 & 4.000 & 4.000 & 0.000 & 1.000 \\ 5.000 & 5.000 & 5.000 & 0.000 & 0.000 \\ 0.000 & 2.000 & 0.000 & 4.000 & 4.000 \\ 0.000 & 0.000 & 0.000 & 5.000 & 5.000 \\ 0.000 & 1.000 & 0.000 & 2.000 & 2.000 \end{bmatrix}$$

Interpretation of SVD

User-to-concept similarity matrix

$$U = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \\ \left[\begin{matrix} 0.149 & 0.033 \\ 0.403 & -0.119 \\ 0.552 & -0.086 \\ 0.672 & -0.198 \\ 0.190 & 0.575 \\ 0.120 & 0.725 \\ 0.095 & 0.287 \end{matrix}\right] & \end{array}$$

Matrix of concepts

$$\Sigma = \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{array}{l} \leftarrow \textit{Science Fiction} \\ \leftarrow \textit{Romance} \end{array}$$

Interpretation of SVD

User-to-concept similarity matrix

$$U = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \\ \end{array}$$

$$U = \begin{bmatrix} 0.149 & 0.033 \\ 0.403 & -0.119 \\ 0.552 & -0.086 \\ 0.672 & -0.198 \\ 0.190 & 0.575 \\ 0.120 & 0.725 \\ 0.095 & 0.287 \end{bmatrix}$$

Matrix of concepts

$$\Sigma = \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{array}{l} \leftarrow \textit{Science Fiction} \\ \leftarrow \textit{Romance} \end{array}$$

Interpretation of SVD

User-to-concept similarity matrix

$$U = \begin{array}{c} \begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \end{array} \\ \begin{bmatrix} 0.149 & 0.033 \\ 0.403 & -0.119 \\ 0.552 & -0.086 \\ 0.672 & -0.198 \\ 0.190 & 0.575 \\ 0.120 & 0.725 \\ 0.095 & 0.287 \end{bmatrix} \end{array}$$

Matrix of concepts

$$\Sigma = \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{array}{l} \leftarrow \textit{Science Fiction} \\ \leftarrow \textit{Romance} \end{array}$$

Interpretation of SVD

User-to-concept similarity matrix

$$
U = \begin{array}{c} \begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \end{array} \\ \begin{bmatrix} 0.149 & 0.033 \\ 0.403 & -0.119 \\ 0.552 & -0.086 \\ 0.672 & -0.198 \\ 0.190 & 0.575 \\ 0.120 & 0.725 \\ 0.095 & 0.287 \end{bmatrix} \end{array}
$$

Projection to latent structures

$$U^T : \mathbb{R}^5 \to \mathbb{R}^2, \qquad U^T : \text{Movies} \to \text{Space of Concepts}$$

Interpretation of SVD

Movie-to-concept similarity matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \end{array} \\
V = & \begin{bmatrix} 0.551 & -0.175 \\ 0.588 & -0.023 \\ 0.551 & -0.175 \\ 0.123 & 0.688 \\ 0.179 & 0.682 \end{bmatrix}
\end{array}
$$

Matrix of concepts

$$
\Sigma = \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{array}{l} \leftarrow \textit{Science Fiction} \\ \leftarrow \textit{Romance} \end{array}
$$

Interpretation of SVD

Movie-to-concept similarity matrix

$$\begin{array}{cc} & \begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \end{array} \\ V = & \begin{bmatrix} 0.551 & -0.175 \\ 0.588 & -0.023 \\ 0.551 & -0.175 \\ 0.123 & 0.688 \\ 0.179 & 0.682 \end{bmatrix} \end{array}$$

Matrix of concepts

$$\Sigma = \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{array}{l} \leftarrow \textit{Science Fiction} \\ \leftarrow \textit{Romance} \end{array}$$

Interpretation of SVD

Movie-to-concept similarity matrix

$$V = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \\ \end{array}$$

$$V = \begin{bmatrix} 0.551 & -0.175 \\ 0.588 & -0.023 \\ 0.551 & -0.175 \\ 0.123 & 0.688 \\ 0.179 & 0.682 \end{bmatrix}$$

Matrix of concepts

$$\Sigma = \begin{bmatrix} 12.574 & 0 \\ 0 & 9.449 \end{bmatrix} \begin{array}{l} \leftarrow \textit{Science Fiction} \\ \leftarrow \textit{Romance} \end{array}$$

Interpretation of SVD

Movie-to-concept similarity matrix

$$V = \begin{array}{c} \\ \end{array}\begin{array}{cc} \textit{Sci-Fi} & \textit{Romance} \end{array}$$

$$V = \begin{bmatrix} 0.551 & -0.175 \\ 0.588 & -0.023 \\ 0.551 & -0.175 \\ 0.123 & 0.688 \\ 0.179 & 0.682 \end{bmatrix}$$

Projection to latent structures

$$V^T : \mathbb{R}^5 \to \mathbb{R}^2, \qquad U^T : \text{Users} \to \text{Space of Concepts}$$

Interpretation of SVD

We have new users

$$N1 = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$N2 = \begin{bmatrix} 0 & 4 & 5 & 0 & 0 \end{bmatrix}$$

Cosine similarity of the user profiles

User profiles $N1$ and $N2$ are orthogonal, then

$$\text{Similarity}\left(N1, N2\right) = \cos\frac{\pi}{2} = 0$$

Interpretation of SVD

Let us map our users into the space of concepts

$$N1 \cdot V = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.551 & -0.175 \\ 0.588 & -0.023 \\ 0.551 & -0.175 \\ 0.123 & 0.688 \\ 0.179 & 0.682 \end{bmatrix} = \begin{bmatrix} 2.755 & -0.875 \end{bmatrix}$$

$$N2 \cdot V = \begin{bmatrix} 0 & 4 & 5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.551 & -0.175 \\ 0.588 & -0.023 \\ 0.551 & -0.175 \\ 0.123 & 0.688 \\ 0.179 & 0.682 \end{bmatrix} = \begin{bmatrix} 5.107 & -0.967 \end{bmatrix}$$

Interpretation of SVD

Let us map the user N1 representation back into movie space

$$N1 \cdot V = \begin{bmatrix} 2.755 & -0.875 \end{bmatrix}$$

$$N1 \cdot V \cdot V^T =$$

$$\begin{bmatrix} 2.755 & -0.875 \end{bmatrix} \begin{bmatrix} 0.551 & 0.588 & 0.551 & 0.123 & 0.179 \\ -0.175 & -0.023 & -0.175 & 0.688 & 0.682 \end{bmatrix} =$$

$$\begin{bmatrix} 1.671 & 1.640 & 1.671 & -0.263 & -0.104 \end{bmatrix}$$

Predictions for user N1

$$\begin{bmatrix} 5 & 4.907 & 5 & 0 & 0 \end{bmatrix}$$

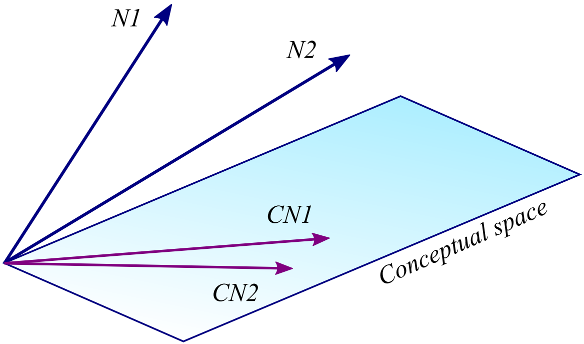
Interpretation of SVD

New user profiles in the space of concepts

$$CN1 = N1 \cdot V = \begin{bmatrix} 2.755 & -0.875 \end{bmatrix}$$

$$CN2 = N2 \cdot V = \begin{bmatrix} 5.107 & -0.967 \end{bmatrix}$$

$$\text{Similarity}\,(CN1, CN2) = 0.993$$



Interpretation of SVD

We have new movie

$$New = \begin{bmatrix} 2 & 3 & 2 & 3 & 2 & 2 & 3 \end{bmatrix}^T$$

$$CNew = U^T \cdot New = \begin{bmatrix} 5.530 \\ -2.405 \end{bmatrix}$$

Cosine similarity

$$SF = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad R = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

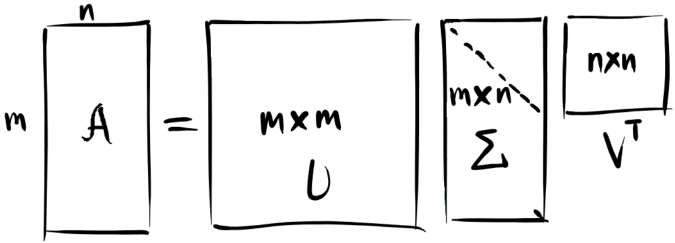
$$\text{Similarity}\left(CNew, SF\right) = 0.917$$

$$\text{Similarity}\left(CNew, R\right) = 0.399$$

History of SVD

Contribution to matrix factorization

- Eugenio Beltrami (1835–1899)
- Camille Jordan (1838–1921)
- James J. Sylvester (1814–1897)
- Erhard Schmidt (1876–1959)
- Hermann Weyl (1885–1955)
- Carl H. Eckart (1902–1973)
- Gale J. Young (1912–1990)



Credits

- Leskovec, Rajaraman, and Ullman, *Mining of Massive Datasets*, Cambridge University Press 40 W. 20 St. New York, NYUnited States.
- *Mining of Massive Datasets*, Stanford University, Lectures 41–50. https://www.youtube.com/playlist?list=PLLssT5z_DsK9JDLcT8T62VtzwyW9LNepV
- Luis Serrano, *How does Netflix recommend movies? Matrix Factorization.* https://www.youtube.com/watch?v=ZspR5PZemcs&t=283s
- Haesun Park and Lars Elden,*Matrix Rank Reduction for Data Analysis and Feature Extraction*, 10.1201/9781420028683.ch14.
- Most pictures taken from Google image search.