Singular Value Decomposition (SVD) and Its Applications

Jiaxing Shen

5 April 2022

Outline

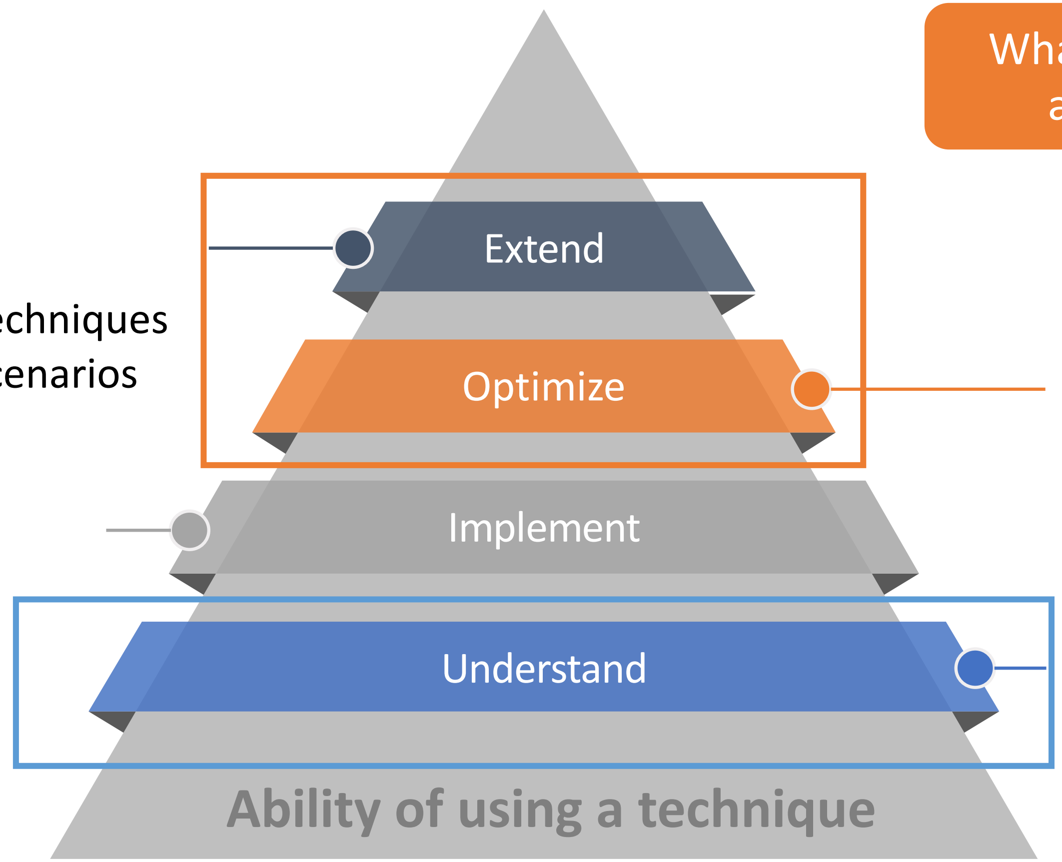
- Why SVD?
- Preliminaries
- Understand SVD
- Popular Applications

Four Ability Levels of Using a Technique

What required in original and novel research

Extend

- How is it related to other techniques
- What are other potential scenarios



Optimize

- How to improve it: efficiency, reliability, distributed environments

Implement

- How to implement
- What are pros of cons

Understand

- What it is
- Where it could be applied

Focus of the talk

Why SVD?

- SVD is the foundation of Recommender Systems that are at the heart of huge companies like Google, YouTube, Amazon, Facebook, Netflix…
- "… it (SVD) is not nearly as famous as it should be." -- Gilbert Strang

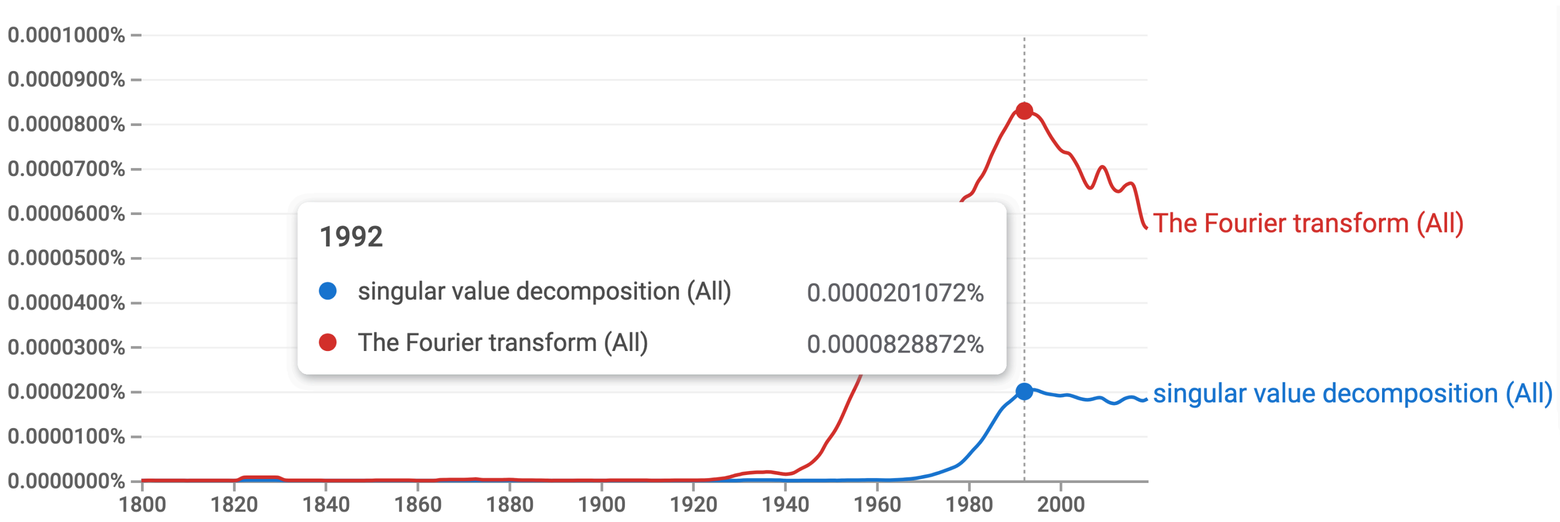
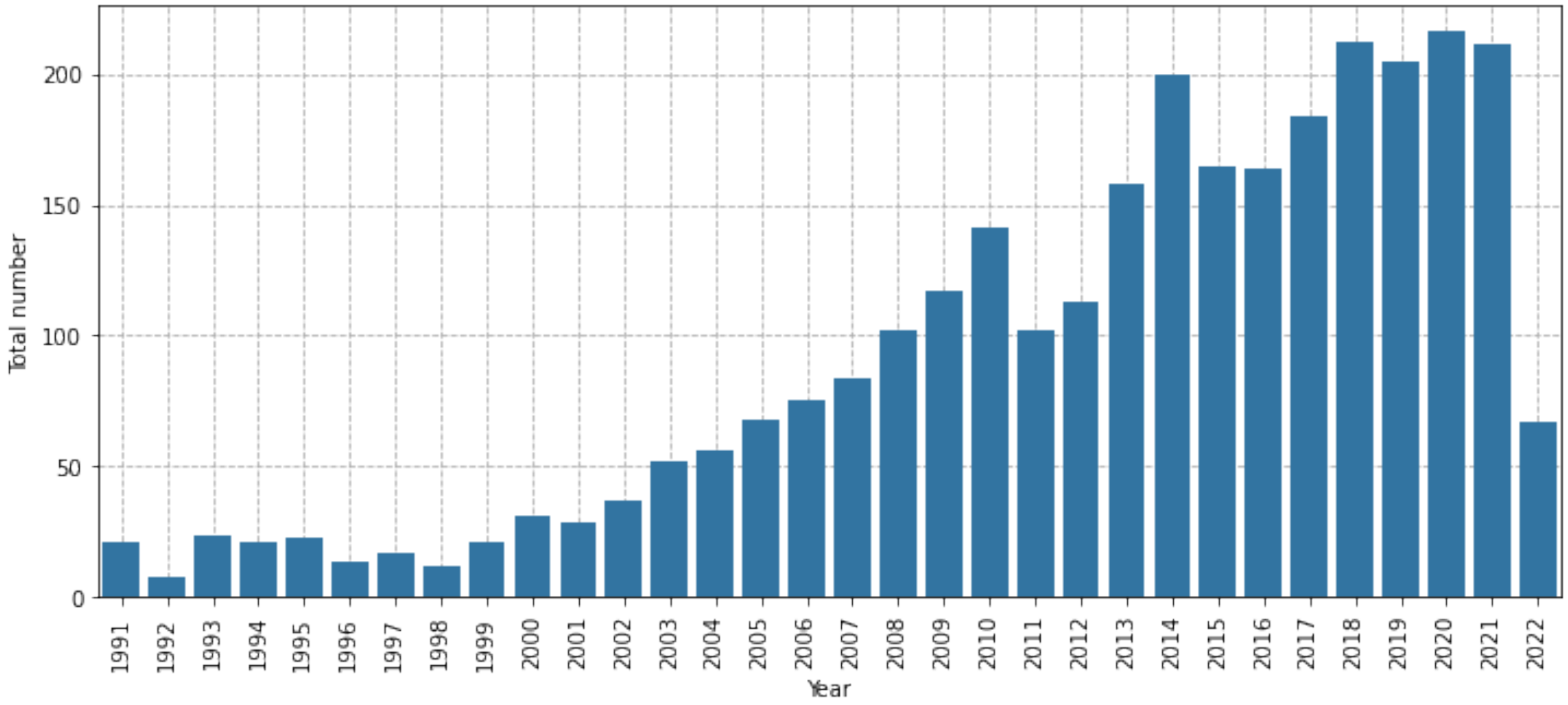


Figure from Google Ngram Viewer

Why SVD?

- SVD is an enduringly popular technique



Data from DBLP by searching “Singular value decomposition” and “SVD”

Why SVD

- Even now still appear in top venues like NeurIPS, ICML, TKDE….

Sami Abu-El-Haija, Hesham Mostafa, Marcel Nassar, Valentino Crespi, Greg Ver Steeg, Aram Galstyan:
Implicit SVD for Graph Representation Learning. NeurIPS 2021: 8419-8431

Aming Wu, Suqi Zhao, Cheng Deng, Wei Liu:
Generalized and Discriminative Few-Shot Object Detection via SVD-Dictionary Enhancement. NeurIPS 2021: 6353-6364

Vasileios Kalantzis, Georgios Kollias, Shashanka Ubaru, Athanasios N. Nikolakopoulos, Lior Horesh, Kenneth L. Clarkson:
Projection techniques to update the truncated SVD of evolving matrices with applications. ICML 2021: 5236-5246

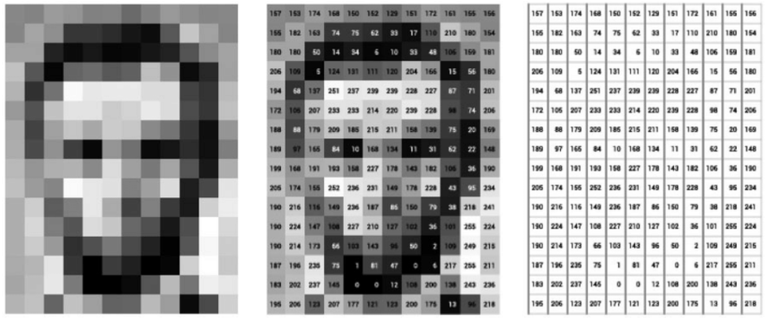
Xiang Li, Shusen Wang, Kun Chen, Zhihua Zhang:
Communication-Efficient Distributed SVD via Local Power Iterations. ICML 2021: 6504-6514

Wenwen Min, Juan Liu, Shihua Zhang:
Group-Sparse SVD Models via \$L_1\$L1- and \$L_0\$L0-norm Penalties and their Applications in Biological Data. IEEE Trans. Knowl. Data Eng. 33(2): 536-550 (2021)

Preliminaries: How to see a matrix?

- As a simple yet useful data structure,

  e.g., an image as a matrix



- As a linear equation

$$\left.\begin{aligned} 2x + y - 3z &= -4 \\ 4x - 2y + z &= 9 \\ 3x + 5y - 2z &= 5 \end{aligned}\right\} \rightarrow \begin{bmatrix} 2 & 1 & -3 & \vdots & -4 \\ 4 & -2 & 1 & \vdots & 9 \\ 3 & 5 & -2 & \vdots & 5 \end{bmatrix}$$

- As a mappings of vectors

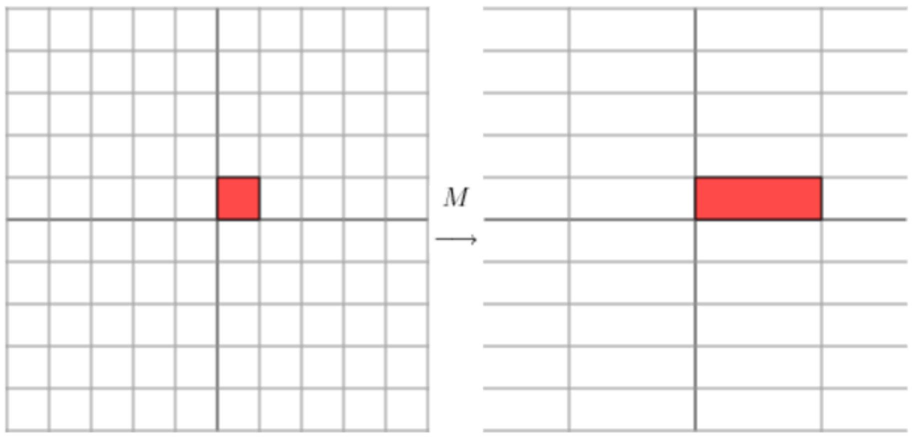
$$A_{[m \times n]} \Leftrightarrow f(.) : \mathbb{R}^{\mathrm{n}} \rightarrow \mathbb{R}^{\mathrm{m}}$$

$$\begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$$

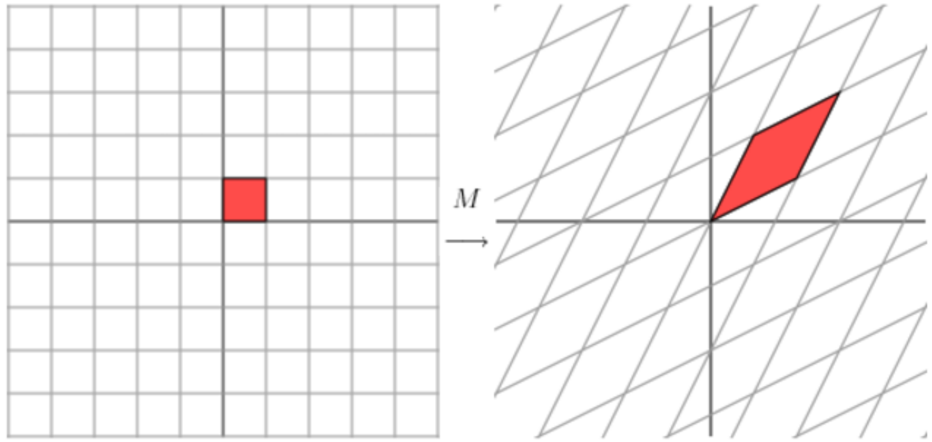
- As a linear transformation

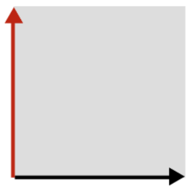
Preliminaries: Matrix as linear transformations

$$M = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

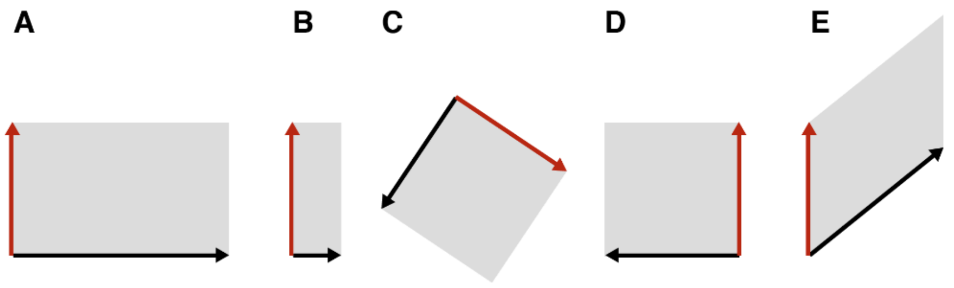


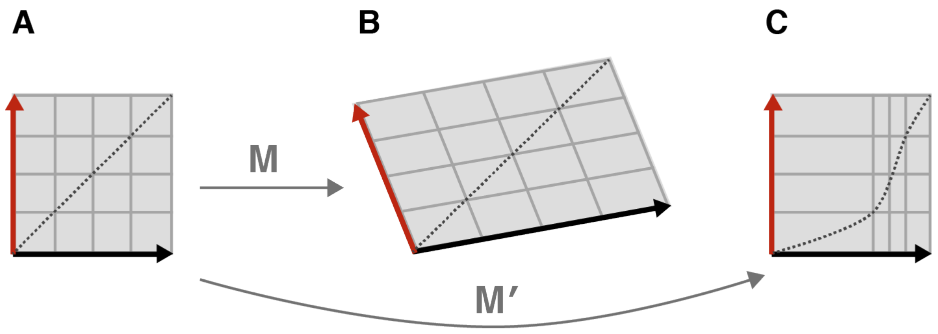
$$M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$





(A) stretched, (B) compressed, (C) rotated, (D) reflected or flipped, and (E) sheared.





M’ is a non-linear transformation!

SVD Theorem

$$A_{[m\times n]} = U_{[m\times m]}\Sigma_{[m\times n]}V^T_{[n\times n]}$$

Orthogonal matrix

$$Q^{-1} = Q^T$$

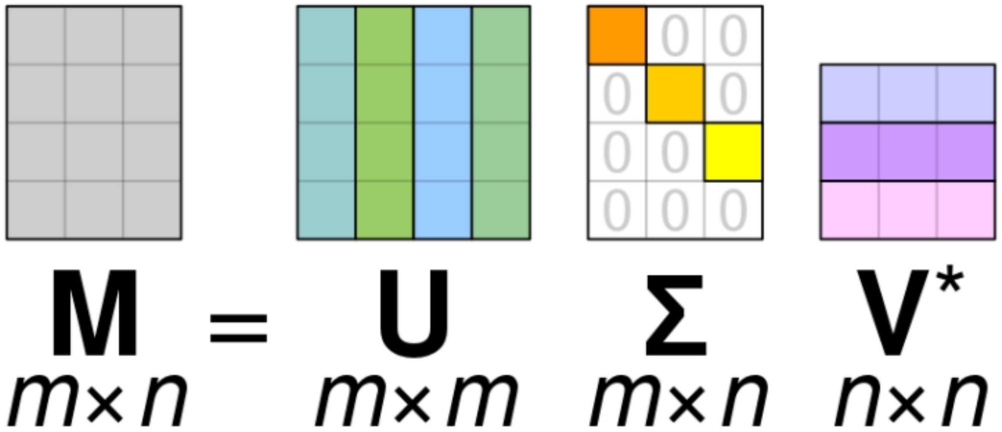
Diagonal matrix

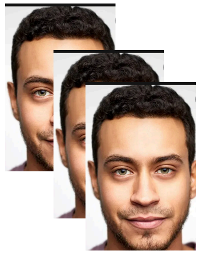
$$Q : q_{ij} = 0 \text{ if } i \neq j$$

A scale transformation

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

$$= \underbrace{\begin{bmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{bmatrix}}_{U} \boxed{\underbrace{\begin{bmatrix} 3.0000 & 0 \\ 0 & 1.0000 \end{bmatrix}}_{\Sigma}} \underbrace{\begin{bmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{bmatrix}}_{V^T}$$





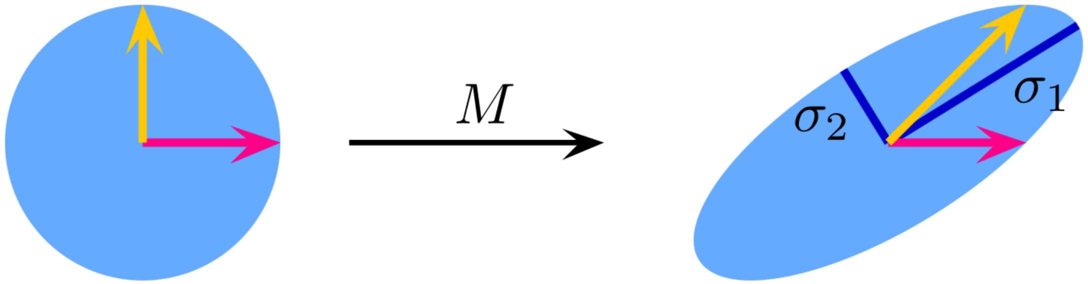
Sigma: ordered by importance of eigenfaces

M is a matrix of million faces, a column vector is a face image

Each columns of **U** (left singular vectors) is a “eignfaces”

Each row of **V** (right singular vectors) is a mixture of **U**s to make a real face

Geometric Explanation of SVD



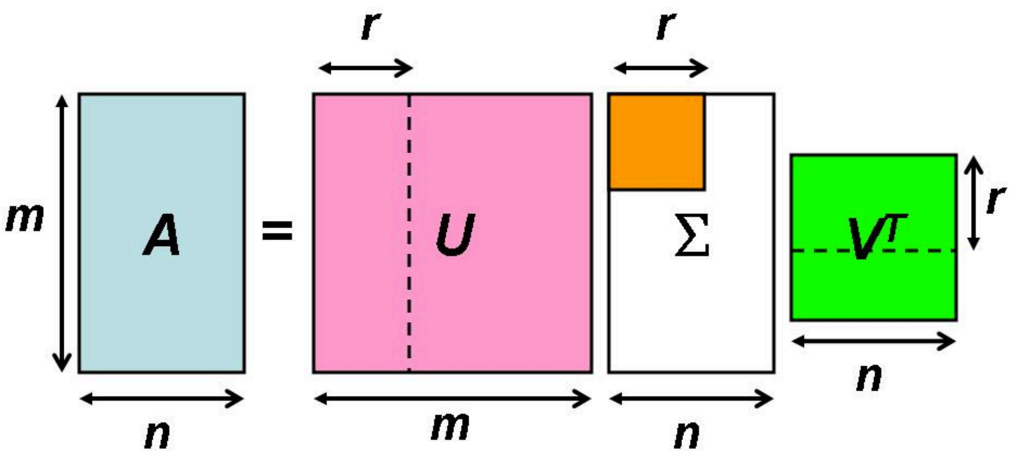
Any linear transformations can be represented by rotation, scaling, and rotation again.

$$M = U \cdot \Sigma \cdot V^*$$

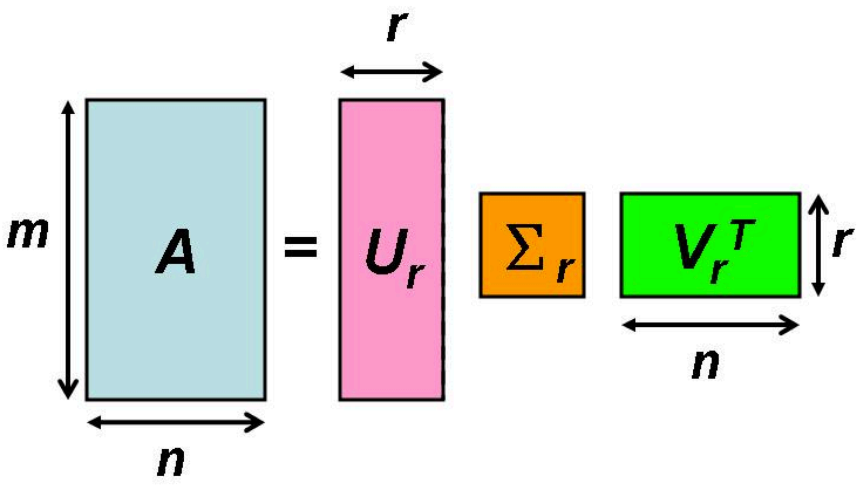
$$A_{[m\times n]} = U_{[m\times m]}\Sigma_{[m\times n]}V^T_{[n\times n]}$$

r is rank of A

$$A_{[m\times n]} = U_{[m\times r]}\Sigma_{[r\times r]}V^T_{[r\times n]}$$



$$A = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$



$$= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T.$$

$$\begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \ddots \end{bmatrix}$$

SVD – Properties

- Always possible to decompose any real matrix $A = U\Sigma V^T$
- Singular values in Sigma are all positive, and sorted in descending order
- The number of nonzero singular values of A equals the rank of A
- Matrix could be approximated using the first r singular values and the corresponding singular vectors
- The sum of the first r singular values over the total sum of singular values indicate the approximation ratio

Popular applications of SVD

- Noise removal
- Data compression
- Dimension reduction
- Latent semantic analysis

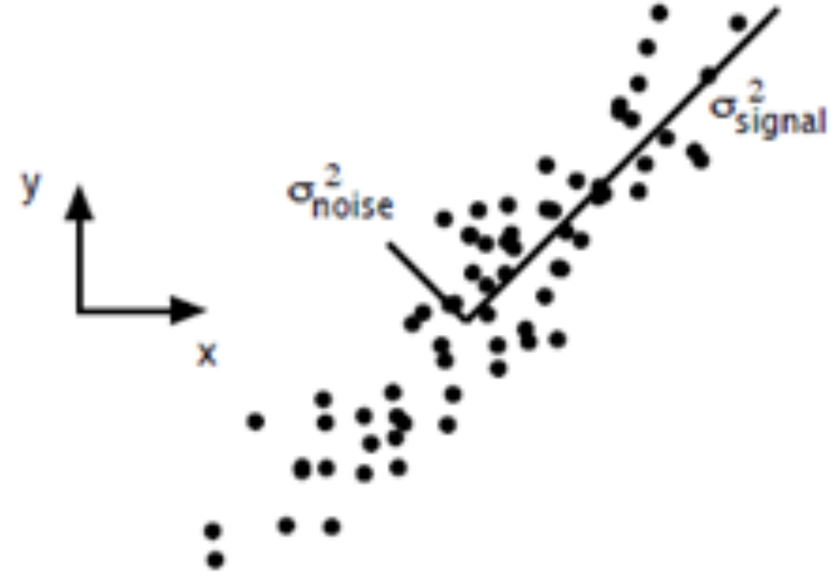
Noise removal

Motivation

- Sensors are vulnerable to noises, sensor readings are thus inaccurate

How

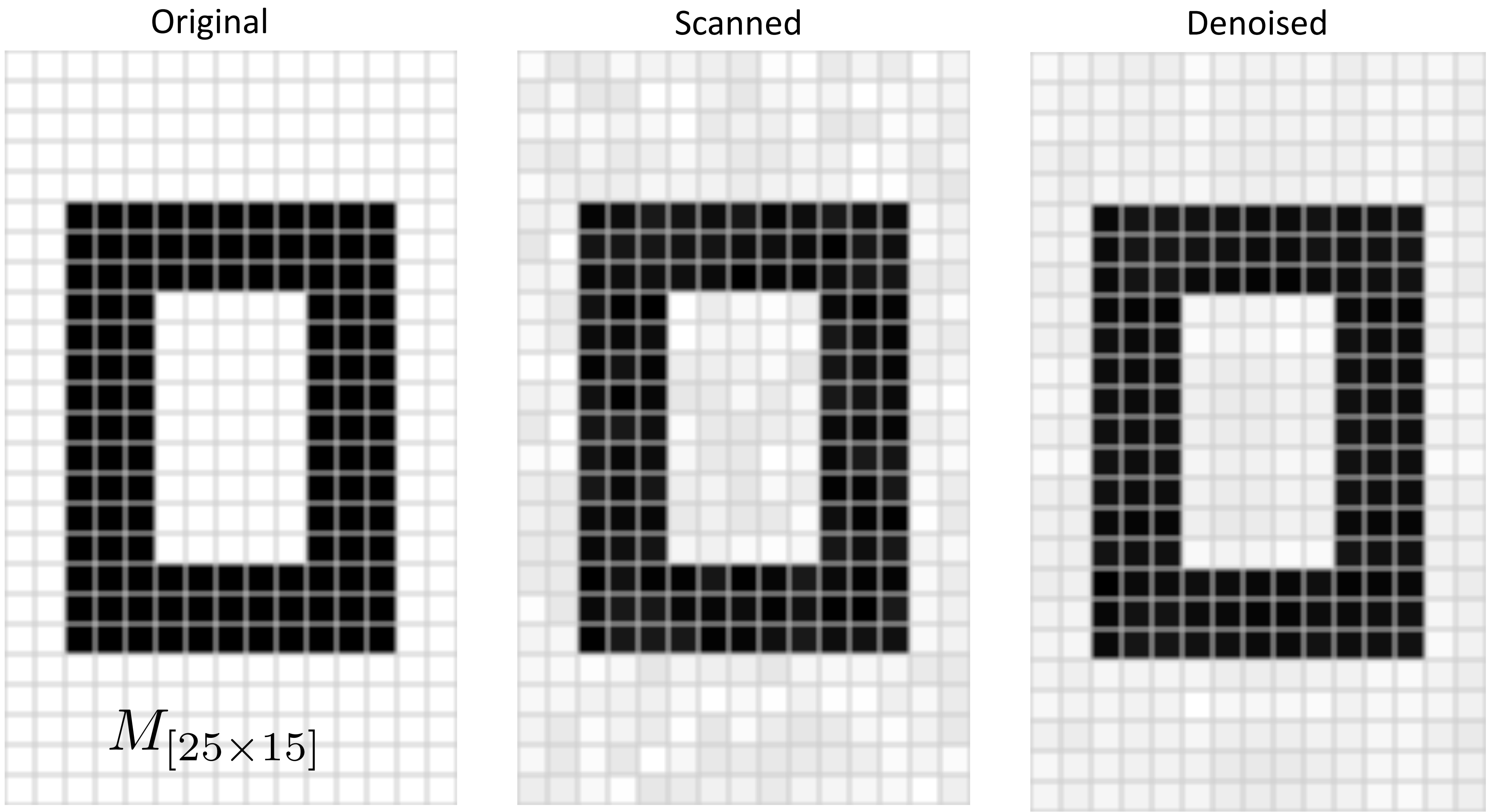
- Set those singular values that are smaller than a threshold to zero



Use a scanner to scan a picture

The scanned picture is with too much noise

$$M_{[25\times 15]} \approx u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + u_3\sigma_3 v_3^T$$



$\sigma_1 = 14.15$
$\sigma_2 = 4.67$
$\sigma_3 = 3.00$

$\sigma_4 = 0.21$
$\sigma_5 = 0.19$
...
$\sigma_{15} = 0.05$

Noise!

Data compression

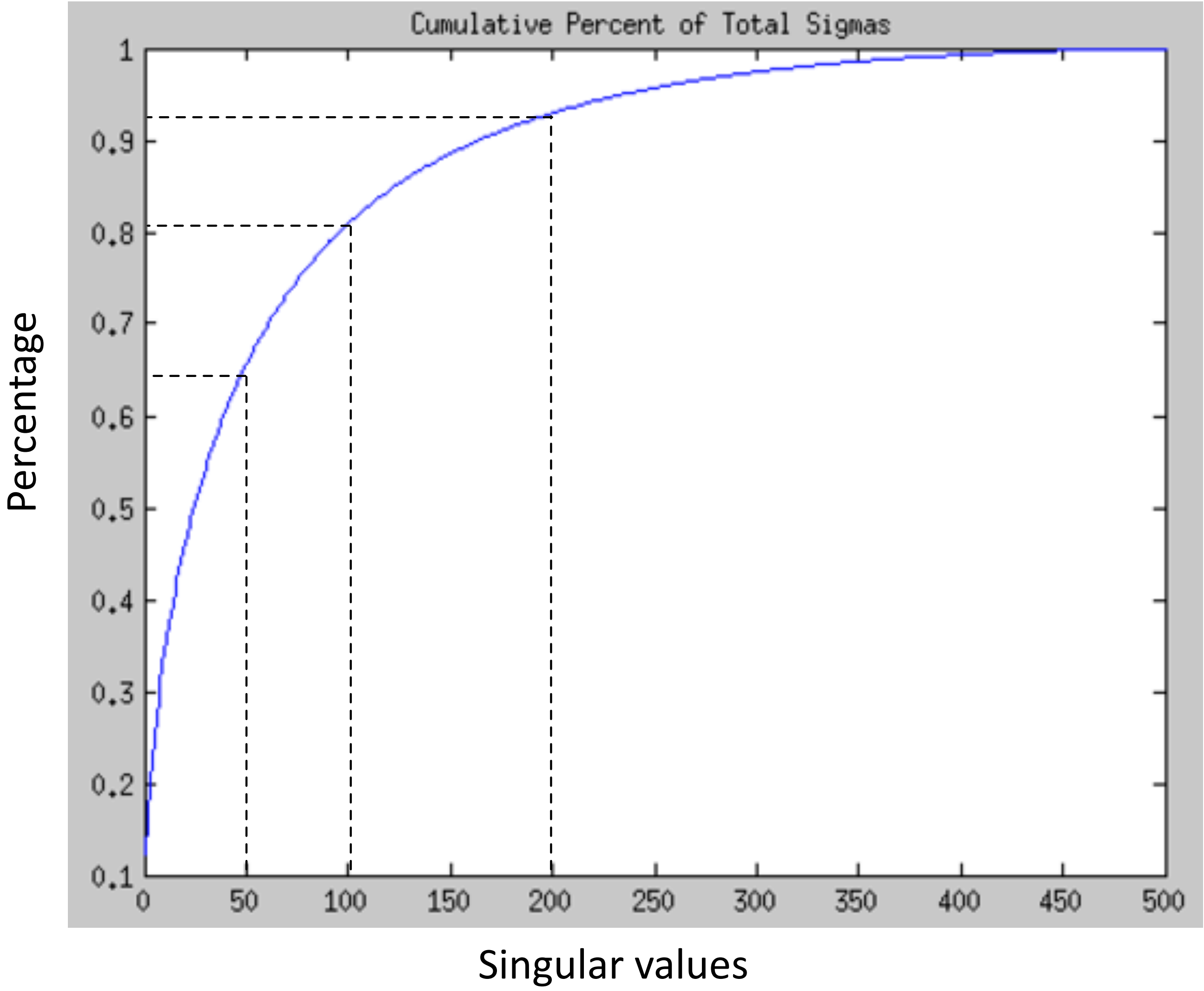
Motivation

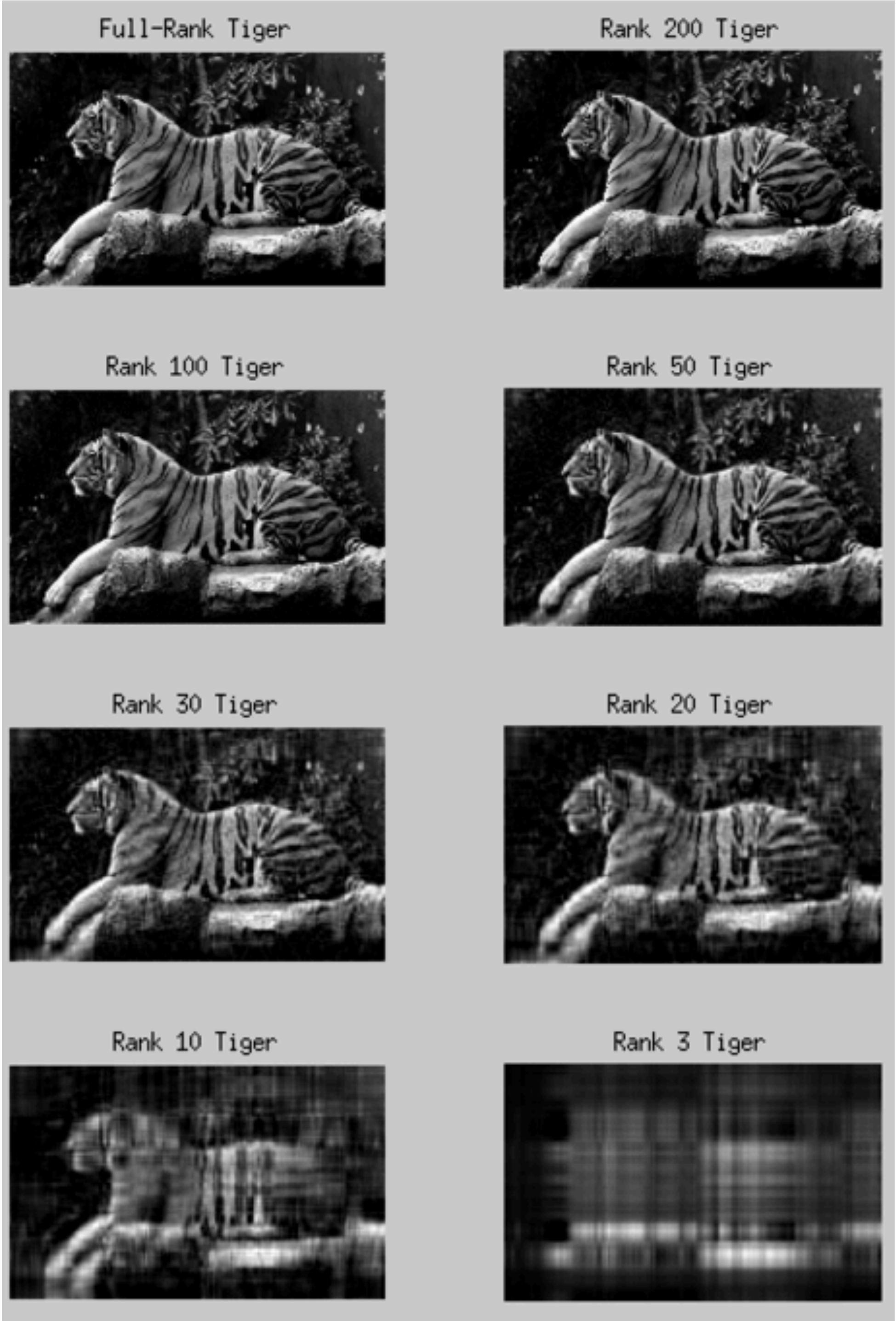
- Resources are limited in IoT devices
  - Storage
  - Bandwidth

How

- Choose *n* largest singular values Σ', *n*-rank approximation $A \approx U\Sigma' V^T$
- The percentage of "information" contained in the approximation matrix is $\dfrac{sum(\Sigma')}{sum(\Sigma)}$

Data compression

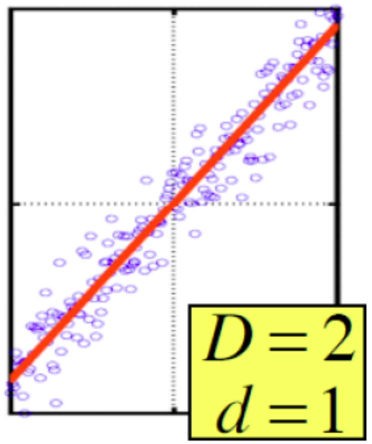


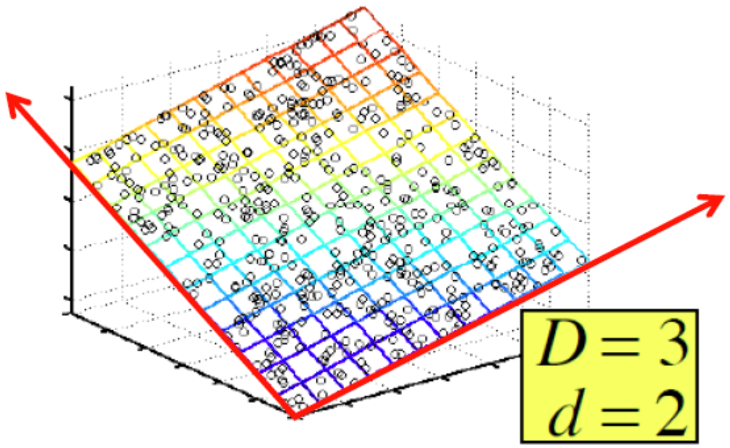


Dimension reduction

Motivation

- Visualize high dimensional data
- Redundant features
- Combine features



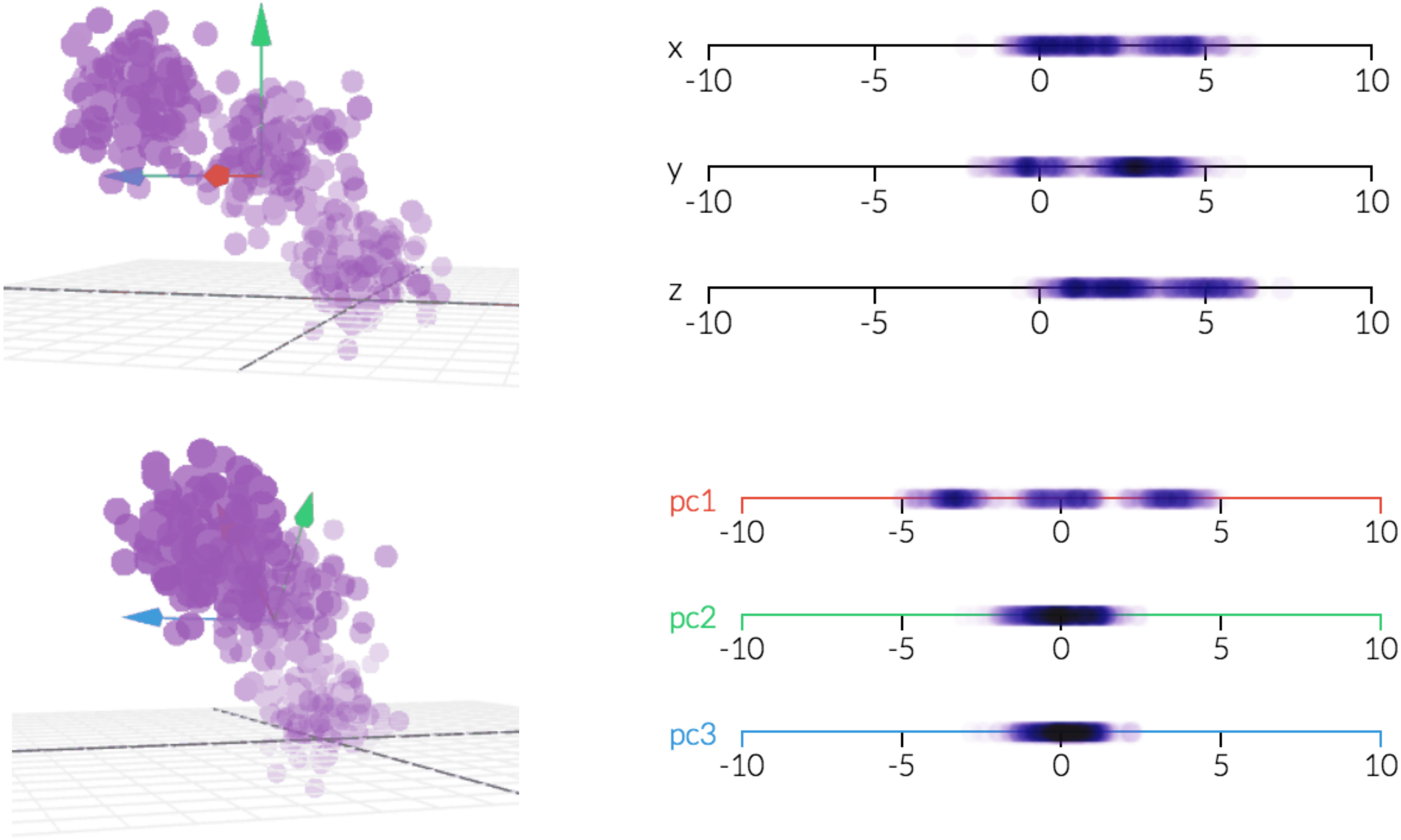


How

- Principal component analysis (PCA)

PCA algorithm

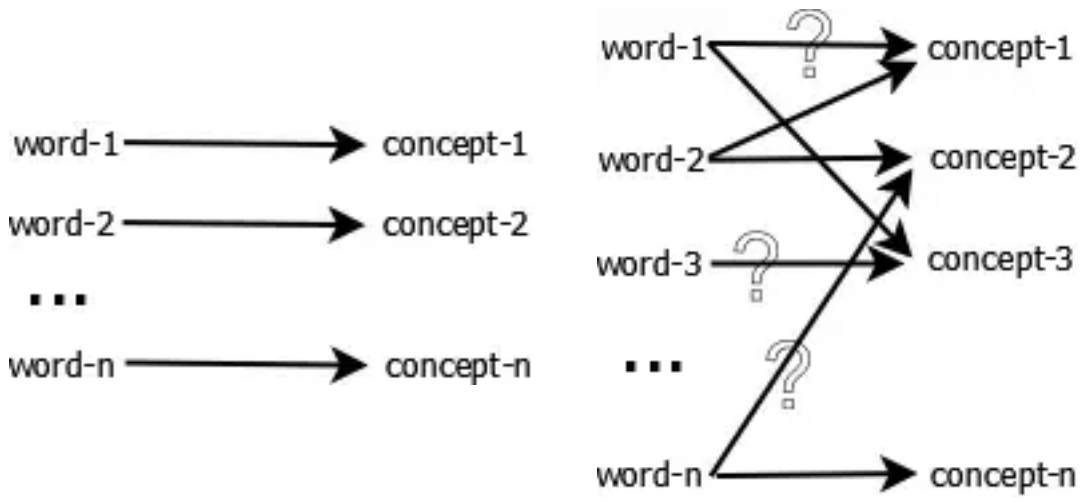
1. Formulate an input matrix A
2. Computer covariance matrix C of A
3. [U S V] = svd(C)


x
-10 -5 0 5 10
y
-10 -5 0 5 10
z
-10 -5 0 5 10
pc1
-10 -5 0 5 10
pc2
-10 -5 0 5 10
pc3
-10 -5 0 5 10

Latent semantic analysis

Motivation

- Polysemy (one word with multiple concepts)
- Synonymy (multiple words with one concept)



LSA or LSI (latent semantic indexing) means analyzing documents to find the underlying meaning or concepts of those documents.

How

- Word matrix (term frequency)
- Fill in missing values
- SVD

Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

	C1	C2	C3
book	0.15	-0.27	0.04
dads	0.24	0.38	-0.09
dummies	0.13	-0.17	0.07
estate	0.18	0.19	0.45
guide	0.22	0.09	-0.46
investing	0.74	-0.21	0.21
market	0.18	-0.30	-0.28
real	0.18	0.19	0.45
rich	0.36	0.59	-0.34
stock	0.25	-0.42	-0.28
value	0.12	-0.14	0.23

*

3.91	0	0
0	2.61	0
0	0	2.00

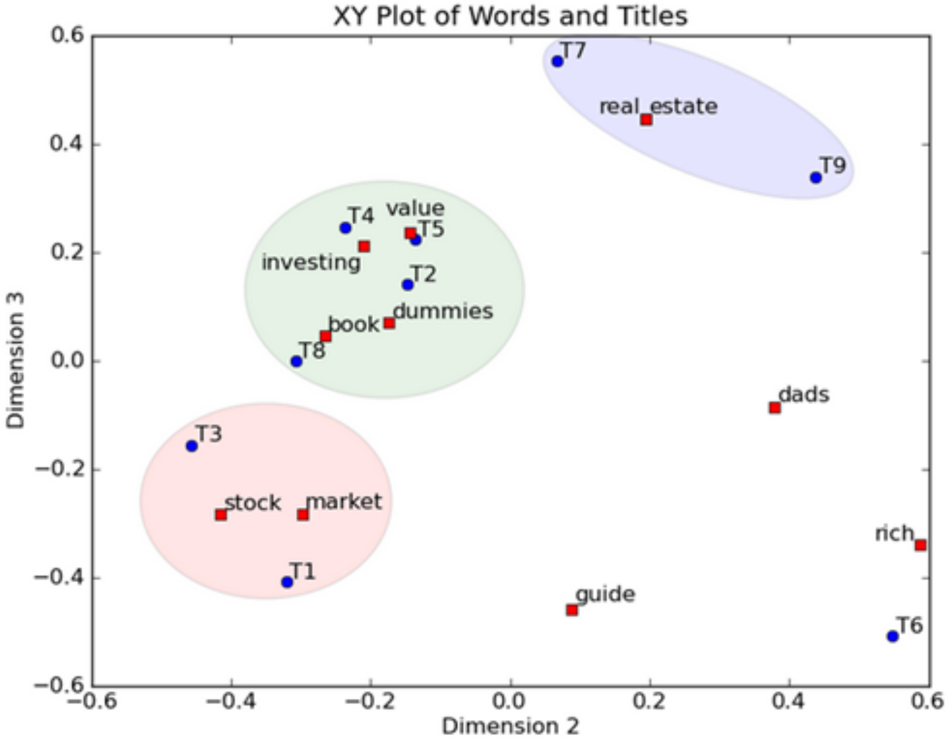
*

	T1	T2	T3	T4	T5	T6	T7	T8	T9
C1	0.35	0.22	0.34	0.26	0.22	0.49	0.28	0.29	0.44
C2	-0.32	-0.15	-0.46	-0.24	-0.14	0.55	0.07	-0.31	0.44
C3	-0.41	0.14	-0.16	0.25	0.22	-0.51	0.55	0.00	0.34

The strength of the concept

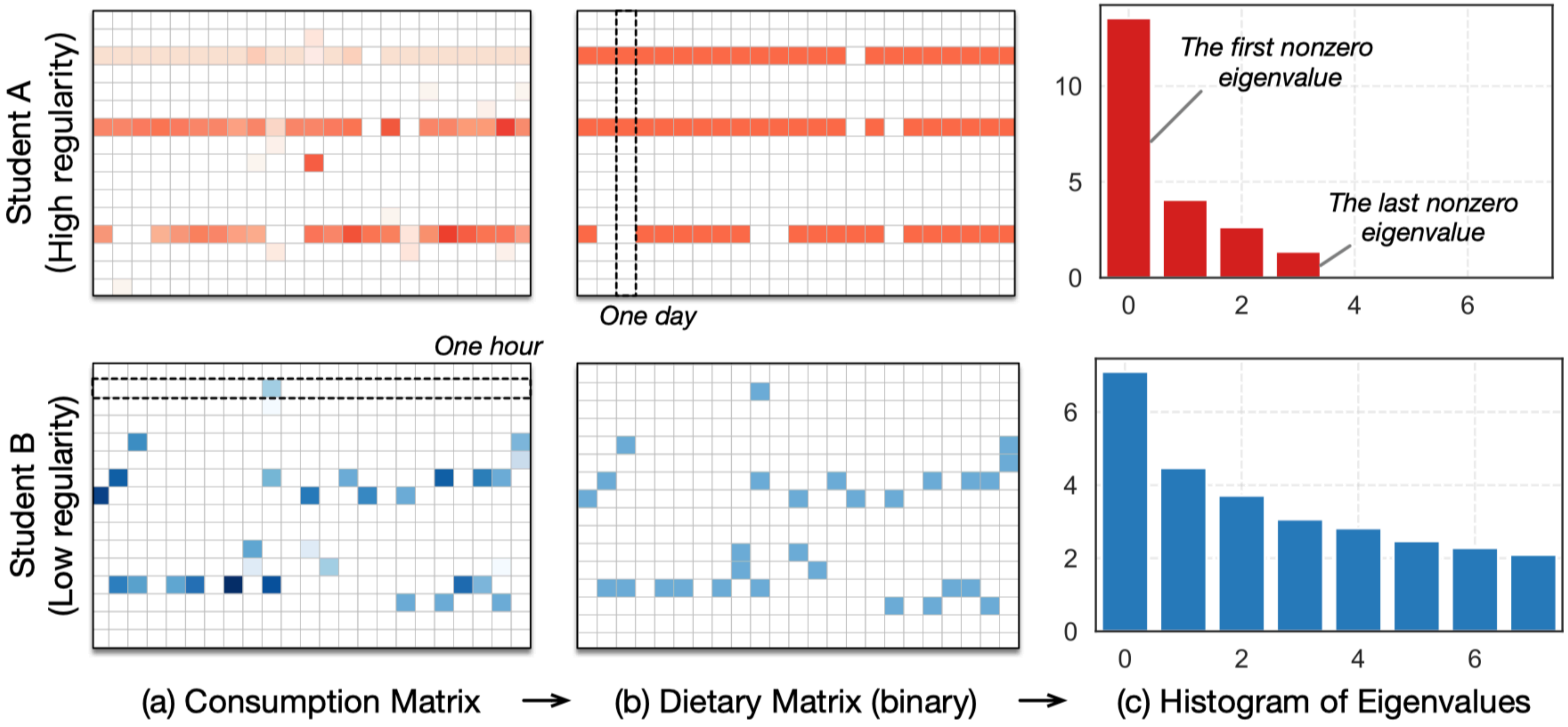
A row represents the relevance or important of articles in the concept

A column represents the relevance or important of books in the concept

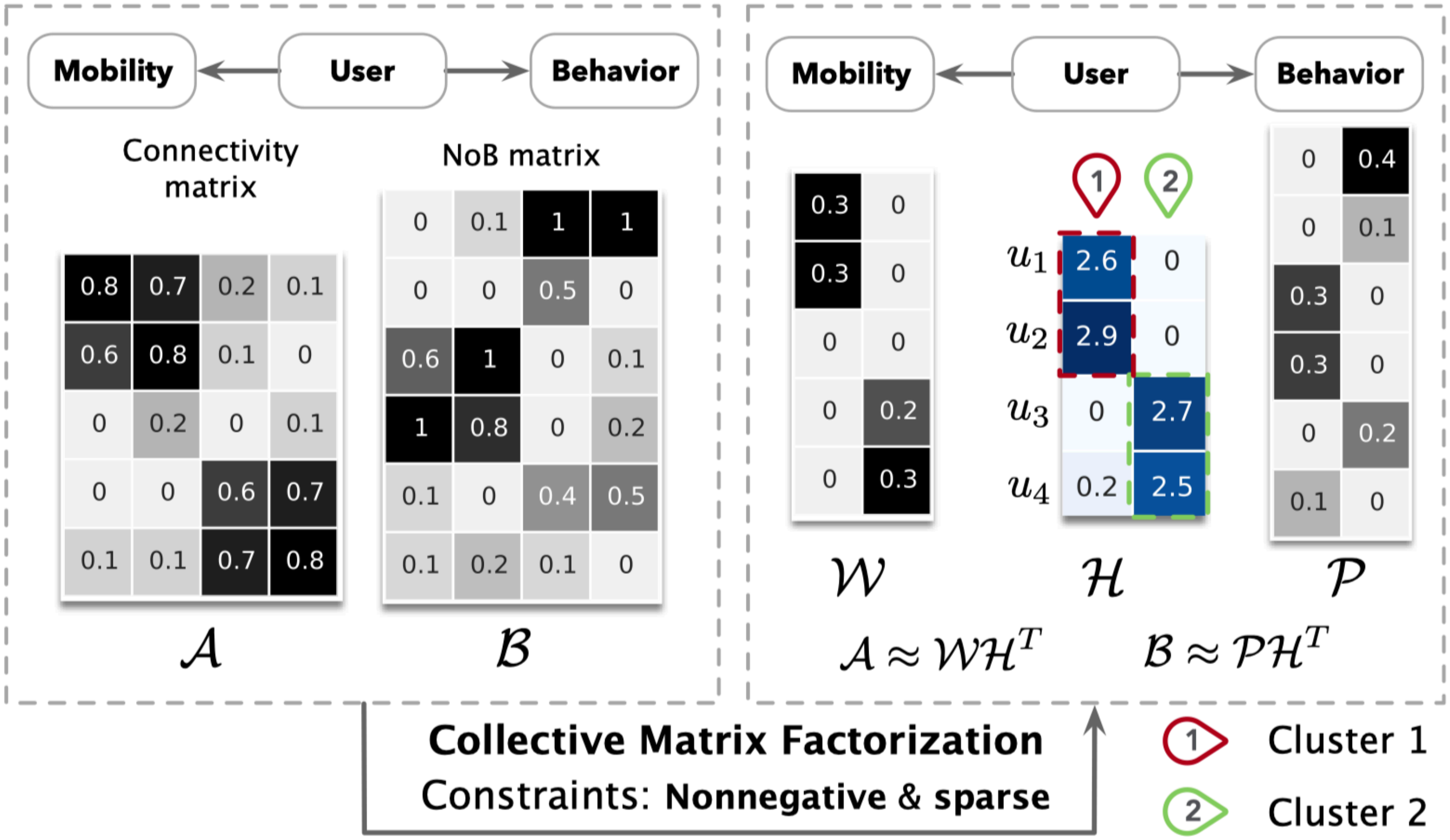


Concept refers to a category of articles, like business, history, or literature.

SVD as a feature extractor



Other matrix factorizations



References

- Gilbert Strang, Linear Algebra and Its Applications. Brooks Cole.
- http://andrew.gibiansky.com/blog/mathematics/cool-linear-algebra-singular-value-decomposition/
- https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/
- http://www.ams.org/samplings/feature-column/fcarc-svd
- https://ccjou.wordpress.com/2009/09/01/%E5%A5%87%E7%95%B0%E5%80%BC%E5%88%86%E8%A7%A3-svd/
- https://www.quora.com/What-is-a-good-explanation-of-Latent-Semantic-Indexing-LSI