

15-826: Multimedia (Databases) and Data Mining

Lecture #16: SVD - part I (definitions)

C. Faloutsos

Must-read Material

- [Numerical Recipes in C](#) ch. 2.6;
- [MM Textbook](#) Appendix D



Outline

Goal: ‘Find **similar / interesting** things’

- Intro to DB
- ➔ • Indexing - similarity search
- ↳ • Data Mining



Indexing - Detailed outline

- primary key indexing
- secondary key / multi-key indexing
- spatial access methods
- fractals
- text
- ➔ • Singular Value Decomposition (SVD)
- multimedia
- ...



SVD - Detailed outline

- 
- Motivation
 - Definition - properties
 - Interpretation
 - Complexity
 - Case studies
 - Additional properties



Problem

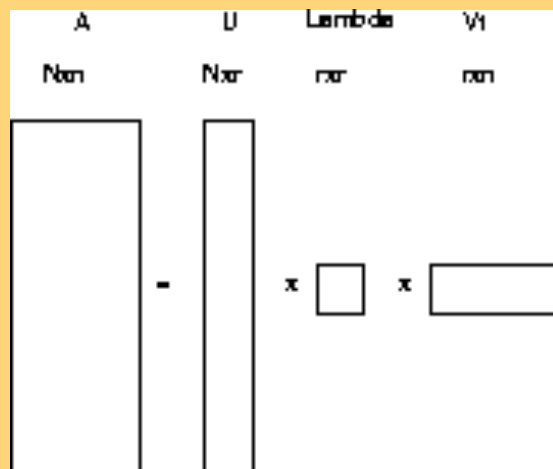
- How to find ‘concepts’ in a set of doc’s?
 - ($\sim$ clusters)





Conclusion

- How to find ‘concepts’ in a set of doc’s?
 - ($\sim$ clusters)
- SVD (= LSI) on the document-term matrix



SVD - Motivation

- problem #1: text - LSI: find ‘concepts’
- problem #2: compression / dim. reduction

SVD - Motivation

- problem #1: text - LSI: find ‘concepts’

term	data	information	retrieval	brain	lung
document					
CS-TR1	1	1	1	0	0
CS-TR2	2	2	2	0	0
CS-TR3	1	1	1	0	0
CS-TR4	5	5	5	0	0
MED-TR1	0	0	0	2	2
MED-TR2	0	0	0	3	3
MED-TR3	0	0	0	1	1

SVD - Motivation

- Customer-product, for recommendation system:

	bread	lettuce	tomatos	beef	chicken
↑	1	1	1	0	0
↓	2	2	2	0	0
↑	1	1	1	0	0
↓	5	5	5	0	0
↑	0	0	0	2	2
↓	0	0	0	3	3
↑	0	0	0	1	1
↓					

SVD - Motivation

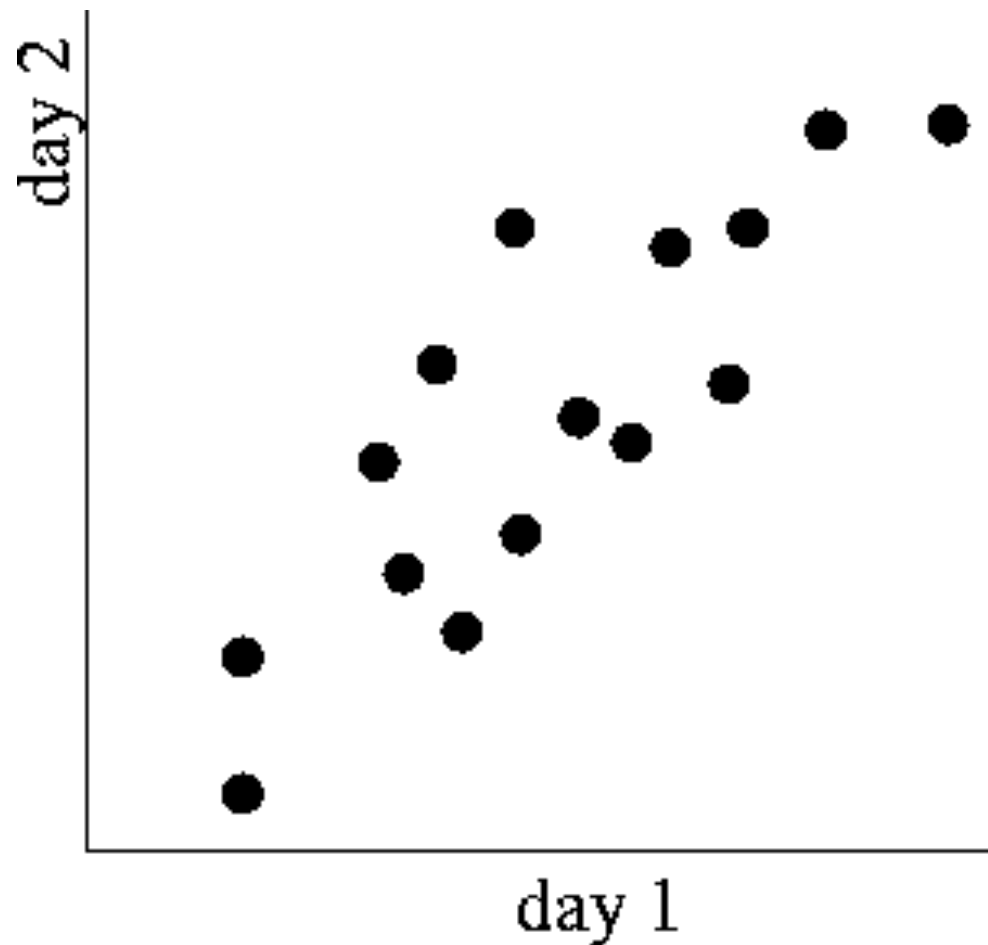
- problem #2: compress / reduce dimensionality

Problem - specs

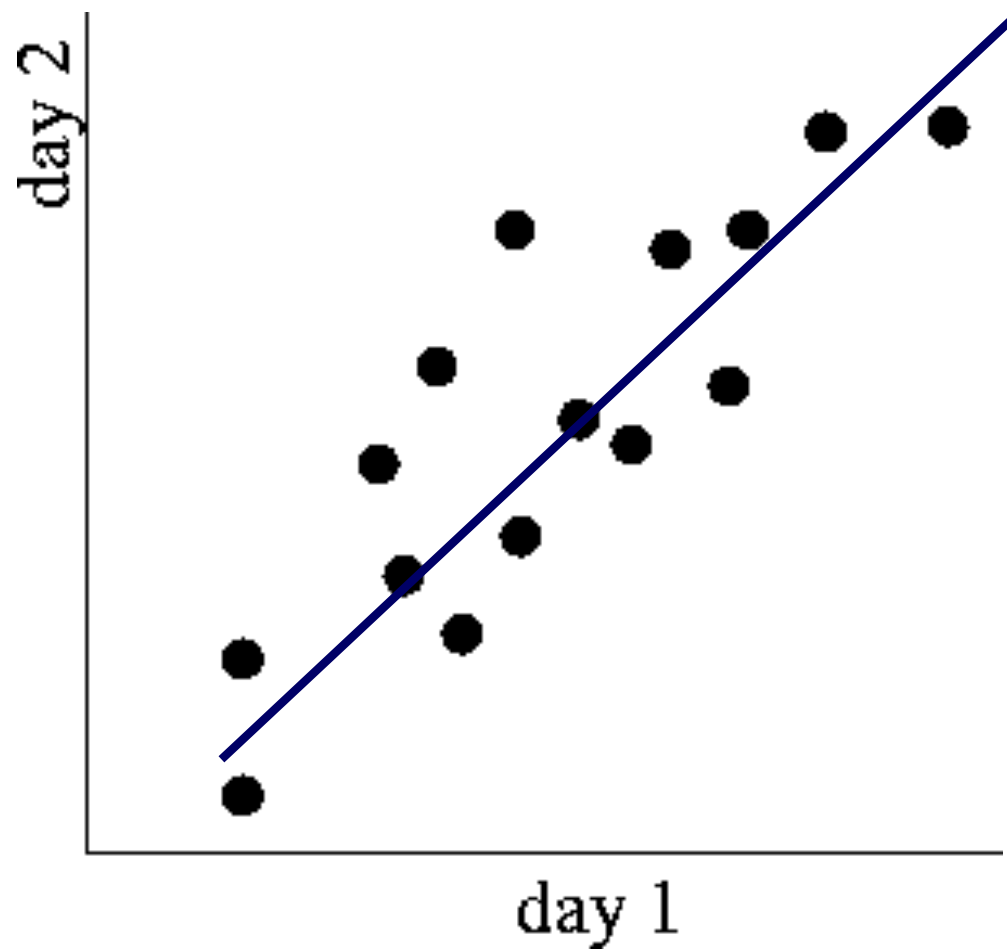
- $\sim 10^6$ rows; $\sim 10^3$ columns; no updates;
- random access to any cell(s) ; small error: OK

customer	day	We	Th	Fr	Sa	Su
		7/10/96	7/11/96	7/12/96	7/13/96	7/14/96
ABC Inc.		1	1	1	0	0
DEF Ltd.		2	2	2	0	0
GHI Inc.		1	1	1	0	0
KLM Co.		5	5	5	0	0
Smith		0	0	0	2	2
Johnson		0	0	0	3	3
Thompson		0	0	0	1	1

SVD - Motivation



SVD - Motivation





SVD - Detailed outline

- Motivation
- ➔ • Definition - properties
- Interpretation
- Complexity
- Case studies
- Additional properties

SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}$$

$3 \times 2 \qquad 2 \times 1$

SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}$$

A diagram illustrating the dimensions of the matrices in the multiplication above. It shows three dimension labels: 3×2 , 2×1 , and 3×1 . A long curved arrow points from the 2 in 3×2 to the 2 in 2×1 . A shorter curved arrow points from the 1 in 2×1 to the 1 in 3×1 . Another curved arrow points from the 3 in 3×2 to the 3 in 3×1 .

SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \end{bmatrix}$$

$3 \times 2 \quad 2 \times 1 \quad 3 \times 1$

SVD - Definition

(reminder: matrix multiplication)

$$\begin{array}{c}
 \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}
 \end{array}$$

$3 \times 2 \quad \quad 2 \times 1 \quad \quad 3 \times 1$

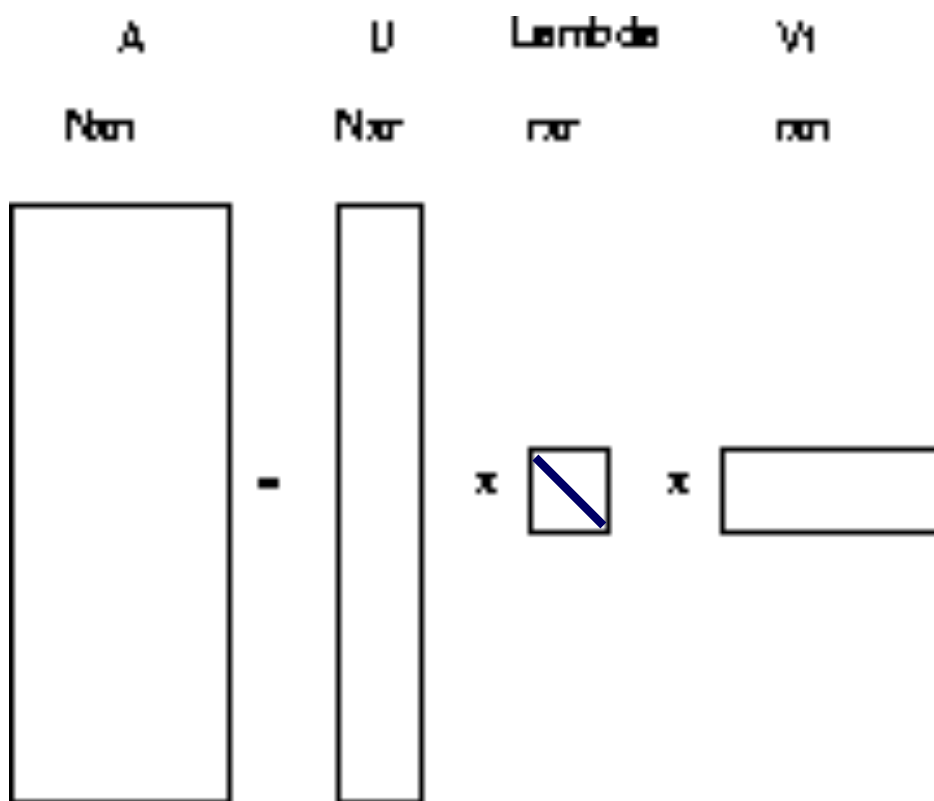
SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

SVD - Definition

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:



SVD - Definition

$$\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} (\mathbf{V}_{[m \times r]})^T$$

- $\mathbf{A}$: $n \times m$ matrix (eg., n documents, m terms)
- $\mathbf{U}$: $n \times r$ matrix (n documents, r concepts)
- $\mathbf{\Lambda}$: $r \times r$ diagonal matrix (strength of each 'concept') (r : rank of the matrix)
- $\mathbf{V}$: $m \times r$ matrix (m terms, r concepts)

SVD - Properties

THEOREM [Press+92]: always possible to decompose matrix $\mathbf{A}$ into $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$, where

- $\mathbf{U}, \mathbf{\Lambda}, \mathbf{V}$: unique (*)
- $\mathbf{U}, \mathbf{V}$: column orthonormal (ie., columns are unit vectors, orthogonal to each other)
 - $\mathbf{U}^T \mathbf{U} = \mathbf{I}; \mathbf{V}^T \mathbf{V} = \mathbf{I}$ ($\mathbf{I}$: identity matrix)
- $\mathbf{\Lambda}$: singular are positive, and sorted in decreasing order

SVD - Example

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

retrieval
inf. ↓
data brain lung

$$\begin{array}{c} \uparrow \\ \text{CS} \\ \downarrow \\ \uparrow \\ \text{MD} \\ \downarrow \end{array}
 \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}
 =
 \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}
 \times
 \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}
 \times
 \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Example

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

retrieval
 data inf. ↓ brain lung

CS-concept
 MD-concept

CS
 MD

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Example

- $A = U \Lambda V^T$ - example: doc-to-concept
similarity matrix

retrieval
CS-concept
MD-concept

data
inf. ↓
brain
lung

↑

CS

↓

↑

MD

↓

[

1

2

1

5

0

0

0

]

=

[

0.18

0.36

0.18

0.90

0

0.53

0

0.27

]

×

[

9.64

0

]

×

[

0.58

0.58

0.58

0

0

]

SVD - Example

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

retrieval
inf. ↓ brain lung

‘strength’ of CS-concept

↑

CS

↓

↑

MD

↓

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

=

$$\begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix}$$

×

$$\begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix}$$

×

$$\begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Example

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

term-to-concept similarity matrix

Diagram illustrating the CS-MD matrix multiplication process:

CS Matrix (7x5):

1	1	1	0	0
2	2	2	0	0
1	1	1	0	0
5	5	5	0	0
0	0	0	2	2
0	0	0	3	3
0	0	0	1	1

MD Matrix (5x2):

0.18	0
0.36	0
0.18	0
0.90	0
0	0.53
0	0.80
0	0.27

CS-concept Matrix (7x2):

9.64	0
0	5.29

Term-to-Concept Similarity Matrix (2x5):

0.58	0.58	0.58	0	0
0	0	0	0.71	0.71

The process involves multiplying the CS matrix by the MD matrix to produce the CS-concept matrix, and then multiplying the CS-concept matrix by the term-to-concept similarity matrix to produce the final term-to-term similarity matrix.

SVD - Example

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

retrieval
data inf. ↓ brain lung

term-to-concept
similarity matrix

CS
↑
↓
↑
MD
↓

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

CS-concept



SVD - Detailed outline

- Motivation
- Definition - properties
- • Interpretation
- Complexity
- Case studies
- Additional properties

SVD - Interpretation #1

‘documents’, ‘terms’ and ‘concepts’ :

- U : document-to-concept similarity matrix
- V : term-to-concept sim. matrix
- Λ : its diagonal elements: ‘strength’ of each concept

SVD – Interpretation #1

‘documents’, ‘terms’ and ‘concepts’ :

Q: if $\mathbf{A}$ is the document-to-term matrix, what is $\mathbf{A}^T \mathbf{A}$?

A:

Q: $\mathbf{A} \mathbf{A}^T$?

A:

SVD – Interpretation #1

‘documents’, ‘terms’ and ‘concepts’ :

Q: if $\mathbf{A}$ is the document-to-term matrix, what is $\mathbf{A}^T \mathbf{A}$?

A: term-to-term ($[m \times m]$) similarity matrix

Q: $\mathbf{A} \mathbf{A}^T$?

A: document-to-document ($[n \times n]$) similarity matrix

SVD properties

- V are the eigenvectors of the *covariance matrix* $A^T A$
- U are the eigenvectors of the *Gram (inner-product) matrix* $A A^T$

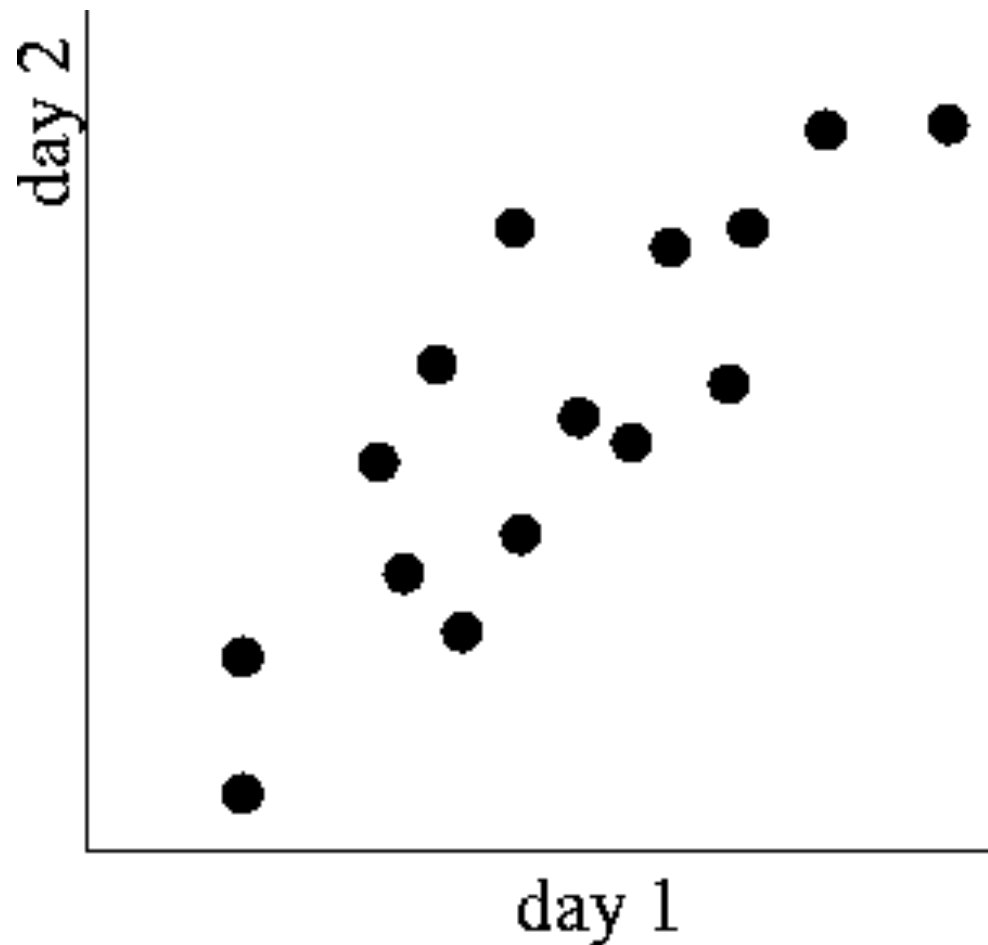
Further reading:

1. Ian T. Jolliffe, *Principal Component Analysis* (2nd ed), Springer, 2002.
2. Gilbert Strang, *Linear Algebra and Its Applications* (4th ed), Brooks Cole, 2005.

SVD - Interpretation #2

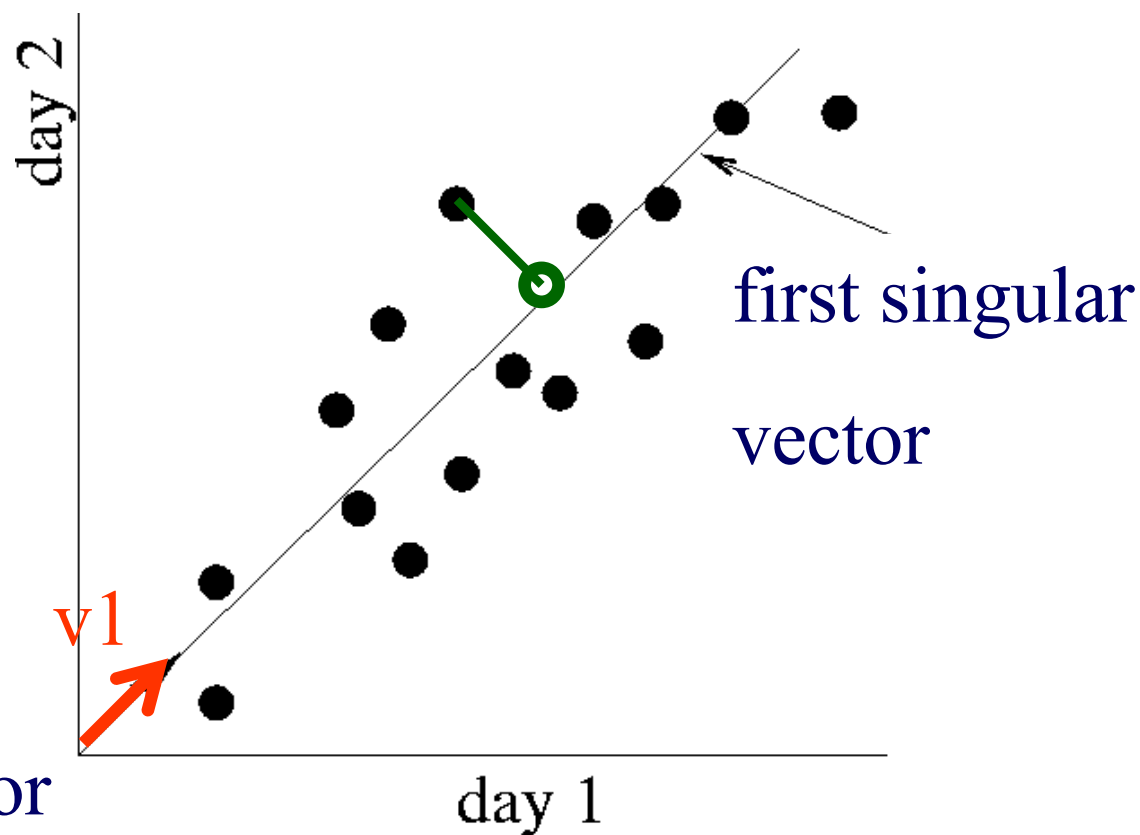
- best axis to project on: (‘best’ = min sum of squares of projection errors)

SVD - Motivation



SVD - interpretation #2

SVD: gives
best axis to project



- minimum RMS error

SVD - Interpretation #2

customer	day	We	Th	Fr	Sa	Su
		7/10/96	7/11/96	7/12/96	7/13/96	7/14/96
ABC Inc.		1	1	1	0	0
DEF Ltd.		2	2	2	0	0
GHI Inc.		1	1	1	0	0
KLM Co.		5	5	5	0	0
Smith		0	0	0	2	2
Johnson		0	0	0	3	3
Thompson		0	0	0	1	1

SVD - Interpretation #2

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

v1

SVD - Interpretation #2

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:

variance ('spread') on the v_1 axis

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #2

- $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$ - example:
 - $\mathbf{U} \mathbf{\Lambda}$ gives the coordinates of the points in the projection axis

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #2

- More details
- Q: how exactly is dim. reduction done?

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #2

- More details
- Q: how exactly is dim. reduction done?
- A: set the smallest singular values to zero:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #2

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #2

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #2

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 0.18 \\ 0.36 \\ 0.18 \\ 0.90 \\ 0 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 9.64 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \end{bmatrix}$$

SVD - Interpretation #2

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

SVD - Interpretation #2

Exactly equivalent:

‘spectral decomposition’ of the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

Preliminaries: Spectral form

$$U \times V^T = u_1 \times v_1^T + u_2 \times v_2^T$$

SVD – Spectral form

Exactly equivalent:
 ‘spectral decomposition’ of the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} | & | \\ u_1 & u_2 \\ | & | \end{bmatrix} \times \begin{bmatrix} \lambda_1 & \emptyset \\ \emptyset & \lambda_2 \end{bmatrix} \times \begin{bmatrix} \text{---} & v_1 & \text{---} \\ \text{---} & v_2 & \text{---} \end{bmatrix}$$

SVD – Spectral form

Exactly equivalent:

‘spectral decomposition’ of the matrix:

$$\begin{array}{c} \updownarrow \\ n \end{array} \begin{array}{c} \leftarrow m \rightarrow \\ \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \end{array} = \lambda_1 \underbrace{u_1}_{\text{column}} \underbrace{v_1^T}_{\text{row}} + \lambda_2 \underbrace{u_2}_{\text{column}} \underbrace{v_2^T}_{\text{row}} + \dots$$

SVD – Spectral form

Exactly equivalent:

‘spectral decomposition’ of the matrix:

$$\begin{array}{c} \updownarrow \\ n \end{array} \begin{array}{c} \leftarrow m \rightarrow \\ \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \end{array} = \begin{array}{c} \leftarrow r \text{ terms} \rightarrow \\ \lambda_1 \quad \begin{array}{c} u_1 \\ \nearrow \\ n \times 1 \end{array} \quad \begin{array}{c} v_1^T \\ \nwarrow \\ 1 \times m \end{array} + \lambda_2 \quad u_2 \quad v_2^T + \dots \end{array}$$

SVD – Spectral form

approximation / dim. reduction:

by keeping the first few terms (Q: how many?)

$$\begin{array}{c} \updownarrow \\ n \end{array} \begin{array}{c} \leftarrow m \rightarrow \\ \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \end{array} = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \dots$$

assume: $\lambda_1 \geq \lambda_2 \geq \dots$

SVD - Spectral form

A (heuristic - [Fukunaga]): keep 80-90% of 'energy' (= sum of squares of λ_i 's)

$$\begin{array}{c} \updownarrow \\ n \end{array} \begin{array}{c} \leftarrow m \rightarrow \\ \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \end{array} = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \dots$$

assume: $\lambda_1 \geq \lambda_2 \geq \dots$

SVD - Spectral form

Also, if there are clear blocks, each corresponds to a ‘concept’/singular-vector-pair

$$\begin{array}{c} \updownarrow n \end{array} \begin{array}{c} \leftarrow m \rightarrow \\ \left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] \end{array} = \underbrace{\lambda_1 \quad u_1 \quad v_1^T}_{} + \underbrace{\lambda_2 \quad u_2 \quad v_2^T}_{} + \dots$$

assume: $\lambda_1 \geq \lambda_2 \geq \dots$

SVD – Spectral form

$$U \times V^T = u_1 \times v_1^T + u_2 \times v_2^T$$



SVD – Spectral form

Arithmetic example:

$$\begin{aligned}
 \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 10 & 2 \\ -10 & -2 \end{bmatrix} &= \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \times [10 \quad 2] + \\
 &\quad \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \times [-10 \quad -2] \\
 &= \begin{bmatrix} -10 & -2 \\ -10 & -2 \\ -10 & -2 \end{bmatrix}
 \end{aligned}$$



SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
 - #1: documents/terms/concepts
 - #2: dim. reduction
 - – #3: picking non-zero, rectangular ‘blobs’
- Complexity
- Case studies
- Additional properties

SVD - Interpretation #3

- finds non-zero ‘blobs’ in a data matrix

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Interpretation #3

- finds non-zero ‘blobs’ in a data matrix

$$\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right] = \left[\begin{array}{cc} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{array} \right] \times \left[\begin{array}{cc} 9.64 & 0 \\ 0 & 5.29 \end{array} \right] \times \left[\begin{array}{ccccc} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{array} \right]$$

SVD - Interpretation #3

- finds non-zero ‘blobs’ in a data matrix =
- ‘communities’ (bi-partite cores, here)

$$\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array} \right]$$

Row 1



Row 4



Row 5



Row 7

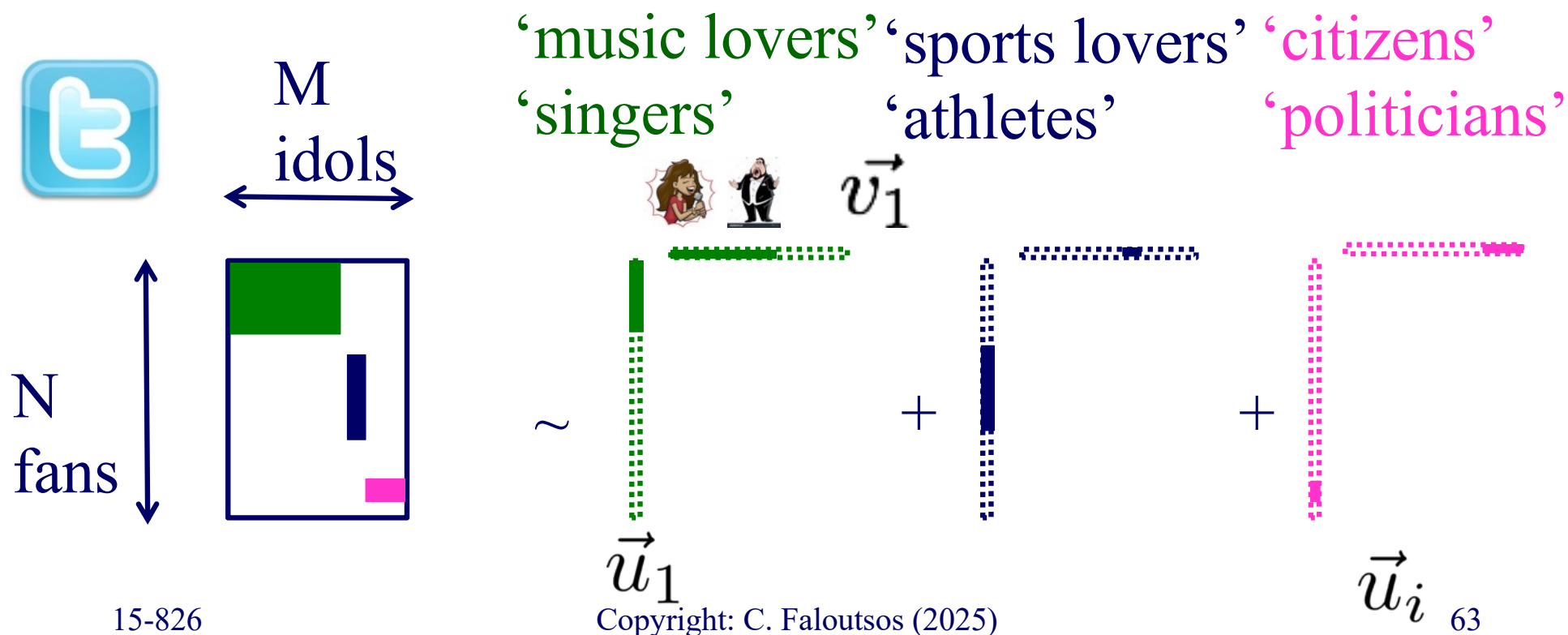


SVD – Interpretation #3

- Another pictorial example
- And its connection to EigenSpokes

SVD – Interpretation #3

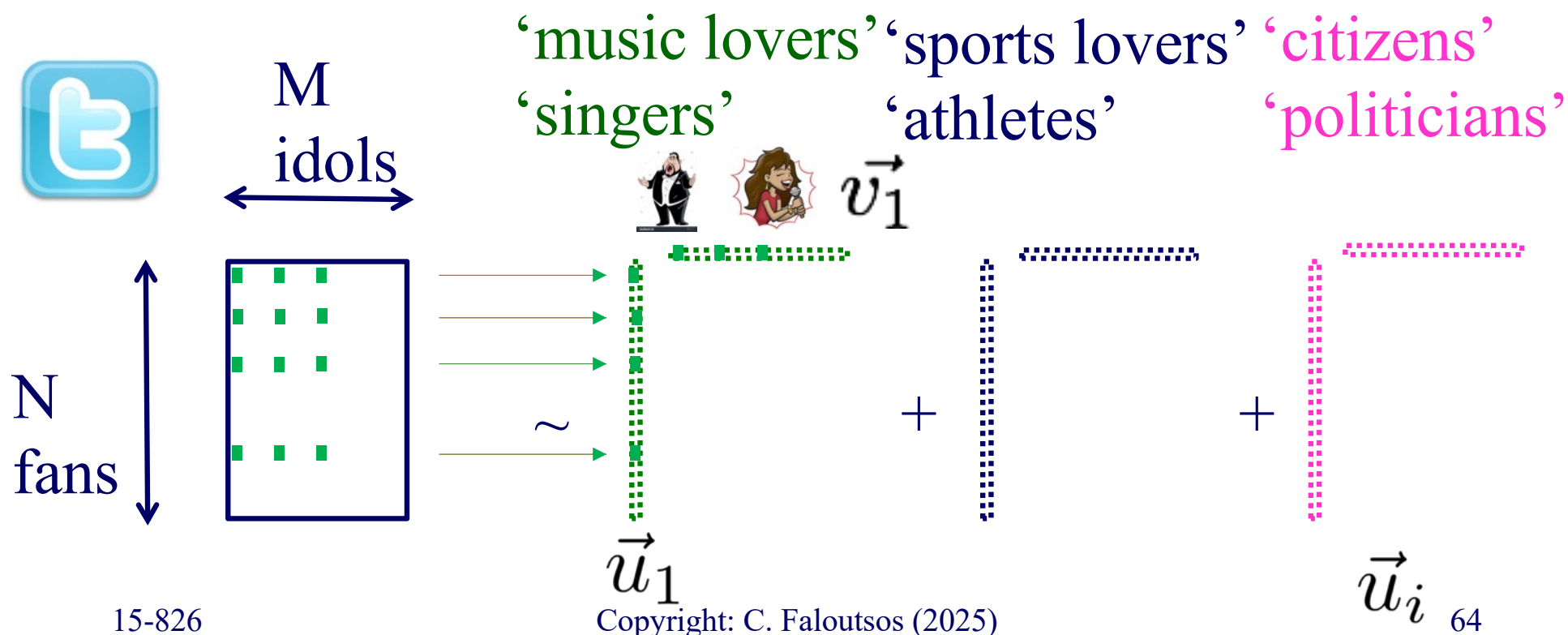
- It finds blocks



SVD – Interpretation #3

- It finds blocks

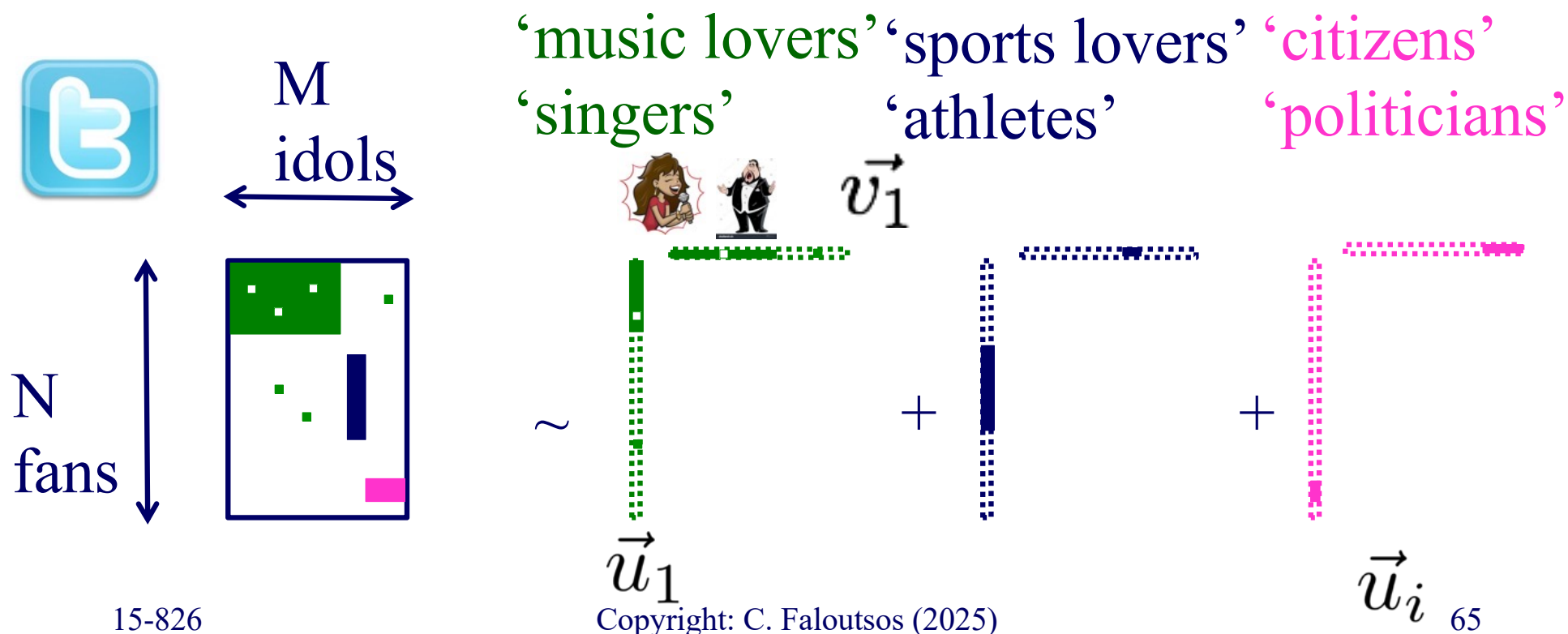
A) Even if shuffled!

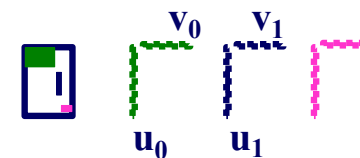


SVD – Interpretation #3

- It finds blocks

B) Even if ‘salt+pepper’ noise



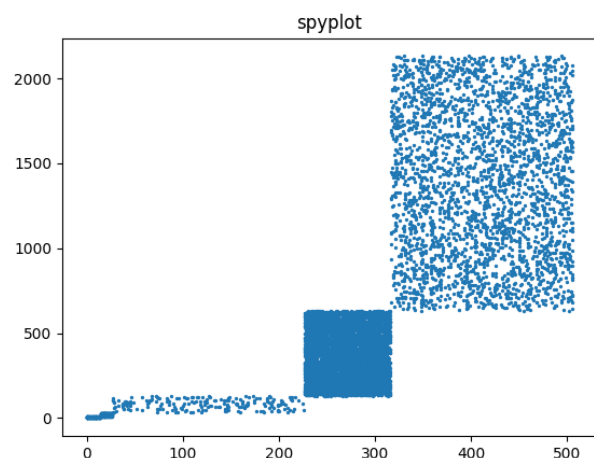


Toy example – EigenSpokes

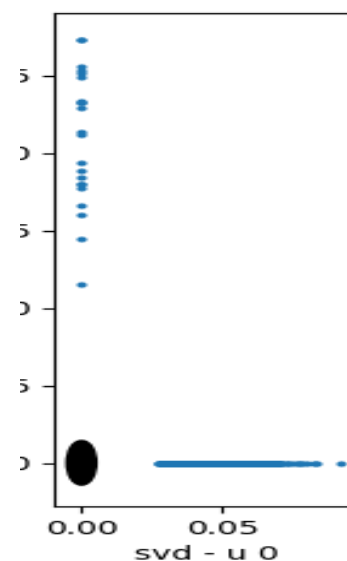
‘idols’



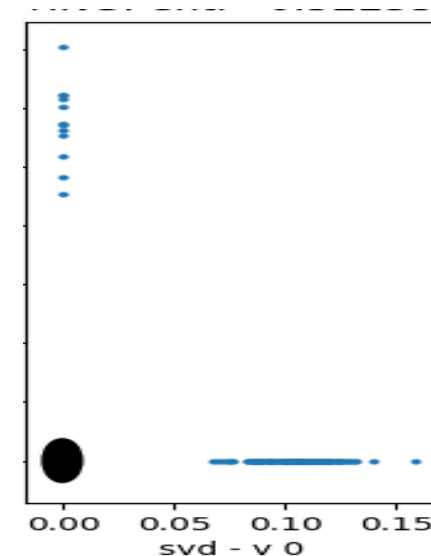
‘fans’



u_1



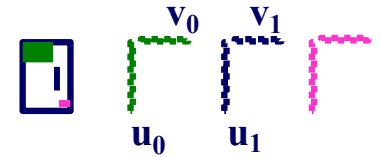
v_1



u_0

v_0

EigenPlots

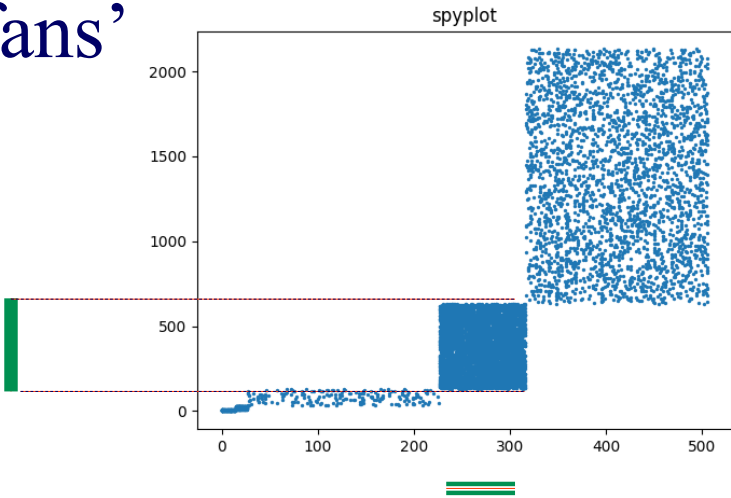


Toy example – EigenSpokes

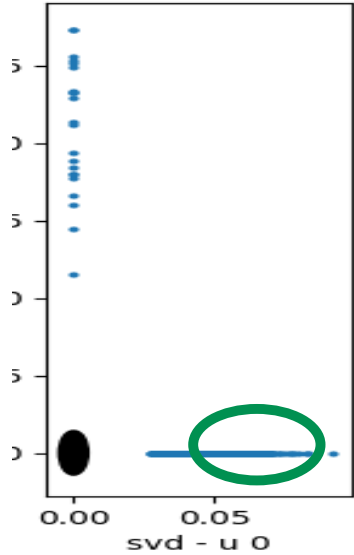
‘idols’



‘fans’

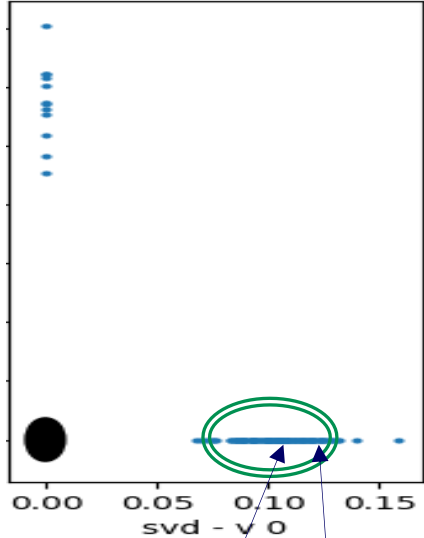


u_1



u_0

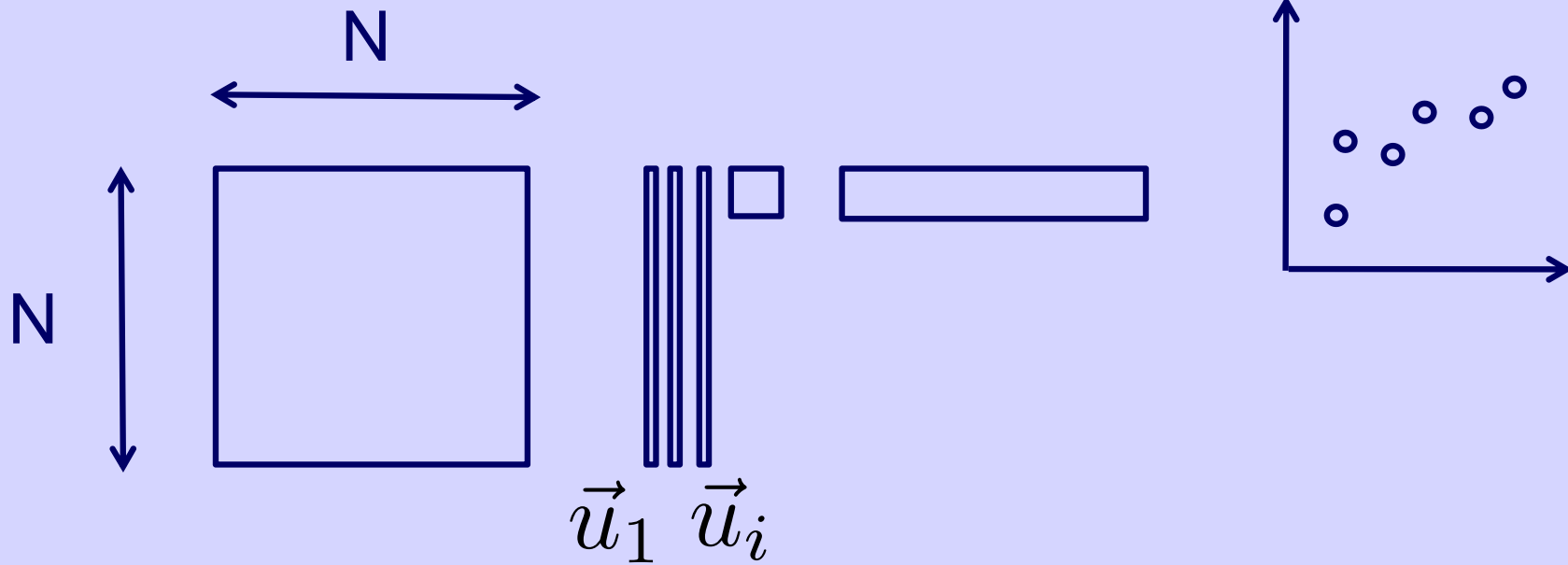
v_1



v_0

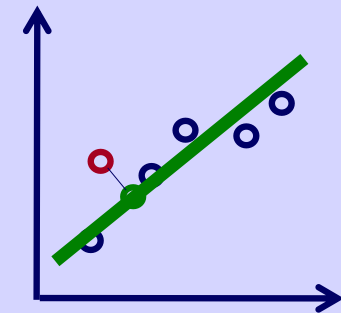
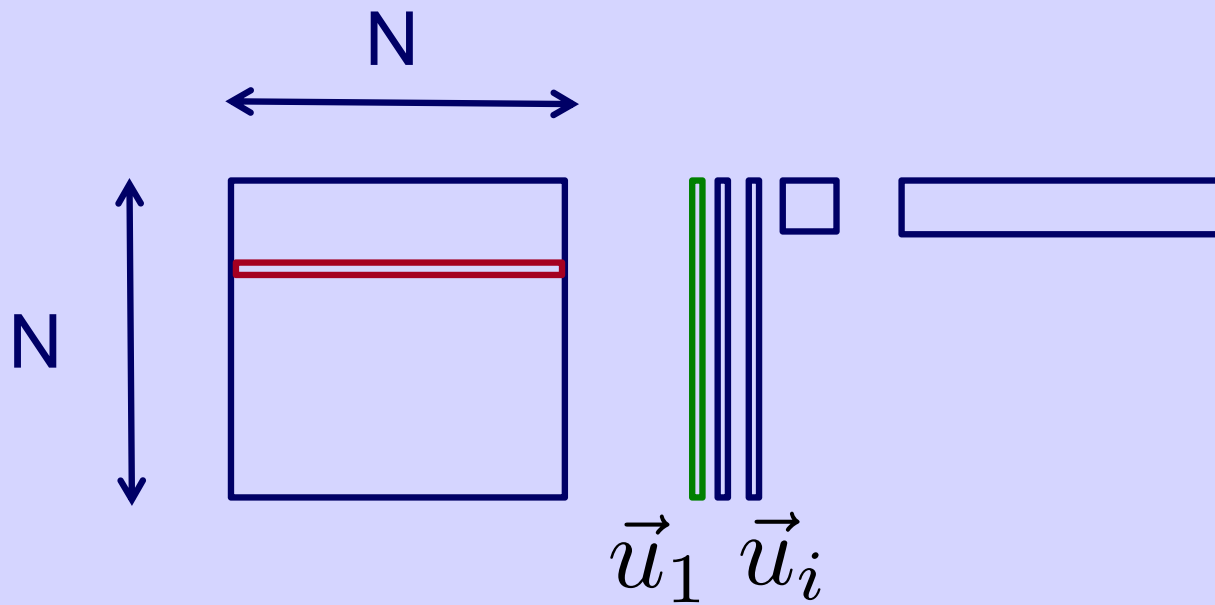
EigenSpokes

- Eigenvectors of adjacency matrix



EigenSpokes

- Eigenvectors of adjacency matrix



SVD - Drill

- Compute SVD by hand
 - On a block-structured, toy matrix
 - Exploiting the properties of SVD

SVD - Drill

- Drill: find the SVD, ‘by inspection’ !
- Q: rank = ??

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} ?? \\ \\ \\ \\ \end{bmatrix} \times \begin{bmatrix} ?? \\ \\ \\ \end{bmatrix} \times \begin{bmatrix} ?? \\ \\ \\ \\ \end{bmatrix}$$

SVD - Drill

- A: rank = 2 (2 linearly independent rows/cols)

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} | & | \\ ?? & ?? \\ | & | \end{bmatrix} \times \begin{bmatrix} ?? & 0 \\ 0 & ?? \end{bmatrix} \times \begin{bmatrix} \text{---} & ?? & \text{---} \\ \text{---} & ?? & \text{---} \end{bmatrix}$$

SVD - Drill

- A: rank = 2 (2 linearly independent rows/cols)

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} ?? & 0 \\ 0 & ?? \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

orthogonal??

SVD - Drill

- column vectors: are orthogonal - but not unit vectors:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix} \times \begin{bmatrix} ?? & 0 \\ 0 & ?? \end{bmatrix} \times \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

SVD - Drill

- and the singular values are:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix} \times \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \times \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

SVD - Drill

- Q: How to check we are correct?

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{bmatrix} \times \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \times \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

SVD - Drill

- A: SVD properties:
 - matrix product should give back matrix A
 - matrix U should be column-orthonormal, i.e., columns should be unit vectors, orthogonal to each other
 - ditto for matrix V
 - matrix Λ should be diagonal, with positive values



SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- ➔ • Complexity
- Case studies
- Additional properties

SVD - Complexity

- $O(n * m * m)$ or $O(n * n * m)$ (whichever is less)
- less work, if we just want singular values
- or if we want first k singular vectors
- or if the matrix is **sparse** [Berry]
- Implemented: in any linear algebra package (LINPACK, matlab, Splus/R, mathematica ...)

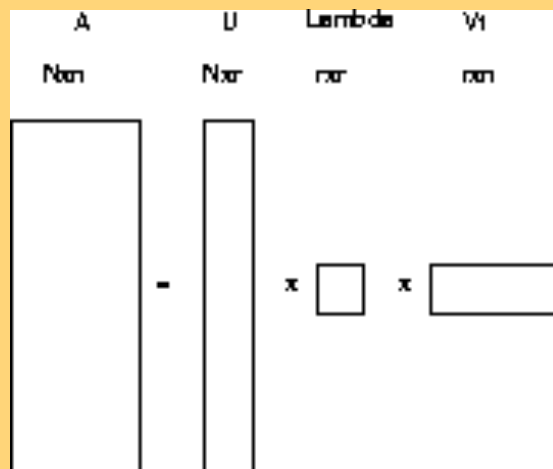
SVD - conclusions so far

- SVD: $A = U \Lambda V^T$: unique (*)
- U : document-to-concept similarities
- V : term-to-concept similarities
- Λ : strength of each concept
- dim. reduction: keep the first few strongest singular values (80-90% of 'energy')
 - SVD: picks up linear correlations
- SVD: picks up non-zero 'blobs'



Conclusion

- How to find ‘concepts’ in a set of doc’s?
 - ($\sim$ clusters)
- SVD (= LSI) on the document-term matrix



References

- Berry, Michael: <http://www.cs.utk.edu/~lsi/>
- Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition, Academic Press.
- Press, W. H., S. A. Teukolsky, et al. (1992) (3rd ed.: 2007): *Numerical Recipes in C*, Cambridge University Press.