CS665: Advanced Data Mining

Lecture#15: SVD-1
U Kang
KAIST

Outline

➡️ ☐ **Motivation**

☐ Definition - properties

☐ Interpretation

☐ Complexity

☐ Conclusion

SVD - Motivation

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

SVD - Motivation

- problem #1: text - LSI: find 'concepts'

term document	data	information	retrieval	brain	lung
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 :

$$
\begin{array}{c}
\text{vegetarians} \\
\text{meat eaters}
\end{array}
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
$$

bread lettuce tomatos beef chicken

SVD - Motivation
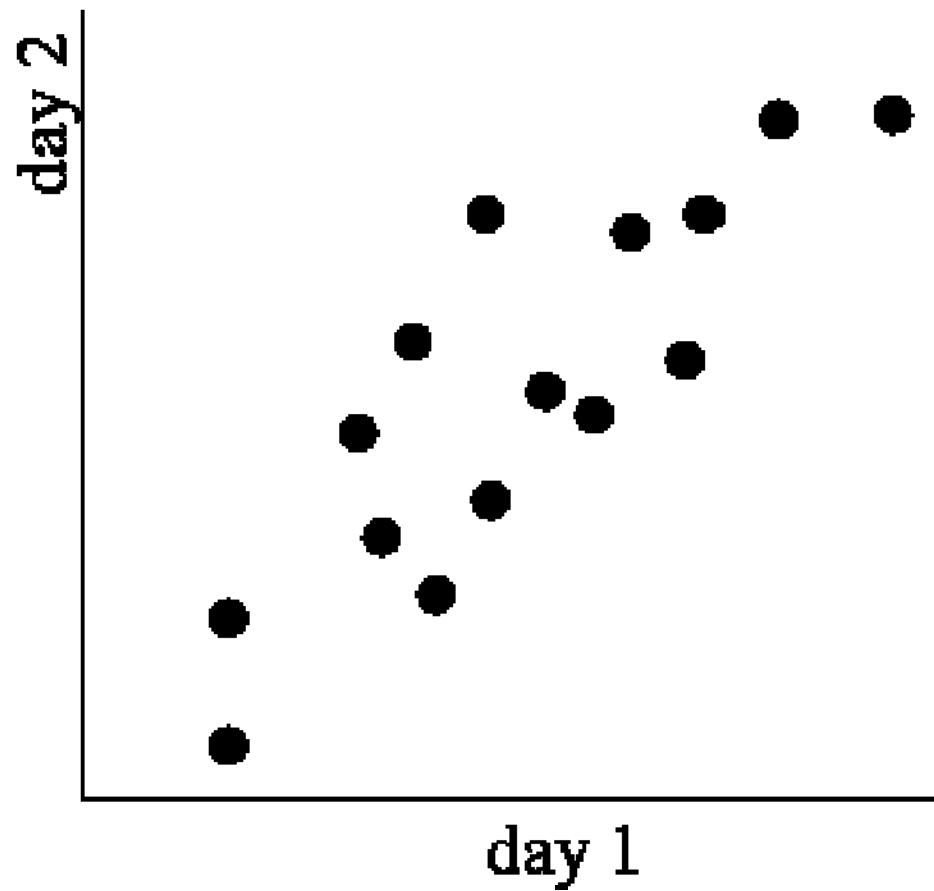
- problem #2: compress / reduce dimensionality

Problem - specs

- ~10**6 rows; ~10**3 columns; no updates;
- random access to any cell(s) ; small error: OK

day customer	Wo 7/10/06	Th 7/11/06	Fr 7/12/06	Sa 7/13/06	Su 7/14/06
ABC Inc.	1	1	1	0	0
DEF Ltd.	2	2	2	0	0
GHT 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
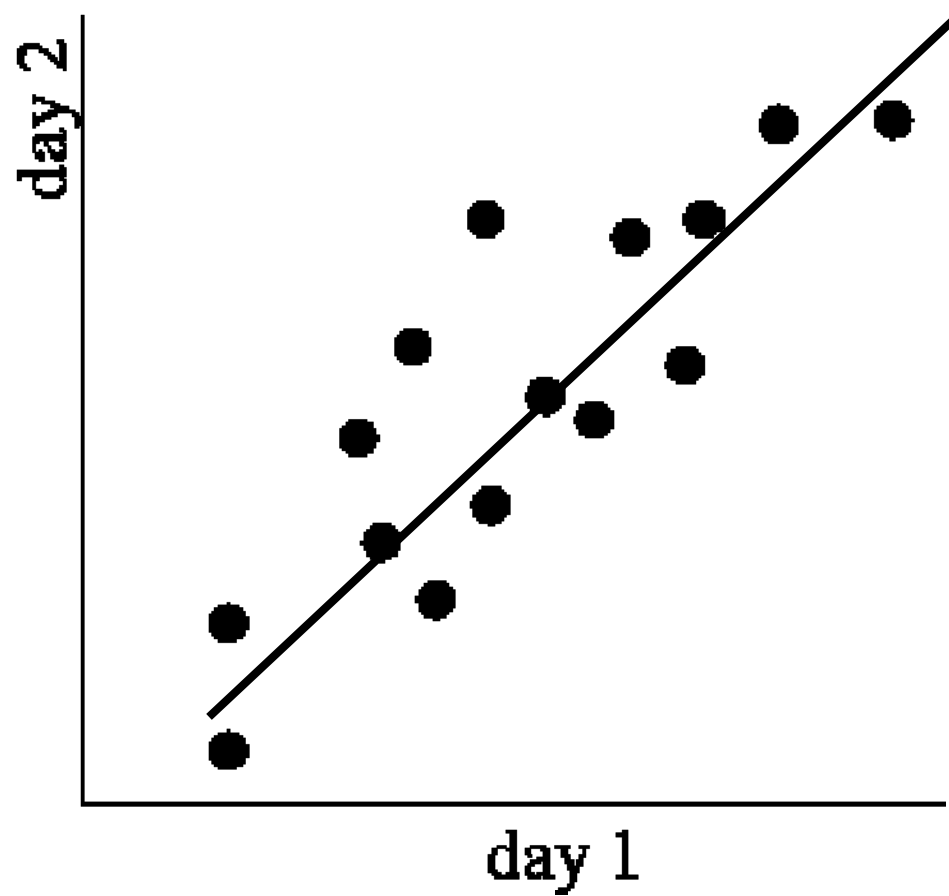
Outline

☑ Motivation

➡ ☐ **Definition - properties**

☐ Interpretation

☐ Complexity

☐ Conclusion

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 x *2*   *2* x *1*

SVD - Definition

(reminder: matrix multiplication)

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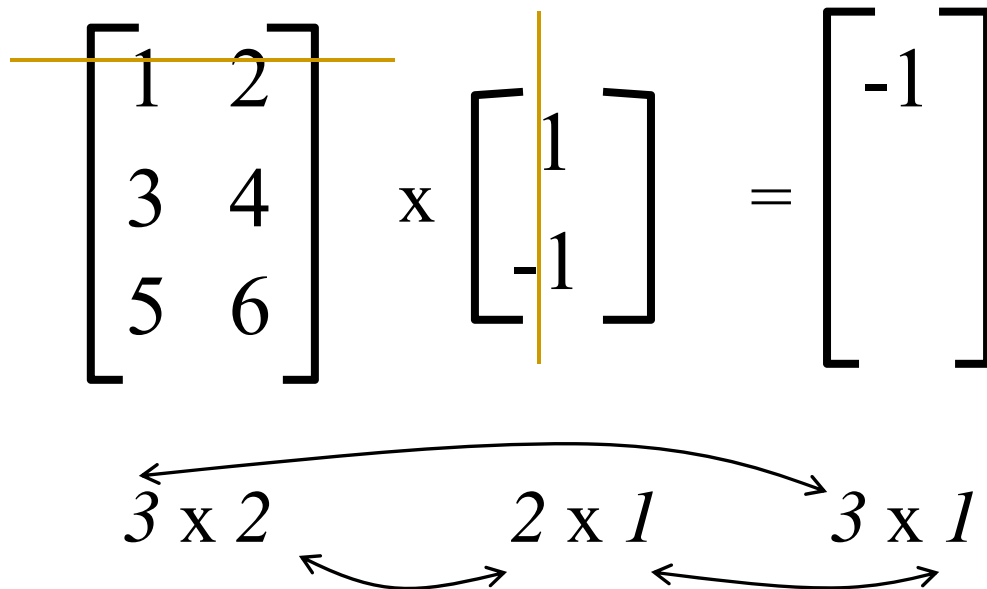
3 x 2   *2 x 1*   *3 x 1*

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 x 2   *2 x 1*   *3 x 1*

SVD - Definition

(reminder: matrix multiplication)

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

3 x 2   *2 x 1*   *3 x 1*

SVD - Definition

(reminder: matrix multiplication)

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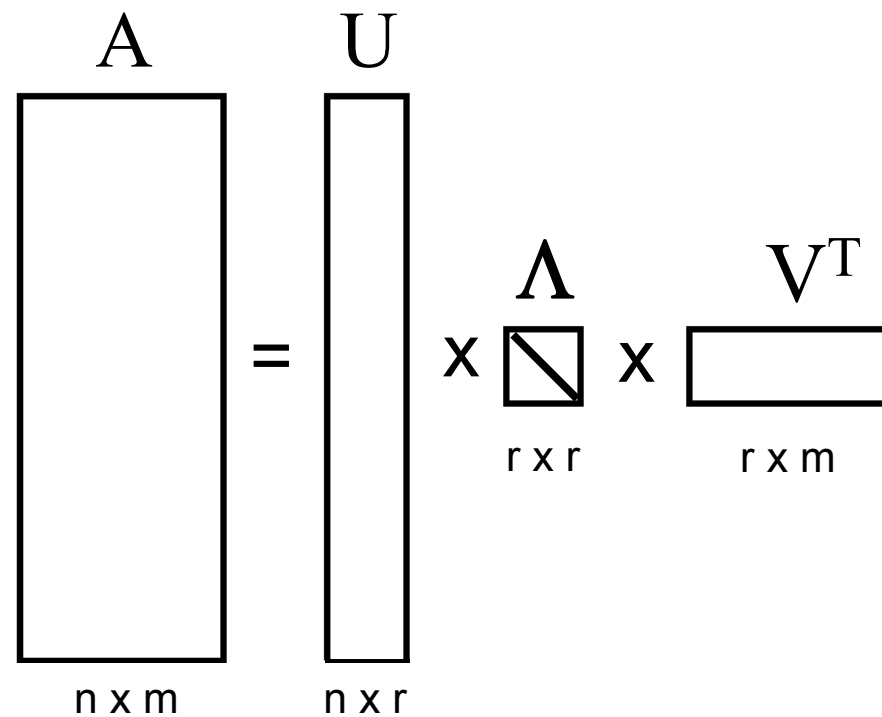
SVD - Definition

$$A_{[n \times m]} = U_{[n \times r]} \Lambda_{[r \times r]} (V_{[m \times r]})^T$$

- **A**: n x m matrix (eg., n documents, m terms)
- **U**: n x r matrix (n documents, r concepts)
- **Λ**: r x r diagonal matrix (strength of each 'concept') (r : rank of the matrix)
- **V**: m x r matrix (m terms, r concepts)

SVD - Definition

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

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)
  - $\mathbf{U}^T\ \mathbf{U} = \mathbf{I}$; $\mathbf{V}^T\ \mathbf{V} = \mathbf{I}$ (**I:** identity matrix)
- $\mathbf{\Lambda}$: singular are positive, and sorted in decreasing order

SVD - Example

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

$$
\begin{array}{c}
\text{data} \quad \text{inf.} \quad \text{retrieval} \quad \text{lung} \\
\text{brain}
\end{array}
$$

$$
\begin{array}{c}
\text{CS} \\
\\
\text{MD}
\end{array}
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\;\mathbf{x}\;
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\;\mathbf{x}\;
$$

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

SVD - Example

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

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

data, inf., retrieval, brain, lung

CS-concept, MD-concept

SVD - Example

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

doc-to-concept
similarity matrix

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

data   inf.   retrieval   brain   lung   CS-concept   MD-concept

SVD - Example

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

'strength' of CS-concept

$$
\begin{array}{c} \text{data} \quad \text{inf.} \quad \text{retrieval} \quad \text{brain} \quad \text{lung} \end{array}
$$

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

SVD - Example

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

term-to-concept
similarity matrix

$$
\begin{array}{c}
\text{retrieval} \\
\text{inf.} \downarrow \text{brain} \quad \text{lung} \\
\text{data}
\end{array}
$$

CS-concept

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

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

SVD - Example

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

term-to-concept
similarity matrix

$$
\begin{array}{c}
\quad \overset{\text{retrieval}}{\underset{\downarrow}{\text{inf.}}} \\
\text{data} \quad \text{brain}^{\text{lung}}
\end{array}
$$

CS-concept

$$
\overset{CS}{\underset{\downarrow}{\uparrow}} \quad
\overset{MD}{\underset{\downarrow}{\uparrow}}
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\ \mathbf{x}\
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\ \mathbf{x}\
\begin{bmatrix}
0.58 & 0.58 & 0.58 & 0 & 0 \\
0 & 0 & 0 & 0.71 & 0.71
\end{bmatrix}
$$

Outline

☑ Motivation

☑ Definition - properties

➡ ☐ **Interpretation**

    #1: documents/terms/concepts

    #2: dim. reduction

    #3: picking non-zero, rectangular clusters

☐ Complexity

☐ Conclusion

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 x m]) similarity matrix

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

A: document-to-document ([n x n]) similarity matrix

SVD properties

- **V** are the eigenvectors of the *covariance matrix* $\mathbf{A}^{\mathrm{T}}\mathbf{A}$

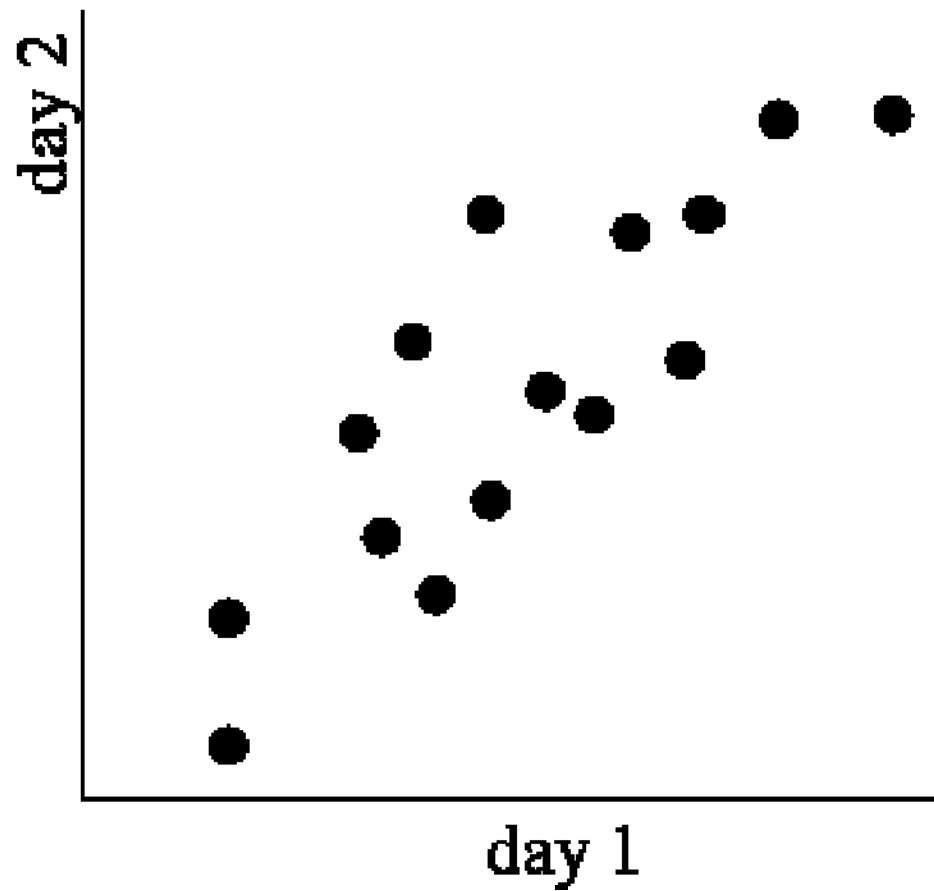- **U** are the eigenvectors of the *Gram (inner-product) matrix* $\mathbf{A}\mathbf{A}^{\mathrm{T}}$

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

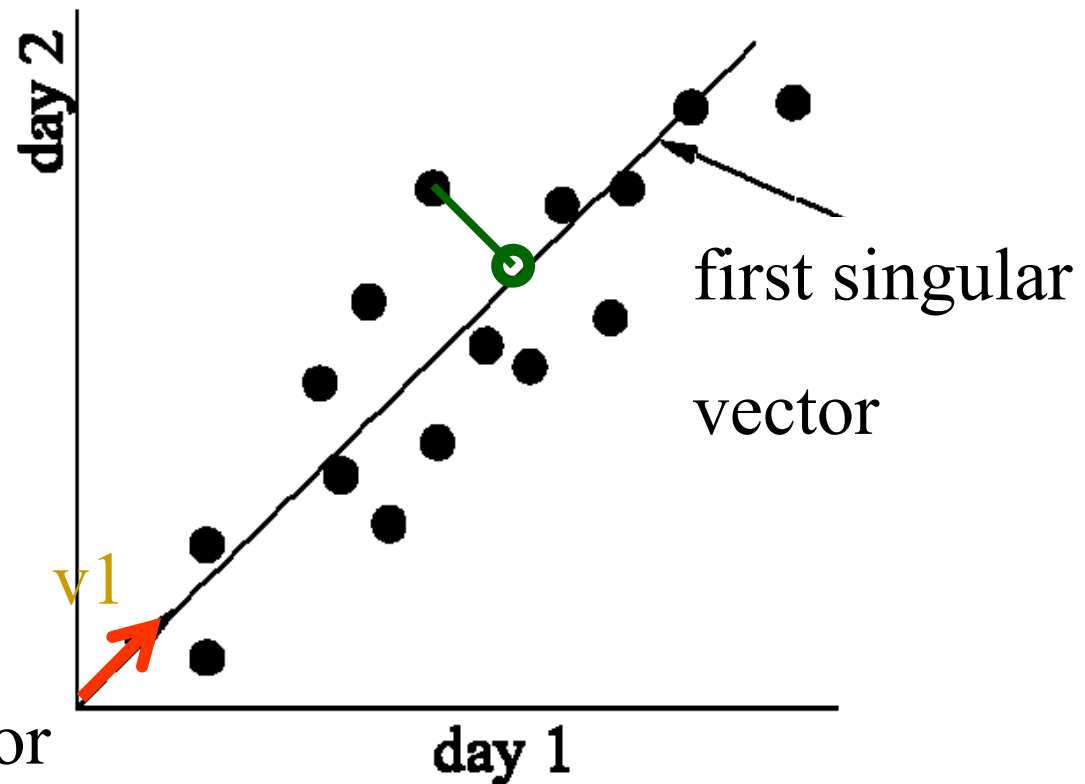- best axis to project on: ('best' = min sum of squares of projection errors)

SVD - Motivation

SVD - interpretation #2

SVD: gives
best axis to project

first singular

vector

v1

- minimum RMS error

SVD - Interpretation #2

day customer	Wc 7/10/96	Th 7/11/96	Fr 7/12/96	Sa 7/13/96	Su 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

- $\mathbf{A} = \mathbf{U} \, \boldsymbol{\Lambda} \, \mathbf{V}^\mathrm{T}$ - 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

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

variance ('spread') on the v1 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

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

  - $\mathbf{U \; \Lambda}$ gives the coordinates of the points in the projection axis

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\
0.90 & 0 \\
0 & 0.53 \\
0 & 0.80 \\
0 & 0.27
\end{bmatrix}
\; \mathrm{x} \;
\begin{bmatrix}
9.64 & 0 \\
0 & 5.29
\end{bmatrix}
\; \mathrm{x}
$$

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

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:

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

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

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

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

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

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

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:

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

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
| & | \\
u_1 & u_2 \\
| & |
\end{bmatrix}
\times
\begin{bmatrix}
\lambda_1 & \varnothing \\
\varnothing & \lambda_2
\end{bmatrix}
\times
\begin{bmatrix}
\underline{\quad} & v_1 & \underline{\quad} \\
\underline{\quad} & v_2 & \underline{\quad}
\end{bmatrix}
$$

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

$$
\begin{array}{c} \xleftarrow{\hspace{1em}} m \xrightarrow{\hspace{1em}} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} = \lambda_1 \quad u_1 \quad v^T_1 + \lambda_2 \quad u_2 \quad v^T_2 + ...
$$

n

Exactly equivalent:

'spectral decomposition' of the matrix:

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
= \lambda_1 \quad u_1 \quad v^T_1 \; + \; \lambda_2 \quad u_2 \quad v^T_2 + \ldots
$$

m

n

r terms

$n \times 1$ $\qquad 1 \times m$

SVD - Interpretation #2

approximation / dim. reduction:
by keeping the first few terms (Q: how many?)

$$
\underset{n}{\overset{\longleftarrow \quad m \quad \longrightarrow}{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}} = \lambda_1 \quad u_1 \quad v^T_1 \; + \; \lambda_2 \quad u_2 \quad v^T_2 + \dots
$$

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

SVD - Interpretation #2

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

$$
n \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \lambda_1 \ u_1 \ v^T_1 + \lambda_2 \ u_2 \ v^T_2 + ...
$$

assume: $\lambda_1 >= \lambda_2 >= ...$

Pictorially: matrix form of SVD

$$\mathbf{A} \approx \mathbf{U\Sigma V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$

- Best rank-k approximation in L2

Pictorially: matrix form of SVD

$$\mathbf{A} \approx \mathbf{U\Sigma V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$
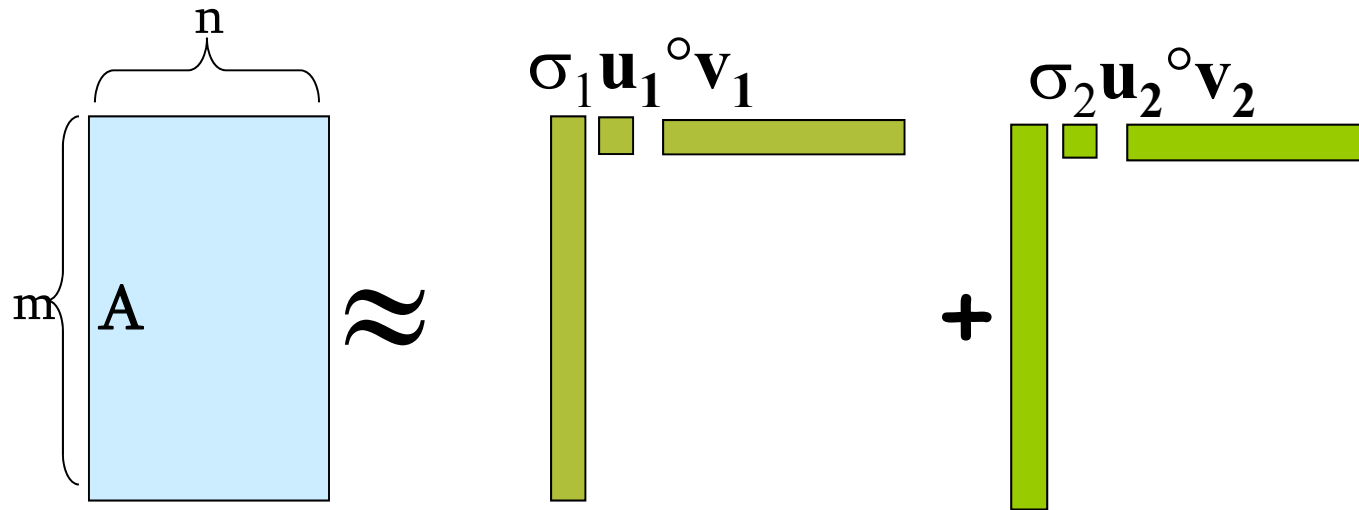
❑ Best rank-k approximation in L2

SVD - Interpretation #3

- finds non-zero clusters 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 clusters 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 clusters in a data matrix =
- 'communities' (bi-partite cores, here)

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

Row 1

Row 4

Row 5

Row 7

Col 1

Col 3

Col 4
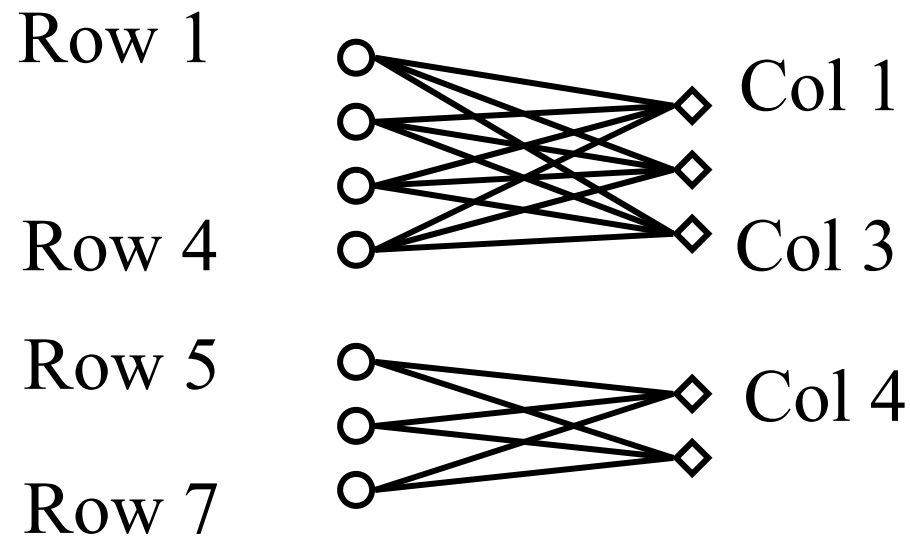
SVD - Interpretation #3

- Excercise: find the SVD, 'by inspection'!
- Q: rank = ??

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

SVD - Interpretation #3

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

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

SVD - Interpretation #3

- 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 - Interpretation #3

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

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
1/sqrt(3) & 0 \\
1/sqrt(3) & 0 \\
1/sqrt(3) & 0 \\
0 & 1/sqrt(2) \\
0 & 1/sqrt(2)
\end{bmatrix}
\text{ x }
\begin{bmatrix}
?? & 0 \\
0 & ??
\end{bmatrix}
\text{ x}
$$

$$
\begin{bmatrix}
1/sqrt(3) & 1/sqrt(3) & 1/sqrt(3) & 0 & 0 \\
0 & 0 & 0 & 1/sqrt(2) & 1/sqrt(2)
\end{bmatrix}
$$

SVD - Interpretation #3

- and the singular values are:

$$
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
1/sqrt(3) & 0 \\
1/sqrt(3) & 0 \\
1/sqrt(3) & 0 \\
0 & 1/sqrt(2) \\
0 & 1/sqrt(2)
\end{bmatrix}
\text{ x }
\begin{bmatrix}
3 & 0 \\
0 & 2
\end{bmatrix}
\text{ x }
$$

$$
\begin{bmatrix}
1/sqrt(3) & 1/sqrt(3) & 1/sqrt(3) & 0 & 0 \\
0 & 0 & 0 & 1/sqrt(2) & 1/sqrt(2)
\end{bmatrix}
$$

SVD - Interpretation #3

- 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 - Interpretation #3

- **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
  - same for matrix **V**
  - matrix Λ should be diagonal, with positive values

Outline

☑ Motivation

☑ Definition - properties

☑ Interpretation

➡ ☐ **Complexity**

☐ Conclusion

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 ...)
  - Hadoop Implementation: HEigen [Kang+ PAKDD 2011]

Outline

- ☑ Motivation
- ☑ Definition - properties
- ☑ Interpretation
- ☑ Complexity
- ➡ ☐ **Conclusion**

SVD - conclusions so far

- SVD: $\mathbf{A} = \mathbf{U} \, \mathbf{\Lambda} \, \mathbf{V^T}$: unique (*)
  - $\mathbf{U}$: document-to-concept similarities
  - $\mathbf{V}$: term-to-concept similarities
  - $\mathbf{\Lambda}$: 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 clusters

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). Numerical Recipes in C, Cambridge University Press.