15-826: Multimedia Databases and Data Mining

Lecture #18: 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
- ...

CMU SCS

SVD - Detailed outline

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

CMU SCS

SVD - Motivation

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

CMU SCS

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

CMU SCS

SVD - Motivation

- Customer-product, for recommendation system:

$$\begin{array}{c|ccccc} & \text{bread} & \text{lettuce} & \text{tomatos} & \text{beef} & \text{chicken} \\ \hline \uparrow & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ \text{vegetarians} & 1 & 1 & 1 & 0 & 0 \\ \downarrow & 5 & 5 & 5 & 0 & 0 \\ \uparrow & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ \text{meat eaters} \downarrow & 0 & 0 & 0 & 1 & 1 \end{array}$$

SVD - Motivation

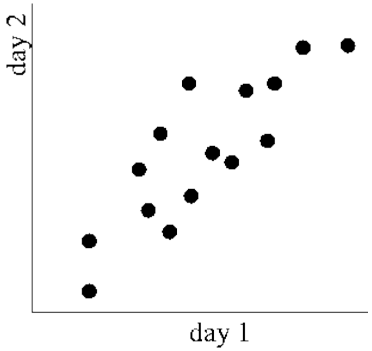
- problem #2: compress / reduce dimensionality

Problem - specs

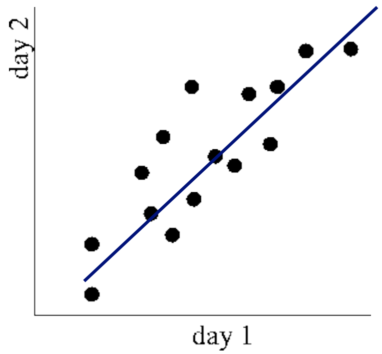
- ~10**6 rows; ~10**3 columns; no updates;
- random access to any cell(s) ; small error: OK

day customer	We 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

SVD - Motivation



SVD - Motivation



SVD - Detailed outline

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

SVD - Definition

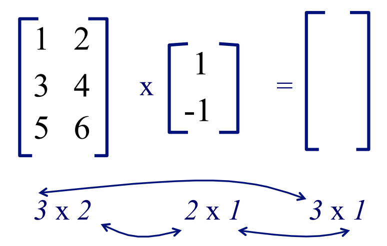
(reminder: matrix multiplication

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

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

SVD - Definition

(reminder: matrix multiplication

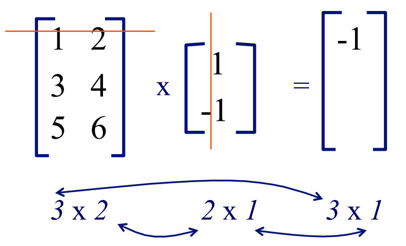


$$\begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix} \text{ x } \begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

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

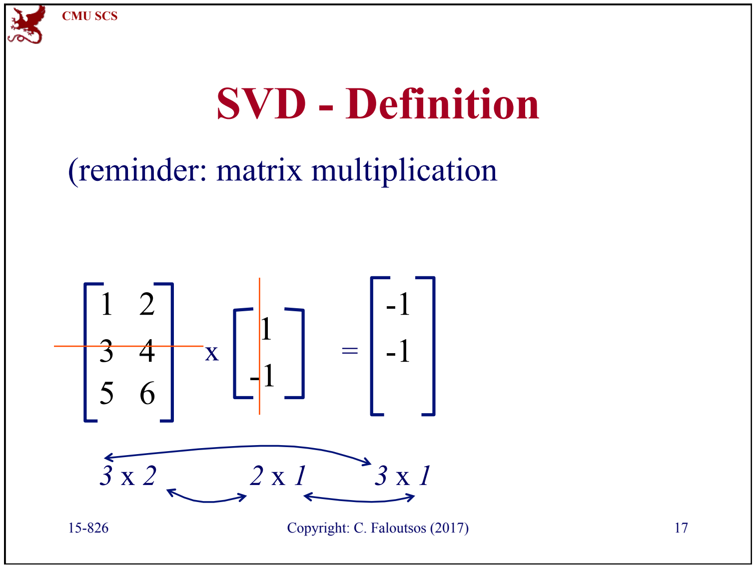
SVD - Definition

(reminder: matrix multiplication



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

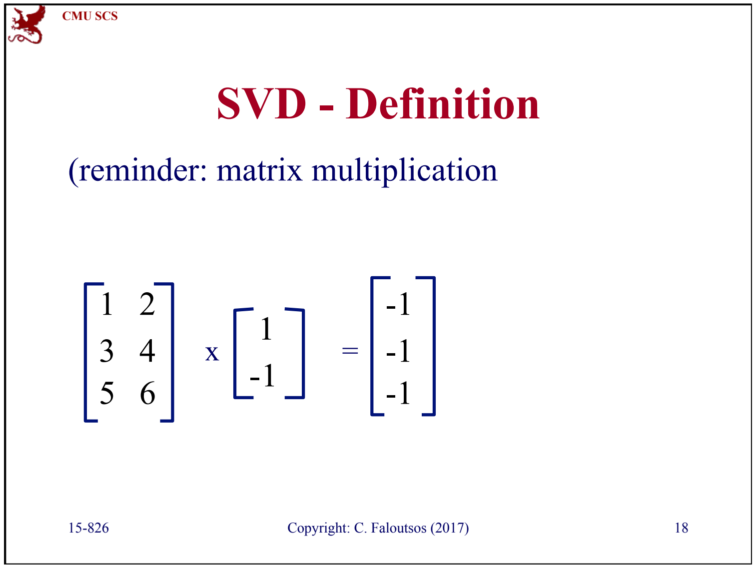
3 x *2*   *2* x *1*   *3* x *1*



SVD - Definition

(reminder: matrix multiplication

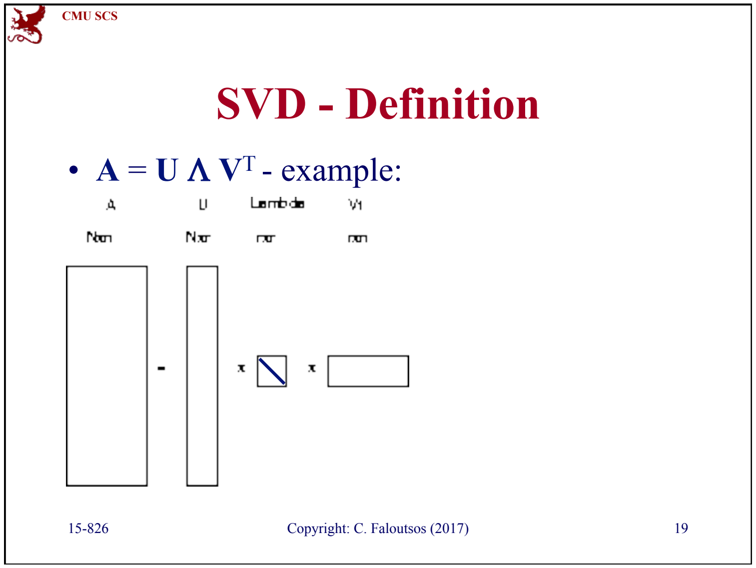
$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \\ \end{bmatrix}$$



SVD - Definition

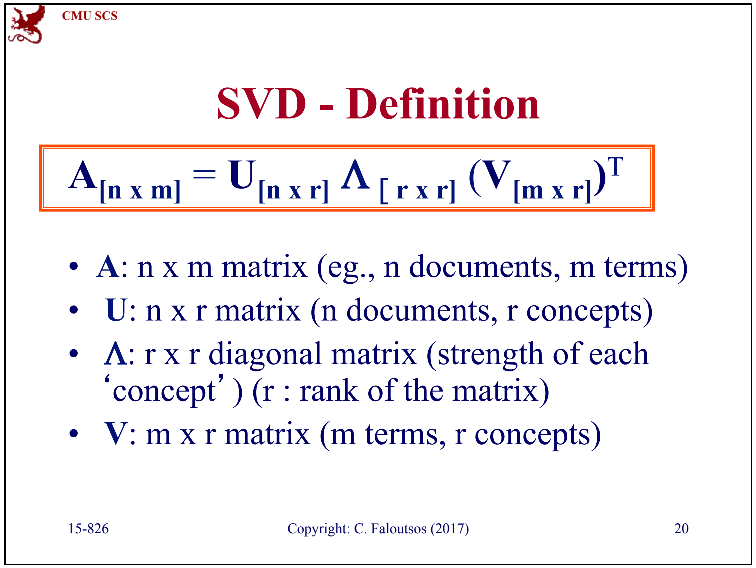
(reminder: matrix multiplication

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



SVD - Definition

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



SVD - Definition

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

- **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)

CMU SCS

SVD - Properties

THEOREM [Press+92]: always possible to decompose matrix **A** into **A** = **U Λ V**^{T} , where

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

CMU SCS

SVD - Example

- **A** = **U Λ V**^{T} - example:

$$\begin{array}{c|ccccc} & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \hline \text{CS} & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 5 & 5 & 5 & 0 & 0 \\ \text{MD} & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ & 0 & 0 & 0 & 1 & 1 \end{array} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \text{x} \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \text{x} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

CMU SCS

SVD - Example

- **A** = **U Λ V**^{T} - example:

CS-concept

MD-concept

$$\begin{array}{c|ccccc} & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \hline \text{CS} & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 5 & 5 & 5 & 0 & 0 \\ \text{MD} & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ & 0 & 0 & 0 & 1 & 1 \end{array} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \text{x} \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \text{x} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

CMU SCS

SVD - Example

- **A** = **U Λ V**^{T} - example:

doc-to-concept similarity matrix

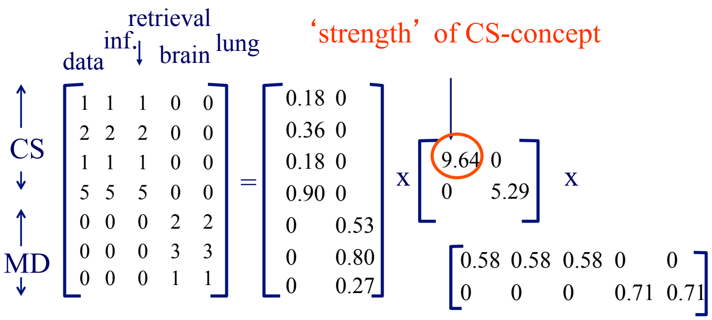
CS-concept

MD-concept

$$\begin{array}{c|ccccc} & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \hline \text{CS} & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 5 & 5 & 5 & 0 & 0 \\ \text{MD} & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ & 0 & 0 & 0 & 1 & 1 \end{array} = \begin{bmatrix} \boxed{0.18} & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \text{x} \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \text{x} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Example

- **A** = **U** **Λ** **V**^{T} - example:

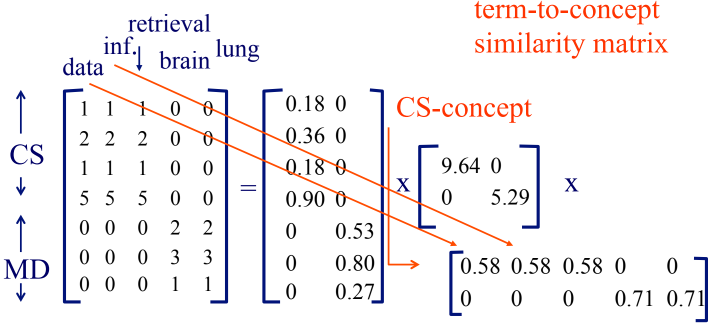


‘strength’ of CS-concept

$$\begin{array}{c|ccccc} & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \hline \text{CS} & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 5 & 5 & 5 & 0 & 0 \\ \text{MD} & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ & 0 & 0 & 0 & 1 & 1 \end{array} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Example

- **A** = **U** **Λ** **V**^{T} - example:



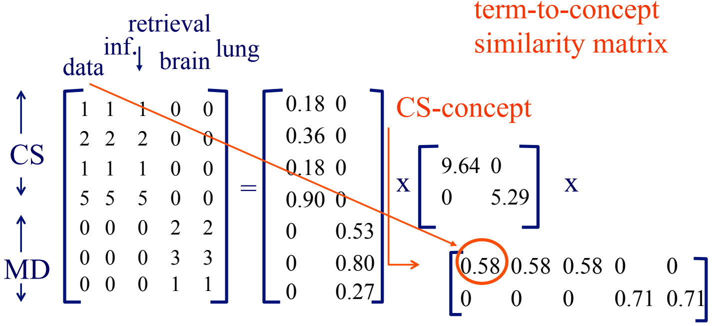
term-to-concept similarity matrix

CS-concept

$$\begin{array}{c|ccccc} & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \hline \text{CS} & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 5 & 5 & 5 & 0 & 0 \\ \text{MD} & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ & 0 & 0 & 0 & 1 & 1 \end{array} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - Example

- **A** = **U** **Λ** **V**^{T} - example:



term-to-concept similarity matrix

CS-concept

$$\begin{array}{c|ccccc} & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \hline \text{CS} & 1 & 1 & 1 & 0 & 0 \\ & 2 & 2 & 2 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 5 & 5 & 5 & 0 & 0 \\ \text{MD} & 0 & 0 & 0 & 2 & 2 \\ & 0 & 0 & 0 & 3 & 3 \\ & 0 & 0 & 0 & 1 & 1 \end{array} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

SVD - 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 **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 **A** is the document-to-term matrix, what is $\mathbf{A}^T\mathbf{A}$?

A: term-to-term ([m x m]) similarity matrix

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

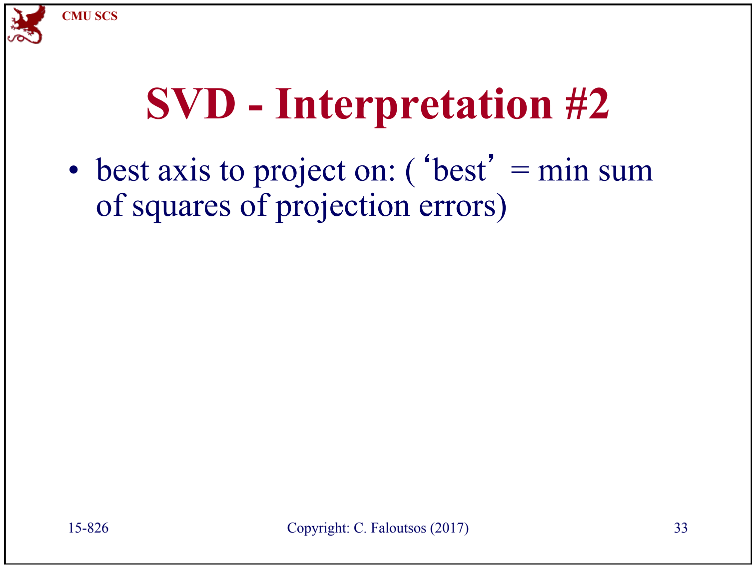
A: document-to-document ([n x n]) similarity matrix

SVD properties

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

Further reading:

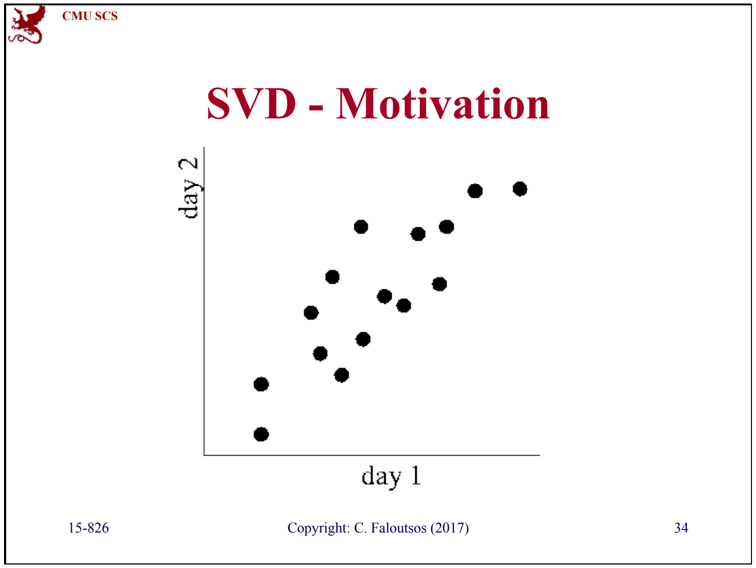
1. Ian T. Jolliffe, *Principal Component Analysis* (2nd ed), Springer, 2002.
2. Gilbert Strang, *Linear Algebra and Its Applications* (4th ed), Brooks Cole, 2005.



CMU SCS

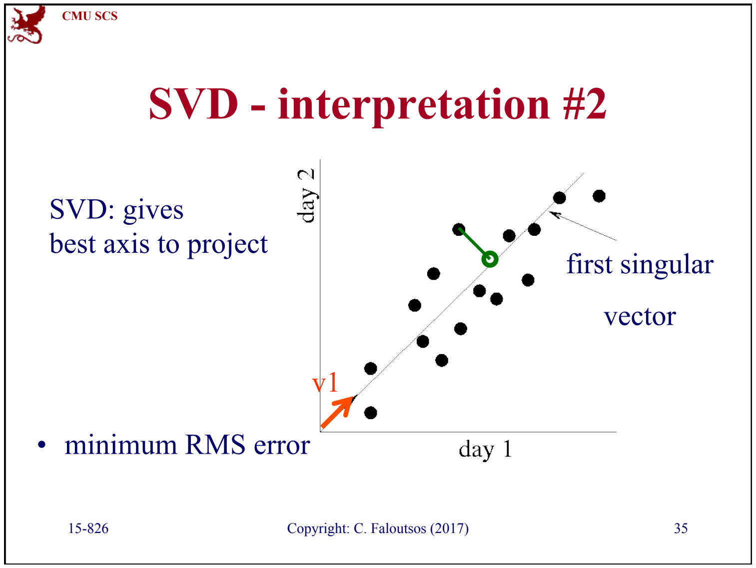
SVD - Interpretation #2

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



CMU SCS

SVD - Motivation

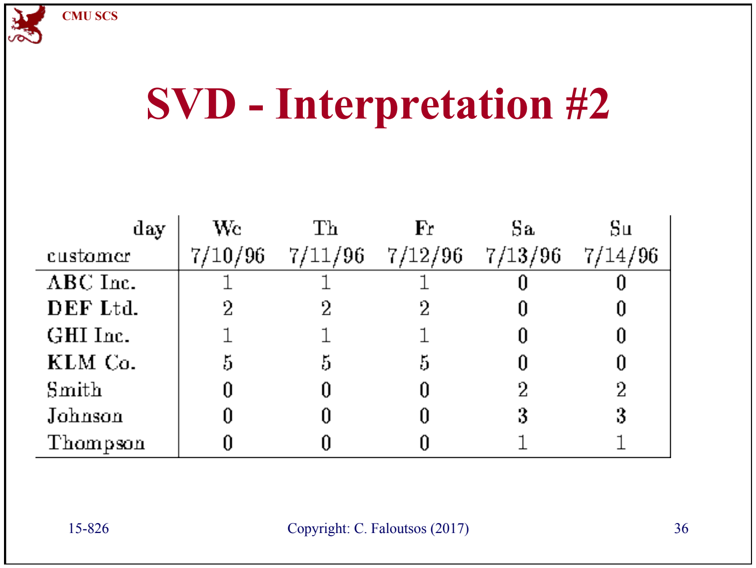


CMU SCS

SVD - interpretation #2

SVD: gives
best axis to project

- minimum RMS error



CMU SCS

SVD - Interpretation #2

day customer	We 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

SVD - Interpretation #2

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

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{bmatrix} \overset{\text{v1}}{\boxed{0.58 \quad 0.58 \quad 0.58 \quad 0 \quad 0}} \\ 0 \quad\;\; 0 \quad\;\; 0 \quad\;\; 0.71 \quad 0.71 \end{bmatrix}$$

SVD - Interpretation #2

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

variance ('spread') on the v1 axis

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

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?

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

CMU SCS

SVD - Interpretation #2

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

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

CMU SCS

SVD - Interpretation #2

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

CMU SCS

SVD - Interpretation #2

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

CMU SCS

SVD - Interpretation #2

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

SVD - Interpretation #2

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

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

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

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

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

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

$$n \updownarrow \overset{\leftarrow\ m\ \rightarrow}{\begin{bmatrix}1&1&1&0&0\\2&2&2&0&0\\1&1&1&0&0\\5&5&5&0&0\\0&0&0&2&2\\0&0&0&3&3\\0&0&0&1&1\end{bmatrix}} = \lambda_1\ u_1\ v^T_1 + \lambda_2\ u_2\ v^T_2 + \ldots$$

CMU SCS

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

$$\begin{array}{c|ccccc|}
 & \multicolumn{5}{c}{m} \\
 & 1 & 1 & 1 & 0 & 0 \\
 & 2 & 2 & 2 & 0 & 0 \\
 & 1 & 1 & 1 & 0 & 0 \\
n & 5 & 5 & 5 & 0 & 0 \\
 & 0 & 0 & 0 & 2 & 2 \\
 & 0 & 0 & 0 & 3 & 3 \\
 & 0 & 0 & 0 & 1 & 1
\end{array} = \overbrace{\lambda_1 \underbrace{u_1}_{n \times 1} \underbrace{v^T_1}_{1 \times m} + \lambda_2 u_2 v^T_2 + \ldots}^{r \text{ terms}}$$

CMU SCS

SVD - Interpretation #2

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

$$\begin{array}{c|ccccc|}
 & \multicolumn{5}{c}{m} \\
 & 1 & 1 & 1 & 0 & 0 \\
 & 2 & 2 & 2 & 0 & 0 \\
 & 1 & 1 & 1 & 0 & 0 \\
n & 5 & 5 & 5 & 0 & 0 \\
 & 0 & 0 & 0 & 2 & 2 \\
 & 0 & 0 & 0 & 3 & 3 \\
 & 0 & 0 & 0 & 1 & 1
\end{array} = \lambda_1 u_1 v^T_1 + \lambda_2 u_2 v^T_2 + \ldots$$

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

CMU SCS

SVD - Interpretation #2

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

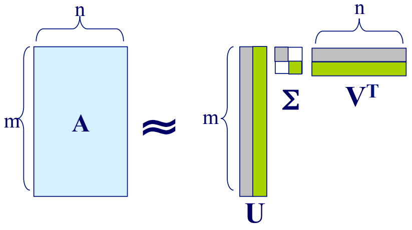
$$\begin{array}{c|ccccc|}
 & \multicolumn{5}{c}{m} \\
 & 1 & 1 & 1 & 0 & 0 \\
 & 2 & 2 & 2 & 0 & 0 \\
 & 1 & 1 & 1 & 0 & 0 \\
n & 5 & 5 & 5 & 0 & 0 \\
 & 0 & 0 & 0 & 2 & 2 \\
 & 0 & 0 & 0 & 3 & 3 \\
 & 0 & 0 & 0 & 1 & 1
\end{array} = \lambda_1 u_1 v^T_1 + \lambda_2 u_2 v^T_2 + \ldots$$

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

CMU SCS

Pictorially: matrix form of SVD

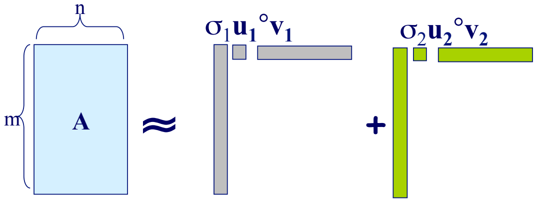
$$\mathbf{A} \approx \mathbf{U\Sigma V}^T = \sum_i \sigma_i \mathbf{u}_i \circ \mathbf{v}_i$$



– Best rank-k approximation in L2

Pictorially: Spectral form of SVD

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



– Best rank-k approximation in L2

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

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

SVD - Interpretation #3

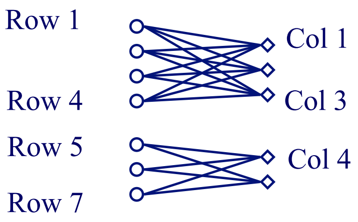
- finds non-zero 'blobs' in a data matrix

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

SVD - Interpretation #3

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

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



SVD - Interpretation #3

- 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 - Interpretation #3

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

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

SVD - Interpretation #3

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

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{\curvearrowleft} \times \begin{bmatrix} ?? & 0 \\ 0 & ?? \end{bmatrix} \times \curvearrowright\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

orthogonal??

CMU SCS

SVD - Interpretation #3

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

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

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

CMU SCS

SVD - Interpretation #3

- and the singular values are:

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

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

CMU SCS

SVD - Interpretation #3

- Q: How to check we are correct?

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

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

CMU SCS

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
  - 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 Λ 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'

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.