Carnegie Mellon

Powerful Tools for Data Mining
Fractals, power laws, SVD

C. Faloutsos
Carnegie Mellon University

Carnegie Mellon

Part I:
Singular Value Decomposition (SVD)
= Eigenvalue decomposition

www.cs.cmu.edu/~christos/

Carnegie Mellon

SVD - outline

- Motivation
- Definition - properties
- Interpretation; more properties
- Complexity
- Applications - success stories
- Conclusions

Carnegie Mellon

SVD - Motivation

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

Carnegie Mellon

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

Carnegie Mellon

SVD - Motivation

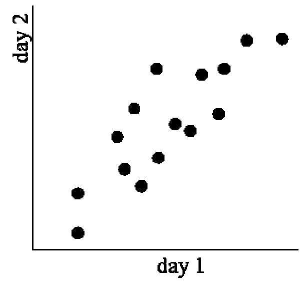
- problem #2: compress / reduce dimensionality

Problem - specs

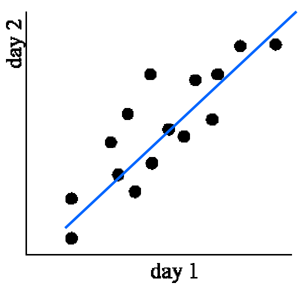
- ~10**6 rows; ~10**3 columns; no updates;
- Compress / extract features

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 - outline

- Motivation
- Definition - properties
- Interpretation; more properties
- Complexity
- Applications - success stories
- Conclusions

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*

Skip

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

Skip

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

Skip

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 - notation

Conventions:

- bold capitals -> matrix (eg. **A**, **U**, **Λ**, **V**)
- bold lower-case -> <u>column</u> vector (eg., **x**, $\mathbf{v}_1$, $\mathbf{u}_3$)
- regular lower-case -> scalars (eg., λ_1 , λ_r)

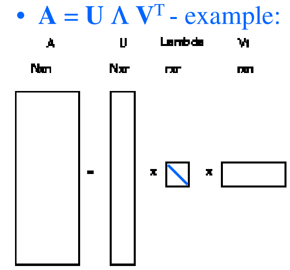
SVD - Definition

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

- **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)

SVD - Definition

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



SVD - Properties

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

- $\mathbf{U}, \Lambda, \mathbf{V}$: 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)
- Λ**:** eigenvalues are positive, and sorted in decreasing order

SVD - Example

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

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

SVD - Example

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

CS-concept

MD-concept

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

SVD - Example

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

doc-to-concept similarity matrix

CS-concept

MD-concept

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

SVD - Example

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

'strength' of CS-concept

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

SVD - Example

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

term-to-concept similarity matrix

CS-concept

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

SVD - Example

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

- Motivation
- Definition - properties
- Interpretation; more properties
- Complexity
- Applications - success stories
- Conclusions

SVD - Interpretation #1

'documents', 'terms' and 'concepts': [Foltz+92]

- **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 $\mathbf{A}^T\,\mathbf{A}$?

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

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

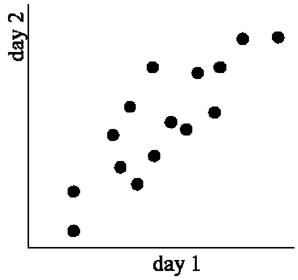
Q: $\mathbf{A}\,\mathbf{A}^T$?

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

SVD - Interpretation #2

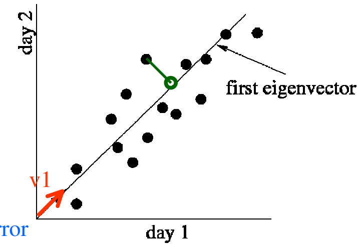
- best axis to project on: ('best' = min sum of squares of projection errors)

SVD - Motivation



SVD - interpretation #2

SVD: gives
best axis to project



- minimum RMS error

SVD - Interpretation #2

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

- **A = U Λ V**^{T} - example:

v1

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

SVD - Interpretation #2

- **A = U Λ V**^{T} - example:

variance ('spread') on the v1 axis

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

SVD - Interpretation #2

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

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

SVD - Interpretation #2

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

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

SVD - Interpretation #2

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

SVD - Interpretation #2

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

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

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

skip

SVD - Interpretation #2

Exactly equivalent:
'*spectral decomposition*' of the matrix:

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

skip

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

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

skip

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{n \times m} = \lambda_1 \mathbf{u}_1 \mathbf{v}^T_1 + \lambda_2 \mathbf{u}_2 \mathbf{v}^T_2 + \ldots$$

skip

SVD - Interpretation #2

Exactly equivalent:
'spectral decomposition' of the matrix:

r terms

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{n \times m} = \lambda_1 \underbrace{\mathbf{u}_1}_{n \times 1} \underbrace{\mathbf{v}^T_1}_{1 \times m} + \lambda_2 \mathbf{u}_2 \mathbf{v}^T_2 + \ldots$$

SVD - Interpretation #2

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

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{n \times m} = \lambda_1 \mathbf{u}_1 \mathbf{v}^T_1 + \lambda_2 \mathbf{u}_2 \mathbf{v}^T_2 + \ldots$$

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

SVD - Interpretation #2

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

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}}_{n \times m} = \lambda_1 \mathbf{u}_1 v^T_1 + \lambda_2 \mathbf{u}_2 \mathbf{v}^T_2 + \ldots$$

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

SVD - outline

- Motivation
- Definition - properties
- Interpretation
  - #1: documents/terms/concepts
  - #2: dim. reduction
  - #3: picking non-zero, rectangular 'blobs'
  - #4: fixed points - graph interpretation
- Complexity
- ...

SVD - Interpretation #3

skip

- 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

skip

- 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

skip

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

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

SVD - Interpretation #3

skip

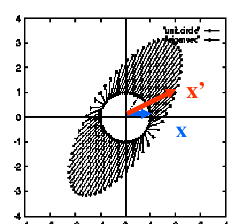
- finds non-zero 'blobs' in a data matrix

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

SVD - Interpretation #4

- **A** as vector transformation

$$\overset{\mathbf{x'}}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}} = \overset{\mathbf{A}}{\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}} \overset{\mathbf{x}}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}$$

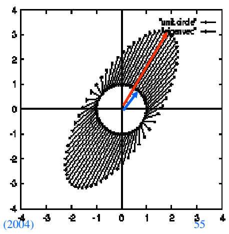


Carnegie Mellon

SVD - Interpretation #4

- if **A** is symmetric, then, by defn., its eigenvectors remain parallel to themselves (‘fixed points’)

$$\begin{array}{ccccc} \lambda_1 & \mathbf{v_1} & & \mathbf{A} & \mathbf{v_1} \\ 3.62 \; * & \begin{bmatrix} 0.52 \\ 0.85 \end{bmatrix} & = & \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} & \begin{bmatrix} 0.52 \\ 0.85 \end{bmatrix} \end{array}$$



Carnegie Mellon

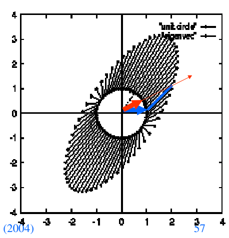
SVD - Interpretation #4

- fixed point operation ($\mathbf{A}^T \mathbf{A}$ is symmetric)
  - $\boxed{\mathbf{A}^T \mathbf{A}\, \mathbf{v}_i = \lambda_i^2\, \mathbf{v}_i}$ (i=1,..., r) [Property 1]
  - eg., **A**: transition matrix of a Markov Chain; then $\mathbf{v}_1$ is related to the steady state probability distribution (= a particle, doing a random walk on the graph)
- convergence: for (almost) any **v’**:
  - $\boxed{(\mathbf{A}^T \mathbf{A})^k\, \mathbf{v}' \sim \text{(constant)}\ \mathbf{v}_1}$ [Property 2]
  - where $\mathbf{v}_1$: strongest ‘right’ eigenvector

Carnegie Mellon

SVD - Interpretation #4

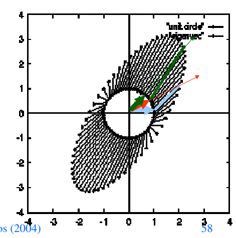
- convergence - illustration:



Carnegie Mellon

SVD - Interpretation #4

- convergence - illustration:



Carnegie Mellon

SVD - Interpretation #4

- $\mathbf{v}_1$ ... $\mathbf{v}_r$: ‘right’ eigenvectors
- symmetrically: $\mathbf{u}_1$ ... $\mathbf{u}_r$: ‘left’ eigenvectors for matrix **A** (= **U** Λ $\mathbf{V}^T$):

$$\mathbf{A}\, \mathbf{v}_i = \lambda_i\, \mathbf{u}_i$$
$$\mathbf{u}_i^{\;T}\, \mathbf{A} = \lambda_i\, \mathbf{v}_i^{\;T}$$

(right eigenvectors -> ‘authority’ scores in Kleinberg’s algo;

left ones -> ‘hub’ scores)

Carnegie Mellon

SVD - outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Applications - success stories
- Conclusions

Carnegie Mellon

SVD - Complexity

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

Carnegie Mellon

SVD - conclusions so far

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

Carnegie Mellon

SVD - conclusions so far

- dim. reduction: keep the first few strongest eigenvalues (80-90% of 'energy')
- SVD: picks up linear correlations
- SVD: picks up non-zero 'blobs'
- $\mathbf{v}_1$: fixed point (-> steady-state prob.)

Carnegie Mellon

SVD - outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Applications - success stories
- Conclusions

Carnegie Mellon

SVD - Case studies

- ...
- Applications - success stories
  - ~~multi-lingual IR; LSI queries~~
  - ~~compression~~
  - PCA - 'ratio rules'
  - Karhunen-Lowe transform
  - Kleinberg/google algorithm

Carnegie Mellon

skip

Case study - LSI

Q1: How to do queries with LSI?

Q2: multi-lingual IR (english query, on spanish text?)

Case study - LSI

Q1: How to do queries with LSI?

Problem: Eg., find documents with ‘data’

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

Case study - LSI

Q1: How to do queries with LSI?

A: map query vectors into ‘concept space’ – how?

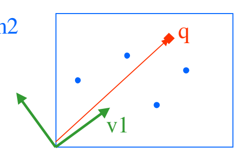
$$
\begin{array}{c} \\ \text{CS} \\ \\ \\ \text{MD} \\ \\ \end{array}
\begin{array}{c}
\begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\
\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 &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

Case study - LSI

Q1: How to do queries with LSI?

A: map query vectors into ‘concept space’ – how?

$$
q^T = \begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}
$$

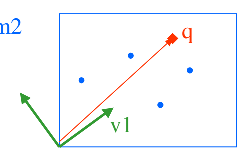


Case study - LSI

Q1: How to do queries with LSI?

A: map query vectors into ‘concept space’ – how?

$$
q^T = \begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}
$$



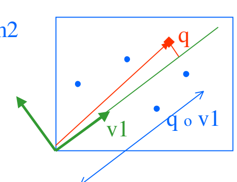
A: inner product (cosine similarity) with each ‘concept’ vector v_i

Case study - LSI

Q1: How to do queries with LSI?

A: map query vectors into ‘concept space’ – how?

$$
q^T = \begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}
$$



A: inner product (cosine similarity) with each ‘concept’ vector v_i

Case study - LSI

compactly, we have:

$$\mathbf{q}^T_{concept} = \mathbf{q}^T \mathbf{V}$$

Eg:

$$
q^T = \begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array}
\underbrace{\begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix}}_{\text{term-to-concept similarities}}
= \begin{array}{c} \text{CS-concept} \\ \downarrow \\ \begin{bmatrix} 0.58 & 0 \end{bmatrix} \end{array}
$$

Skip

Case study - LSI

Drill: how would the document (‘information’, ‘retrieval’) handled by LSI?

Skip

Case study - LSI

Drill: how would the document (‘information’, ‘retrieval’) handled by LSI? A: SAME:

$\mathbf{d}^T_{\mathbf{concept}} = \mathbf{d}^T\ \mathbf{V}$

Eg:

$$\mathbf{d}^T = \begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} = \begin{array}{c} \text{CS-concept} \\ \downarrow \\ \begin{bmatrix} 1.16 & 0 \end{bmatrix} \end{array}$$

term-to-concept similarities

Skip

Case study - LSI

Observation: document (‘information’, ‘retrieval’) will be retrieved by query (‘data’), although it does not contain ‘data’!!

$$\begin{array}{lccc} & \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} & & \text{CS-concept} \\ \mathbf{d}^T = & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix} & \dashrightarrow & \begin{bmatrix} 1.16 & 0 \end{bmatrix} \\ \mathbf{q}^T = & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} & \dashrightarrow & \begin{bmatrix} 0.58 & 0 \end{bmatrix} \end{array}$$

Skip

Case study - LSI

Q1: How to do queries with LSI?

➡ Q2: multi-lingual IR (english query, on spanish text?)

Skip

Case study - LSI

- Problem:
  - given many documents, translated to both languages (eg., English and Spanish)
  - answer queries across languages

Skip

Case study - LSI

- Solution: ~ LSI

	data	inf.	retrieval	brain	lung	datos	informacion			
CS	1	1	1	0	0	1	1	1	0	0
CS	2	2	2	0	0	1	2	2	0	0
CS	1	1	1	0	0	1	1	1	0	0
CS	5	5	5	0	0	5	5	4	0	0
MD	0	0	0	2	2	0	0	0	2	2
MD	0	0	0	3	3	0	0	0	2	3
MD	0	0	0	1	1	0	0	0	1	1

Case study - LSI

- Solution: ~ LSI

$$
\begin{array}{c|ccccc:ccccc}
 & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} & \text{informacion} & \text{datos} & & & \\
\hline
\text{CS} & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
\text{CS} & 2 & 2 & 2 & 0 & 0 & 1 & 2 & 2 & 0 & 0 \\
\text{CS} & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
\text{CS} & 5 & 5 & 5 & 0 & 0 & 5 & 5 & 4 & 0 & 0 \\
\text{MD} & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 0 & 2 & 2 \\
\text{MD} & 0 & 0 & 0 & 3 & 3 & 0 & 0 & 0 & 2 & 3 \\
\text{MD} & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1
\end{array}
$$

SVD - outline - Case studies

- ...
- Applications - success stories
  - multi-lingual IR; LSI queries
  - compression
  - PCA - 'ratio rules'
  - Karhunen-Lowe transform
  - Kleinberg/google algorithm

Case study: compression

[Korn+97]

Problem:

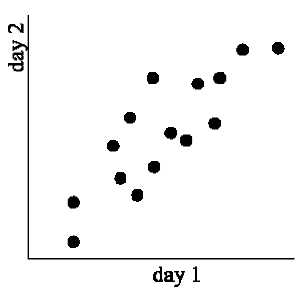
- given a matrix
- compress it, but maintain 'random access'

Problem - specs

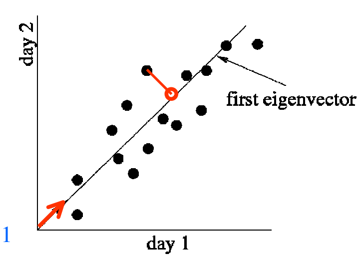
- ~10**6 rows; ~10**3 columns; no updates;
- random access to any cell(s) ; small error: OK

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

Idea



SVD - reminder

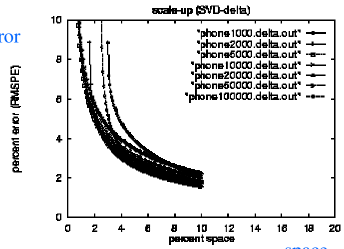


- space savings: 2:1
- minimum RMS error

Performance - scaleup

Skip

error



SVDD.

space

SVD - outline - Case studies

- ...
- Applications - success stories
  - multi-lingual IR; LSI queries
  - compression
  - PCA - 'ratio rules'
  - Karhunen-Lowe transform
  - Kleinberg/google algorithm

PCA - 'Ratio Rules'

[Korn+00]

Typically: '*Association Rules*' (eg.,

{bread, milk} -> {butter}

But:

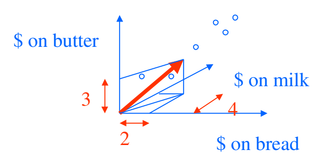
- which set of rules is 'better'?
- how to reconstruct missing/corrupted values?

PCA - 'Ratio Rules'

Idea: try to find 'concepts':

- Eigenvectors dictate rules about ratios:

bread:milk:butter = 2:4:3



PCA - 'Ratio Rules'

[Korn+00]

Typically: '*Association Rules*' (eg.,

{bread, milk} -> {butter}

But:

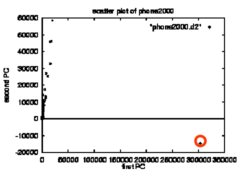
- ✓ how to reconstruct missing/corrupted values?
- ✓ which set of rules is 'better'?

PCA - 'Ratio Rules'

Moreover: visual data mining, for free:

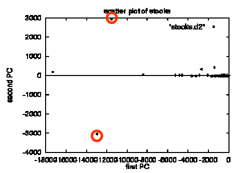
PCA - Ratio Rules

- no Gaussian clusters; Zipf-like distribution



(a) 'phone2000'

phonecalls

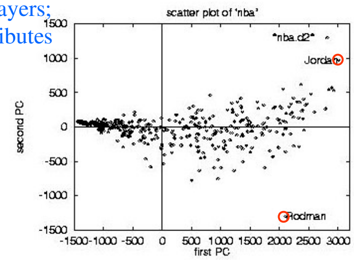


(b) 'stocks'

stocks

PCA - Ratio Rules

NBA dataset
~500 players;
~30 attributes



PCA - Ratio Rules

- PCA: get eigenvectors v1, v2, ...
- ignore entries with small abs. value
- try to interpret the rest

PCA - Ratio Rules

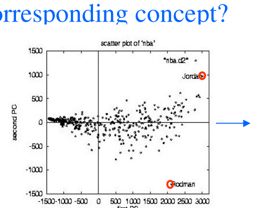
NBA dataset - **V** matrix (term to 'concept' similarities)

field	RR_1	RR_2	RR_3
minutes played	.808	$-.4$	
field goals			
goal attempts			
points	.406	.199	
total rebounds		$-.489$	.602
assists			$-.486$
steals			$-.07$

v1

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?



v1

Ratio Rules - example

- RR1: minutes:points = 2:1
- corresponding concept?
- A: 'goodness' of player

Carnegie Mellon

Ratio Rules - example

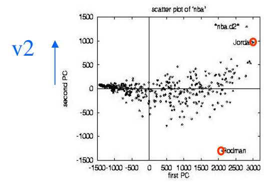
- RR2: points: rebounds negatively correlated(!)

field	RR_1	RR_2	RR_3
minutes played	.808	$-.4$	
field goals			
goal attempts			
points	.406	.199	
total rebounds		$-.489$	.602
assists			$-.486$
steals			$-.07$

Carnegie Mellon

Ratio Rules - example

- RR2: points: rebounds negatively correlated(!) - concept?



Carnegie Mellon

Ratio Rules - example

- RR2: points: rebounds negatively correlated(!) - concept?
- A: position: offensive/defensive

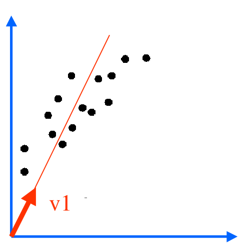
Carnegie Mellon

SVD - Case studies

- ...
- Applications - success stories
  - multi-lingual IR; LSI queries
  - compression
  - PCA - ‘ratio rules’
  - Karhunen-Lowe transform
  - Kleinberg/google algorithm

Carnegie Mellon

K-L transform

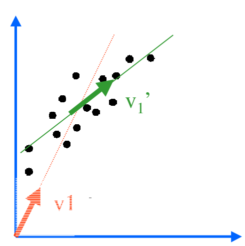


[Duda & Hart]; [Fukunaga]

A subtle point:
SVD will give vectors that go through the origin

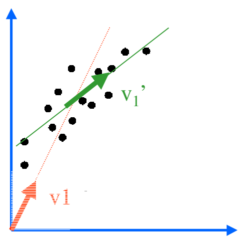
Carnegie Mellon

K-L transform



A subtle point:
SVD will give vectors that go through the origin
Q: how to find v_1’ ?

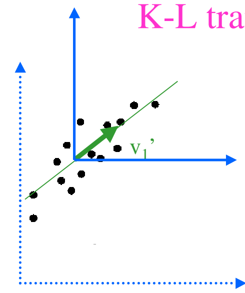
K-L transform



A subtle point:
SVD will give vectors that go through the origin
Q: how to find v_1’ ?

A: ‘centered’ PCA, ie., move the origin to center of gravity

K-L transform



A subtle point:
SVD will give vectors that go through the origin
Q: how to find v_1’ ?

A: ‘centered’ PCA, ie., move the origin to center of gravity and THEN do SVD

SVD - outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Applications - success stories
- Conclusions

SVD - Case studies

- ...
- Applications - success stories
  - multi-lingual IR; LSI queries
  - compression
  - PCA - ‘ratio rules’
  - Karhunen-Lowe transform
  - Kleinberg/google algorithm

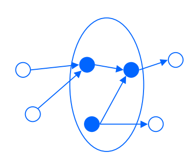
Kleinberg’s algorithm

- Problem dfn: given the web and a query
- find the most ‘authoritative’ web pages for this query

Step 0: find all pages containing the query terms
Step 1: expand by one move forward and backward

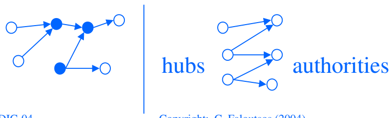
Kleinberg’s algorithm

- Step 1: expand by one move forward and backward



Kleinberg's algorithm

- on the resulting graph, give high score (= 'authorities') to nodes that many important nodes point to
- give high importance score ('hubs') to nodes that point to good 'authorities')



Kleinberg's algorithm

observations

- recursive definition!
- each node (say, '*i*'-th node) has both an authoritativeness score a_i and a hubness score h_i

Kleinberg's algorithm

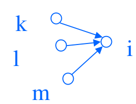
Let **A** be the adjacency matrix:

the (i,j) entry is 1 if the edge from i to j exists

Let **h** and **a** be [n x 1] vectors with the 'hubness' and 'authoritativiness' scores.

Then:

Kleinberg's algorithm



Then:

$$a_i = h_k + h_l + h_m$$

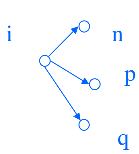
that is

a_i = Sum (h_j) over all j that (j,i) edge exists

or

$$\mathbf{a} = \mathbf{A}^T \mathbf{h}$$

Kleinberg's algorithm



symmetrically, for the 'hubness':

$$h_i = a_n + a_p + a_q$$

that is

h_i = Sum (q_j) over all j that (i,j) edge exists

or

$$\mathbf{h} = \mathbf{A}\,\mathbf{a}$$

Kleinberg's algorithm

In conclusion, we want vectors **h** and **a** such that:

$$\mathbf{h} = \mathbf{A}\,\mathbf{a}$$

$$\mathbf{a} = \mathbf{A}^T \mathbf{h}$$

That is:

$$\mathbf{a} = \mathbf{A}^T\mathbf{A}\,\mathbf{a}$$

Carnegie Mellon

Kleinberg's algorithm

a is a right- eigenvector of the adjacency matrix **A** (by dfn!)

Starting from random **a'** and iterating, we'll eventually converge

(Q: to which of all the eigenvectors? why?)

MDIC-04 Copyright: C. Faloutsos (2004) 115

Carnegie Mellon

Kleinberg's algorithm

(Q: to which of all the eigenvectors? why?)

A: to the one of the strongest eigenvalue, because of [Property 2]:

$$(\mathbf{A}^T \mathbf{A})^k \mathbf{v'} \sim \text{(constant) } \mathbf{v}_1$$

MDIC-04 Copyright: C. Faloutsos (2004) 116

Carnegie Mellon

Kleinberg's algorithm - results

Eg., for the query 'java':

0.328 www.gamelan.com

0.251 java.sun.com

0.190 www.digitalfocus.com ("the java developer")

MDIC-04 Copyright: C. Faloutsos (2004) 117

Carnegie Mellon

Kleinberg's algorithm - discussion

- 'authority' score can be used to find 'similar pages' (how?)
- closely related to 'citation analysis', social networks / 'small world' phenomena

MDIC-04 Copyright: C. Faloutsos (2004) 118

Carnegie Mellon

google/page-rank algorithm

- closely related: imagine a particle randomly moving along the edges (*)
- compute its steady-state probabilities

(*) with occasional random jumps

MDIC-04 Copyright: C. Faloutsos (2004) 119

Carnegie Mellon

SVD - outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Applications - success stories
- Conclusions

MDIC-04 Copyright: C. Faloutsos (2004) 120

Carnegie Mellon

SVD - conclusions

SVD: a **valuable** tool - two major properties:

- (A) it gives the *optimal* low-rank approximation (= least squares best hyper-plane)
  - it finds 'concepts' = principal components = features; (linear) correlations
  - dim. reduction; visualization
  - (solves over- and under-constraint problems)

MDIC-04 Copyright: C. Faloutsos (2004) 121

Carnegie Mellon

SVD - conclusions - cont'd

- (B) eigenvectors: fixed points

$$\mathbf{A}^{\mathbf{T}}\, \mathbf{A}\, \mathbf{v}_{\mathbf{i}} = \lambda_i\, \mathbf{v}_{\mathbf{i}}$$

  - steady-state prob. in Markov Chains / graphs
  - hubs/authorities

MDIC-04 Copyright: C. Faloutsos (2004) 122

Carnegie Mellon

Conclusions - in detail:

- given a (document-term) matrix, it finds 'concepts' (LSI)
- ... and can reduce dimensionality (KL)
- ... and can find rules (PCA; RatioRules)

MDIC-04 Copyright: C. Faloutsos (2004) 123

Carnegie Mellon

Conclusions cont'd

- ... and can find fixed-points or steady-state probabilities (Kleinberg / google/ Markov Chains)
- (... and can solve optimally over- and under-constraint linear systems - see Appendix, (Eq. C1) - Moore/Penrose pseudo-inverse)

MDIC-04 Copyright: C. Faloutsos (2004) 124

Carnegie Mellon

References

- Berry, Michael: http://www.cs.utk.edu/~lsi/
- Brin, S. and L. Page (1998). *Anatomy of a Large-Scale Hypertextual Web Search Engine*. 7th Intl World Wide Web Conf.
- Chen, C. M. and N. Roussopoulos (May 1994). *Adaptive Selectivity Estimation Using Query Feedback*. Proc. of the ACM-SIGMOD , Minneapolis, MN.

MDIC-04 Copyright: C. Faloutsos (2004) 125

Carnegie Mellon

References

- Duda, R. O. and P. E. Hart (1973). *Pattern Classification and Scene Analysis*. New York, Wiley.
- Faloutsos, C. (1996). *Searching Multimedia Databases by Content*, Kluwer Academic Inc.
- Foltz, P. W. and S. T. Dumais (Dec. 1992). "*Personalized Information Delivery: An Analysis of Information Filtering Methods*." Comm. of ACM (CACM) 35(12): 51-60.

MDIC-04 Copyright: C. Faloutsos (2004) 126

Carnegie Mellon

References

- Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, Academic Press.
- Jolliffe, I. T. (1986). *Principal Component Analysis*, Springer Verlag.
- Kleinberg, J. (1998). *Authoritative sources in a hyperlinked environment*. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.

Carnegie Mellon

References

- Korn, F., H. V. Jagadish, et al. (May 13-15, 1997). *Efficiently Supporting Ad Hoc Queries in Large Datasets of Time Sequences*. ACM SIGMOD, Tucson, AZ.
- Korn, F., A. Labrinidis, et al. (1998). *Ratio Rules: A New Paradigm for Fast, Quantifiable Data Mining*. VLDB, New York, NY.

Carnegie Mellon

References

- Korn, F., A. Labrinidis, et al. (2000). "*Quantifiable Data Mining Using Ratio Rules*." VLDB Journal 8(3-4): 254-266.
- Press, W. H., S. A. Teukolsky, et al. (1992). *Numerical Recipes in C*, Cambridge University Press.
- Strang, G. (1980). *Linear Algebra and Its Applications*, Academic Press.

Carnegie Mellon

Appendix: SVD properties

- (A): obvious
- (B): less obvious
- (C): least obvious (and most powerful!)

Carnegie Mellon

Properties - by defn.:

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

A(1): $\mathbf{U}^{T}_{[r \times n]} \mathbf{U}_{[n \times r]} = \mathbf{I}_{[r \times r]}$ (identity matrix)

A(2): $\mathbf{V}^{T}_{[r \times n]} \mathbf{V}_{[n \times r]} = \mathbf{I}_{[r \times r]}$

A(3): $\Lambda^{k} = \text{diag}(\lambda_1^{k}, \lambda_2^{k}, \ldots \lambda_r^{k})$ (k: ANY real number)

A(4): $\mathbf{A}^{\mathbf{T}} = \mathbf{V}\, \mathbf{\Lambda}\, \mathbf{U}^{\mathbf{T}}$

Carnegie Mellon

Less obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V}^{\mathbf{T}}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A}^{T})_{[m \times n]} = ??$

Less obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V^T}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A^T})_{[m \times n]} = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^T$

symmetric; Intuition?

Less obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V^T}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A^T})_{[m \times n]} = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^T$

symmetric; Intuition?

'document-to-document' similarity matrix

B(2): symmetrically, for '**V**'

$(\mathbf{A^T})_{[m \times n]} \mathbf{A}_{[n \times m]} = \mathbf{V} \mathbf{\Lambda}^2 \mathbf{V}^T$

Intuition?

Less obvious properties

A: term-to-term similarity matrix

B(3): $((\mathbf{A^T})_{[m \times n]} \mathbf{A}_{[n \times m]})^k = \mathbf{V} \mathbf{\Lambda}^{2k} \mathbf{V}^T$

and

B(4): $(\mathbf{A}^T \mathbf{A})^k \sim \mathbf{v}_1 \lambda_1^{2k} \mathbf{v}_1^T$ for k>>1

where

$\mathbf{v}_1$: [m x 1] first column (eigenvector) of **V**

λ_1: strongest eigenvalue

Less obvious properties

B(4): $(\mathbf{A}^T \mathbf{A})^k \sim \mathbf{v}_1 \lambda_1^{2k} \mathbf{v}_1^T$ for k>>1

B(5): $(\mathbf{A}^T \mathbf{A})^k \mathbf{v'} \sim \text{(constant)}\ \mathbf{v}_1$

ie., for (almost) any **v'**, it converges to a vector parallel to $\mathbf{v}_1$

Thus, useful to compute first eigenvector/value (as well as the next ones, too...)

Less obvious properties - repeated:

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V^T}_{[r \times m]}$

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A^T})_{[m \times n]} = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^T$

B(2): $(\mathbf{A^T})_{[m \times n]} \mathbf{A}_{[n \times m]} = \mathbf{V} \mathbf{\Lambda}^2 \mathbf{V}^T$

B(3): $((\mathbf{A^T})_{[m \times n]} \mathbf{A}_{[n \times m]})^k = \mathbf{V} \mathbf{\Lambda}^{2k} \mathbf{V}^T$

B(4): $(\mathbf{A}^T \mathbf{A})^k \sim v_1 \lambda_1^{2k} v_1^T$

B(5): $(\mathbf{A}^T \mathbf{A})^k \mathbf{v'} \sim \text{(constant)}\ \mathbf{v}_1$

Least obvious properties

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]} \mathbf{\Lambda}_{[r \times r]} \mathbf{V^T}_{[r \times m]}$

C(1): $\mathbf{A}_{[n \times m]} \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$

let $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{(-1)} \mathbf{U}^T \mathbf{b}$

if under-specified, $\mathbf{x}_0$ gives 'shortest' solution

if over-specified, it gives the 'solution' with the smallest least squares error

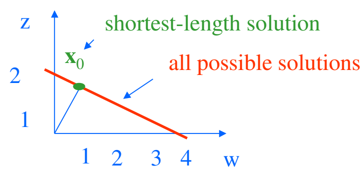
(see Num. Recipes, p. 62)

Carnegie Mellon

Least obvious properties

Illustration: under-specified, eg

$[1\ 2]\ [w\ z]^T = 4$ (ie, $1\ w + 2\ z = 4$)

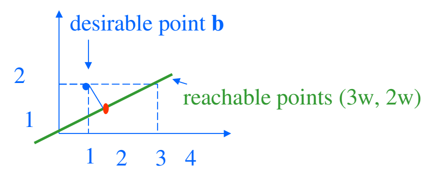


Carnegie Mellon

Least obvious properties

Illustration: over-specified, eg

$[3\ 2]^T\ [w] = [1\ 2]^T$ (ie, $3\ w = 1;\ 2\ w = 2$)



Carnegie Mellon

Least obvious properties - cont'd

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]}\ \mathbf{\Lambda}_{[r \times r]}\ \mathbf{V^T}_{[r \times m]}$

C(2): $\mathbf{A}_{[n \times m]}\ \mathbf{v_1}_{[m \times 1]} = \lambda_1\ \mathbf{u_1}_{[n \times 1]}$

where $\mathbf{v_1}$, $\mathbf{u_1}$ the first (column) vectors of **V**, **U.** ($\mathbf{v_1}$ == right-eigenvector)

C(3): symmetrically: $\mathbf{u_1}^T \mathbf{A} = \lambda_1 \mathbf{v_1}^T$

$\mathbf{u_1}$ == left-eigenvector

Therefore:

Carnegie Mellon

Least obvious properties - cont'd

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]}\ \mathbf{\Lambda}_{[r \times r]}\ \mathbf{V^T}_{[r \times m]}$

$\boxed{\text{C(4): } \mathbf{A^T\ A\ v_1} = \lambda_1^2\ \mathbf{v_1}}$

(**fixed point** - the usual dfn of eigenvector for a symmetric matrix)

Carnegie Mellon

Least obvious properties - altogether

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]}\ \mathbf{\Lambda}_{[r \times r]}\ \mathbf{V^T}_{[r \times m]}$

C(1): $\mathbf{A}_{[n \times m]}\ \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$

then, $\mathbf{x_0} = \mathbf{V}\ \mathbf{\Lambda}^{(-1)}\ \mathbf{U}^T\ \mathbf{b}$**:** shortest, actual or least-squares solution

C(2): $\mathbf{A}_{[n \times m]}\ \mathbf{v_1}_{[m \times 1]} = \lambda_1\ \mathbf{u_1}_{[n \times 1]}$

C(3): $\mathbf{u_1}^T \mathbf{A} = \lambda_1 \mathbf{v_1}^T$

C(4): $\mathbf{A^T\ A\ v_1} = \lambda_1^2\ \mathbf{v_1}$

Carnegie Mellon

Properties - conclusions

A(0): $\mathbf{A}_{[n \times m]} = \mathbf{U}_{[n \times r]}\ \mathbf{\Lambda}_{[r \times r]}\ \mathbf{V^T}_{[r \times m]}$

B(5): $(\mathbf{A}^T \mathbf{A})^k\ \mathbf{v'} \sim$ (constant) $\mathbf{v}_1$

C(1): $\mathbf{A}_{[n \times m]}\ \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$

then, $\mathbf{x_0} = \mathbf{V}\ \mathbf{\Lambda}^{(-1)}\ \mathbf{U}^T\ \mathbf{b}$**:** shortest, actual or least-squares solution

C(4): $\mathbf{A^T\ A\ v_1} = \lambda_1^2\ \mathbf{v_1}$