

CMU SCS

## 15-826: Multimedia Databases and Data Mining

*SVD - part I (definitions)*  
C. Faloutsos

CMU SCS

## Outline

Goal: 'Find similar / interesting things'

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

15-826 Copyright: C. Faloutsos (2006) 2

CMU SCS

## Indexing - Detailed outline

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

15-826 Copyright: C. Faloutsos (2006) 3

CMU SCS

## SVD - Detailed outline

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

15-826 Copyright: C. Faloutsos (2006) 4

CMU SCS

## SVD - Motivation

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

15-826 Copyright: C. Faloutsos (2006) 5

CMU SCS

## SVD - Motivation

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

document	term	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

15-826 Copyright: C. Faloutsos (2006) 6

CMU SCS

## SVD - Motivation

- problem #2: compress / reduce dimensionality

15-826 Copyright: C. Faloutsos (2006) 7

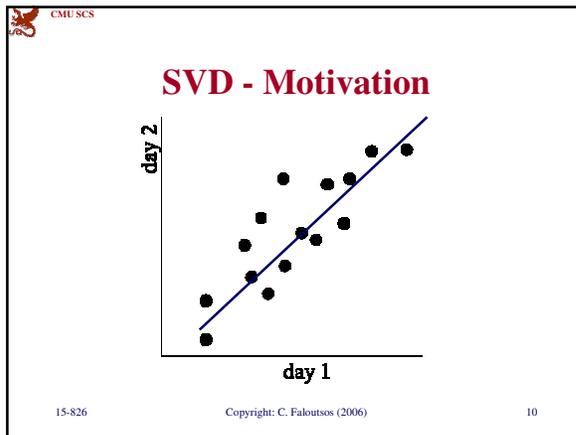
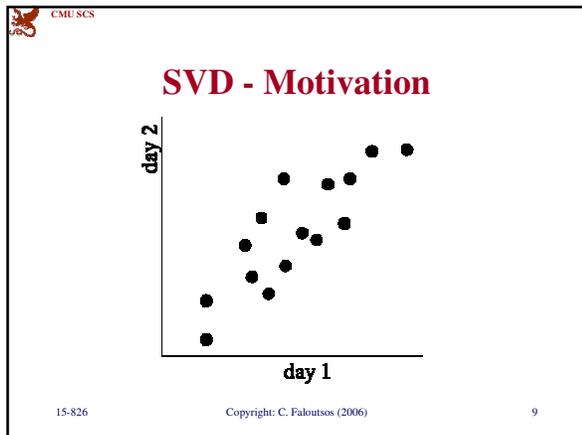
CMU SCS

## Problem - specs

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

customer	day	We	Th	Fr	Sa	Su
	7/10/06	7/11/06	7/12/06	7/13/06	7/14/06	
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	

15-826 Copyright: C. Faloutsos (2006) 8



CMU SCS

## SVD - Detailed outline

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

15-826 Copyright: C. Faloutsos (2006) 11

CMU SCS

## SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \phantom{0} \\ \phantom{0} \\ \phantom{0} \end{bmatrix}$$

$3 \times 2$        $2 \times 1$

15-826 Copyright: C. Faloutsos (2006) 12

CMU SCS

### SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \phantom{0} \\ \phantom{0} \\ \phantom{0} \end{bmatrix}$$

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

15-826      Copyright: C. Faloutsos (2006)      13

CMU SCS

### SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ \phantom{0} \\ \phantom{0} \end{bmatrix}$$

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

15-826      Copyright: C. Faloutsos (2006)      14

CMU SCS

### SVD - Definition

(reminder: matrix multiplication)

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \phantom{0} \end{bmatrix}$$

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

15-826      Copyright: C. Faloutsos (2006)      15

CMU SCS

### SVD - Definition

(reminder: matrix multiplication)

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

15-826      Copyright: C. Faloutsos (2006)      16

CMU SCS

### SVD - Definition

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

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

15-826      Copyright: C. Faloutsos (2006)      17

CMU SCS

### SVD - Definition

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

A	U	Lambda	V <sup>T</sup>
n × m	n × r	r × r	r × m
<div style="border: 1px solid black; width: 40px; height: 80px; margin: 0 auto;"></div>	<div style="border: 1px solid black; width: 20px; height: 80px; margin: 0 auto;"></div>	<div style="border: 1px solid black; width: 20px; height: 20px; margin: 0 auto; display: flex; align-items: center; justify-content: center;"> <span style="font-size: 10px;">λ</span> </div>	<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
=			
=			
=			
=			

15-826      Copyright: C. Faloutsos (2006)      18



### SVD - Example

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

retrieval  
inf. ↓ brain lung

data

CS

MD

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

term-to-concept similarity matrix

CS-concept

15-826 Copyright: C. Faloutsos (2006) 25

### SVD - Detailed outline

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

15-826 Copyright: C. Faloutsos (2006) 26

### 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

15-826 Copyright: C. Faloutsos (2006) 27

### SVD - Interpretation #2

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

15-826 Copyright: C. Faloutsos (2006) 28

### SVD - Motivation

15-826 Copyright: C. Faloutsos (2006) 29

### SVD - interpretation #2

SVD: gives best axis to project

- minimum RMS error

15-826 Copyright: C. Faloutsos (2006) 30

CMU SCS

## SVD - Interpretation #2

customer	day	We 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
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

15-826 Copyright: C. Faloutsos (2006) 31

CMU SCS

## SVD - Interpretation #2

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

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

*v1*

15-826 Copyright: C. Faloutsos (2006) 32

CMU SCS

## SVD - Interpretation #2

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

variance ('spread') on the  $v_1$  axis

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

15-826 Copyright: C. Faloutsos (2006) 33

CMU SCS

## SVD - Interpretation #2

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

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

15-826 Copyright: C. Faloutsos (2006) 34

CMU SCS

## SVD - Interpretation #2

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

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

15-826 Copyright: C. Faloutsos (2006) 35

CMU SCS

## SVD - Interpretation #2

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

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

15-826 Copyright: C. Faloutsos (2006) 36

**SVD - Interpretation #2**

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

15-826 Copyright: C. Faloutsos (2006) 37

**SVD - Interpretation #2**

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

15-826 Copyright: C. Faloutsos (2006) 38

**SVD - Interpretation #2**

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

15-826 Copyright: C. Faloutsos (2006) 39

**SVD - Interpretation #2**

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

15-826 Copyright: C. Faloutsos (2006) 40

**SVD - Interpretation #2**

Exactly equivalent:  
‘spectral decomposition’ of the matrix:

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

15-826 Copyright: C. Faloutsos (2006) 41

**SVD - Interpretation #2**

Exactly equivalent:  
‘spectral decomposition’ of the matrix:

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

15-826 Copyright: C. Faloutsos (2006) 42

**SVD - Interpretation #2**

Exactly equivalent:  
‘spectral decomposition’ of the matrix:

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

15-826 Copyright: C. Faloutsos (2006) 43

**SVD - Interpretation #2**

Exactly equivalent:  
‘spectral decomposition’ of the matrix:

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

← r terms →

15-826 Copyright: C. Faloutsos (2006) 44

**SVD - Interpretation #2**

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

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

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

15-826 Copyright: C. Faloutsos (2006) 45

**SVD - Interpretation #2**

A (heuristic - [Fukunaga]): keep 80-90% of  
‘energy’ (= sum of squares of  $\lambda_i$ ’s)

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

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

15-826 Copyright: C. Faloutsos (2006) 46

**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

15-826 Copyright: C. Faloutsos (2006) 47

**SVD - Interpretation #3**

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

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

15-826 Copyright: C. Faloutsos (2006) 48

**SVD - Interpretation #3**

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

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

15-826 Copyright: C. Faloutsos (2006) 49

**SVD - Interpretation #3**

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

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

15-826 Copyright: C. Faloutsos (2006) 50

**SVD - Interpretation #3**

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

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

15-826 Copyright: C. Faloutsos (2006) 51

**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??

15-826 Copyright: C. Faloutsos (2006) 52

**SVD - Interpretation #3**

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

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

15-826 Copyright: C. Faloutsos (2006) 53

**SVD - Interpretation #3**

- and the singular values are:

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

15-826 Copyright: C. Faloutsos (2006) 54

CMU SCS

## SVD - Interpretation #3

- Q: How to check we are correct?

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

15-826 Copyright: C. Faloutsos (2006) 55

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  **$\Lambda$**  should be diagonal, with positive values

15-826 Copyright: C. Faloutsos (2006) 56

CMU SCS

## SVD - Detailed outline

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

15-826 Copyright: C. Faloutsos (2006) 57

CMU SCS

## 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, mathematica ...)

15-826 Copyright: C. Faloutsos (2006) 58

CMU SCS

## 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
- $\Lambda$** : 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 'blobs'

15-826 Copyright: C. Faloutsos (2006) 59

CMU SCS

## 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.

15-826 Copyright: C. Faloutsos (2006) 60