

CSE 6242 / CX 4242

Apr 3, 2014

# **Text Analytics (Text Mining)**

LSI (uses SVD), Visualization

**Duen Horng (Polo) Chau**  
Georgia Tech

# **Singular Value Decomposition (SVD):**

## **Motivation**

### **Problem #1:**

Text - LSI uses SVD find “concepts”

### **Problem #2:**

Compression / dimensionality reduction

# SVD - Motivation

Problem #1: text - LSI: find “concepts”

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

# SVD - Motivation

Customer-product, for recommendation system:

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

# SVD - Motivation

- problem #2: compress / reduce dimensionality

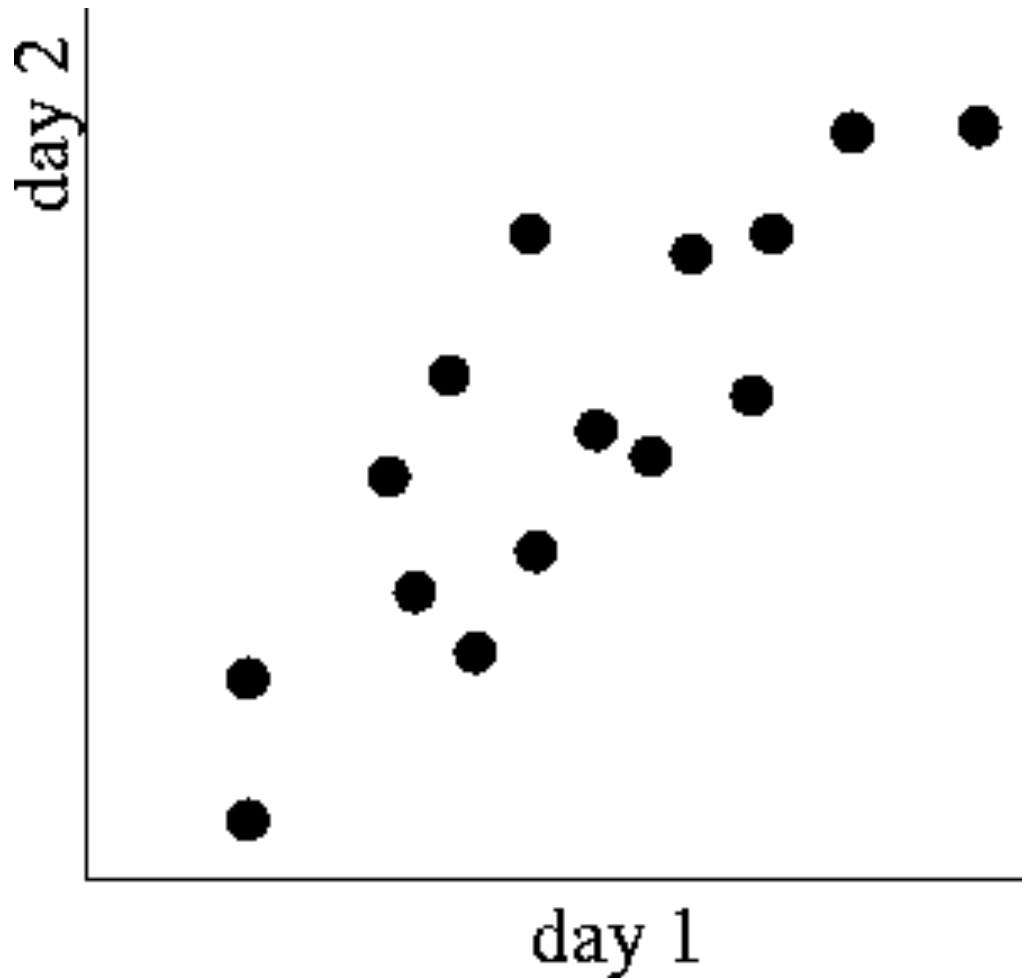
# Problem - Specification

~10<sup>6</sup> rows; ~10<sup>3</sup> columns; no updates;

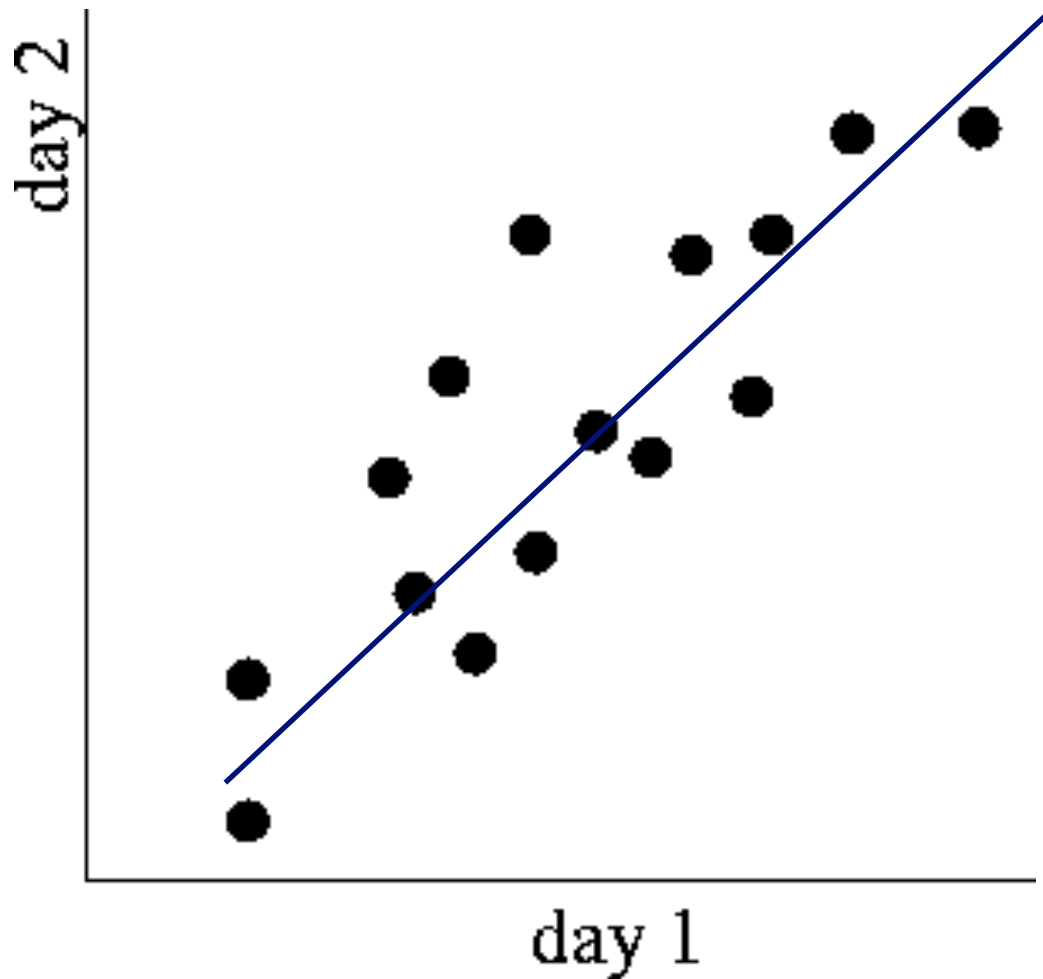
Random access to any cell(s); small error: OK

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

# SVD - Motivation



# SVD - Motivation





# SVD - Definition

(reminder: matrix multiplication)

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

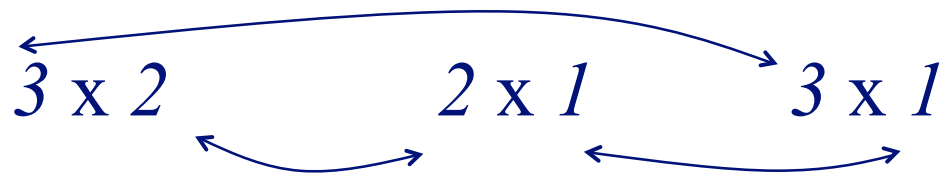
$3 \times 2$

$2 \times 1$

# SVD - Definition

(reminder: matrix multiplication)

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



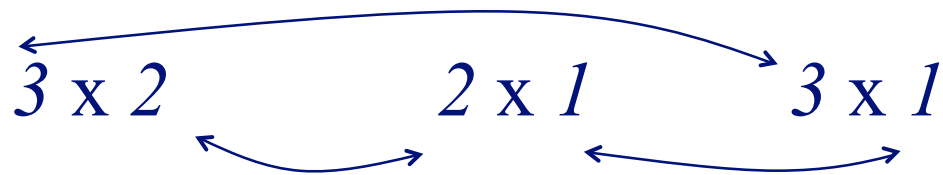
$3 \times 2$        $2 \times 1$        $3 \times 1$

# SVD - Definition

(reminder: matrix multiplication)

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

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

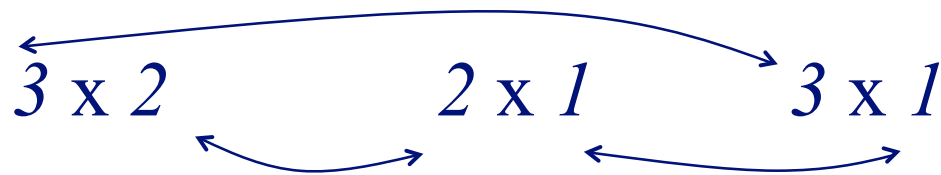


# SVD - Definition

(reminder: matrix multiplication)

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

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



# SVD - Definition

(reminder: matrix multiplication)

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

# SVD - Definition

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

**A: n x m matrix**

e.g., n documents, m terms

**U: n x r matrix**

e.g., n documents, r concepts

**$\mathbf{\Lambda}$ : r x r diagonal matrix**

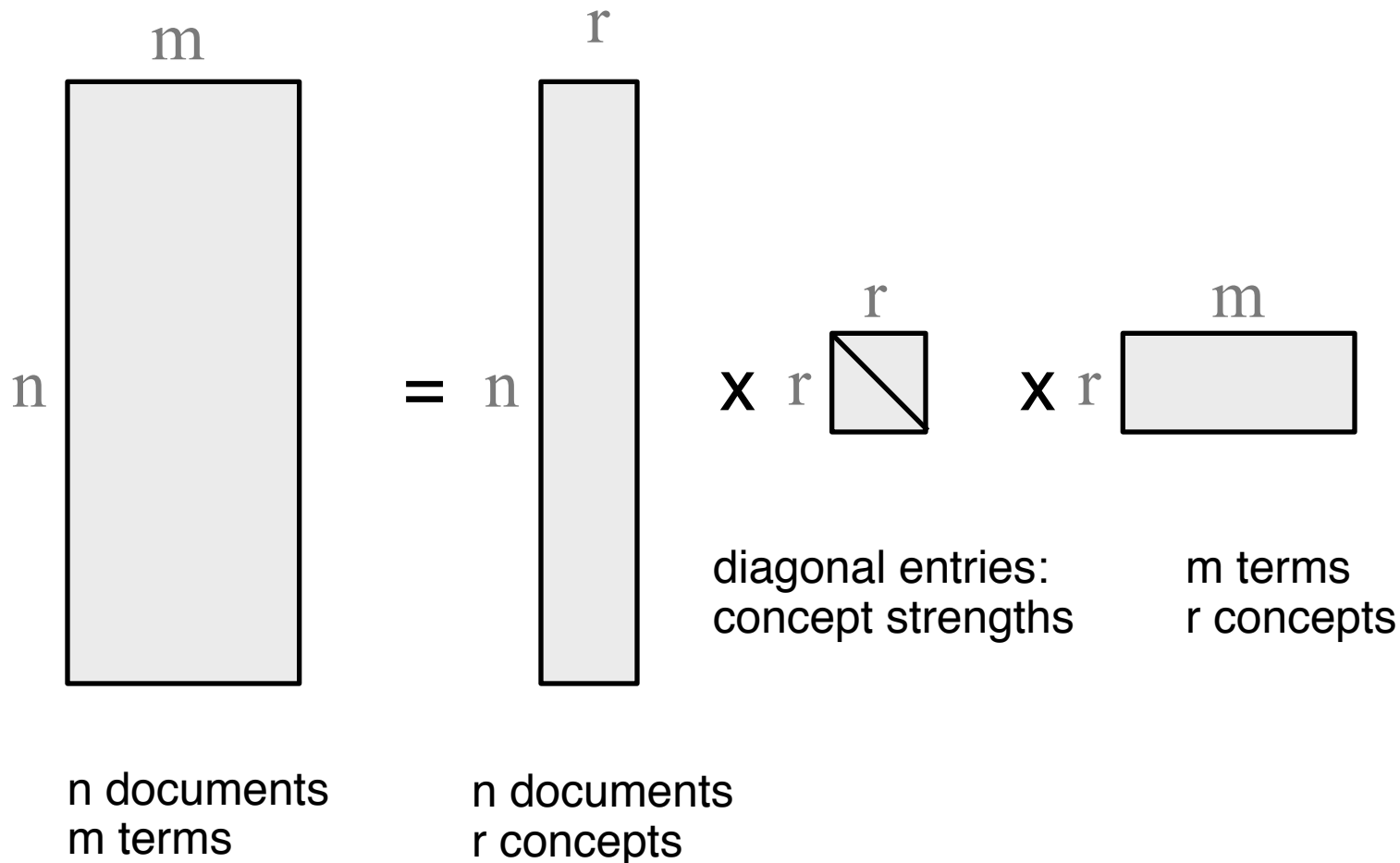
r : rank of the matrix; strength of each 'concept'

**V: m x r matrix**

e.g., m terms, r concepts

# SVD - Definition

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



# SVD - Properties

**THEOREM** [Press+92]:

**always possible to decompose** matrix  $\mathbf{A}$  into

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

$\mathbf{U}$ ,  $\mathbf{\Lambda}$ ,  $\mathbf{V}$ : **unique**, most of the time

$\mathbf{U}$ ,  $\mathbf{V}$ : column **orthonormal**

i.e., columns are unit vectors, orthogonal to each other

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \quad (\mathbf{I}: \text{identity matrix})$$

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}$$

$\mathbf{\Lambda}$ : diagonal matrix with non-negative diagonal entries, sorted in decreasing order



# SVD - Example

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

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

# SVD - Example

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

Diagram illustrating the SVD decomposition of matrix  $\mathbf{A}$  into  $\mathbf{U}$ ,  $\mathbf{\Lambda}$ , and  $\mathbf{V}^T$ .

**Matrix  $\mathbf{A}$  (Dimensions: 7 rows by 5 columns):**

- Columns are labeled: data, inf., retrieval, brain, lung.
- Rows are grouped by vertical arrows:
  - CS (Conceptual Space) for the first 4 rows (data, inf., retrieval, brain).
  - MD (Motoric Dimension) for the last 3 rows (lung, brain, data).

**SVD Decomposition:**

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

**Matrix  $\mathbf{U}$  (Dimensions: 7 rows by 2 columns):**

- Columns are labeled: CS-concept (indicated by an orange arrow to the first column) and MD-concept (indicated by an orange arrow to the second column).

**Matrix  $\mathbf{\Lambda}$  (Dimensions: 2 rows by 2 columns):**

**Matrix  $\mathbf{V}^T$  (Dimensions: 5 rows by 5 columns):**

# SVD - Example

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

doc-to-concept  
similarity matrix

retrieval CS-concept MD-concept

data inf. brain lung

CS

MD

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

# SVD - Example

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

retrieval  
inf. ↓ brain lung

‘strength’ of CS-concept

↑

CS

↓

↑

MD

↓

1	1	1	0	0	=	0.18	0	×	9.64	0	×	0.58	0.58	0.58	0	0
2	2	2	0	0		0.36	0		0	5.29		0	0	0	0.71	0.71
1	1	1	0	0		0.18	0		0							
5	5	5	0	0		0.90	0		0							
0	0	0	2	2		0	0.53		0							
0	0	0	3	3		0	0.80		0							
0	0	0	1	1		0	0.27		0							

data

# SVD - Example

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

term-to-concept  
similarity matrix

retrieval  
inf. ↓ brain lung

data

CS

MD

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

CS-concept

# SVD - Example

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

term-to-concept  
similarity matrix

retrieval  
inf. ↓  
data brain lung

CS  
↑  
↓  
↑  
MD  
↓

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

CS-concept

# SVD - Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:

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

# SVD – Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:

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

A:

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

A:



# SVD – Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:

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

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

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

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

# SVD properties

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

$$\mathbf{X}^T \mathbf{X} = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T)^T (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) = \mathbf{V} \mathbf{\Sigma}^2 \mathbf{V}^T$$

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

$$\mathbf{X} \mathbf{X}^T = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T)^T = \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^T$$

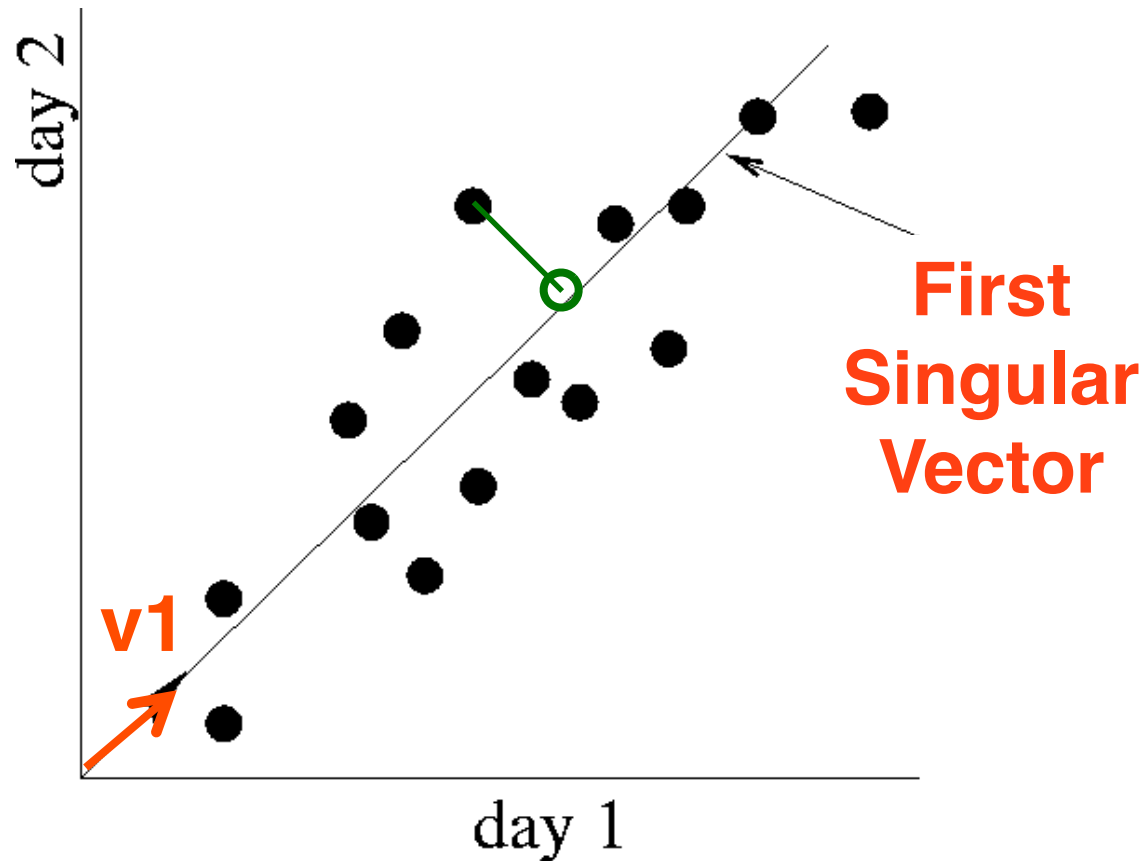
Thus, SVD is closely related to PCA, and can be numerically more stable.  
For more info, see:

<http://math.stackexchange.com/questions/3869/what-is-the-intuitive-relationship-between-svd-and-pca>  
Ian T. Jolliffe, *Principal Component Analysis* (2<sup>nd</sup> ed), Springer, 2002.  
Gilbert Strang, *Linear Algebra and Its Applications* (4<sup>th</sup> ed), Brooks Cole, 2005.

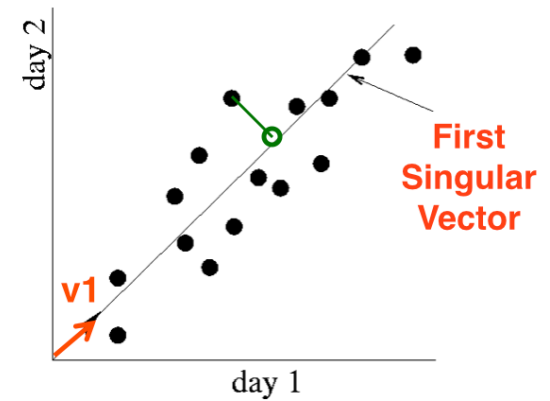
# SVD - Interpretation #2

best axis to project on

(‘best’ = min sum of squares of projection errors)



# SVD - Interpretation #2



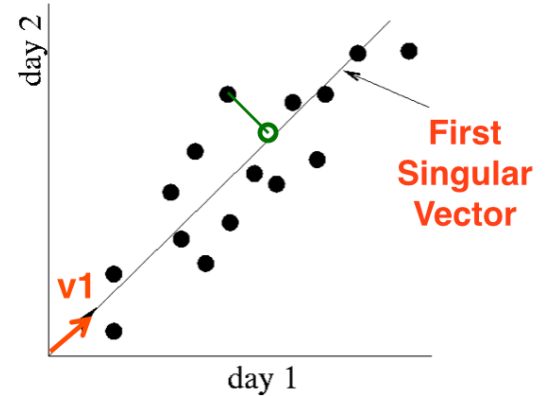
- $A = U \Lambda V^T$  - example:

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

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

The value 9.64 in the diagonal matrix is circled in orange, with an arrow pointing to it from the text 'variance ('spread') on the  $v_1$  axis'. The label  $v_1$  is placed to the right of the third matrix. The entire equation is labeled with 'x' between the matrices.

# SVD - Interpretation #2



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

–  $\mathbf{U} \mathbf{\Lambda}$  gives the **coordinates** of the points in the projection axis

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

# SVD - Interpretation #2

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

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

# SVD - Interpretation #2

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

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

# SVD - Interpretation #2

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



# SVD - Interpretation #2

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

The image shows a matrix decomposition with several elements crossed out with orange lines, indicating they are zero or negligible. 
 In the first matrix, the last three rows are non-zero. 
 In the second matrix, the first two columns are crossed out, leaving only the third column (0.53, 0.80, 0.27). 
 In the third matrix, the second column is crossed out, leaving only the first column (9.64, 0). 
 In the fourth matrix, the first two rows are crossed out, leaving only the last three rows (0.58, 0.58, 0.58 in the first row and 0, 0, 0 in the second row).

# SVD - Interpretation #2

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

# SVD - Interpretation #2

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

# SVD - Interpretation #3

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

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

# SVD - Interpretation #3

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

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

# SVD - Interpretation #3

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

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

Row 1



Row 4



Row 5



Row 7



# **SVD algorithm**

- Numerical Recipes in C (free)

# SVD - Interpretation #3

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

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



# 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)

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



orthogonal??

$$\mathbf{x} \begin{bmatrix} ?? & 0 \\ 0 & ?? \end{bmatrix} \mathbf{x}$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

# SVD - Interpretation #3

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

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

# SVD - Interpretation #3

- and the singular values are:

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

# SVD - Interpretation #3

- Q: How to check we are correct?

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

# SVD - Interpretation #3

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

# SVD - Complexity

$O(n*m*m)$  or  $O(n*n*m)$  (whichever is less)

Faster version, if just want singular values  
or if we want first  $k$  singular vectors  
or if the matrix is sparse [Berry]

No need to write your own!

Available in most linear algebra packages  
(LINPACK, matlab, Splus/R,  
mathematica ...)

# 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.



# 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'

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

# Case study - LSI

Q1: How to do queries with LSI?

A: map query vectors into 'concept space' – how?

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

# Case study - LSI

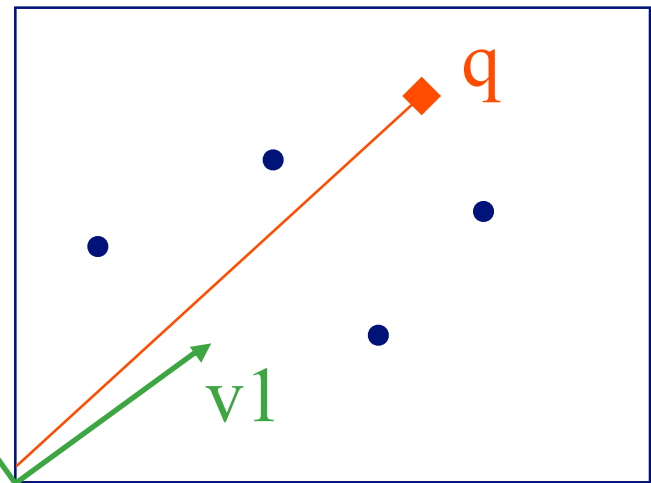
Q1: How to do queries with LSI?

A: map query vectors into 'concept space' – how?

$$q = \begin{matrix} & \text{data} & \text{inf} & \text{retrieval} & \text{brain} & \text{lung} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

term2

v2



term1

# Case study - LSI

Q1: How to do queries with LSI?

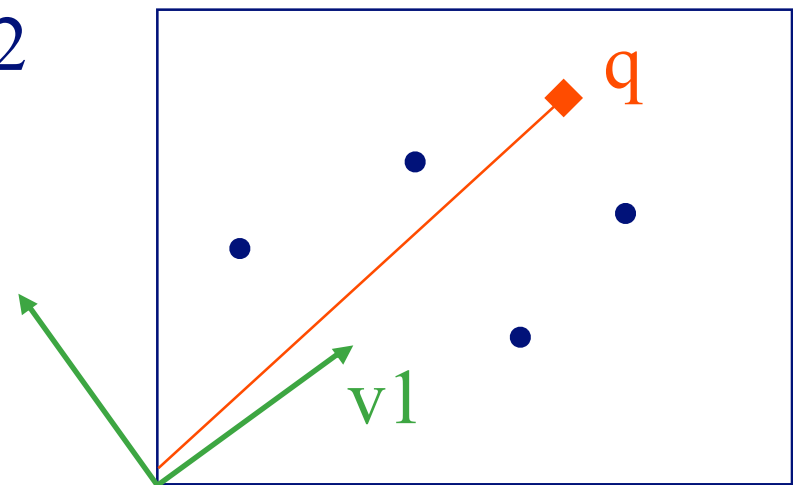
A: map query vectors into 'concept space' – how?

$$q = \begin{matrix} & \begin{matrix} \text{data} & \text{inf} & \text{retrieval} & \text{brain} & \text{lung} \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

A: inner product  
(cosine similarity)  
with each 'concept' vector  $v_i$

term2

$v_2$



term1

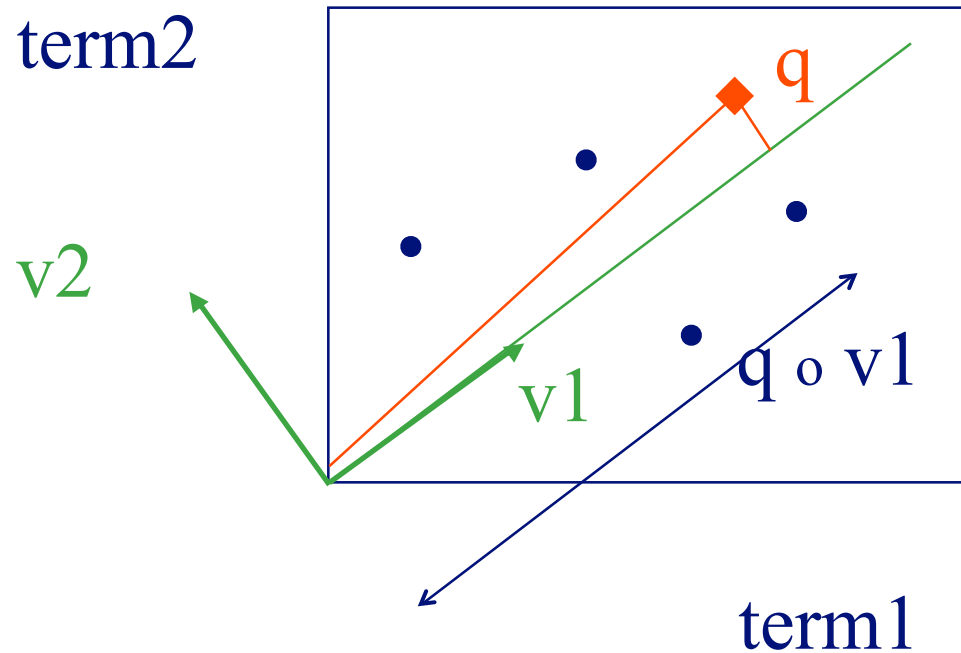
# Case study - LSI

Q1: How to do queries with LSI?

A: map query vectors into 'concept space' – how?

$$q = \begin{matrix} & \text{data} & \text{inf} & \text{retrieval} & \text{brain} & \text{lung} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

A: inner product  
(cosine similarity)  
with each 'concept' vector  $v_i$



# Case study - LSI

compactly, we have:

$$q V = q_{\text{concept}}$$

Eg:

$$q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} \text{data} & \text{inf} & \text{retrieval} & \text{brain} & \text{lung} \\ & \downarrow & & & \end{matrix} \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} \begin{matrix} \text{CS-concept} \\ \downarrow \\ 0.58 & 0 \end{matrix}$$

term-to-concept similarities

# **Case study - LSI**

Drill: how would the document ('information', 'retrieval') be handled by LSI?



# Case study - LSI

Drill: how would the document ('information', 'retrieval') be handled by LSI? **A: SAME:**

$$d_{\text{concept}} = d V$$

Eg:  $d = \begin{bmatrix} \text{data} & \text{inf} & \text{retrieval} & \text{brain} & \text{lung} \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$

$$d = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

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

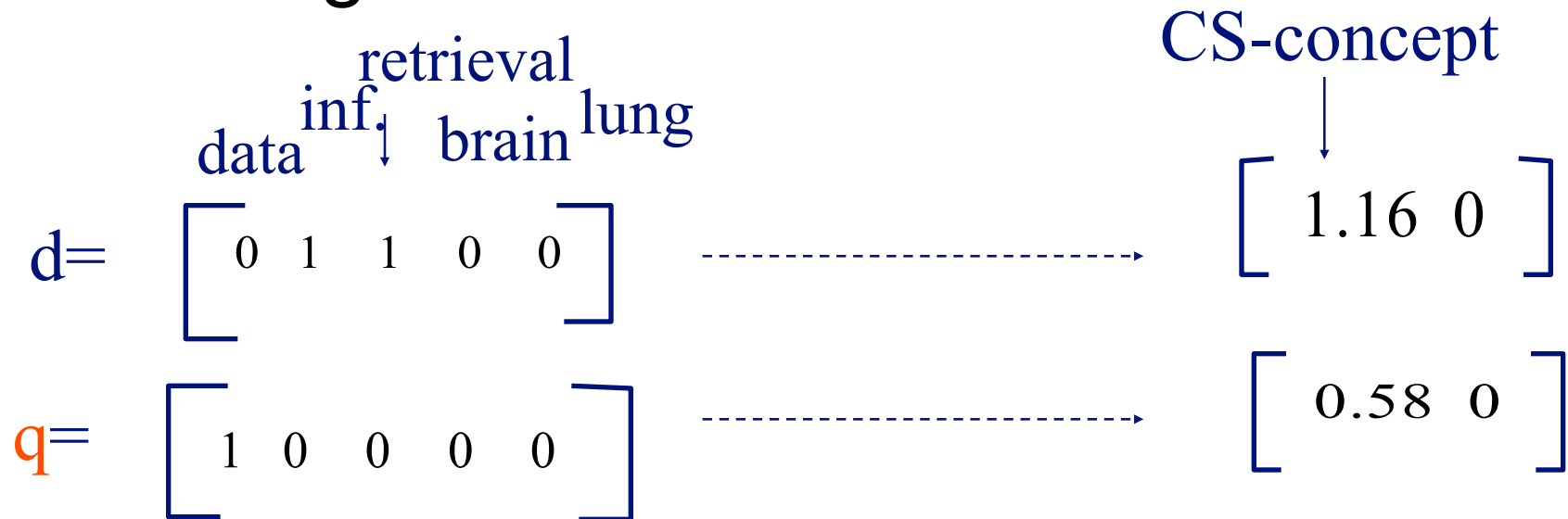
term-to-concept  
similarities

CS-concept

$$= \begin{bmatrix} 1.16 & 0 \end{bmatrix}$$

# Case study - LSI

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



# Case study - LSI

Q1: How to do queries with LSI?

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

# Case study - LSI

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

# Case study - LSI

- Solution:  $\sim$  LSI

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

# Switch Gear to **Text Visualization**

What comes up to your mind?

What visualization have you seen before?

# Word/Tag Cloud (still popular?)

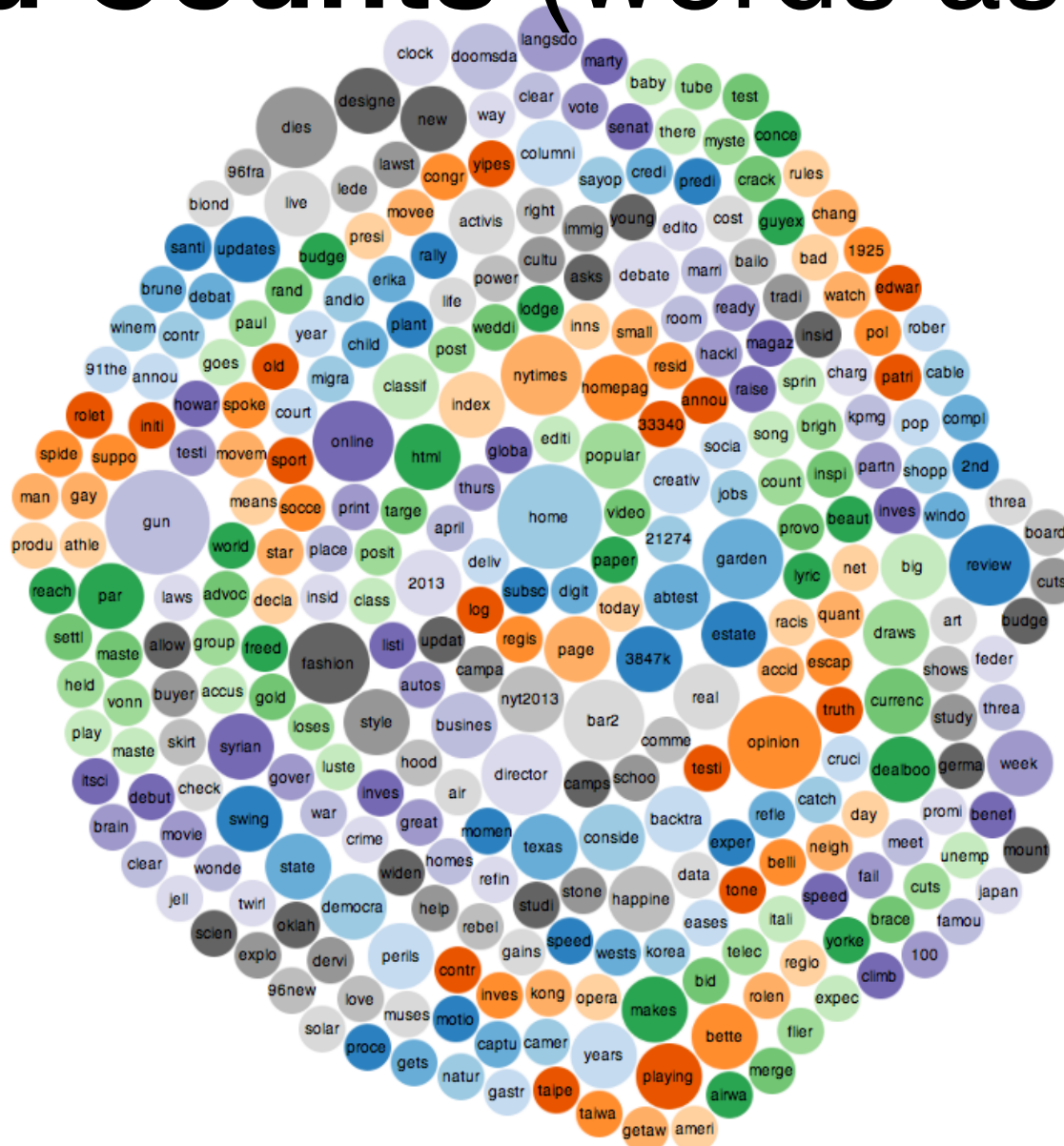


<http://www.wordle.net>

Twitter | [Tweeps](#) | [Wikipedia](#) | [Custom](#)

Keyword:

# Word Counts (words as bubbles)



<http://www.infocaptor.com/bubble-my-page>



# Word Tree

## word tree

We

☐ reverse tree ☐ one phrase per line

Shift-click to make that word the root.



substitute spectacle for politics, or treat name-calling as reasoned debate. We must act, we must act knowing that our work will be imperfect. We must act, knowing that today's victories will be only partial, and that it will be up to those who stand here in four years, and forty years, and four hundred years hence to advance the timeless spirit once conferred to us in a spare Philadelphia hall.

My fellow Americans, the oath I have sworn before you today, like the one recited by others who serve in this Capitol, was an oath to God and country, not party or faction - and we must faithfully execute that pledge during the duration of our service. But the words I spoke today are not so different from the oath that is taken each time a soldier signs up for duty, or an immigrant realizes her dream. My oath is not so

# Phrase Net

Visualize pairs of words that satisfy a particular pattern, e.g., X and Y

