CSE 6242 A / CS 4803 DVA
Apr 11, 2013

Text Analytics (Text Mining)

LSI (SVD), Visualization

Duen Horng (Polo) Chau
Georgia Tech

Some lectures are partly based on materials by
Professors Guy Lebanon, Jeffrey Heer, John Stasko, Christos Faloutsos, Le Song

SVD - Motivation

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

SVD - Motivation

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

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

SVD - Motivation

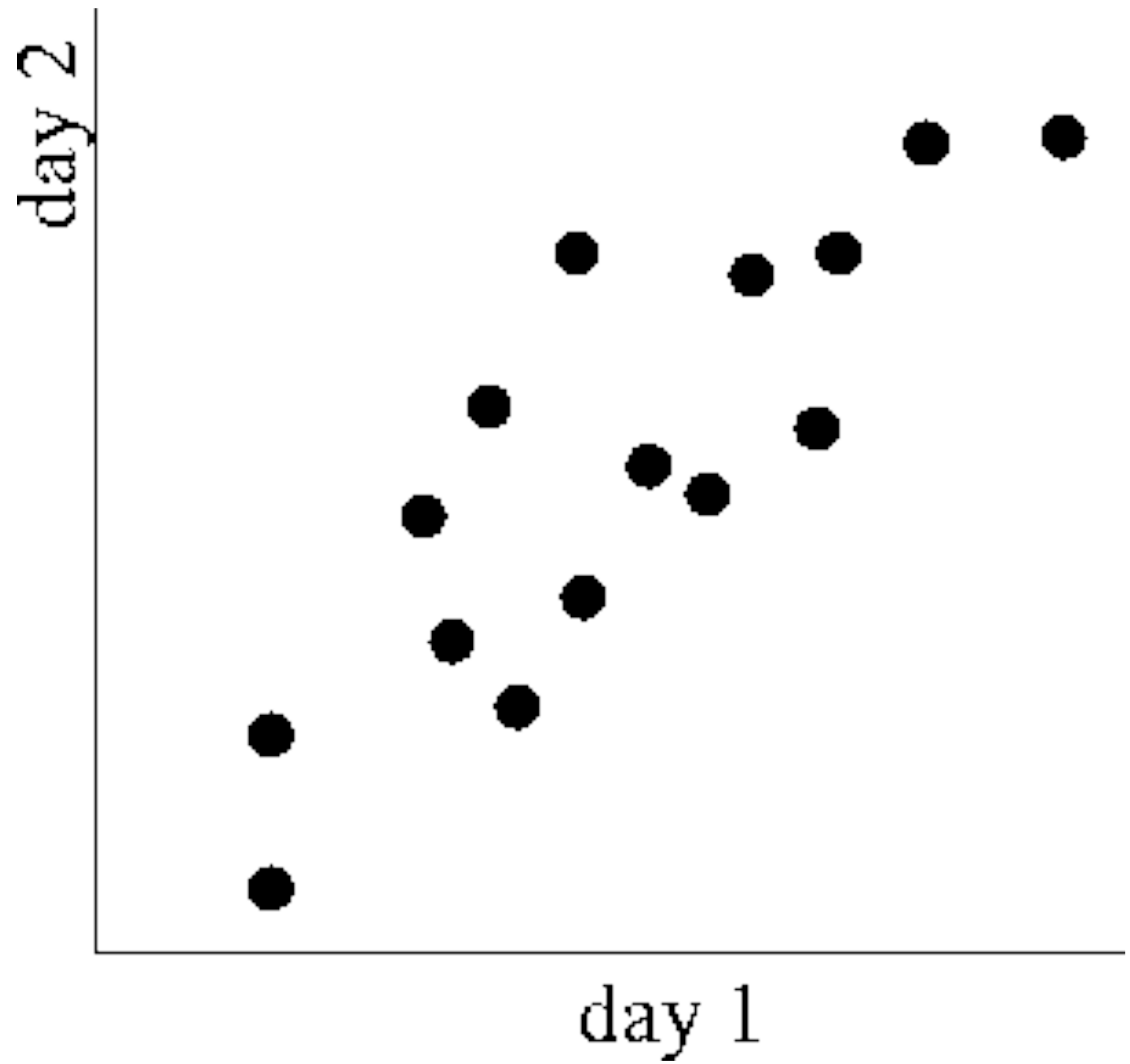
- Customer-product, for recommendation system:

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

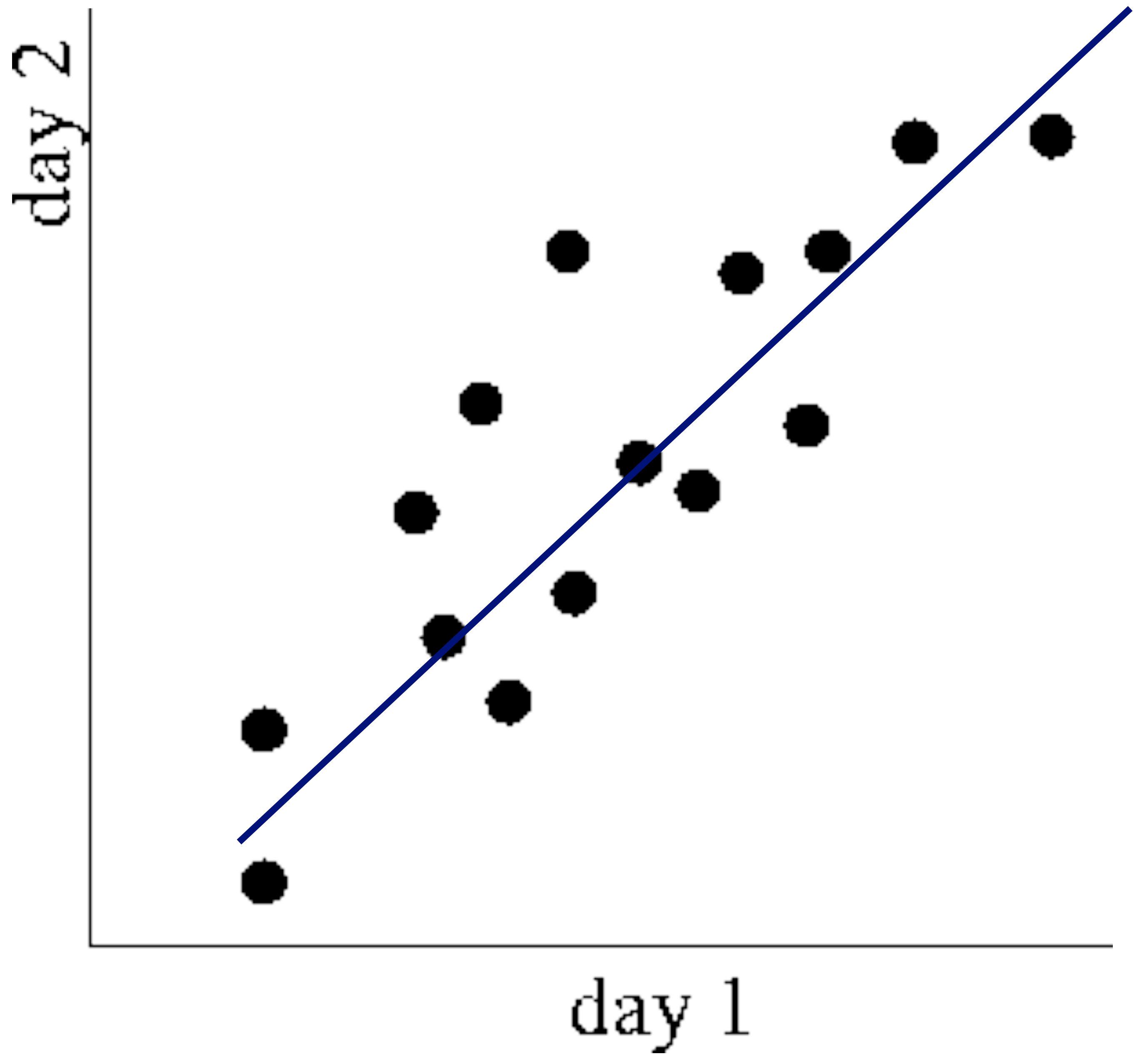
SVD - Motivation

- problem #2: compress / reduce dimensionality (mostly skipped)

SVD - Motivation



SVD - Motivation



SVD - Definition

(reminder: matrix multiplication

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

$\mathit{3}$ x $\mathit{2}$ $\qquad$ $\mathit{2}$ x $\mathit{1}$

SVD - Definition

(reminder: matrix multiplication

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

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

SVD - Definition

(reminder: matrix multiplication

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

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

SVD - Definition

(reminder: matrix multiplication

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

3 x 2 $\quad$ 2 x 1 $\quad$ 3 x 1

SVD - Definition

(reminder: matrix multiplication

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

SVD - Definition

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

A: n x m matrix
e.g., n documents, m terms

U: n x r matrix
e.g., n documents, r concepts

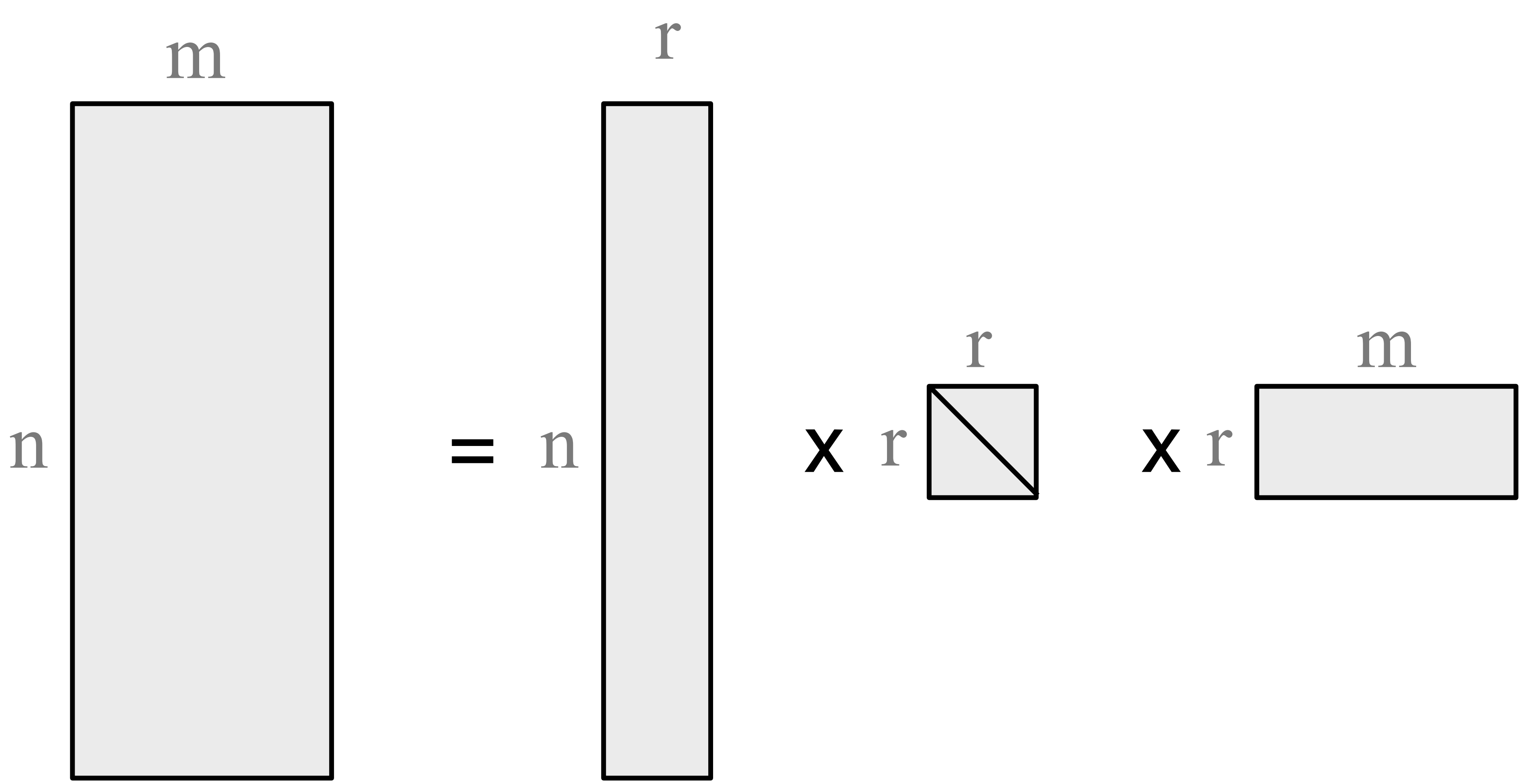
Λ: r x r diagonal matrix
r : rank of the matrix; strength of each ‘concept’

V: m x r matrix

e.g., m terms, r concepts

SVD - Definition

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



SVD - Properties

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

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

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

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

$$\mathbf{U}^T\,\mathbf{U} = \mathbf{I}$$

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

(**I**: identity matrix)

Λ: singular are positive, and sorted in decreasing order

SVD - Example

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

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

SVD - Example

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

doc-to-concept similarity matrix

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

SVD - Example

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

‘strength’ of CS-concept

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

SVD - Example

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

term-to-concept similarity matrix

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

CS-concept

SVD - Example

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

term-to-concept similarity matrix

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

CS-concept

SVD - Interpretation #1

'documents', 'terms' and 'concepts':

- **U**: document-to-concept similarity matrix
- **V**: term-to-concept sim. matrix
- Λ: its diagonal elements: 'strength' of each concept

SVD – Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:

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

A:

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

A:

SVD – Interpretation #1

'documents', 'terms' and 'concepts':

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

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

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

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

SVD properties

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

$$\mathbf{X}^{\mathsf{T}}\mathbf{X} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}\right)^{\mathsf{T}}\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}\right) = \mathbf{V}\boldsymbol{\Sigma}^{2}\mathbf{V}^{\mathsf{T}}$$

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

$$\mathbf{X}\mathbf{X}^{\mathsf{T}} = \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}\right)\left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{T}}\right)^{\mathsf{T}} = \mathbf{U}\boldsymbol{\Sigma}^{2}\mathbf{U}^{\mathsf{T}}$$

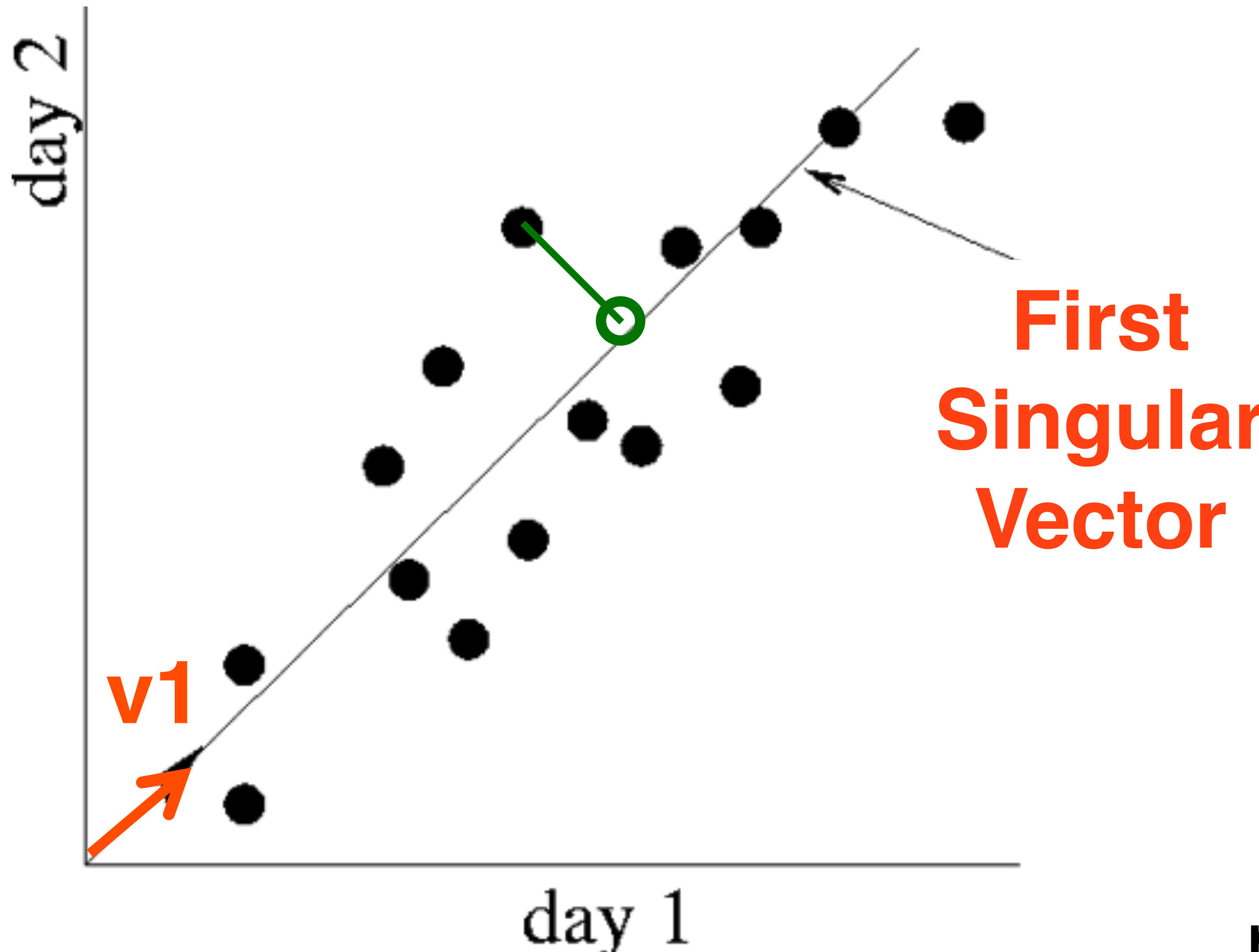
Thus, SVD is closely related to PCA, and can be numerically more stable.
For more info, see:

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

SVD - Interpretation #2

best axis to project on

('best' = min sum of squares of projection errors)



min RMS error

SVD - Interpretation #2

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

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

$$\begin{array}{c} \hfill \text{v1} \\ \begin{bmatrix} \boxed{0.58 \quad 0.58 \quad 0.58 \quad 0 \quad 0} \\ 0 \quad 0 \quad 0 \quad 0.71 \quad 0.71 \end{bmatrix} \end{array}$$

SVD - Interpretation #2

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

variance ('spread') on the v1 axis

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

SVD - Interpretation #2

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

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

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 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

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

SVD - Interpretation #2

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

SVD - Interpretation #2

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

SVD - Interpretation #2

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

SVD - Interpretation #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

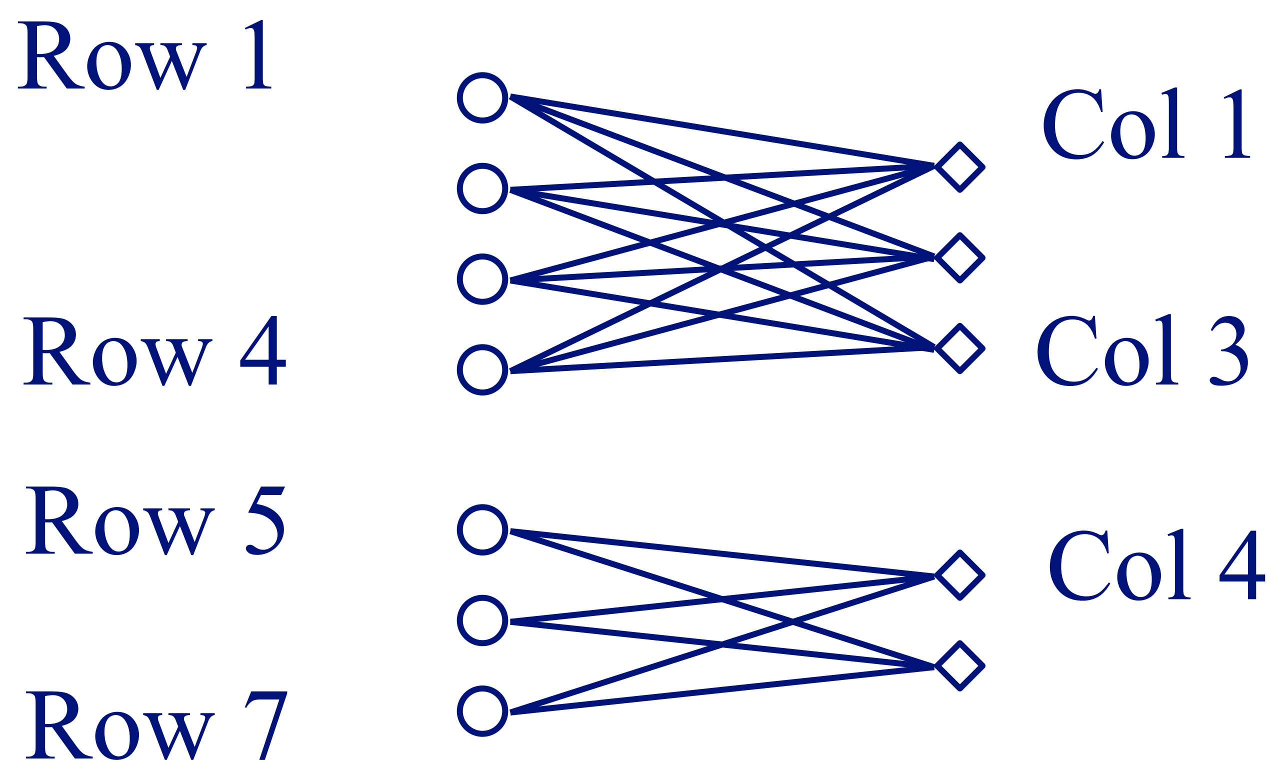
- finds non-zero 'blobs' in a data matrix

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

SVD - Interpretation #3

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

$$\left[\begin{array}{ccc|cc} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right]$$



SVD algorithm

- Numerical Recipes in C (free)

SVD - Interpretation #3

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

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

SVD - Interpretation #3

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

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

SVD - Interpretation #3

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

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

orthogonal??

SVD - Interpretation #3

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

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

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

SVD - Interpretation #3

- and the singular values are:

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

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

SVD - Interpretation #3

- Q: How to check we are correct?

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

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

SVD - Interpretation #3

- A: SVD properties:
  - matrix product should give back matrix $\mathbf{A}$
  - matrix $\mathbf{U}$ should be column-orthonormal, i.e., columns should be unit vectors, orthogonal to each other
  - ditto for matrix $\mathbf{V}$
  - matrix Λ should be diagonal, with positive 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'

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

Case study - LSI

Q1: How to do queries with LSI?

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

$$
\begin{array}{c|ccccc}
 & \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\
\hline
\text{CS} & & & & & \\
\text{MD} & & & & &
\end{array}
$$

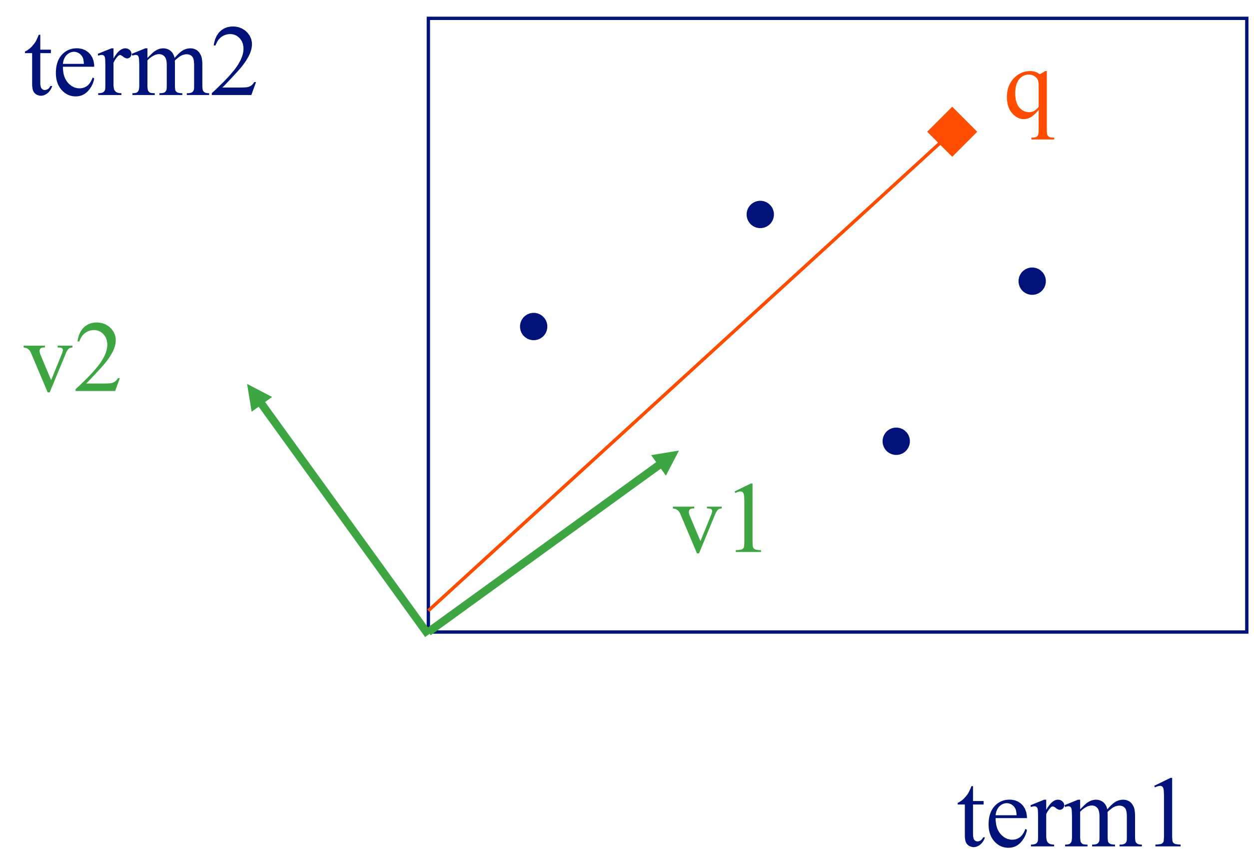
$$
\begin{array}{c}
\text{CS} \\ \\ \\ \\ \text{MD} \\ \\ \\
\end{array}
\begin{bmatrix}
1 & 1 & 1 & 0 & 0 \\
2 & 2 & 2 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 \\
5 & 5 & 5 & 0 & 0 \\
0 & 0 & 0 & 2 & 2 \\
0 & 0 & 0 & 3 & 3 \\
0 & 0 & 0 & 1 & 1
\end{bmatrix}
=
\begin{bmatrix}
0.18 & 0 \\
0.36 & 0 \\
0.18 & 0 \\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

Case study - LSI

Q1: How to do queries with LSI?

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

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

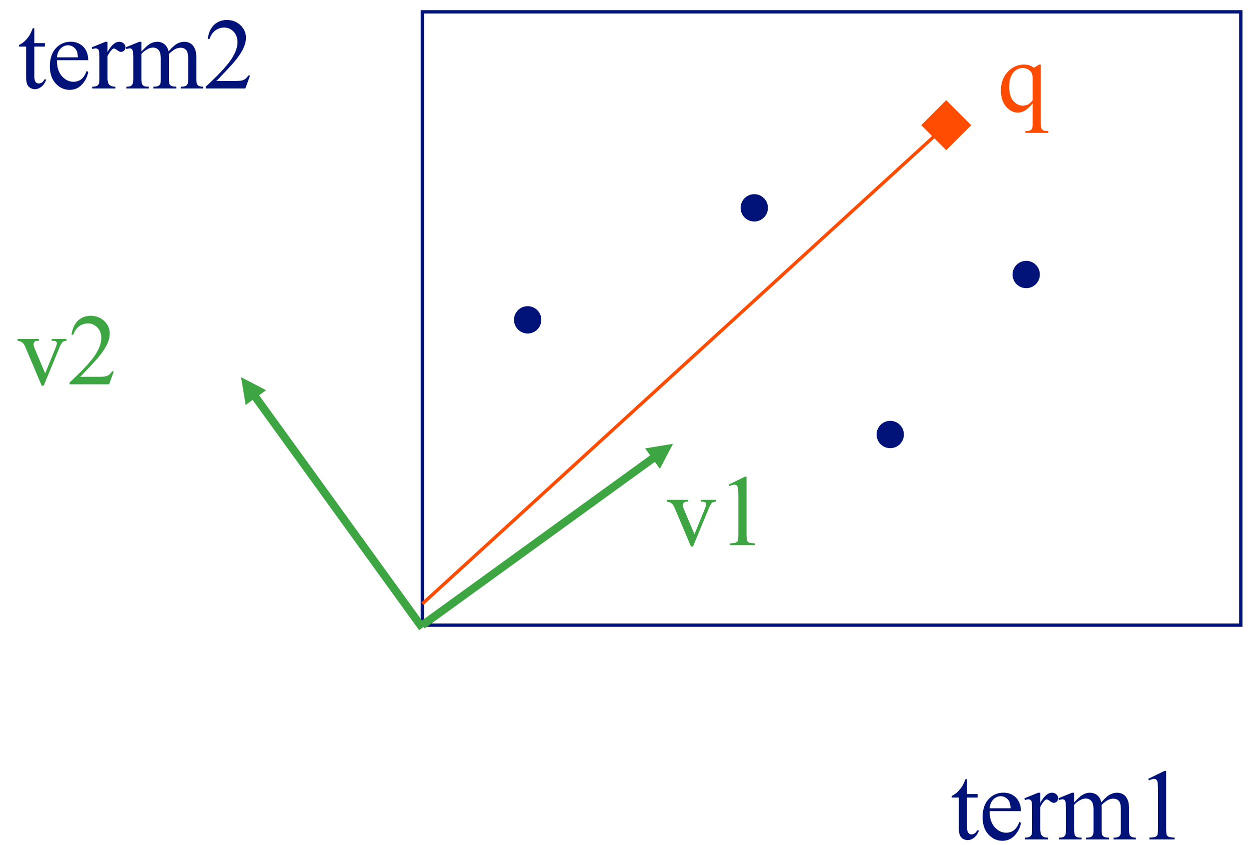


Case study - LSI

Q1: How to do queries with LSI?

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

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



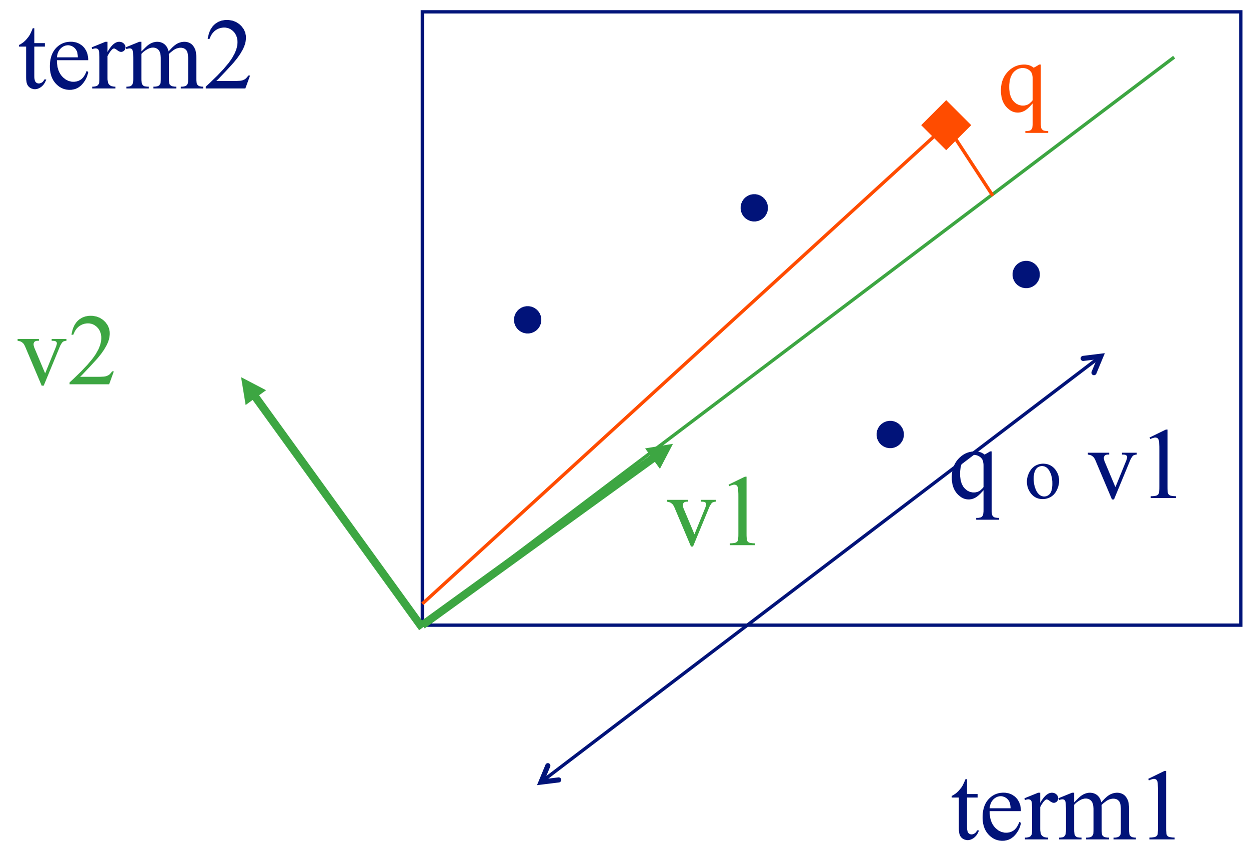
A: inner product
(cosine similarity)
with each ‘concept’ vector v_i

Case study - LSI

Q1: How to do queries with LSI?

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

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



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

Case study - LSI

compactly, we have:

$$q\,\mathbf{V} = q_{concept}$$

Eg:

$$q = \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \end{array}$$

$$q = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} = \begin{array}{c} \text{CS-concept} \\ \begin{bmatrix} 0.58 & 0 \end{bmatrix} \end{array}$$

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:

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

Eg:

$$d = \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \\ \end{array}$$
$$d = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} = \begin{bmatrix} 1.16 & 0 \end{bmatrix}$$

CS-concept

term-to-concept similarities

Case study - LSI

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

$$
d = \begin{array}{c} \begin{array}{ccccc} \text{data} & \text{inf.} & \text{retrieval} & \text{brain} & \text{lung} \end{array} \\ \left[\begin{array}{ccccc} 0 & 1 & 1 & 0 & 0 \end{array} \right] \end{array} \dashrightarrow \begin{array}{c} \begin{array}{cc} \text{CS-concept} & \end{array} \\ \left[\begin{array}{cc} 1.16 & 0 \end{array} \right] \end{array}
$$

$$
q = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \end{array} \right] \dashrightarrow \left[\begin{array}{cc} 0.58 & 0 \end{array} \right]
$$

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: ~ LSI

	data	inf.	retrieval	brain	lung	datos	informacion			
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
MD	0	0	0	2	2	0	0	0	2	2
	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**

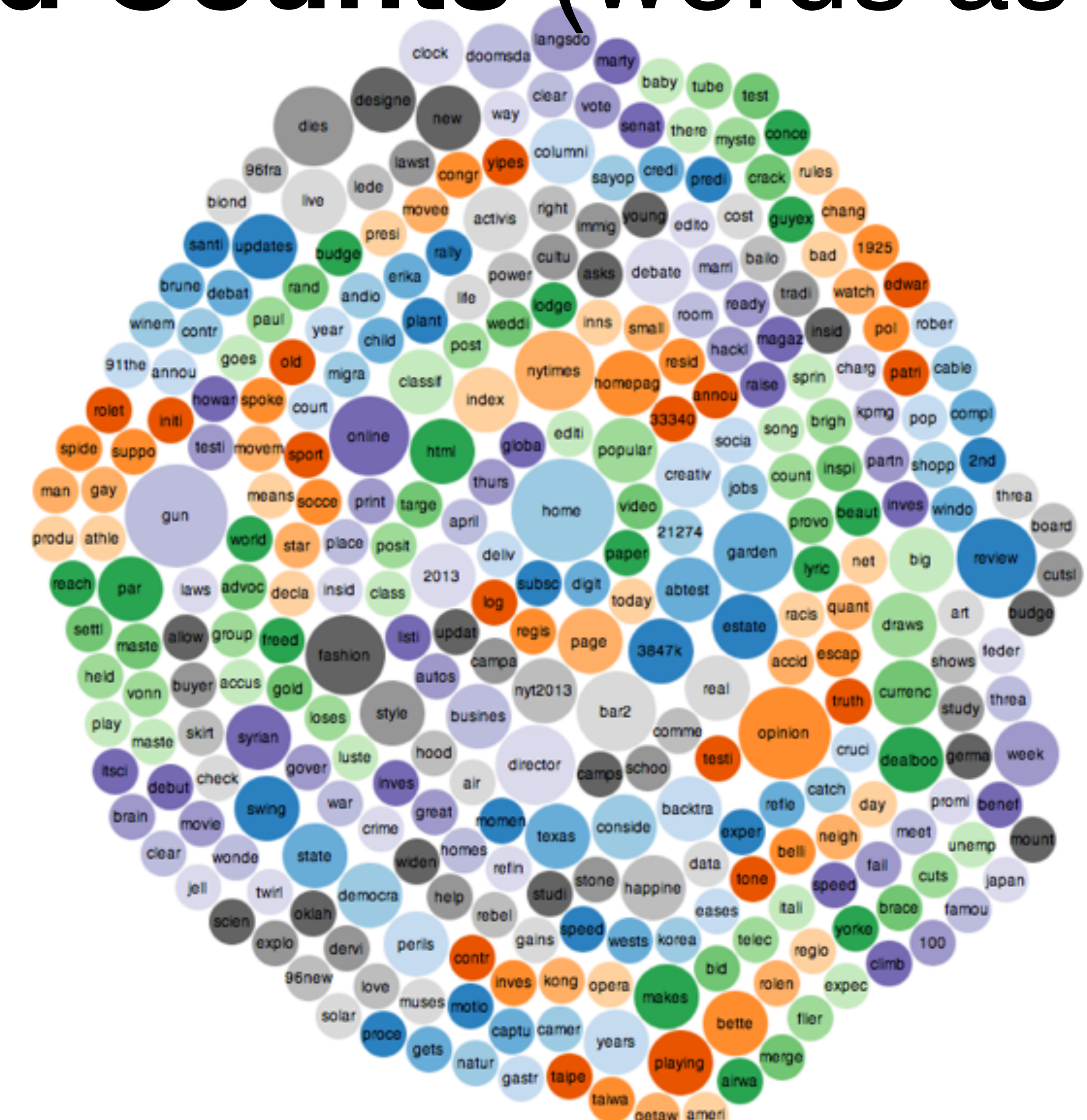
What comes up to your mind?

What visualization have you seen before?

Word/Tag Cloud (still popular?)

http://www.wordle.net

Word Counts (words as bubbles)



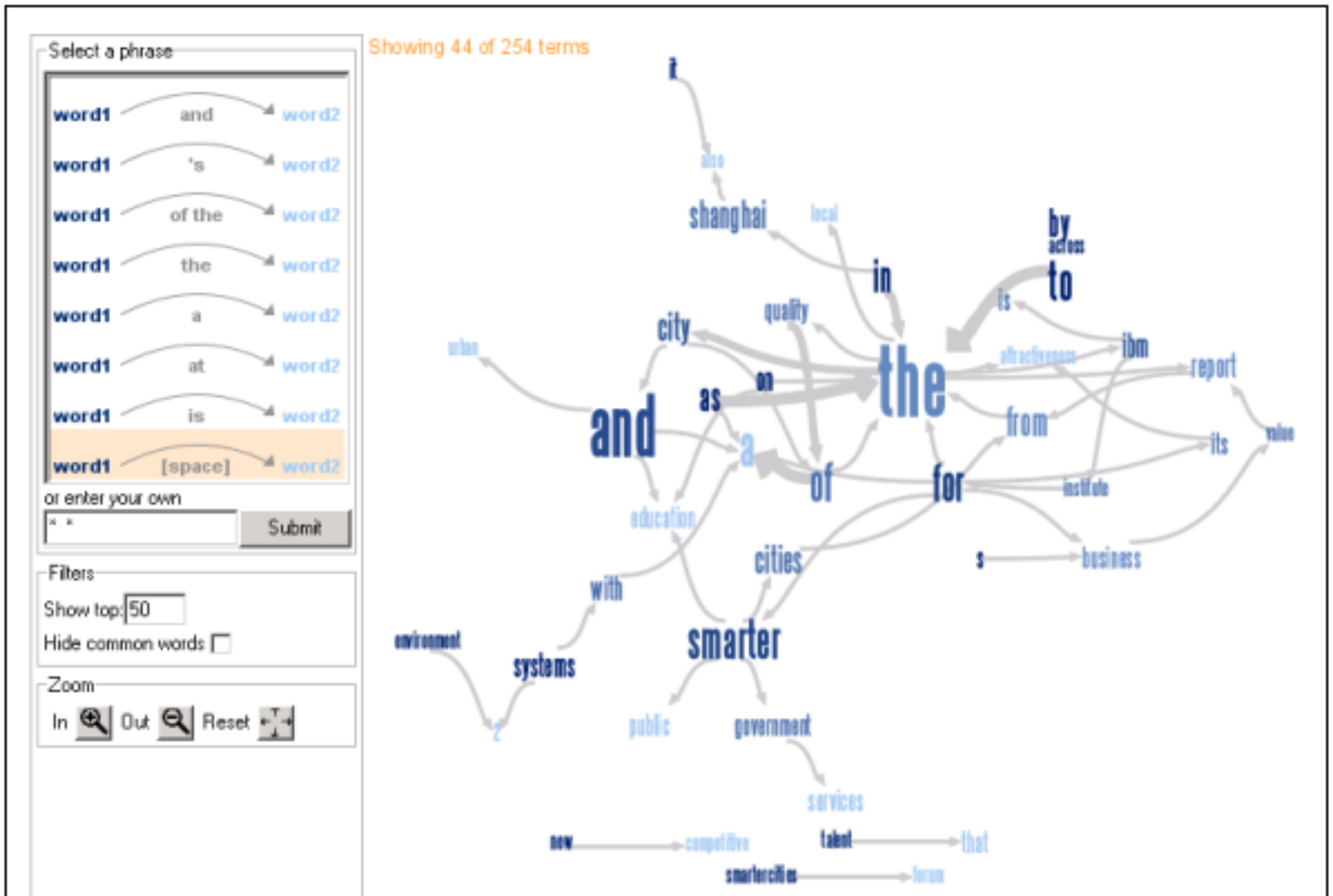
http://www.infocaptor.com/bubble-my-page

Word Tree

http://www.jasondavies.com/wordtree/

Phrase Net

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



http://www-958.ibm.com/software/data/cognos/manyeyes/page/Phrase_Net.html