ASSUMPTION

Docs are similar if they have ~~the~~ ~~SAME~~ words
$\underbrace{}_{\text{SIMILAR}}$

SIMILAR

Words are similar if they occur in ~~the~~ ~~SAME~~ docs.

SIMILAR

$\boxed{\text{CIRCULARITY}}$

Circularity.

- Factor analysis. — LSA
- Expectation Maximization — PLSA
- PageRank style approaches.
  - SIMRANK

PageRank

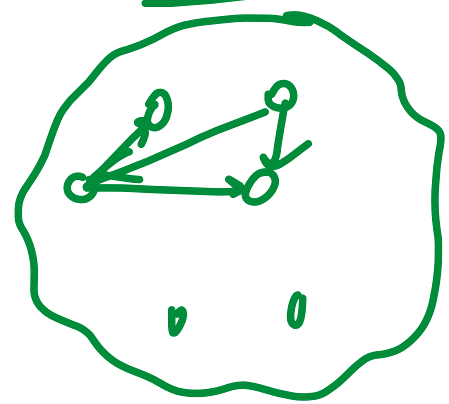
$$P(d|q) \propto \underbrace{p(q|d)}\,\boxed{p(d)}$$

PageRank.

A

A page is important if it is pointed to by several important pages.

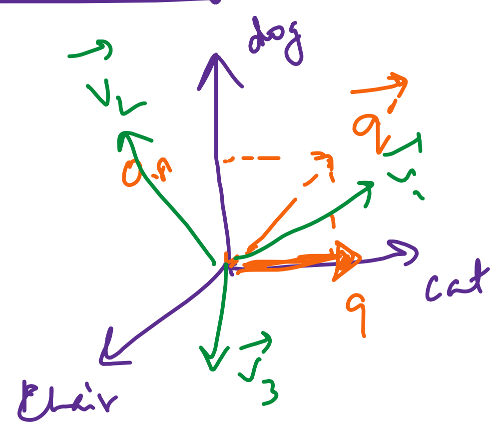
B



Latent Semantic Analysis.

COVARIANCE MATRIX

Factor Analysis.



words: cat, dog, chair

$$M \rightarrow \begin{array}{cc} & \begin{array}{ccc} \text{cat} & \text{dog} & \text{chair} \end{array} \\ \begin{array}{c} \text{cat} \\ \text{dog} \\ \text{chair} \end{array} & \begin{bmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Principal Component Analysis

query $\begin{array}{ccc} [1 & 0 & 0] \\ \text{cat} & \text{dog} & \text{chair} \end{array}$

Stretching

rotating.

$$M\vec{q} = \vec{q}'$$

operator

$$M\vec{d} = \vec{d}'$$

each concept is a linear combination of terms

$\boxed{\text{UNIMODE}}$ FACTOR ANALYSIS.

rotation & stretching

$\downarrow$

vector sum of stretches.

$M\vec{q}$

$\downarrow$

$\vec{v}_1, \vec{v}_2, \vec{v}_3$

<u>eigenvectors</u>

$$\left.\begin{aligned} M\vec{v}_1 &= \lambda_1\vec{v}_1 \\ M\vec{v}_2 &= \lambda_2\vec{v}_2 \\ M\vec{v}_3 &= \lambda_3\vec{v}_3 \end{aligned}\right\}$$

$\Bigg\{$ <u>dimensionality</u> <u>reduction.</u>

$$\vec{q} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3\vec{v}_3$$

$$\vec{q'} = M\vec{q} = \alpha_1 M\vec{v}_1 + \alpha_2 M\vec{v}_2 + \alpha_3 M\vec{v}_3 \quad \times$$

$$= \underset{1.00}{\alpha_1\lambda_1}\vec{v}_1 + \underset{0.8}{\alpha_2\lambda_2}\vec{v}_2 + \alpha_3\lambda_3\vec{v}_3 \; \cdots\cdots$$

$\underline{0.00003} \longrightarrow 0$

Latent dimensions.

$$
\begin{array}{ll|ccc}
 & & \multicolumn{3}{c}{\overbrace{\hspace{8em}}^{\text{taste}}} \\
 & & \text{sour} & \text{semi-sweet} & \text{sweet} \\
\text{Colour} & \text{green} & 0.8 & & 0.1 \\
 & \text{yellow} & & 0.9 & \\
 & \text{red} & 0.1 & & 0.8
\end{array}
$$

$$
\begin{array}{l|c}
 & \overbrace{\hspace{8em}}^{\text{words}} \\
\text{docs} \downarrow & \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}
\end{array}
$$

unripe ripe

$$
\begin{array}{c|ccc}
 & \text{ur} & \text{sr} & \text{r.} \\
\text{g} & & & \\
\text{y} & & & \\
\text{r} & & &
\end{array}
\qquad
\begin{array}{c|ccc}
 & \text{sour} & \text{semi-s} & \text{sweet} \\
\text{ur} & & & \\
\text{sr} & & & \\
\text{r} & & &
\end{array}
$$

colour × ripeness

ripeness × taste

concept

term

$m \times n$

doc $\begin{bmatrix} \quad \end{bmatrix} \sim \begin{bmatrix} \quad \end{bmatrix}$ $m \times n$

Param

$k\overbrace{(m+n)}$

Synopsis $\begin{bmatrix} \quad \end{bmatrix}$

$\underbrace{m \times k + k \times n}$

k

nodes.

Latent Semantic Indexing

Search using LSI - an example

Retrieve documents related to 'terrorism'

Level of exactness required - **Medium**

LSI Engine

...an act of terrorism may have caused a russian passenger plane crash during a flight from siberia ...80%

...the bodies of four international journalists have been found and were identified by colleagues
tuesday a day after their convoy was ambushed in a narrow mountain pass on the road to the
afghan capital kabul... 15.71%

Does not contain the word "terrorism"

...in the absence of an al-qaida heir attention is already turning to the obvious candidate for demonization saddam hussein hawks are rallying for a war against the ultimate terrorist ... 8.44%

(TRDDC)

Latent Semantic Analysis

Starting point : a term document matrix

Document #	filtering	indexing	clustering	decompose	sediments	purification	oscillation	matrix	differential
1	1	1	1	0	0	0	0	1	0
2	1	0	1	0	0	0	0	0	0
3	1	1	1	0	0	0	0	0	0
4	0	0	0	1	1	1	0	0	0
5	0	0	0	1	0	1	0	0	0
6	1	0	0	1	1	1	0	0	0
7	0	0	0	0	0	0	1	1	1
8	0	0	0	0	0	0	1	0	1
9	0	0	0	1	0	0	1	1	1

Objective : Determine the best set of underlying factors or concepts, that best explain the relationship between terms and documents

Assumption : There is some underlying or latent structure in word usage that is partially obscured by variability in word choice

Procedure : Singular Value Decomposition

Key Idea : Factor Analysis

A bit of Linear Algebra

- In Linear Algebra, a matrix like the term-document matrix can be viewed as an Operator

  - The matrix can act upon a vector (when it is multiplied with that vector), and relocate it to a different position.

$$M = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}$$

can act on the vector

$$\vec{A} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

and move it to a new location given by $M\ \vec{A}$:

$$M\ \vec{A} = \begin{pmatrix} 3 \\ 0 \\ 5 \end{pmatrix}$$

The action of a matrix can be viewed as a combination of translation and rotation of vector

Goal: charactering a matrix formally in terms of its properties that govern its action on vectors

The concept of eigenvectors does precisely that

Eigenvectors

We consider all vectors $\vec{x}$ that, when acted on by M, stretch themselves to a different location $\lambda \vec{x}$, where λ is a scalar, but do not undergo any rotation. Thus

$$M \vec{x} = \lambda \vec{x} \qquad (1)$$

The vectors satisfying (1) are called eigenvectors, and each of these eigenvectors is associated with a corresponding value of λ referred to as an eigenvalue.

$(M - \lambda I) \vec{x} = 0$, where I is an identity matrix of dimensions matching M; this is called the characteristic equation.

in our example, we have the following three eigenvectors

$$v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{and } v_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

associated with the eigenvalues $\lambda_1 = 1$, $\lambda_2 = 1$ and $\lambda_3 = 4$ respectively.

Why are eigenvectors important?

We now study the effect of M on any arbitrary vector $\vec{x}$

$$\vec{x} = \begin{pmatrix} 4 \\ -1 \\ 4 \end{pmatrix}$$

We can express $\vec{x}$ as a linear combination of v_1, v_2 and v_3.

The revised position $M\,\vec{x}$ is now given by

$$\begin{aligned} M\,\vec{x} &= M\,(1v_1 + 2v_2 + 3v_3) \\ &= Mv_1 + 2Mv_2 + 3Mv_3 \\ &= \lambda_1 v_1 + 2\lambda_2 v_2 + 3\lambda_3 v_3 \end{aligned}$$

The interesting aspect of this rewrite is that we can see that the total effect of M on $\vec{x}$ is expressed as a weighted combination of effects due to each eigenvector.

Eigenvectors having very small eigenvalues associated with them have a small effect on the operation of M on $\vec{x}$.

$\boxed{\text{Critical Intuition behind SVD/LSI}}$

Matrix Diagonalization Theorem

For a given square real valued $m \times m$ matrix M with linearly independent eigenvectors, we can obtain a factorization

$$M = U \Lambda U^{-1}$$

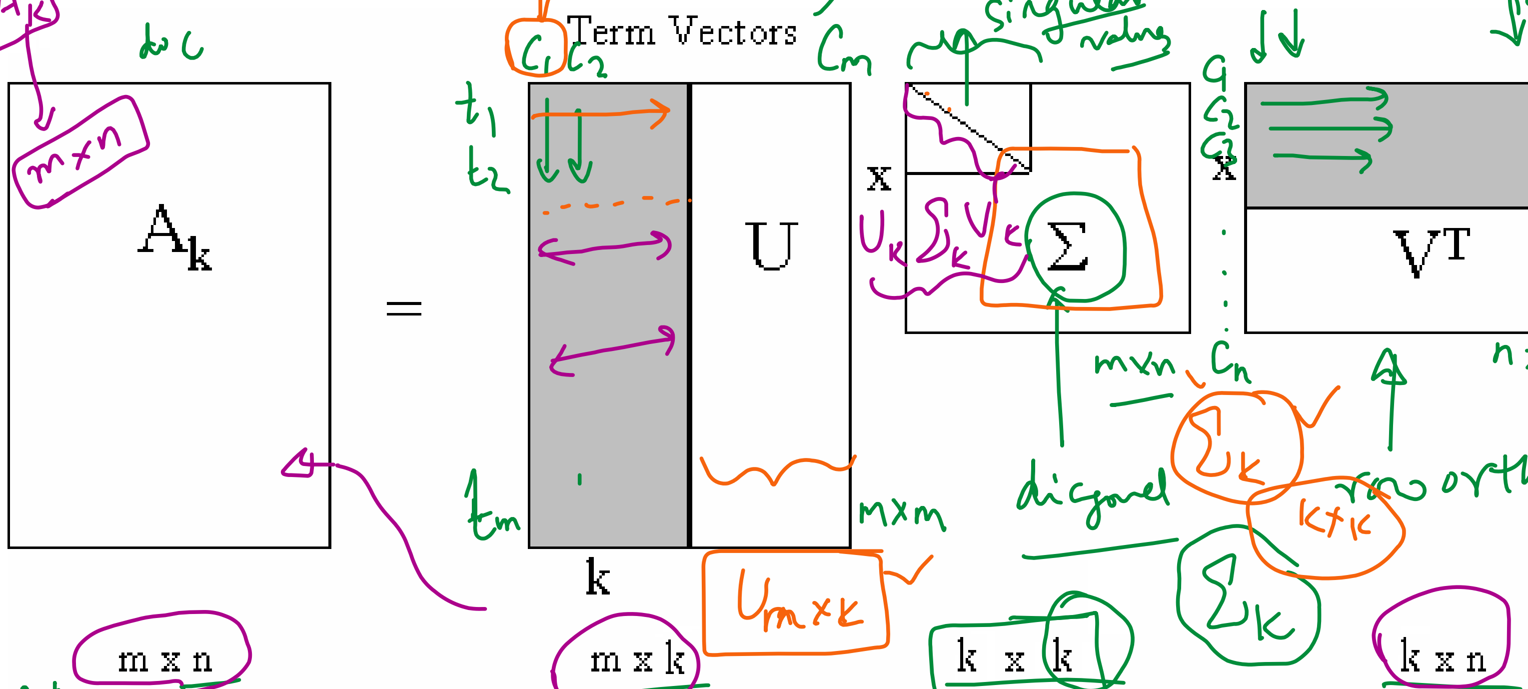
such that the columns of U are the eigenvectors of M, and Λ is a diagonal matrix whose diagonal elements are eigenvalues of M arranged in decreasing order. This result is due to the Matrix Diagonalization Theorem. This result applies to square matrices, but not to rectangular ones like the term-document matrix.

$A_k \approx U_{m \times k} \times \Sigma_{k \times k} \times V^T_{k \times n}$

Singular Value Decomposition

From unimode to two-mode factor analysis

$A \to A_k$



Column orthogonal matrix

U & V are column orthogonal

singular value

diagonal

m : no. of terms

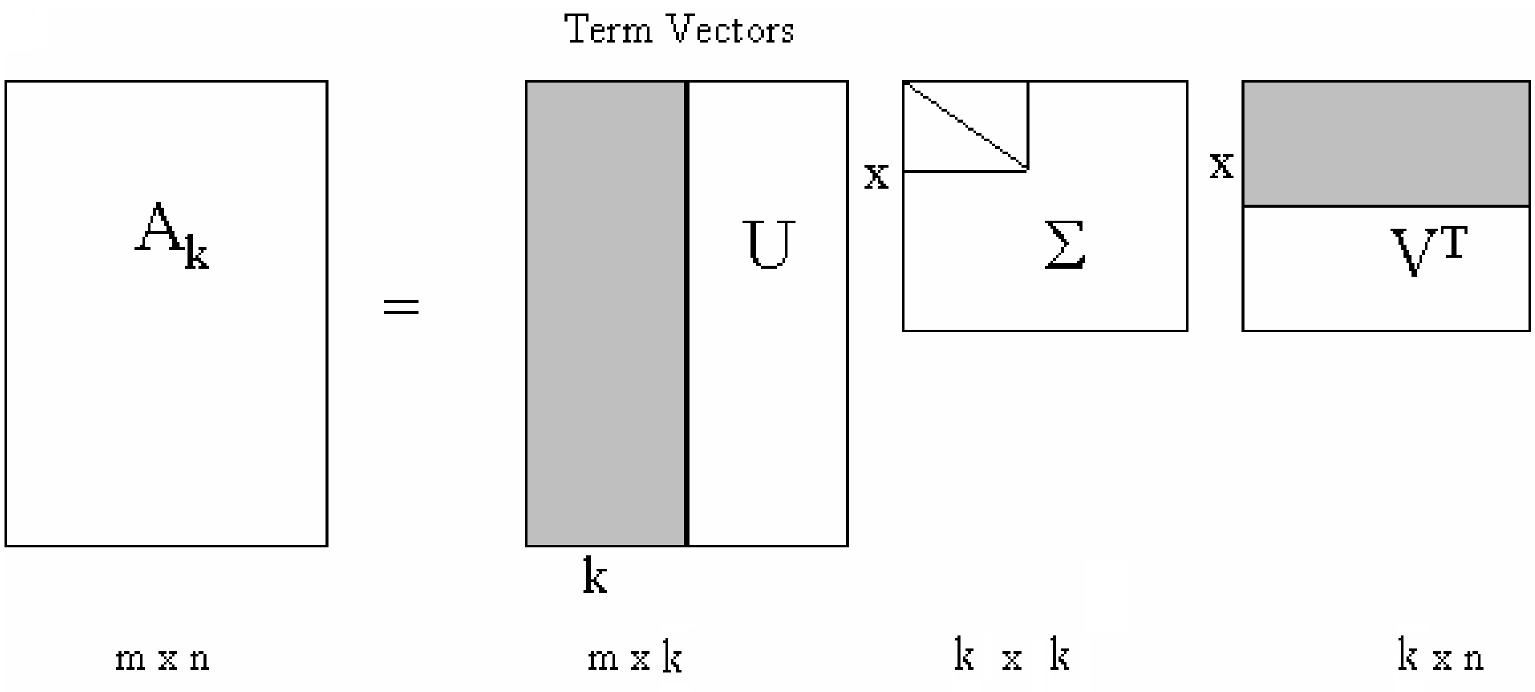
n : no. of docs.

A_k is the reduced rank approximation of the original matrix A

Adapted from [Deerwester et al]

Singular Value Decomposition

From unimode to two-mode factor analysis



A_k is the reduced rank approximation of the original matrix A

Adapted from [Deerwester et al]

Connection of SVD to Matrix Diagonalization

$$\begin{array}{c|cccc} & d_1 & d_2 & d_3 & d_4 \\ \hline t_1 & & & & \\ t_2 & & A & & \\ t_3 & & & & \end{array}$$

3×4

$A^T A$ $\quad$ $A A^T$

$d \times d$ $\quad$ $t \times t$

eigenvalues.

U $\quad$ V

Singular value of A $= \sqrt{\text{eigenvalues of } AA^T \text{ or } A^TA}$

Example

k=3

IR Chemistry Maths

Document #	filtering	indexing	clustering	decompose	sediments	purification	oscillation	matrix	differential
1	1	1	1	0	0	0	0	1	0
2	1	0	1	0	0	0	0	0	0
3	1	1	1	0	0	0	0	0	0
4	0	0	0	1	1	1	0	0	0
5	0	0	0	1	0	1	0	0	0
6	1	0	0	1	1	1	0	0	0
7	0	0	0	0	0	0	1	1	1
8	0	0	0	0	0	0	1	0	1
9	0	0	0	1	0	0	1	1	1

Document #	filtering	indexing	clustering	decompose	sediments	purification	oscillation	matrix	differential
1	1.17	0.83	1.11	-0.00	-0.00	-0.06	0.16	0.59	0.17
2	0.89	0.53	0.72	0.00	0.06	0.04	-0.1	0.22	-0.1
3	1.07	0.73	0.98	-0.05	0.04	-0.01	-0.11	0.32	-0.11
4	0.24	-0.12	-0.14	1.09	0.74	1.01	0.00	-0.04	0.00
5	0.16	-0.1	-0.11	0.82	0.54	0.74	0.04	-0.00	0.04
6	0.67	0.15	0.23	1.16	0.82	1.09	-0.06	0.06	-0.06
7	-0.02	0.07	0.05	0.23	-0.12	-0.15	1.02	0.84	1.017
8	-0.13	-0.03	-0.07	0.18	-0.08	-0.09	0.74	0.56	0.74
9	0.04	0.02	-0.02	0.67	0.15	0.23	1.11	0.89	1.11

Example 2

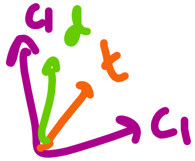
Technical Memo Example

Titles

c1:	*Human* machine *interface* for Lab ABC *computer* applications
c2:	A *survey of user* opinion of *computer system response time*
c3:	The *EPS user interface* management *system*
c4:	*System* and *human system* engineering testing of *EPS*
c5:	Relation of *user*-perceived *response time* to error measurement
m1:	The generation of random, binary, unordered *trees*
m2:	The intersection *graph* of paths in *trees*
m3:	*Graph minors* IV: Widths of *trees* and well-quasi-ordering
m4:	*Graph minors: A survey*

Terms	Documents								
	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Example 2



docs.

terms

$$A = \begin{bmatrix}
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}$$

txt

c1 c2

$$U = \begin{bmatrix}
0.22 & -0.11 & 0.29 & -0.41 & -0.11 & -0.34 & 0.52 & -0.06 & -0.41 \\
0.20 & -0.07 & 0.14 & -0.55 & 0.28 & 0.50 & -0.07 & -0.01 & -0.11 \\
0.24 & 0.04 & -0.16 & -0.59 & -0.11 & -0.25 & -0.30 & 0.06 & 0.49 \\
0.40 & 0.06 & -0.34 & 0.10 & 0.33 & 0.38 & 0.00 & 0.00 & 0.01 \\
0.64 & -0.17 & 0.36 & 0.33 & -0.16 & -0.21 & -0.17 & 0.03 & 0.27 \\
0.27 & 0.11 & -0.43 & 0.07 & 0.08 & -0.17 & 0.28 & -0.02 & -0.05 \\
0.27 & 0.11 & -0.43 & 0.07 & 0.08 & -0.17 & 0.28 & -0.02 & -0.05 \\
0.30 & -0.14 & 0.33 & 0.19 & 0.11 & 0.27 & 0.03 & -0.02 & -0.17 \\
0.21 & 0.27 & -0.18 & -0.03 & -0.54 & 0.08 & -0.47 & -0.04 & -0.58 \\
0.01 & 0.49 & 0.23 & 0.03 & 0.59 & -0.39 & -0.29 & 0.25 & -0.23 \\
0.04 & 0.62 & 0.22 & 0.00 & -0.07 & 0.11 & 0.16 & -0.68 & 0.23 \\
0.03 & 0.45 & 0.14 & -0.01 & -0.30 & 0.28 & 0.34 & 0.68 & 0.18
\end{bmatrix}$$

$$\Sigma = \begin{bmatrix}
3.34 & & & & & & & & \\
& 2.54 & & & & & & & \\
& & 2.35 & & & & & & \\
& & & 1.64 & & & & & \\
& & & & 1.50 & & & & \\
& & & & & 1.31 & & & \\
& & & & & & 0.85 & & \\
& & & & & & & 0.56 & \\
& & & & & & & & 0.36
\end{bmatrix}$$

c1 c2

dxt

$$V = \begin{array}{c} d_1 \\ d_2 \\ \\ \\ \\ \\ \\ \\ \\ \end{array}
\begin{bmatrix}
0.20 & -0.06 & 0.11 & -0.95 & 0.05 & -0.08 & 0.18 & -0.01 & -0.06 \\
0.61 & 0.17 & -0.50 & -0.03 & -0.21 & -0.26 & -0.43 & 0.05 & 0.24 \\
0.46 & -0.03 & 0.21 & 0.04 & 0.38 & 0.72 & -0.24 & 0.01 & 0.02 \\
0.54 & -0.23 & 0.57 & 0.27 & -0.21 & -0.37 & 0.26 & -0.02 & -0.08 \\
0.28 & 0.11 & -0.51 & 0.15 & 0.33 & 0.03 & 0.67 & -0.06 & -0.26 \\
0.00 & 0.19 & 0.10 & 0.02 & 0.39 & -0.30 & -0.34 & 0.45 & -0.62 \\
0.01 & 0.44 & 0.19 & 0.02 & 0.35 & -0.21 & -0.15 & -0.76 & 0.02 \\
0.02 & 0.62 & 0.25 & 0.01 & 0.15 & 0.00 & 0.25 & 0.45 & 0.5[illegible] \\
0.08 & 0.53 & 0.08 & -0.03 & -0.60 & 0.36 & -0.04 & -0.07 & -0.45
\end{bmatrix}$$

Example 2

U_2

0.22	−0.11
0.20	−0.07
0.24	0.04
0.40	0.06
0.64	−0.17
0.27	0.11
0.27	0.11
0.30	−0.14
0.21	0.27
0.01	0.49
0.04	0.62
0.03	0.45

Σ_2

3.34	
	2.54

V_2'

0.20	0.61	0.46	0.54	0.28	0.00	0.02	0.02	0.08
−0.06	0.17	−0.13	−0.23	0.11	0.19	0.44	0.62	0.53

$A_2 = U_2 \Sigma_2 V_2' =$

0.16	0.40	0.38	0.47	0.18	−0.05	−0.12	−0.16	−0.09
0.14	0.37	0.33	0.40	0.16	−0.03	−0.07	−0.10	−0.04
0.15	0.51	0.36	0.41	0.24	0.02	0.06	0.09	0.12
0.26	0.84	0.61	0.70	0.39	0.03	0.08	0.12	0.19
0.45	1.23	1.05	1.27	0.56	−0.07	−0.15	−0.21	−0.05
0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
0.16	0.58	0.38	0.42	0.28	0.06	0.13	0.19	0.22
0.22	0.55	0.51	0.63	0.24	−0.07	−0.14	−0.20	−0.11
0.10	0.53	0.23	0.21	0.27	0.14	0.31	0.44	0.42
−0.06	0.23	−0.14	−0.27	0.14	0.24	0.55	0.77	0.66
−0.06	0.34	−0.15	−0.30	0.20	0.31	0.69	0.98	0.85
−0.04	0.25	−0.10	−0.21	0.15	0.22	0.50	0.71	0.62

A =

1	0	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
0	1	1	0	1	0	0	0	0
0	1	1	2	0	0	0	0	0
0	1	0	0	1	0	0	0	0
0	1	0	0	1	0	0	0	0
0	0	1	1	0	0	0	0	0
0	1	0	0	0	0	0	0	1
0	0	0	0	0	1	1	1	0
0	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	1	1

The “dual mode” advantage

Things we can do

A_k

$t \times d$

- Compute document similarities

$A_k A_k'$

- Compute term similarities

$A_k' A_k$

- Compute relevance of a term to a document

Term-term similarities

	filtering	indexing	clustering	decompose	sediments	purification	oscillation	matrix	differential
filtering	4	2	3	1	1	1	0	1	0
indexing	2	2	2	0	0	0	0	1	0
clustering	3	2	3	0	0	0	0	1	0
decompose	1	0	0	4	2	3	1	1	1
sediments	1	0	0	2	2	2	0	0	0
purification	1	0	0	3	2	3	0	0	0
oscillation	0	0	0	1	0	0	3	2	3
matrix	1	1	1	1	0	0	2	3	2
differential	0	0	0	1	0	0	3	2	3

	filtering	indexing	clustering	decompose	sediments	purification	oscillation	matrix	differential
filtering	3.86	2.28	3.10	1.12	0.93	1.07	-0.13	1.21	-0.11
indexing	2.28	1.56	2.09	-0.05	0.04	-0.07	0.06	0.91	0.07
clustering	3.10	2.09	2.80	-0.04	0.10	-0.03	-0.04	1.13	-0.03
decompose	1.12	-0.05	-0.04	3.74	2.26	3.08	1.08	0.90	1.08
sediments	0.93	0.04	0.10	2.26	1.56	2.10	-0.05	0.03	-0.05
purification	1.07	-0.07	-0.03	3.08	2.10	2.84	-0.01	0.02	-0.01
oscillation	-0.13	0.06	-0.04	1.08	-0.05	-0.01	2.87	2.29	2.87
matrix	1.21	0.91	1.13	0.90	0.03	0.02	2.29	2.32	2.30
differential	-0.11	0.07	-0.03	1.08	-0.05	-0.01	2.87	2.30	2.87

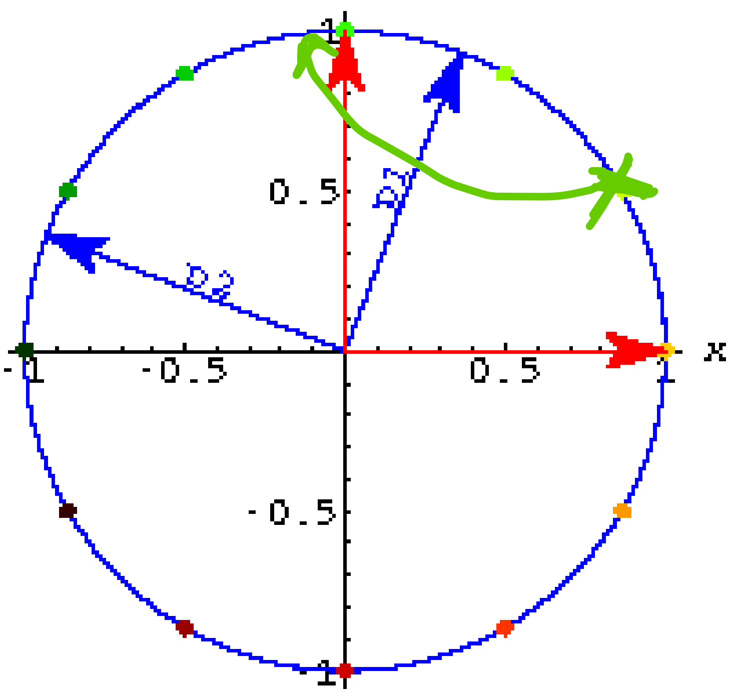
Refer this link for Geometry of SVD

http://websites.uwlax.edu/twill/svd/perpframes/index.html

$\boxed{A\vec{x}}$

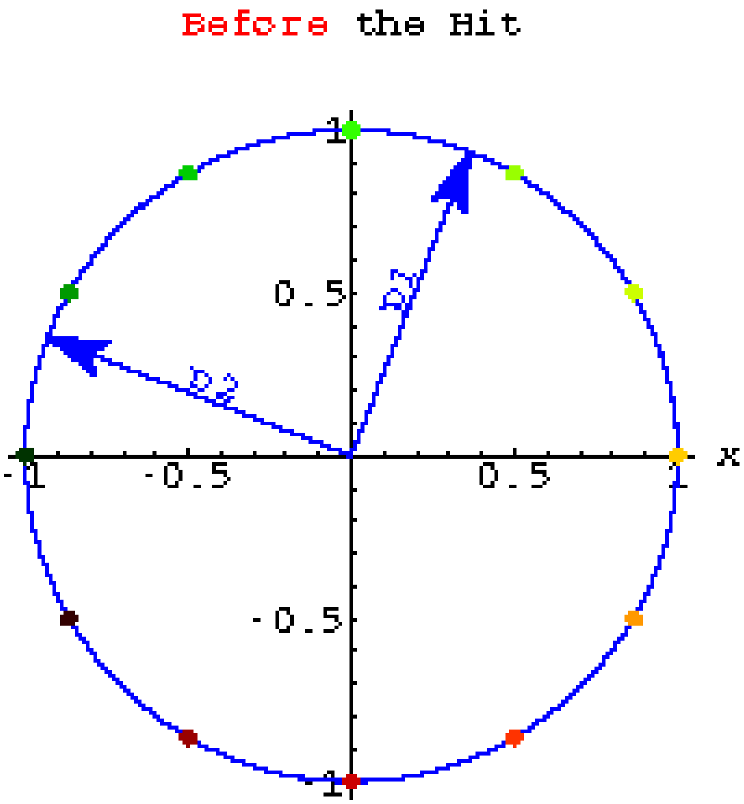
Hanger (Column Orthogonal)

Before the Hit



$$\begin{array}{cc} \vec{p}_1 & \vec{p}_2 \\ \multicolumn{2}{c}{\underbrace{\left(\begin{array}{c|c} 0.36 & -0.93 \\ 0.93 & 0.36 \end{array}\right)}_{U}} \end{array}$$

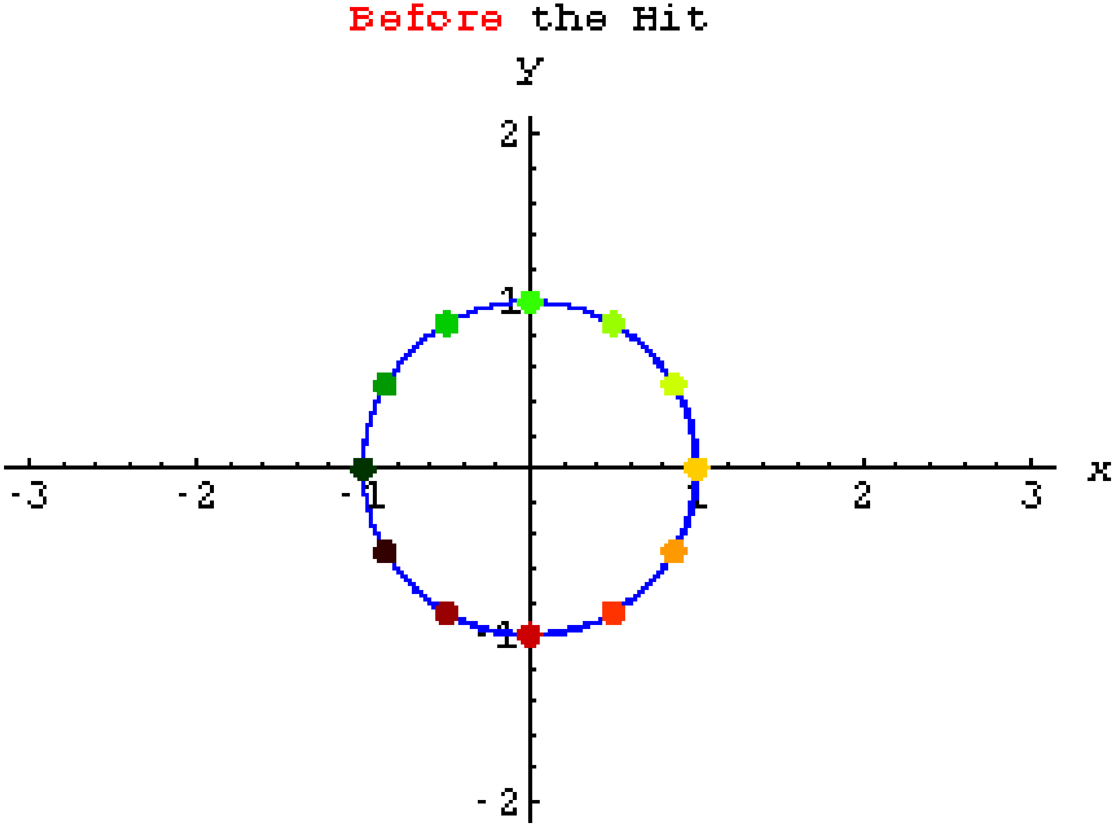
Aligner



$$\begin{pmatrix} 0.36 & 0.93 \\ -0.93 & 0.36 \end{pmatrix} \quad \vec{x}$$

V^T

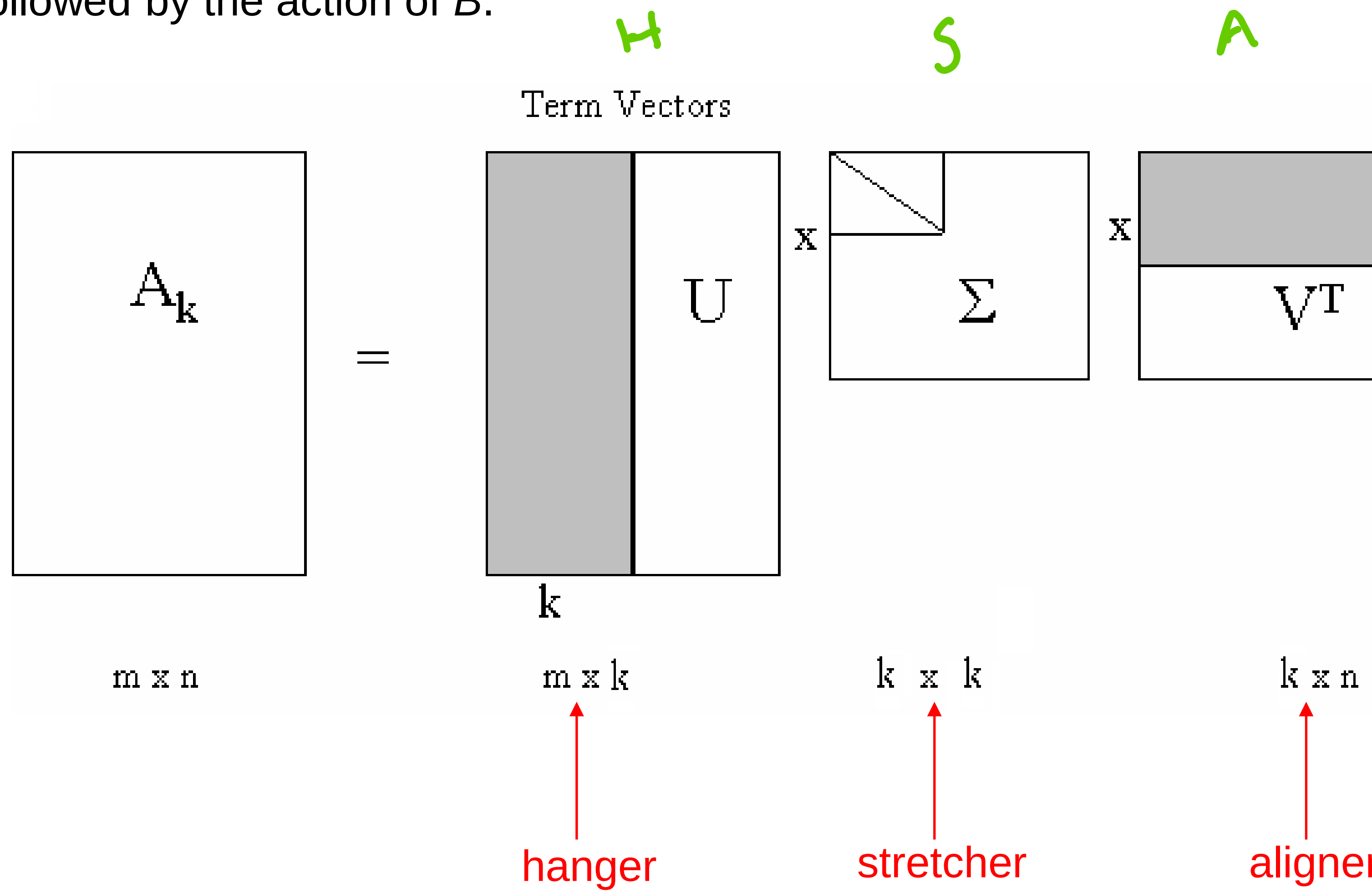
Stretcher

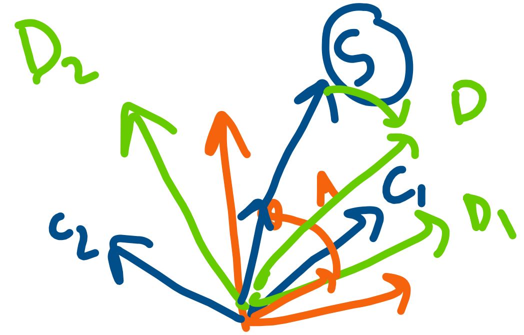


$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

SVD = H$_{\text{anger}}$S$_{\text{tretcher}}$A$_{\text{ligner}}$

Key point: The action of a matrix product: *A=B C* is the same as the action of the *C* followed by the action of *B*.





TERM
CONCEPT

$$[\quad][\quad] \to [\quad]$$

$f \times t$

$\cancel{d}\ t$

$$d\ \cancel{t}\ \left[\ \cancel{T}\ \right]\left[\ q\ \right] = H\ S\ A\ q$$

$t \times d$ $\quad$ $d \times 1$

$\underline{V^T}$

$\boxed{d \times t \quad t \times 1}$

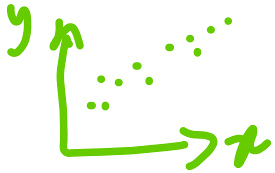
LSA

- Idea (Deerwester et al):

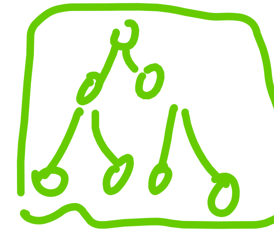
"We would like a representation in which a set of terms, which by itself is incomplete and unreliable evidence of the relevance of a given document, is replaced by some other set of entities which are more reliable indicants. We take advantage of the implicit higher-order (or latent) structure in the association of terms and documents to reveal such relationships."

ML ↔ Info Theory

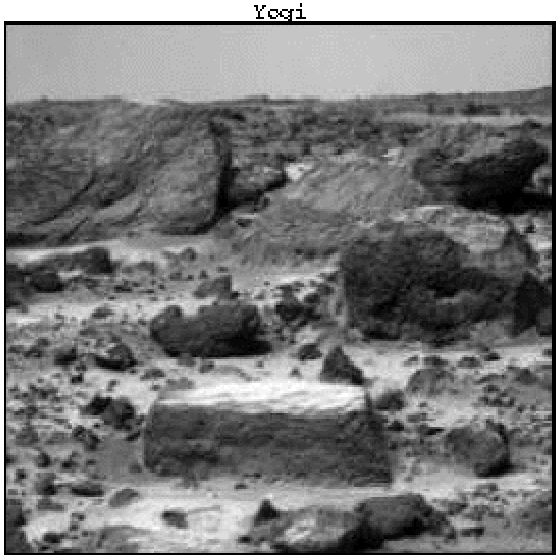
SVD for compression

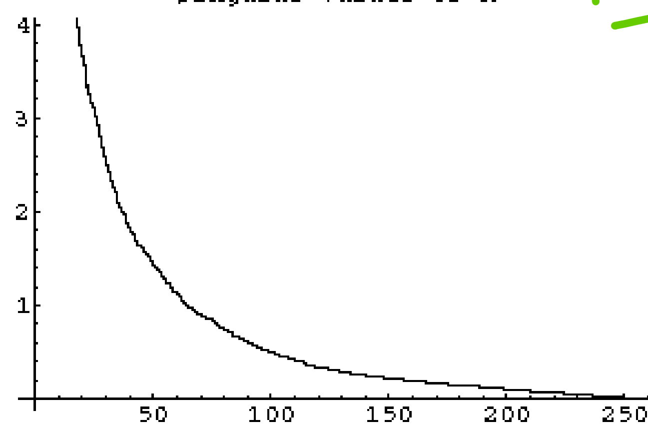


$y = \underline{a + bx}$



$\boxed{\text{McKay}}$





To send the matrix *M* you need to send 256 x 264 = 67584 numbers.

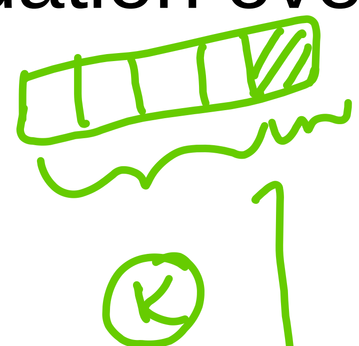
To send the rank 36 approximation to *M* you need only send
the first 36 singular values,
the first 36 hanger vectors, each of which has 256 entries,
the first 36 aligner vectors, each of which has 264 entries.
So in total you need to send only 36(1+256+264)=18756 numbers

http://websites.uwlax.edu/twill/svd/compression/index.html

The Right Choice of Dimensions?

- Reduce dimensionality to:
  - Fewer dimensions, more “collapsing of axes”, better recall, worse precision
  - More dimensions, less collapsing, worse recall, better precision

- Cross validation over training data?

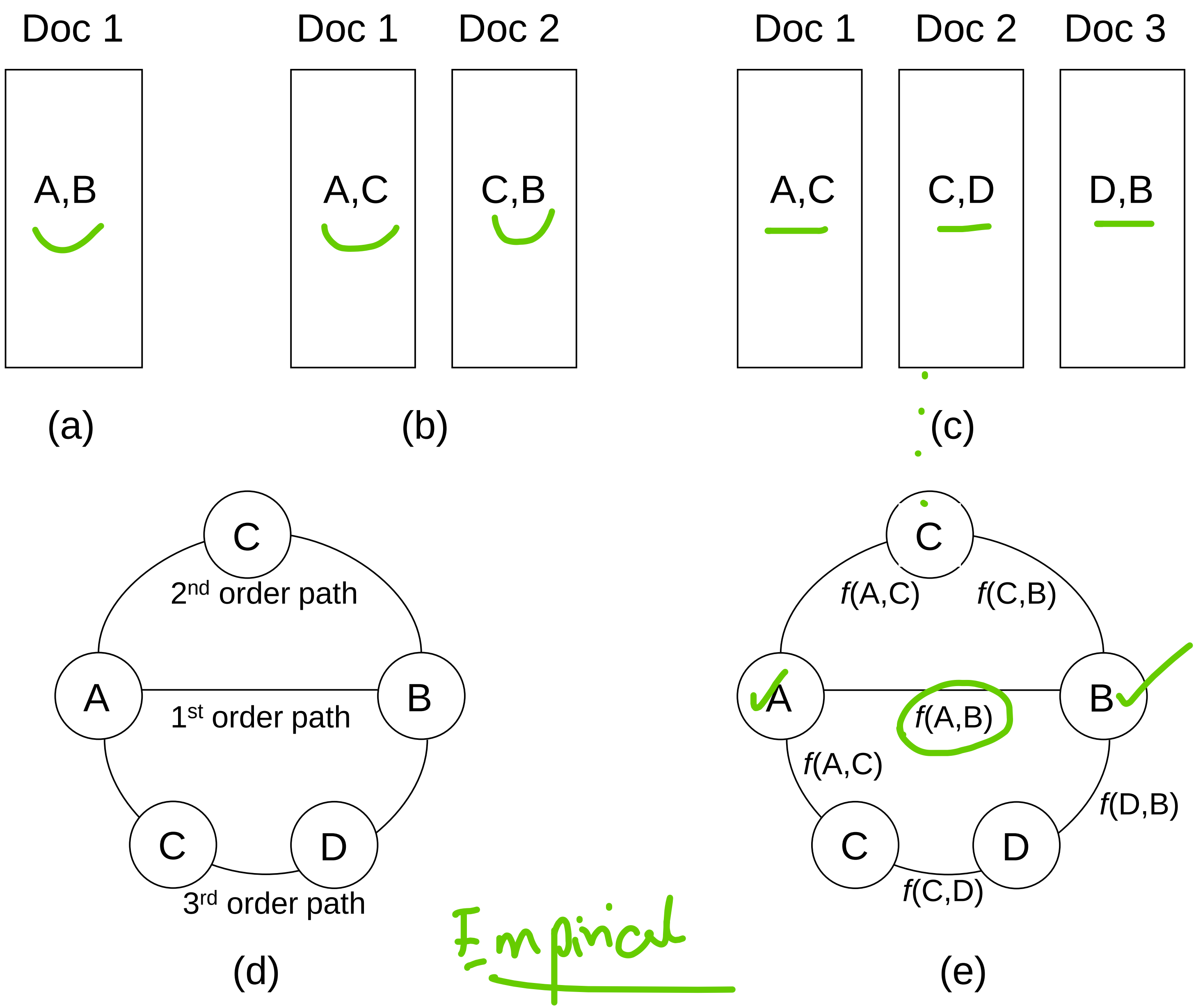


Mapping Queries to New Space

$$\hat{q} = \underbrace{q' U_k \Sigma_k^{-1}}$$

[Is LSA better for synonymy
or polysemy?]

LSA and Orders of Co-Occurrence

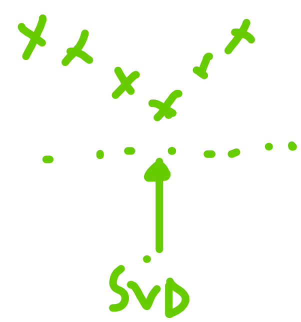


SVD : an important property

It can be shown that $\hat{M}$ is the best k-rank approximation to M in the least-squares sense. The quality of an approximation M_A is measured by the Frobenius Norm of the "discrepancy" matrix $X = M - M_A$, which is given by:

$$\| X \|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^{\;2}}.$$

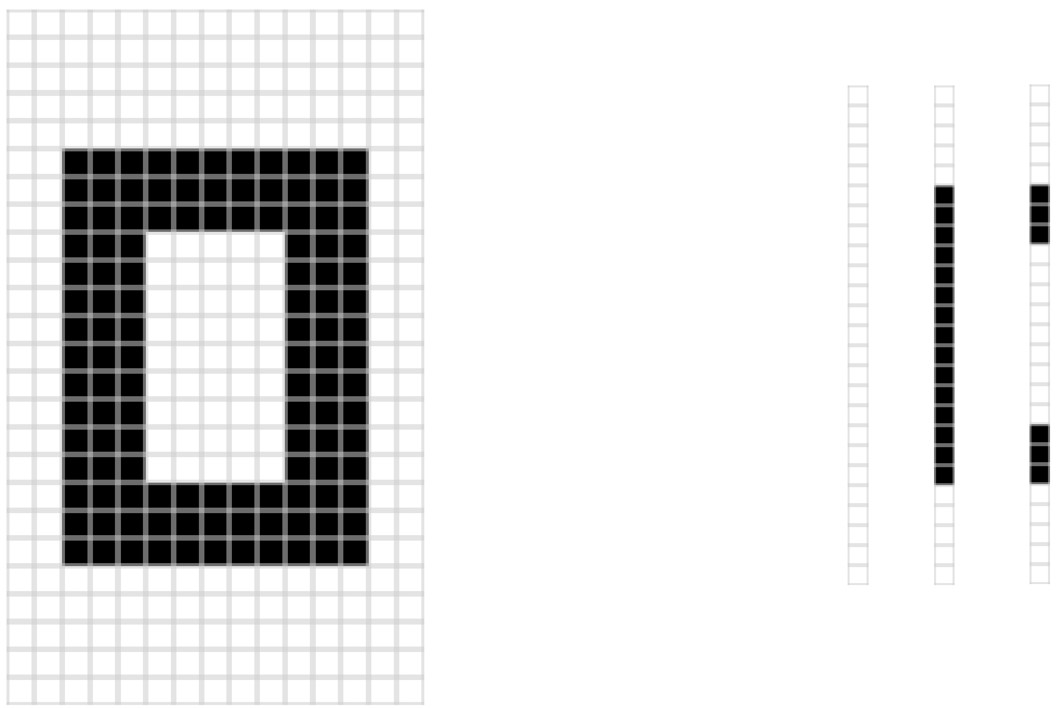
A A_k



SVD and rank reduction

Intuition : Distort a matrix to generate an approximation that has lower rank than the original

Refer: http://www.ams.org/publicoutreach/feature-column/fcarc-svd



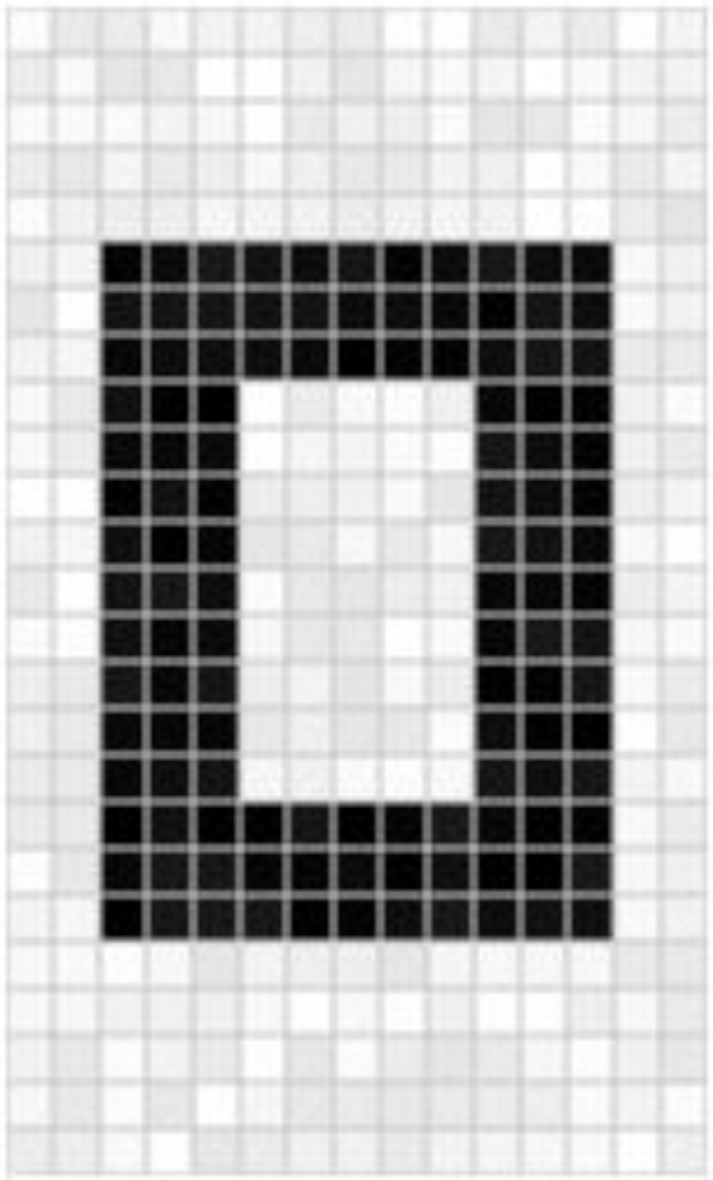
$$
M = \begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&0&0&0&0&0&0&0&0&1&1\\
1&1&0&0&0&1&1&1&1&1&0&0&0&1&1\\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

$\sigma_1 = 14.72$

$\sigma_2 = 5.22$

$\sigma_3 = 3.31$

$M = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T + u_3\sigma_3 v_3^T$



$$\sigma_1 = 14.15$$

$$\sigma_2 = 4.67$$

$$\sigma_3 = 3.00$$

$$\sigma_4 = 0.21$$

$$\sigma_5 = 0.19$$

$$\ldots$$

$$\sigma_{15} = 0.05$$

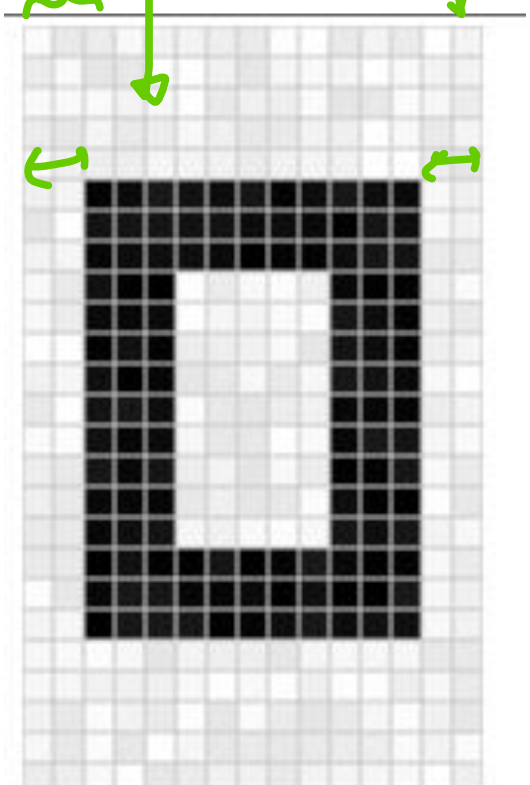
$$M \approx \mathbf{u}_1\sigma_1 \mathbf{v}_1^T + \mathbf{u}_2\sigma_2 \mathbf{v}_2^T + \mathbf{u}_3\sigma_3 \mathbf{v}_3^T$$

$t_1 = 0.9\, c_1 + 0.05\, c_2 + 0.35\, c_3$

$\vdots$

$t_1 \; t_2 \qquad t_M$

d_1



$d_N \qquad c_2 \quad c_3 \quad c_2$

$\sigma_1 = 14.15$

$\sigma_2 = 4.67$

$\sigma_3 = 3.00$

$\sigma_4 = 0.21$

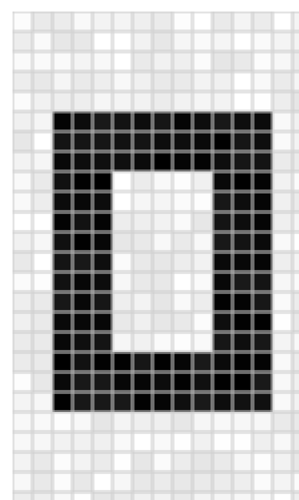
$\sigma_5 = 0.19$

$\dots$

$\sigma_{15} = 0.05$

$$M \approx \mathbf{u}_1\sigma_1\, \mathbf{v}_1^T + \mathbf{u}_2\sigma_2\, \mathbf{v}_2^T + \mathbf{u}_3\sigma_3\, \mathbf{v}_3^T$$

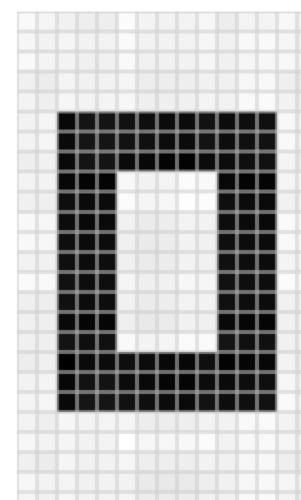
Noisy image



$c_1 \; c_1$

Improved image

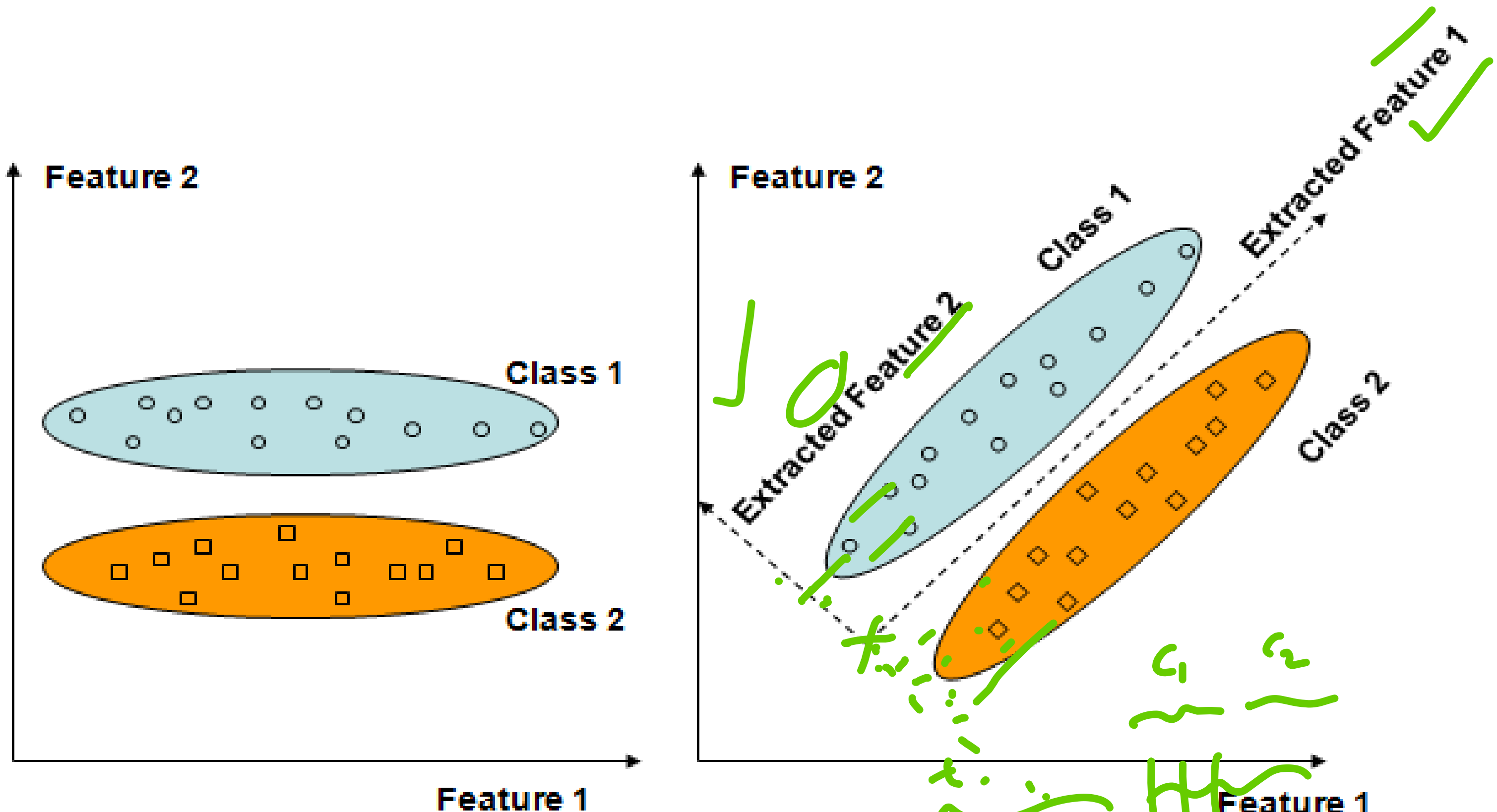
$c_1 \; c_1$



Update Techniques

- Recomputation
- Folding In
- SVD Update

Is LSI good for classification?



Sprinkling

d_1

c_1 c_2

LSA Applications

- Information Retrieval : Term Weighting
- Cross language Retrieval
- Other matrices: Matching people instead of documents

How would you compute SVD of a 3x4 matrix by hand?

References

- Indexing by Latent Semantic Analysis by Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, Richard Harshman