


**15-826: Multimedia Databases
and Data Mining**

Lecture #20: SVD - part III (more case studies)
C. Faloutsos



Must-read Material

- [Textbook](#) Appendix D
- Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.
- Brin, S. and L. Page (1998). Anatomy of a Large-Scale Hypertextual Web Search Engine. 7th Intl World Wide Web Conf.

15-826 Copyright: C. Faloutsos (2009) 2



Outline

Goal: 'Find similar / interesting things'

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

15-826 Copyright: C. Faloutsos (2009) 3

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 (2009) 4

CMU SCS

SVD - Detailed outline

- Motivation
- Definition - properties
- Interpretation
- Complexity
- Case studies
- ➔ SVD properties
- More case studies
- Conclusions

15-826 Copyright: C. Faloutsos (2009) 5

CMU SCS

SVD - detailed outline

- ...
- Case studies
- ➔ SVD properties
- more case studies
 - google/Kleinberg algorithms
 - query feedbacks
- Conclusions

15-826 Copyright: C. Faloutsos (2009) 6

CMU SCS

SVD - Other properties - summary

- can produce orthogonal basis (obvious) (who cares?)
- can solve over- and under-determined linear problems (see C(1) property)
- can compute 'fixed points' (= 'steady state prob. in Markov chains') (see C(4) property)

15-826 Copyright: C. Faloutsos (2009) 7

CMU SCS

SVD -outline of properties

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

15-826 Copyright: C. Faloutsos (2009) 8

CMU SCS

Properties - by defn.:

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

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

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

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

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

15-826 Copyright: C. Faloutsos (2009) 9

CMU SCS

Less obvious properties

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

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

15-826 Copyright: C. Faloutsos (2009) 10

CMU SCS

Less obvious properties

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

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

15-826 Copyright: C. Faloutsos (2009) 11

CMU SCS

Less obvious properties

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

B(1): $\mathbf{A}_{[n \times m]} (\mathbf{A}^T)_{[m \times n]} = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^T$
 symmetric; Intuition?
 'document-to-document' similarity matrix

B(2): symmetrically, for 'V'

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

15-826 Copyright: C. Faloutsos (2009) 12

CMU SCS

Less obvious properties

A: term-to-term similarity matrix

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

B(4): $(A^T A)^k \sim v_1 \lambda_1^{2k} v_1^T$ for $k \gg 1$
 where

- v_1 : $[m \times 1]$ first column (singular-vector) of V
- λ_1 : strongest singular value

15-826 Copyright: C. Faloutsos (2009) 13

CMU SCS

Proof of (B4)?

15-826 Copyright: C. Faloutsos (2009) 14

CMU SCS

Less obvious properties

B(4): $(A^T A)^k \sim v_1 \lambda_1^{2k} v_1^T$ for $k \gg 1$
 B(5): $(A^T A)^k v' \sim (\text{constant}) v_1$
 ie., for (almost) any v' , it converges to a vector parallel to v_1
 Thus, useful to compute first singular vector/value (as well as the next ones, too...)

15-826 Copyright: C. Faloutsos (2009) 15

CMU SCS

Proof of (B5)?

15-826 Copyright: C. Faloutsos (2009) 16

CMU SCS

Less obvious properties - repeated:

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

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

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

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

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

B(5): $(\mathbf{A}^T \mathbf{A})^k \mathbf{v}' \sim (\text{constant}) v_1$

15-826 Copyright: C. Faloutsos (2009) 17

CMU SCS

Least obvious properties

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

C(1): $\mathbf{A}_{[n \times m]} \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$
 let $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{(-1)} \mathbf{U}^T \mathbf{b}$
 if under-specified, \mathbf{x}_0 gives 'shortest' solution
 if over-specified, it gives the 'solution' with the smallest least squares error

(see Num. Recipes, p. 62)

15-826 Copyright: C. Faloutsos (2009) 18

CMU SCS

Least obvious properties

Illustration: under-specified, eg
 $[1 \ 2] [w \ z]^T = 4$ (ie, $1 w + 2 z = 4$)

15-826 Copyright: C. Faloutsos (2009) 19

CMU SCS

Verify formula:

$\mathbf{A} = [1 \ 2] \quad \mathbf{b} = [4]$
 $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$
 $\mathbf{U} = ??$
 $\mathbf{\Lambda} = ??$
 $\mathbf{V} = ??$
 $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{b}$

15-826 Copyright: C. Faloutsos (2009) 20

CMU SCS

Verify formula:

$\mathbf{A} = [1 \ 2] \quad \mathbf{b} = [4]$
 $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$
 $\mathbf{U} = [1]$
 $\mathbf{\Lambda} = [\text{sqrt}(5)]$
 $\mathbf{V} = [1/\text{sqrt}(5) \quad 2/\text{sqrt}(5)]^T$
 $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{b}$

15-826 Copyright: C. Faloutsos (2009) 21

CMU SCS

Verify formula:

$\mathbf{A} = [1 \ 2] \quad \mathbf{b} = [4]$
 $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$
 $\mathbf{U} = [1]$
 $\mathbf{\Lambda} = [\sqrt{5}]$
 $\mathbf{V} = [1/\sqrt{5} \quad 2/\sqrt{5}]^T$
 $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{b} = [1/5 \quad 2/5]^T [4]$
 $= [4/5 \quad 8/5]^T : w = 4/5, z = 8/5$

15-826 Copyright: C. Faloutsos (2009) 22

CMU SCS

Verify formula:

Show that $w = 4/5, z = 8/5$ is

(a) A solution to $1 \cdot w + 2 \cdot z = 4$ and

(b) Minimal (wrt Euclidean norm)

15-826 Copyright: C. Faloutsos (2009) 23

CMU SCS

Verify formula:

Show that $w = 4/5, z = 8/5$ is

(a) A solution to $1 \cdot w + 2 \cdot z = 4$ and
 A: easy

(b) Minimal (wrt Euclidean norm)
 A: $[4/5 \quad 8/5]$ is perpendicular to $[2 \quad -1]$

15-826 Copyright: C. Faloutsos (2009) 24

CMU SCS

Least obvious properties – cont'd

Illustration: over-specified, eg
 $[3 \ 2]^T [w] = [1 \ 2]^T$ (ie, $3w = 1; 2w = 2$)

15-826 Copyright: C. Faloutsos (2009) 25

CMU SCS

Verify formula:

$\mathbf{A} = [3 \ 2]^T \quad \mathbf{b} = [1 \ 2]^T$
 $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$
 $\mathbf{U} = ??$
 $\mathbf{\Lambda} = ??$
 $\mathbf{V} = ??$
 $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{b}$

15-826 Copyright: C. Faloutsos (2009) 26

CMU SCS

Verify formula:

$\mathbf{A} = [3 \ 2]^T \quad \mathbf{b} = [1 \ 2]^T$
 $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$
 $\mathbf{U} = [\frac{3}{\sqrt{13}} \ \frac{2}{\sqrt{13}}]^T$
 $\mathbf{\Lambda} = [\sqrt{13}]$
 $\mathbf{V} = [1]$
 $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{b} = [\frac{7}{13}]$

15-826 Copyright: C. Faloutsos (2009) 27

CMU SCS

Verify formula:

$$[3 \ 2]^T [7/13] = [1 \ 2]^T$$

$$[21/13 \ 14/13]^T \rightarrow \text{'red point'}$$

15-826 Copyright: C. Faloutsos (2009) 28

CMU SCS

Verify formula:

$$[3 \ 2]^T [7/13] = [1 \ 2]^T$$

$$[21/13 \ 14/13]^T \rightarrow \text{'red point' - perpendicular?}$$

15-826 Copyright: C. Faloutsos (2009) 29

CMU SCS

Verify formula:

$$A: [3 \ 2] \cdot ([1 \ 2] - [21/13 \ 14/13]) =$$

$$[3 \ 2] \cdot [-8/13 \ 12/13] = [3 \ 2] \cdot [-2 \ 3] = 0$$

15-826 Copyright: C. Faloutsos (2009) 30

CMU SCS

Least obvious properties - cont'd

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

C(2): $\mathbf{A}_{[n \times m]} \mathbf{v}_1_{[m \times 1]} = \lambda_1 \mathbf{u}_1_{[n \times 1]}$
 where $\mathbf{v}_1, \mathbf{u}_1$ the first (column) vectors of \mathbf{V}, \mathbf{U} . (\mathbf{v}_1 == right-singular-vector)

C(3): symmetrically: $\mathbf{u}_1^T \mathbf{A} = \lambda_1 \mathbf{v}_1^T$
 \mathbf{u}_1 == left-singular-vector

Therefore:

15-826 Copyright: C. Faloutsos (2009) 31

CMU SCS

Least obvious properties - cont'd

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

C(4): $\mathbf{A}^T \mathbf{A} \mathbf{v}_1 = \lambda_1^2 \mathbf{v}_1$

(fixed point - the defn of eigenvector for a symmetric matrix)

15-826 Copyright: C. Faloutsos (2009) 32

CMU SCS

Least obvious properties - altogether

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

C(1): $\mathbf{A}_{[n \times m]} \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$
 then, $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{(-1)} \mathbf{U}^T \mathbf{b}$: shortest, actual or least-squares solution

C(2): $\mathbf{A}_{[n \times m]} \mathbf{v}_1_{[m \times 1]} = \lambda_1 \mathbf{u}_1_{[n \times 1]}$

C(3): $\mathbf{u}_1^T \mathbf{A} = \lambda_1 \mathbf{v}_1^T$

C(4): $\mathbf{A}^T \mathbf{A} \mathbf{v}_1 = \lambda_1^2 \mathbf{v}_1$

15-826 Copyright: C. Faloutsos (2009) 33

CMU SCS

Properties - conclusions

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

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

C(1): $\mathbf{A}_{[n \times m]} \mathbf{x}_{[m \times 1]} = \mathbf{b}_{[n \times 1]}$
 then, $\mathbf{x}_0 = \mathbf{V} \mathbf{\Lambda}^{(-1)} \mathbf{U}^T \mathbf{b}$: shortest, actual or least-squares solution

C(4): $\mathbf{A}^T \mathbf{A} \mathbf{v}_1 = \lambda_1^2 \mathbf{v}_1$

15-826 Copyright: C. Faloutsos (2009) 34

CMU SCS


SVD - detailed outline

- ...
- Case studies
- SVD properties
- more case studies
- ➔ - Kleinberg/google algorithms
- query feedbacks
- Conclusions

15-826 Copyright: C. Faloutsos (2009) 35

CMU SCS

Kleinberg's algo (HITS)



Kleinberg, Jon (1998). *Authoritative sources in a hyperlinked environment*. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.

15-826 Copyright: C. Faloutsos (2009) 36

CMU SCS

Kleinberg's algorithm

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

Step 0: find all pages containing the query terms

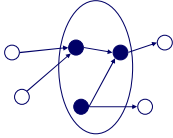
Step 1: expand by one move forward and backward

15-826 Copyright: C. Faloutsos (2009) 37

CMU SCS

Kleinberg's algorithm

- Step 1: expand by one move forward and backward

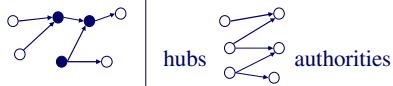


15-826 Copyright: C. Faloutsos (2009) 38

CMU SCS

Kleinberg's algorithm

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



15-826 Copyright: C. Faloutsos (2009) 39

CMU SCS

Kleinberg's algorithm

observations

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

15-826 Copyright: C. Faloutsos (2009) 40

CMU SCS

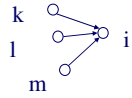
Kleinberg's algorithm

Let E be the set of edges and A be the adjacency matrix:
 the (i,j) is 1 if the edge from i to j exists
 Let h and a be $[n \times 1]$ vectors with the 'hubness' and 'authoritativeness' scores.
 Then:

15-826 Copyright: C. Faloutsos (2009) 41

CMU SCS

Kleinberg's algorithm



Then:

$$a_i = h_k + h_l + h_m$$

that is

$$a_i = \text{Sum}(h_j) \text{ over all } j \text{ that } (j,i) \text{ edge exists}$$

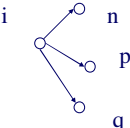
or

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

15-826 Copyright: C. Faloutsos (2009) 42

CMU SCS

Kleinberg's algorithm



symmetrically, for the 'hubness':

$$h_i = a_n + a_p + a_q$$

that is

$$h_i = \text{Sum}(q_j) \text{ over all } j \text{ that } (i,j) \text{ edge exists}$$

or

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

15-826 Copyright: C. Faloutsos (2009) 43

CMU SCS

Kleinberg's algorithm

In conclusion, we want vectors \mathbf{h} and \mathbf{a} such that:

$$\mathbf{h} = \mathbf{A} \mathbf{a} \quad \|\cdot\| = \square \|\cdot\|$$

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

Recall properties:

C(2): $\mathbf{A}_{[n \times m]} \mathbf{v}_1_{[m \times 1]} = \lambda_1 \mathbf{u}_1_{[n \times 1]}$

C(3): $\mathbf{u}_1^T \mathbf{A} = \lambda_1 \mathbf{v}_1^T$

15-826 Copyright: C. Faloutsos (2009) 44

CMU SCS

Kleinberg's algorithm

In short, the solutions to

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

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

are the left- and right- singular-vectors of the adjacency matrix \mathbf{A} .

Starting from random \mathbf{a}' and iterating, we'll eventually converge

(Q: to which of all the singular-vectors? why?)

15-826 Copyright: C. Faloutsos (2009) 45

CMU SCS

Kleinberg's algorithm

(Q: to which of all the singular-vectors?
why?)

A: to the ones of the strongest singular-value,
because of property B(5):

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

15-826 Copyright: C. Faloutsos (2009) 46

CMU SCS

Kleinberg's algorithm - results

Eg., for the query 'java':

0.328 www.gamelan.com

0.251 java.sun.com

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

15-826 Copyright: C. Faloutsos (2009) 47

CMU SCS

Kleinberg's algorithm - discussion

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

15-826 Copyright: C. Faloutsos (2009) 48

CMU SCS

SVD - detailed outline

- ...
- Case studies
- SVD properties
- more case studies
 - ➔ – Kleinberg/google algorithms
 - query feedbacks
- Conclusions

15-826 Copyright: C. Faloutsos (2009) 49

CMU SCS

PageRank (google)



• Brin, Sergey and Lawrence Page (1998). *Anatomy of a Large-Scale Hypertextual Web Search Engine*. 7th Intl World Wide Web Conf.

Larry
Page

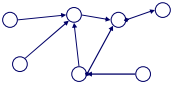
Sergey
Brin

15-826 Copyright: C. Faloutsos (2009) 50

CMU SCS

Problem: PageRank

Given a directed graph, find its most interesting/central node



A node is important, if it is connected with important nodes (recursive, but OK!)

15-826 Copyright: C. Faloutsos (2009) 51

CMU SCS

Problem: PageRank - solution

Given a directed graph, find its most interesting/central node

Proposed solution: Random walk; spot most 'popular' node (-> steady state prob. (ssp))

A node has high ssp, if it is connected with high ssp nodes (recursive, but OK!)

15-826 Copyright: C. Faloutsos (2009) 52

CMU SCS

(Simplified) PageRank algorithm

- Let **A** be the adjacency matrix;
- let **B** be the transition matrix: transpose, column-normalized - then

From **B**

		1		
1			1	
	1/2			1/2
				1/2
	1/2			

$$\begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix} = \begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix}$$

15-826 Copyright: C. Faloutsos (2009) 53

CMU SCS

(Simplified) PageRank algorithm

- B p = p**

B

		1		
1			1	
	1/2			1/2
				1/2
	1/2			

$$\begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix} = \begin{bmatrix} p1 \\ p2 \\ p3 \\ p4 \\ p5 \end{bmatrix}$$

15-826 Copyright: C. Faloutsos (2009) 54

CMU SCS

(Simplified) PageRank algorithm

- $\mathbf{B} \mathbf{p} = \mathbf{1} * \mathbf{p}$
- thus, \mathbf{p} is the **eigenvector** that corresponds to the highest eigenvalue ($=1$, since the matrix is column-normalized)
- Why does such a \mathbf{p} exist?
 - \mathbf{p} exists if \mathbf{B} is $n \times n$, nonnegative, irreducible [Perron–Frobenius theorem]

15-826 Copyright: C. Faloutsos (2009) 55

CMU SCS

(Simplified) PageRank algorithm

- In short: imagine a particle randomly moving along the edges
- compute its steady-state probabilities (ssp)

Full version of algo: with occasional random jumps
 Why? To make the matrix irreducible

15-826 Copyright: C. Faloutsos (2009) 56

CMU SCS

Full Algorithm

- With probability $1-c$, fly-out to a random node
- Then, we have

$$\mathbf{p} = c \mathbf{B} \mathbf{p} + (1-c)/n \mathbf{1} \Rightarrow$$

$$\mathbf{p} = (1-c)/n [\mathbf{I} - c \mathbf{B}]^{-1} \mathbf{1}$$

$\begin{bmatrix} | \\ | \\ | \end{bmatrix}$

$\begin{bmatrix} \square \\ \square \\ \square \end{bmatrix}$

$\begin{bmatrix} | \\ | \\ | \end{bmatrix}$

15-826 Copyright: C. Faloutsos (2009) 57

CMU SCS

Parenthesis: intuition behind eigenvectors

15-826 Copyright: C. Faloutsos (2009) 58

CMU SCS

Formal definition

If A is a $(n \times n)$ square matrix
 (λ, x) is an **eigenvalue/eigenvector** pair
of A if

$Ax = \lambda x$

CLOSELY related to singular values:

15-826 Copyright: C. Faloutsos (2009) 59

CMU SCS

Property #1: Eigen- vs singular-values

if

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

then $A = (B^T B)$ is symmetric and

$$C(4): B^T B v_i = \lambda_i^2 v_i$$

ie, v_1, v_2, \dots : eigenvectors of $A = (B^T B)$

15-826 Copyright: C. Faloutsos (2009) 60

CMU SCS

Property #2

- If $A_{[n \times n]}$ is a real, symmetric matrix
- Then it has n real eigenvalues

(if A is not symmetric, some eigenvalues may be complex)

15-826 Copyright: C. Faloutsos (2009) 61

CMU SCS

Property #3

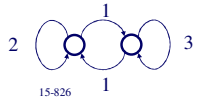
- If $A_{[n \times n]}$ is a real, symmetric matrix
- Then it has n real eigenvalues
- And they agree with its n singular values, except possibly for the sign

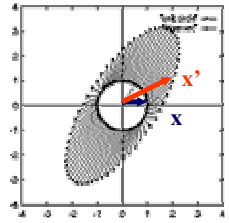
15-826 Copyright: C. Faloutsos (2009) 62

CMU SCS

Intuition

- A as vector transformation

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




15-826 Copyright: C. Faloutsos (2009) 63

CMU SCS

Intuition

- By defn., eigenvectors remain parallel to themselves ('fixed points')

$$\lambda_1 \mathbf{v}_1 = \mathbf{A} \mathbf{v}_1$$

$$3.62 * \begin{bmatrix} 0.52 \\ 0.85 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0.52 \\ 0.85 \end{bmatrix}$$

15-826 Copyright: C. Faloutsos (2009)

CMU SCS

Convergence

- Usually, fast:

15-826 Copyright: C. Faloutsos (2009)

CMU SCS

Convergence

- Usually, fast:

15-826 Copyright: C. Faloutsos (2009)

CMU SCS

Convergence

- Usually, fast:
- depends on ratio $\lambda_1 : \lambda_2$

15-826 Copyright: C. Faloutsos (2009)

CMU SCS

Kleinberg/google - conclusions

SVD helps in graph analysis:

- hub/authority scores: strongest left- and right-singular-vectors of the adjacency matrix
- random walk on a graph: steady state probabilities are given by the strongest eigenvector of the transition matrix

15-826 Copyright: C. Faloutsos (2009) 68

CMU SCS

SVD - detailed outline

- ...
- Case studies
- SVD properties
- more case studies
 - google/Kleinberg algorithms
 - query feedbacks
- Conclusions

15-826 Copyright: C. Faloutsos (2009) 69

CMU SCS

Query feedbacks


[Chen & Roussopoulos, sigmod 94]
 Sample problem:
 estimate selectivities (e.g., 'how many movies were made between 1940 and 1945?')
 for query optimization,
 LEARNING from the query results so far!!

15-826 Copyright: C. Faloutsos (2009) 70

CMU SCS

Query feedbacks

Idea #1: consider a function for the CDF (cummulative distr. function), eg., 6-th degree polynomial (or splines, or anything else)

count, so far 
 year

15-826 Copyright: C. Faloutsos (2009) 71

CMU SCS

Query feedbacks

For example
 $F(x) = \# \text{ movies made until year 'x'}$
 $= a_1 + a_2 * x + a_3 * x^2 + \dots a_7 * x^6$

15-826 Copyright: C. Faloutsos (2009) 72

CMU SCS

Query feedbacks

GREAT idea #2: adapt your model, as you see the actual counts of the actual queries

count, so far

original estimate

actual

year

15-826 Copyright: C. Faloutsos (2009) 73

CMU SCS

Query feedbacks

count, so far

original estimate

actual

year

15-826 Copyright: C. Faloutsos (2009) 74

CMU SCS

Query feedbacks

count, so far

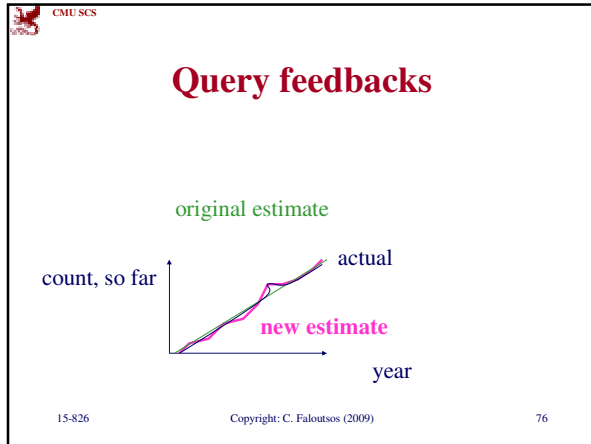
original estimate

actual

a query

year

15-826 Copyright: C. Faloutsos (2009) 75



Query feedback

Eventually, the problem becomes:

- estimate the parameters a_1, \dots, a_7 of the model
- to minimize the least squares errors from the real answers so far.

Formally:

Query feedback

Formally, with n queries and 6-th degree polynomials:

$$\begin{bmatrix} x_{11} & x_{12} & & & & & x_{17} \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ x_{n1} & x_{n2} & & & & & x_{n7} \end{bmatrix}
 \begin{bmatrix} a_1 \\ a_2 \\ & \\ & \\ & \\ a_7 \end{bmatrix}
 =
 \begin{bmatrix} b_1 \\ b_2 \\ & \\ & \\ & \\ b_n \end{bmatrix}$$

CMU SCS

Query feedbacks

where $x_{i,j}$ such that $\text{Sum}(x_{i,j} * a_i) =$ our estimate for the # of movies and b_j : the actual

X11	X12		X17
Xn1	Xn2		Xn7

a1
a2
a7

=

b1
b2
bn

15-826 Copyright: C. Faloutsos (2009) 79

CMU SCS

Query feedbacks

For example, for query 'find the count of movies during (1920-1932)':

$$a_1 + a_2 * 1932 + a_3 * 1932**2 + \dots$$

$$-$$

$$(a_1 + a_2 * 1920 + a_3 * 1920**2 + \dots)$$

X11	X12		X17
Xn1	Xn2		Xn7

a1
a2
a7

=

b1
b2
bn

15-826 Copyright: C. Faloutsos (2009) 80

CMU SCS

Query feedbacks

And thus $X11 = 0$; $X12 = 1932-1920$, etc

$$a_1 + a_2 * 1932 + a_3 * 1932**2 + \dots$$

$$-$$

$$(a_1 + a_2 * 1920 + a_3 * 1920**2 + \dots)$$

X11	X12		X17
Xn1	Xn2		Xn7

a1
a2
a7

=

b1
b2
bn

15-826 Copyright: C. Faloutsos (2009) 81

CMU SCS

Query feedbacks

In matrix form:

$$\mathbf{X} \mathbf{a} = \mathbf{b}$$

1st query

X11	X12		X17
Xn1	Xn2		Xn7

=

a1
a2
a7

b1
b2
bn

15-826 Copyright: C. Faloutsos (2009) 82

CMU SCS

Query feedbacks

In matrix form:

$$\mathbf{X} \mathbf{a} = \mathbf{b}$$

and the least-squares estimate for \mathbf{a} is

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

according to property C(1)
(let $\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$)

15-826 Copyright: C. Faloutsos (2009) 83

CMU SCS

Query feedbacks - enhancements

The solution

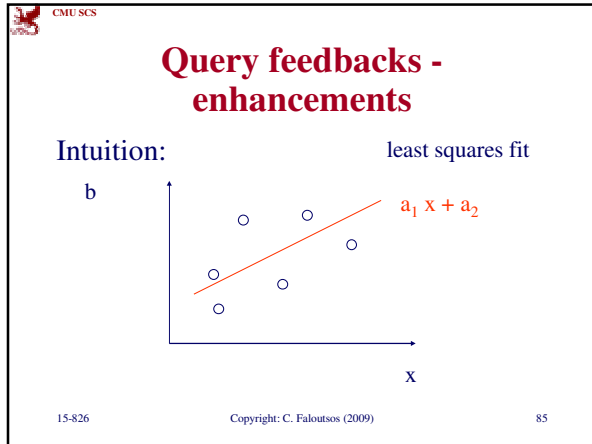
$$\mathbf{a} = \mathbf{V} \mathbf{\Lambda}^{(-1)} \mathbf{U}^T \mathbf{b}$$

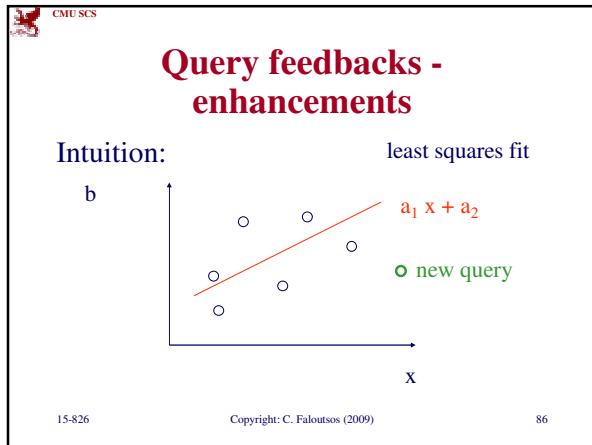
works, but needs expensive SVD each time a new query arrives

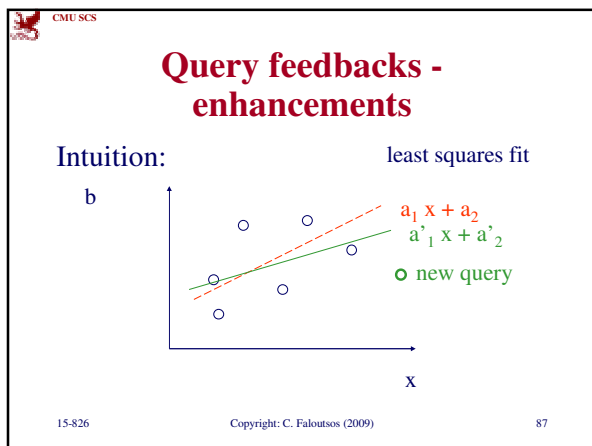
GREAT Idea #3: Use 'Recursive Least Squares', to adapt \mathbf{a} incrementally.

Details: in paper - intuition:

15-826 Copyright: C. Faloutsos (2009) 84







CMU SCS

Query feedbacks - enhancements

the new coefficients can be quickly computed from the old ones, plus statistics in a (7x7) matrix
(no need to know the details, although the RLS is a brilliant method)

15-826 Copyright: C. Faloutsos (2009) 88

CMU SCS

Query feedbacks - enhancements

GREAT idea #4: 'forgetting' factor - we can even down-play the weight of older queries, since the data distribution might have changed.
(comes for 'free' with RLS...)

15-826 Copyright: C. Faloutsos (2009) 89

CMU SCS

Query feedbacks - conclusions

SVD helps find the Least Squares solution, to adapt to query feedbacks
(RLS = Recursive Least Squares is a great method to incrementally update least-squares fits)

15-826 Copyright: C. Faloutsos (2009) 90

CMU SCS

SVD - detailed outline

- ...
- Case studies
- SVD properties
- more case studies
 - google/Kleinberg algorithms
 - query feedbacks
- ➔ • Conclusions

15-826 Copyright: C. Faloutsos (2009) 91

CMU SCS

Conclusions

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


15-826 Copyright: C. Faloutsos (2009) 92

CMU SCS

Conclusions cont'd

- ... and can find fixed-points or steady-state probabilities (google/ Kleinberg/ Markov Chains)
- ... and can solve optimally over- and under-constraint linear systems (least squares / query feedbacks)


15-826 Copyright: C. Faloutsos (2009) 93

 CMU SCS

References

- Brin, S. and L. Page (1998). Anatomy of a Large-Scale Hypertextual Web Search Engine. 7th Intl World Wide Web Conf.
- Chen, C. M. and N. Roussopoulos (May 1994). Adaptive Selectivity Estimation Using Query Feedback. Proc. of the ACM-SIGMOD , Minneapolis, MN.

15-826 Copyright: C. Faloutsos (2009) 94

 CMU SCS

References cont'd

- Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.
- Press, W. H., S. A. Teukolsky, et al. (1992). Numerical Recipes in C, Cambridge University Press.

15-826 Copyright: C. Faloutsos (2009) 95
