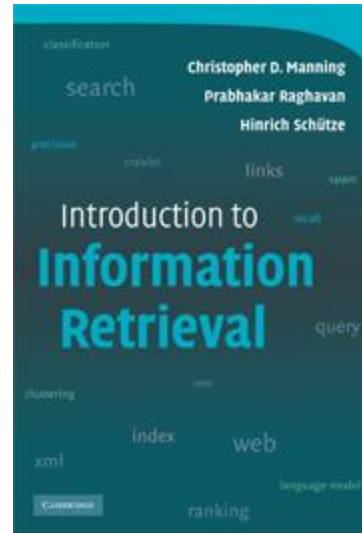


Information Retrieval and Organisation



Chapter 18

Matrix Decomposition & Latent Semantic Indexing

Dell Zhang

Birkbeck, University of London



Singular Value Decomposition

[SVD Tutorial by iMetaSearch](#)

Matrix

- Multiplication, Transposition, ...

$$\begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} \mathbf{X} \begin{bmatrix} 0 & 3 & 0 \\ 2 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 1*0 + 4*2 & 1*3 + 4*1 & 1*0 + 4*4 \\ 0*0 + 3*2 & 0*3 + 3*1 & 0*0 + 3*4 \end{bmatrix}$$
$$= \begin{bmatrix} 8 & 7 & 16 \\ 6 & 3 & 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 4 & 3 \end{bmatrix}$$

\mathbf{A}^T

Example: Golf Scores

Hole	Par	Phil	Tiger	Vijay
1	4	4	4	4
2	5	5	5	5
3	3	3	3	3
4	4	4	4	4
5	4	4	4	4
6	4	4	4	4
7	4	4	4	4
8	3	3	3	3
9	5	5	5	5

SVD: One Factor

Phil	Tiger	Vijay
4	4	4
5	5	5
3	3	3
4	4	4
4	4	4
4	4	4
4	4	4
3	3	3
5	5	5

=

HoleDifficulty
4
5
3
4
4
4
4
3
5

*

PlayerAbility		
Phil	Tiger	Vijay
1	1	1

SVD: One Factor

Phil	Tiger	Vijay
4	4	4
5	5	5
3	3	3
4	4	4
4	4	4
4	4	4
4	4	4
3	3	3
5	5	5

=

HoleDifficulty
0.33
0.41
0.25
0.33
0.33
0.33
0.33
0.25
0.41

*

ScaleFactor
21.07

*

PlayerAbility		
Phil	Tiger	Vijay
0.58	0.58	0.58

SVD: More Factors

Hole	Par	Phil	Tiger	Vijay
1	4	4	4	5
2	5	4	5	5
3	3	3	3	2
4	4	4	5	4
5	4	4	4	4
6	4	3	5	4
7	4	4	4	3
8	3	2	4	4
9	5	5	5	5

SVD: More Factors

Phil	Tiger	Vijay
3.95	4.64	4.34
4.27	5.02	4.69
2.42	2.85	2.66
3.97	4.67	4.36
3.64	4.28	4.00
3.69	4.33	4.05
3.33	3.92	3.66
3.08	3.63	3.39
4.55	5.35	5.00

 $=$

HoleDifficulty
4.34
4.69
2.66
4.36
4.00
4.05
3.66
3.39
5.00

 $*$

PlayerAbility		
Phil	Tiger	Vijay
0.91	1.07	1.00

SVD: More Factors

Phil	Tiger	Vijay
0.05	-0.64	0.66
-0.28	-0.02	0.31
0.58	0.15	-0.66
0.03	0.33	-0.36
0.36	-0.28	0.00
-0.69	0.67	-0.05
0.67	0.08	-0.66
-1.08	0.37	0.61
0.45	-0.35	0.00

 $=$

HoleDifficulty2
-0.18
-0.38
0.80
0.15
0.35
-0.67
0.89
-1.29
0.44

 $*$

PlayerAbility2		
Phil	Tiger	Vijay
0.82	-0.20	-0.53

SVD: Full Factorization

Phil	Tiger	Vijay
4	4	5
4	5	5
3	3	2
4	5	4
4	4	4
3	5	4
4	4	3
2	4	4
5	5	5

=

HoleDifficulty 1-3		
4.34	-0.18	-0.90
4.69	-0.38	-0.15
2.66	0.80	0.40
4.36	0.15	0.47
4.00	0.35	-0.29
4.05	-0.67	0.68
3.66	0.89	0.33
3.39	-1.29	0.14
5.00	0.44	-0.36

*

PlayerAbility 1-3		
Phil	Tiger	Vijay
0.91	1.07	1.00
0.82	-0.20	-0.53
-0.21	0.76	-0.62

SVD: Full Factorization

Phil	Tiger	Vijay
4	4	5
4	5	5
3	3	2
4	5	4
4	4	4
3	5	4
4	4	3
2	4	4
5	5	5

 $=$

HoleDifficulty 1-3		
0.35	0.09	-0.64
0.38	0.19	-0.10
0.22	-0.40	0.28
0.36	-0.08	0.33
0.33	-0.18	-0.20
0.33	0.33	0.48
0.30	-0.44	0.23
0.28	0.64	0.10
0.41	-0.22	-0.25

 $*$

ScaleFactor 1-3		
21.07	0	0
0	2.01	0
0	0	1.42

 $*$

PlayerAbility 1-3		
Phil	Tiger	Vijay
0.53	0.62	0.58
-0.82	0.20	0.53
-0.21	0.76	-0.62



Latent Semantic Indexing

[LSI Tutorial by iMetaSearch](#)

Word \rightarrow Concept

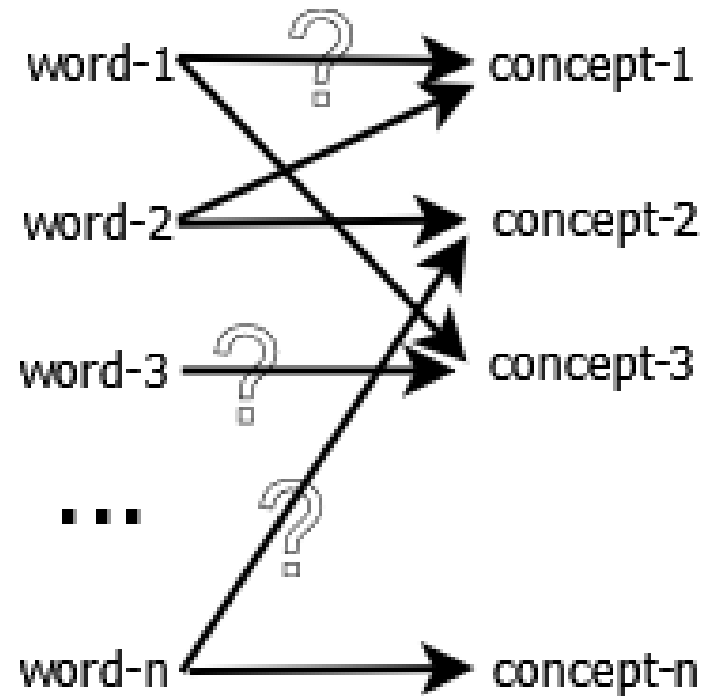
word-1 \longrightarrow concept-1

word-2 \longrightarrow concept-2

...

word-n \longrightarrow concept-n

Word \rightarrow Concept



Word → Concept

- For example, the word “bank”
 - when used together with mortgage, loans, and rates, it probably means *a financial institution*
 - when used together with lures, casting, and fish probably means *a stream or river bank*

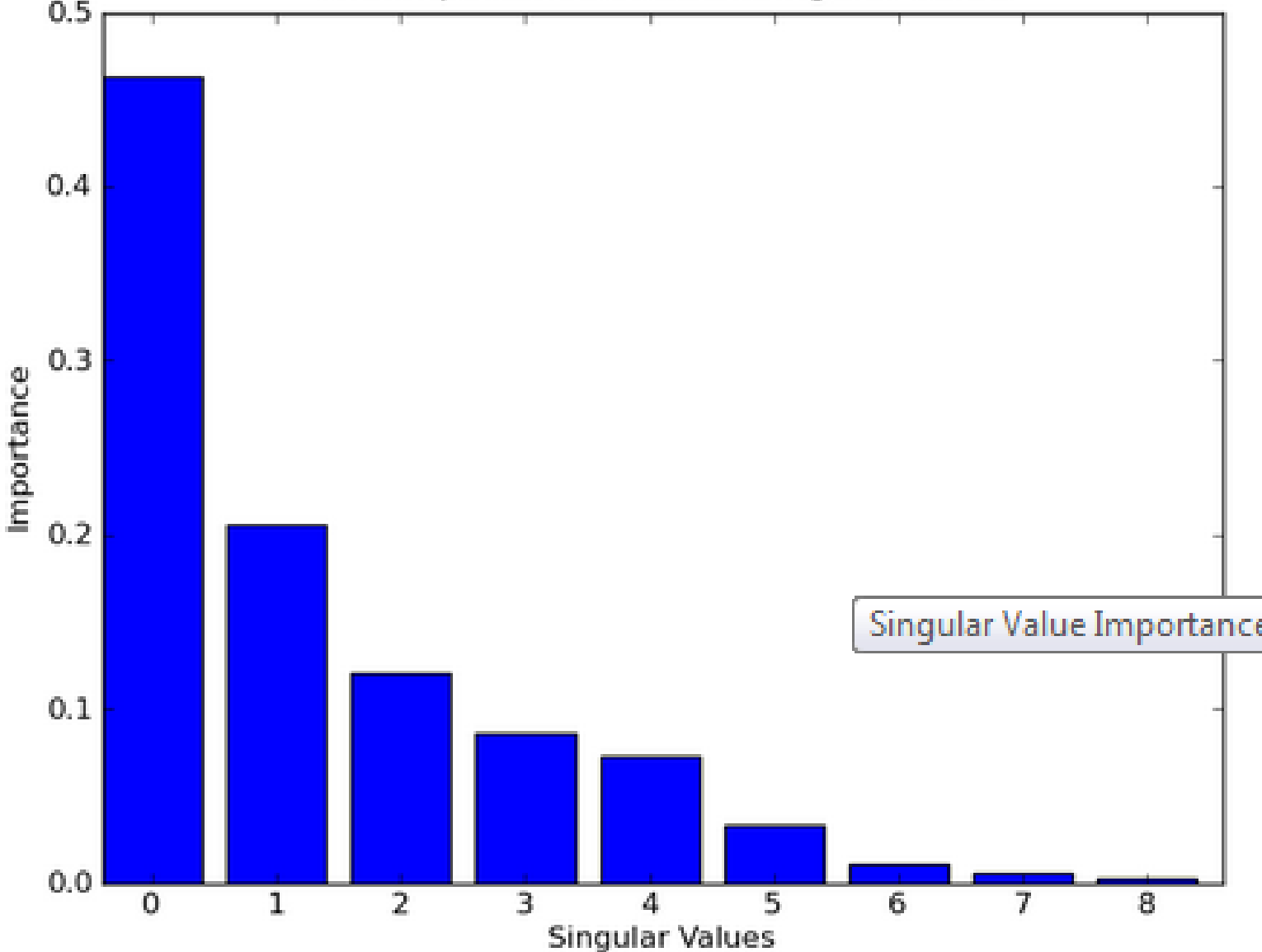
Example: Book Titles

1. The Neatest Little Guide to Stock Market Investing
2. Investing For Dummies, 4th Edition
3. The Little Book of Common Sense Investing: The Only Way to Guarantee Your Fair Share of Stock Market Returns
4. The Little Book of Value Investing
5. Value Investing: From Graham to Buffett and Beyond
6. Rich Dad's Guide to Investing: What the Rich Invest in, That the Poor and the Middle Class Do Not!
7. Investing in Real Estate, 5th Edition
8. Stock Investing For Dummies
9. Rich Dad's Advisors: The ABC's of Real Estate Investing: The Secrets of Finding Hidden Profits Most Investors Miss

Term-Document Count Matrix

Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

The Importance of Each Singular Value



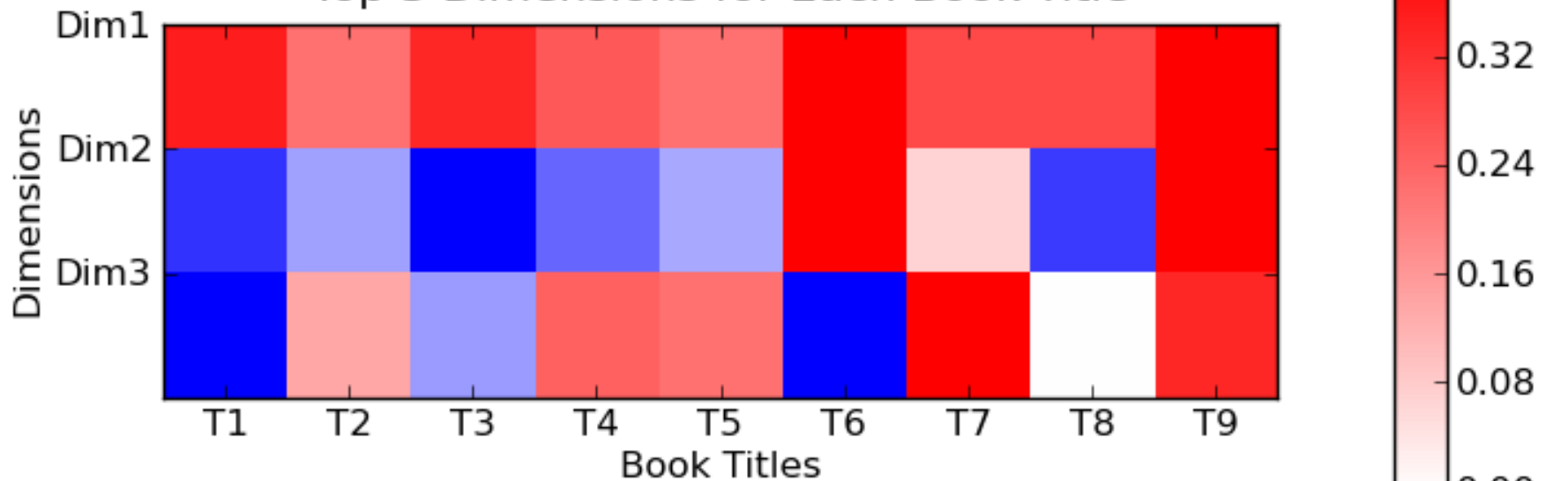
Latent Semantic Indexing

book	0.15	-0.27	0.04
dads	0.24	0.38	-0.09
dummies	0.13	-0.17	0.07
estate	0.18	0.19	0.45
guide	0.22	0.09	-0.46
investing	0.74	-0.21	0.21
market	0.18	-0.30	-0.28
real	0.18	0.19	0.45
rich	0.36	0.59	-0.34
stock	0.25	-0.42	-0.28
value	0.12	-0.14	0.23

3.91	0	0
0	2.61	0
0	0	2.00

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.35	0.22	0.34	0.26	0.22	0.49	0.28	0.29	0.44
-0.32	-0.15	-0.46	-0.24	-0.14	0.55	0.07	-0.31	0.44
-0.41	0.14	-0.16	0.25	0.22	-0.51	0.55	0.00	0.34

Top 3 Dimensions for Each Book Title



Dim2	Titles
red	6-7, 9
blue	1-5, 8

Dim2	Dim3	Titles
red	red	7, 9
red	blue	6
blue	red	2, 4-5, 8
blue	blue	1, 3

XY Plot of Words and Titles

