



DEEP
LEARNING
INSTITUTE



PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

Lecture 20.3 - SVD: Dimensionality Reduction, and Other Uses



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

SVD - Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:

U: document-concept similarity matrix

V: term-concept similarity matrix

Λ : diagonal elements: **concept “strengths”**

SVD - Interpretation #1

‘documents’, ‘terms’ and ‘concepts’:

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

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

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

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

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

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

SVD is closely related to PCA, and can be numerically more stable.

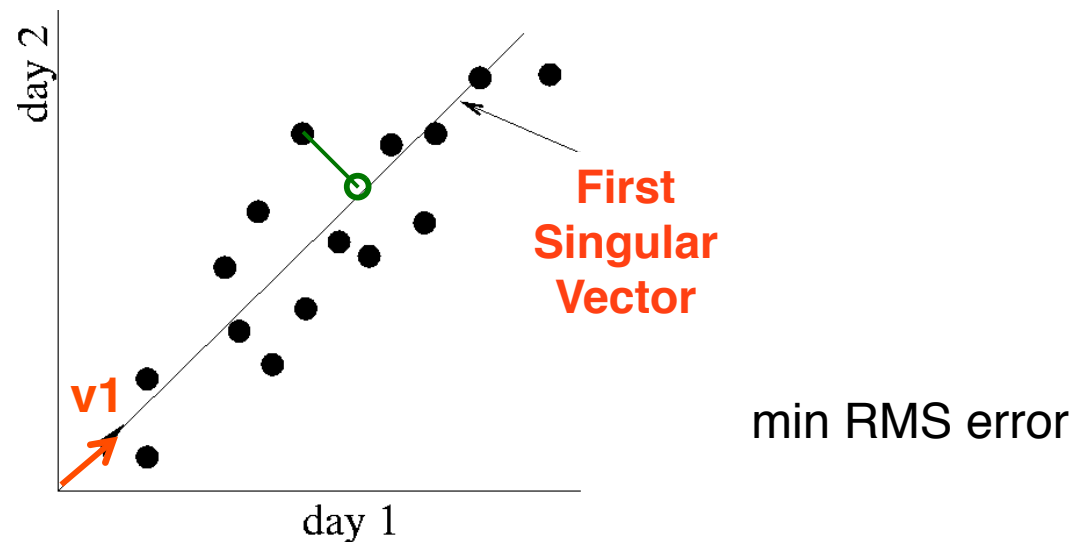
For more info, see:

<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

Find the best axis to project on.

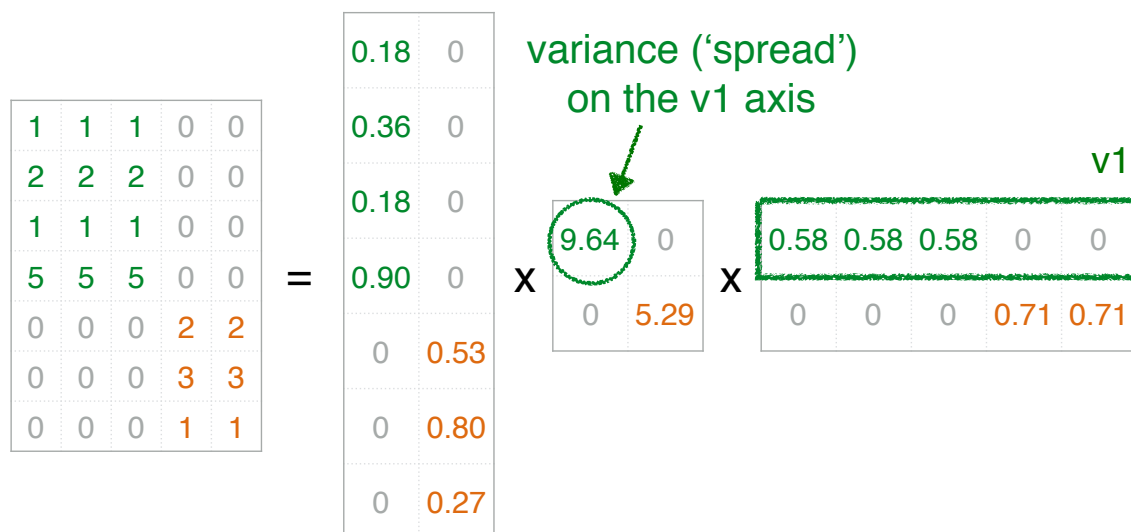
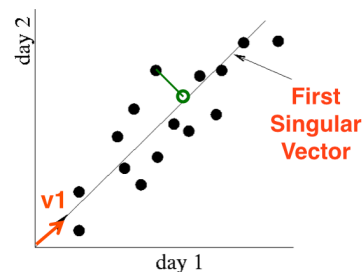
('best' = min sum of squares of projection errors)



Beautiful visualization explaining PCA:
<http://setosa.io/ev/principal-component-analysis/>

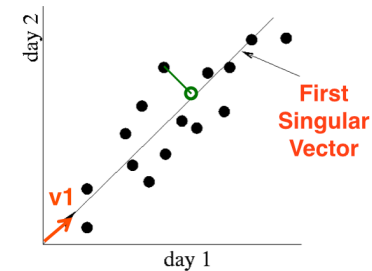
SVD - Interpretation #2

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



SVD - Interpretation #2

$U \Lambda$ gives the **coordinates** of the points in the projection axis



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

$v1$

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:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.9 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

The value 5.9 in the middle matrix is highlighted with a red 'X', indicating it is the smallest singular value being set to zero for dimensionality reduction.

SVD - Interpretation #2

More details

Q: how exactly is dim. reduction done?

A: set the smallest singular values to zero:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.49 \end{bmatrix} \times \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

The diagram illustrates the SVD decomposition of a 7x5 matrix. The original matrix is shown on the left. The decomposition is represented as the product of three matrices: a 7x2 matrix of singular values, a 2x2 matrix of singular values, and a 2x5 matrix of right singular vectors. Red boxes and 'X' marks indicate the process of dimensionality reduction by setting the smallest singular values to zero. In the first matrix, the second column (0.53, 0.80, 0.27) is boxed and marked with a red 'X'. In the second matrix, the second element (5.49) is boxed and marked with a red 'X'. In the third matrix, the last two columns (0.71, 0.71) are boxed and marked with a red 'X'.

SVD - Interpretation #2

More details

Q: how exactly is dim. reduction done?

A: set the smallest singular values to zero:

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

SVD - Interpretation #2

More details

Q: how exactly is dim. reduction done?

A: set the smallest singular values to zero:

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

~

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

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 ...)

Case Study

How to do queries with LSI?

Case Study

How to do queries with LSI?

For example, how to find documents with 'data'?

The diagram illustrates the LSI query process. It starts with a matrix of document counts for 'data', 'info', 'retrieval', 'brain', and 'lung' across 8 documents. Green arrows indicate 'CS docs' (top 5 documents) and orange arrows indicate 'MD docs' (bottom 3 documents). The matrix is then multiplied by a vector of cosine similarities for the 'data' term, which is then multiplied by a matrix of cosine similarities for the other terms.

	data	info	retrieval	brain	lung
1	1	1	1	0	0
2	2	2	2	0	0
3	1	1	1	0	0
4	5	5	5	0	0
5	0	0	0	2	2
6	0	0	0	3	3
7	0	0	0	1	1
8	0	0	0	0	0

=

0.18	0
0.36	0
0.18	0
0.90	0
0	0.53
0	0.80
0	0.27

x

9.64	0
0	5.29

x

0.58	0.58	0.58	0	0
0	0	0	0.71	0.71

Case Study

How to do queries with LSI?

For example, how to find documents with 'data'?

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

The diagram illustrates the LSI query process. It shows a matrix of document counts for concepts (data, info, retrieval, brain, lung) across 7 documents. This is multiplied by a matrix of concept vectors (0.18, 0.36, 0.18, 0.90, 0, 0, 0) and a matrix of document vectors (9.64, 0, 0, 5.29). The result is a vector of document scores (0.53, 0.80, 0.27).

	data	info	retrieval	brain	lung
1	1	1	1	0	0
2	2	2	2	0	0
3	1	1	1	0	0
4	5	5	5	0	0
5	0	0	0	2	2
6	0	0	0	3	3
7	0	0	0	1	1

=

0.18	0
0.36	0
0.18	0
0.90	0
0	0.53
0	0.80
0	0.27

x

9.64	0
0	5.29

x

0.58	0.58	0.58	0	0
0	0	0	0.71	0.71

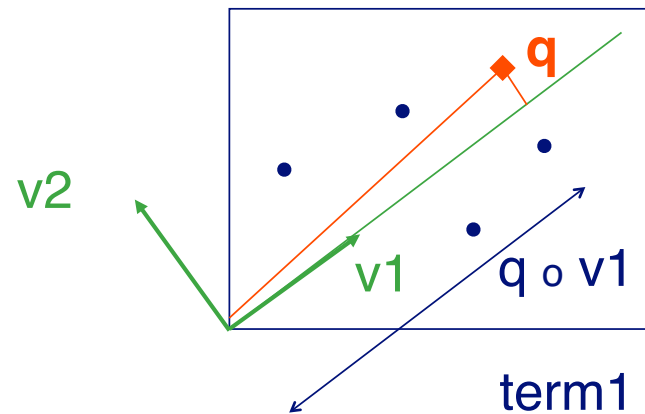
Case Study

How to do queries with LSI?

For example, how to find documents with 'data'?

A: map query vectors into 'concept space', using **inner product** (cosine similarity) with each 'concept' vector v_i

$$\mathbf{q} = \begin{array}{c} \text{data} \\ \text{info} \\ \text{retrieval} \\ \text{brain} \\ \text{lung} \end{array} \begin{array}{|c|c|c|c|c|} \hline 1 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

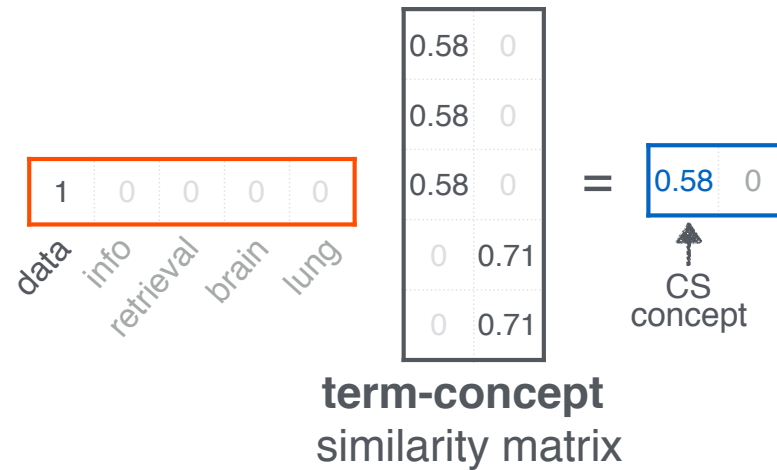


Case Study

How to do queries with LSI?

Compactly, we have:

$$\mathbf{q} \mathbf{V} = \mathbf{q}_{\text{concept}}$$



Case Study

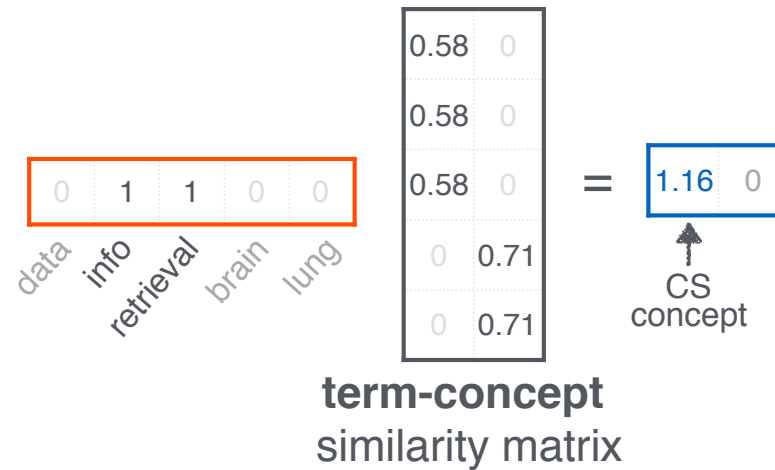
**How would the document
(‘information’, ‘retrieval’) be handled?**

Case Study

How would the document (‘information’, ‘retrieval’) be handled?

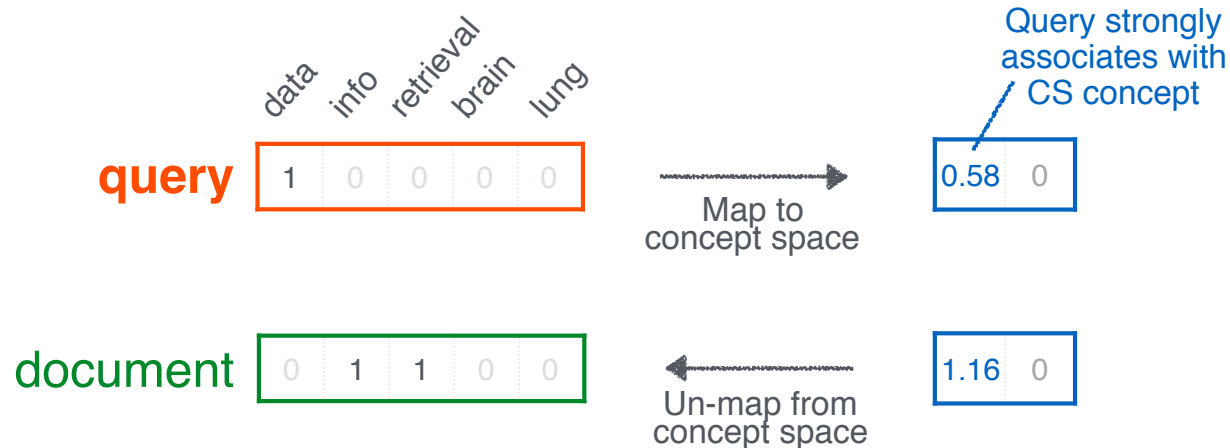
SAME!

$$\mathbf{d} \mathbf{V} = \mathbf{d}_{\text{concept}}$$



Case Study Observation

Document ('information', 'retrieval') will be retrieved by **query** ('data'), even though it does not contain 'data'!!





DEEP
LEARNING
INSTITUTE



PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

Thank You