DLI Accelerated Data Science Teaching Kit

Lecture 20.3 - SVD: Dimensionality Reduction, and Other Uses

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

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* $\mathbf{A}^\mathsf{T}\mathbf{A}$

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

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

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

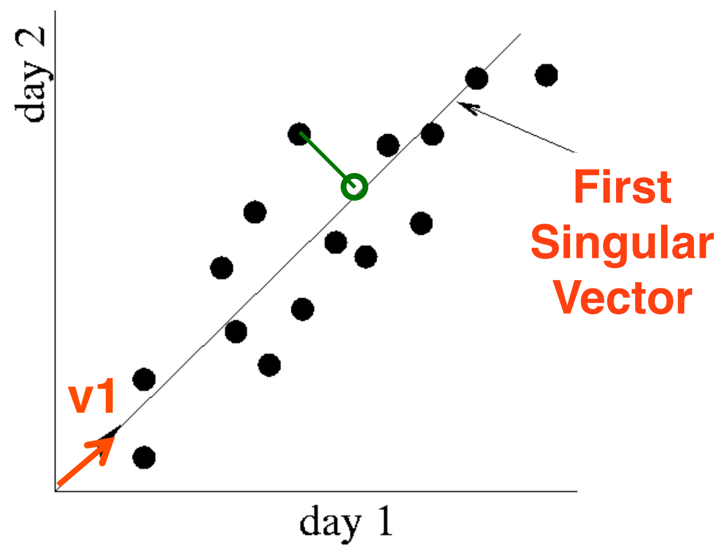
SVD is closely related to PCA, and can be numerically more stable.
For more info, see:

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.

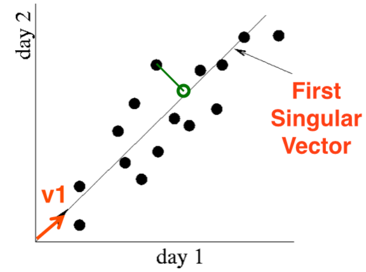
('best' = min sum of squares of projection errors)



min RMS error

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

SVD - Interpretation #2



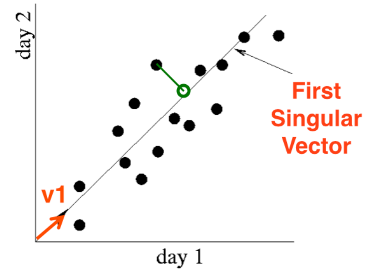
$$A = U \Lambda V^T$$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{array}{c} \text{v1} \\ \begin{bmatrix} \boxed{0.58 \quad 0.58 \quad 0.58 \quad 0 \quad 0} \\ 0 \quad 0 \quad 0 \quad 0.71 \quad 0.71 \end{bmatrix} \end{array}$$

variance ('spread') on the v1 axis

SVD - Interpretation #2

U Λ gives the **coordinates** of the points in the projection axis



$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.18 & 0 \\ 0.36 & 0 \\ 0.18 & 0 \\ 0.90 & 0 \\ 0 & 0.53 \\ 0 & 0.80 \\ 0 & 0.27 \end{bmatrix} \times \begin{bmatrix} 9.64 & 0 \\ 0 & 5.29 \end{bmatrix} \times \begin{array}{r} \text{v1} \\ \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix} \end{array}$$

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:

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

SVD - Interpretation #2

More details

Q: how exactly is dim. reduction done?

A: set the smallest singular values to zero:

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

SVD - Interpretation #2

More details

Q: how exactly is dim. reduction done?

A: set the smallest singular values to zero:

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

SVD - Interpretation #2

More details

Q: how exactly is dim. reduction done?

A: set the smallest singular values to zero:

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

SVD - 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’?

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

Case Study

How to do queries with LSI?

For example, how to find documents with ‘data’?
A: map query vectors into ‘concept space’ – how?

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

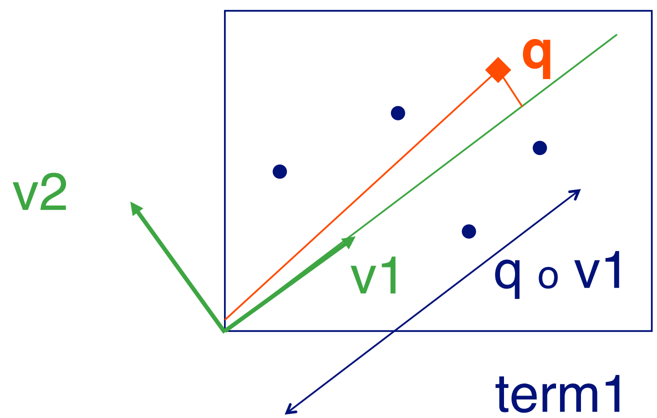
Case Study

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**

q =

data	info	retrieval	brain	lung
1	0	0	0	0



Case Study

How to do queries with LSI?

Compactly, we have:

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

$$\begin{array}{ccccc} \text{data} & \text{info} & \text{retrieval} & \text{brain} & \text{lung} \\ \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.58 & 0 \\ 0.58 & 0 \\ 0.58 & 0 \\ 0 & 0.71 \\ 0 & 0.71 \end{bmatrix} = \begin{bmatrix} 0.58 & 0 \end{bmatrix}$$

CS concept

term-concept similarity matrix

Case Study

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

Case Study

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

SAME!

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

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

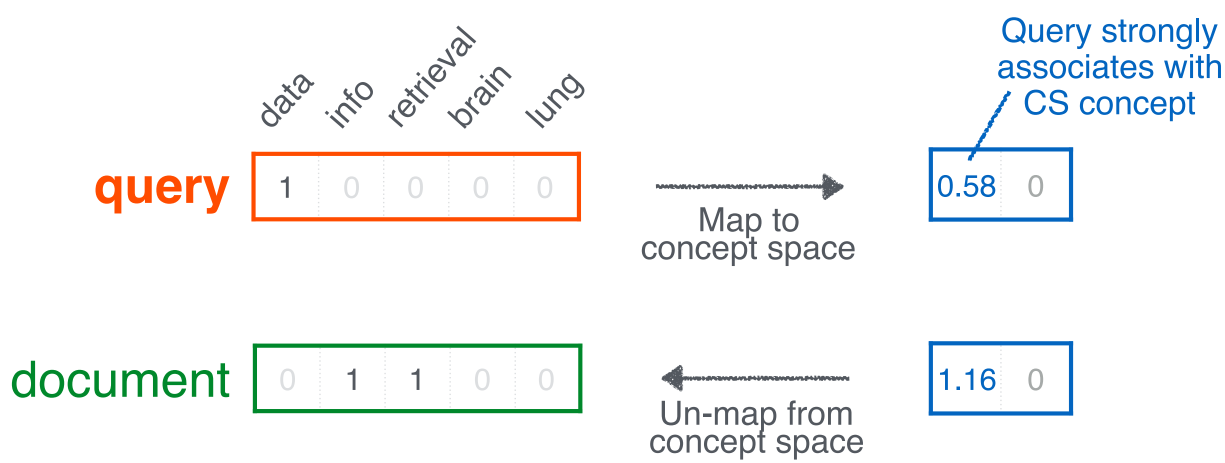
CS concept

term-concept similarity matrix

Case Study

Observation

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



DLI Accelerated Data Science Teaching Kit

Thank You