LSI (Latent Semantic Indexing)

September 12, 2023

Overview

LSI: Latent Semantic Indexing

- TF-IDF matrix is the basis for computing the similarity between documents.
- Can we transform this matrix, so that we get a better measure of similarity between documents (and words)?

	Antony&Cleopatra	*JuliusCaesar*	*TheTempest*	*Hamlet*	*Othello*	*Macbeth*
Antony	5.25	3.18	0	0	0	0.35
Brutus	1.21	6.1	0	1	0	0
Caesar	8.59	2.54	0	1.51	0.25	0
Calpurnia	0	1.54	0	0	0	0
Cleopatra	2.85	0	0	0	0	0
mercy	1.51	0	1.9	0.12	5.25	0.88
worser	1.37	0	0.11	4.15	0.25	1.95

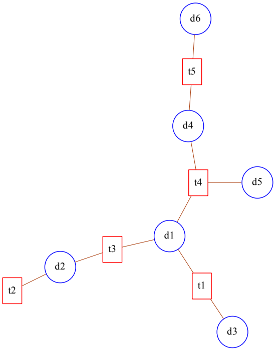
The Problem With Term-doc (Doc-Term) Matrix

Example from the IR book.

Let C=

	d1	d2	d3	d4	d5	d6
ship	1	0	1	0	0	0
boat	0	1	0	0	0	0
ocean	1	1	0	0	0	0
wood	1	0	0	1	1	0
tree	0	0	0	1	0	1

- This is a standard term-document matrix. We use a non-weighted matrix here to simplify the example.
- d2 and d3 have zero similarity
- Even though then contain semantically similar words



Problem with vector space model

Synonym

- Different terms may have identical or similar meanings (weaker: words indicating the same topic).
- No associations between words are made in the vector space representation.

Term-document matrices are very large

- But the number of topics that people talk about is small (in some sense)
- Clothes, movies, politics, ...
- Can we represent the term-document space by a lower dimensional latent space?

Overview of LSI (Latent Semantic Indexing)

- Decompose the term-document matrix into a product of matrices.
- Use Singular value decomposition (SVD).
- $A = U\Sigma V^T$ (where A = term-document matrix)
- We will then use the SVD to compute a new, improved term-document matrix A'.
- We'll get better similarity values out of A' (compared to A).
- Using SVD for this purpose is called latent semantic indexing or LSI.

Vector and Matrix Notations

Length

If $v = (v_1, v_2, \ldots, v_n)$ is a vector, its length $|v|$ is defined as

$$|v| = \sqrt{\sum_{1}^{n} v_i^2} \tag{1}$$

Inner product

Given two vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$. The inner product of x and y is

$$x \cdot y = \sum_{1}^{n} x_i y_i \tag{2}$$

Orthonormal Vectors

Orthogonal vector

- Two vectors are orthogonal to each other if their inner product equals zero.
- In two dimensional space, this is equivalent to saying that the vectors are perpendicular, or that the only angle between them is a 90 angle.

For example, the vectors [2, 1,-2, 4] and [3,-6, 4, 2] are orthogonal because

$$(2, 1, -2, 4) \cdot (3, -6, 4, 2) = 2 \times 3 + 1 \times (-6) - 2 \times 4 + 4 \times 2 = 0 \tag{3}$$

Normal vector

A normal vector (or unit vector) is a vector of length 1.
Any vector with an initial length greater than 0 can be normalized by dividing each component in it by the vector's length.

Orthonormal Vectors

Orthogonal and normal vectors

Identity Matrix I

The identity matrix (denoted as I) is a square matrix with entries on the diagonal equal to 1 and all other entries equal zero.

A matrix A is orthogonal if $AA^T = A^T A = I$.

Eigenvalues and Eigenvectors

Given an $m \times m$ square matrix S. An eigenvector v is a nonzero vector that satisfes the equation

$$Sv = \lambda v \tag{4}$$

λ is a scalar, called an eigenvalue.
$S = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$

$$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 7 \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{5}$$

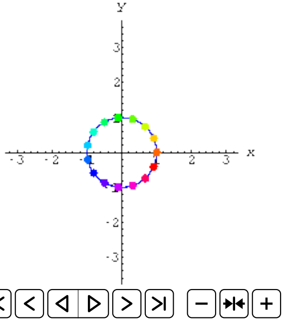
Normalized solution:

$$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix} = 7 \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix} \tag{6}$$

- Intuition: v is unchanged by S (except for scaling)
- Example: stationary distribution of a Markov chain
- There are at most m distinct solutions. can be complex even though S is real.

Matrix action

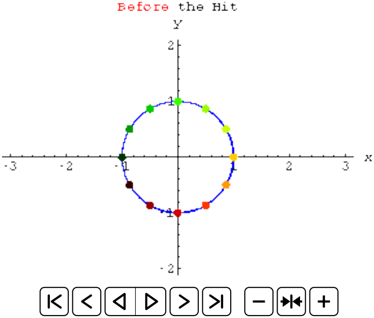
- The plot shows some color coded points on the unit circle.
- Move your mouse over the plot to see what happens when the matrix hits the points on the unit circle.
- You can see that the matrix stretches and rotates the unit circle–that's Matrix Action.
  $M = \begin{pmatrix} 1 & 3 \\ -3 & 2 \end{pmatrix}$
- e.g., $M\binom{1}{0} = \binom{1}{-3}$



(view the animation using PDFReader.)

Matrix action-Strecher

- When you look at that action you can see why it's natural to call a diagonal matrix a "stretcher" matrix.
- The diagonal matrix stretches in the x direction by a factor of "a" and in the y direction by a factor of "b".
- You can verify this by hand using the column way to multiply a matrix times a vector:
- $M = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$
- e.g., $M\binom{1}{0} = \binom{3}{0}$



Another example for eigenvectors

$$S = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

its eigenvalues are 30, 20, 1, and the corresponding eigenvectors are

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$Sv_1 = \lambda_1 v_1$$
$$Sv_2 = \lambda_2 v_2$$
$$Sv_3 = \lambda_3 v_3$$

$$S(v_1, v_2, v_3) = (\lambda_1 v_1, \lambda_2 v_2, \lambda_3 v_3) \qquad = (v_1, v_2, v_3) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

$$SU = U\Sigma$$
$$S = U\Sigma U^{-1}$$

Find solutions for eigenvalues and eigen vectors

$$Sv = \lambda v \tag{7}$$

$$(S - \lambda I)v = 0 \tag{8}$$

$$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} - \lambda I = \begin{pmatrix} 3-\lambda & 2 \\ 2 & 6-\lambda \end{pmatrix} \tag{9}$$

The determinant is set 0, i.e.,

$$(3-\lambda)(6-\lambda) - 4 = 0$$

There are two solutions:

$$\lambda = 7, \lambda = 2$$

For the principal eigenvalue 7, the corresponding eigenvector can be solved from

$$\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 7 \begin{pmatrix} x \\ y \end{pmatrix} \tag{10}$$

The normalized principal eigenvector is $\begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}$ The eigenvector corresponding to eigenvalue 2 is $\begin{pmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{pmatrix}$.

Matrix of eigenvectors

For an $n \times n$ matrix, there are n eigenvectors. Let U is an $n \times n$ matrix consists of n column eigenvectors. U is orthonormal, i.e.,

$$UU^T = U^T U = I$$

For example,

$$U = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix}.$$

$$UU^T = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix} \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Finding eigen pairs by power iteration

- Recall the 'power method' for PageRank–calculate the principal eigenvector for stochastic matrix.
- the method can be extended for other eigenvectors
- when the principal eigenvector is obtained, modify the matrix so that in the new matrix, the principle eigenvector is the second eigenvector

More details can be found in MMD.

$$M = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$$

Iteration 1: $\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$

Iteration 2: $\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 0.530 \\ 0.848 \end{pmatrix} = \begin{pmatrix} 3.286 \\ 6.148 \end{pmatrix}$

Iteration 3: $\begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 0.471 \\ 0.882 \end{pmatrix} = \begin{pmatrix} ... \\ ... \end{pmatrix}$

Eigenvalues and eigenvectors for column stochastic matrix

Recall the PageRank algorithm $Mr = r$.

$$\begin{pmatrix} .5 & .2 \\ .5 & .8 \end{pmatrix} - \lambda I = \begin{pmatrix} .5-\lambda & .2 \\ .5 & .8-\lambda \end{pmatrix}$$

The determinant is set 0, i.e.,

$$(.5-\lambda)(.8-\lambda) - .1 = 0$$

There are two solutions:

$$\lambda = 1, \lambda = 0.3$$

For the principal eigenvalue 1, the corresponding eigenvector can be solved from

$$\begin{pmatrix} 0.5 & 0.2 \\ 0.5 & 0.8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{11}$$

The normalized principal eigenvector is $\begin{pmatrix} 2/\sqrt{29} \\ 5/\sqrt{29} \end{pmatrix}$

Singular value decomposition

Given an $m \times n$ matrix A,

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

Example:

- U: document-to-concept similarity matrix
- V: term-to-concept similarity matrix
- Σ: its diagonal elements: "strength" of each concept

Eigen diagonal decomposition

Matrix diagonalization theorem

Let S be a square $(n \times n)$ real-valued square matrix with n linearly independent eigenvectors there is an eigen decomposition

$$S = U \Sigma U^{-1} \tag{12}$$

- Columns of U are the eigenvectors of S;
- Diagonal elements of Σ are eigenvalues of S

$$\begin{aligned} S\left(v_1, v_2, v_3\right) = \left(\lambda_1 v_1, \lambda_2 v_2, \lambda_3 v_3\right) \qquad &= \left(v_1, v_2, v_3\right) \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \\ SU &= U\Sigma \\ S &= U \Sigma U^{-1} \end{aligned} \tag{13}$$

Symmetric Eigen Decomposition

Perron-Frobenius Theorem

If S is symmetric non-negative matrix, there is an unique eigen decomposition

$$S = U\Sigma U^T \tag{14}$$

- U is orthogonal;
- eigenvalues are non-negative;
- $U^{-1} = U^T$
- Columns of U are normalized eigenvectors
- Columns are orthogonal.
- (everything is real number)

Singular Value Decomposition

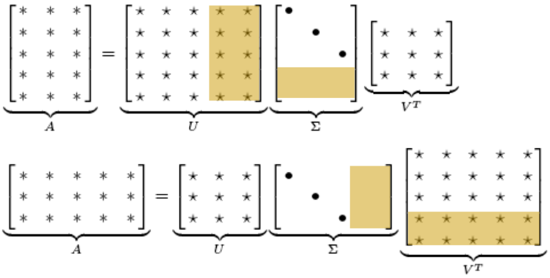
THEOREM [Press+92]

Always possible to decompose matrix A into

$$A = U \Sigma V^T, \tag{15}$$

where

- $A : m \times n; U : m \times r; \Sigma : r \times r; V : n \times r;$
- U, V: column orthonormal (ie., columns are unit vectors, orthogonal to each other)
- $U^T U = I; V^T V = I$ (I: identity matrix)
- Σ: singular are positive, and sorted in decreasing order



SVD Interpretation: documents, terms and concepts

How to compute U, V, Σ?
Note that U and V are orthonormal, so $UU^T = I, VV^T = I$.

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

$$A^T AV = V\Sigma^2 V^T V$$
$$A^T AV = V\Sigma^2$$

V is the eigenvectors of $A^T A$, Σ is the square root of the eigenvalues.
Similarly,

$$AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T = U\Sigma^2 U^T$$
$$(AA^T)U = U\Sigma^2$$

- Why $r \times r$?
- decided by the ‘rank’ r of the matrix A.
- there are r number of non-zero eigenvalues.

Documents, Terms, and Concepts

- Q: if A is the document-to-term matrix, what is $A^T A$?
- term-to-term ([m x m]) similarity matrix
- Q: AA^T ?
- document-to-document ([n x n]) similarity matrix

$$A = \begin{pmatrix} 2 & 0 & 8 & 6 & 0 \\ 1 & 6 & 0 & 1 & 7 \\ 5 & 0 & 7 & 4 & 0 \\ 7 & 0 & 8 & 5 & 0 \\ 0 & 10 & 0 & 0 & 7 \end{pmatrix}$$

$$AA^T = \begin{pmatrix} 104 & 8 & 90 & 108 & 0 \\ 8 & 87 & 9 & 12 & 109 \\ 90 & 9 & 90 & 111 & 0 \\ 108 & 12 & 111 & 138 & 0 \\ 0 & 109 & 0 & 0 & 149 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 79 & 6 & 107 & 68 & 7 \\ 6 & 136 & 0 & 6 & 112 \\ 107 & 0 & 177 & 116 & 0 \\ 68 & 6 & 116 & 78 & 7 \\ 7 & 112 & 0 & 7 & 98 \end{pmatrix}$$

Term-doc Matrix

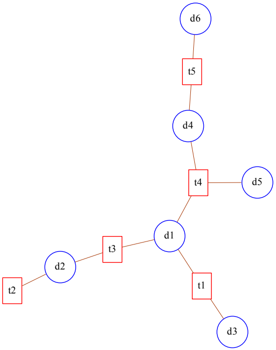
Example from the IR book.

Let C=

	d1	d2	d3	d4	d5	d6
ship	1	0	1	0	0	0
boat	0	1	0	0	0	0
ocean	1	1	0	0	0	0
wood	1	0	0	1	1	0
tree	0	0	0	1	0	1

- This is a standard term-document matrix. We use a non-weighted matrix here to simplify the example.
- Use Matlab to calculate the decomposition

[U, S, V]=svd(C)



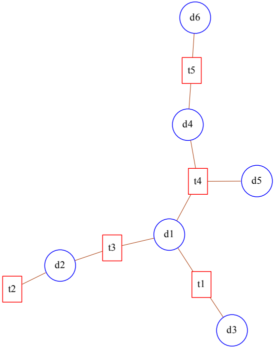
$C * C'$ and $C^T * C$

term-term similarity:

$$C * C^T = \begin{pmatrix} 2 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}$$

doc-doc similarity:

$$C^T * C = \begin{pmatrix} 3 & 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 2 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$



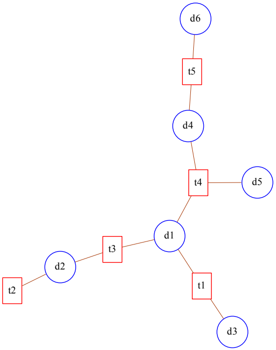
$C = U\Sigma V^T$

$$C = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} 0.44 & -0.29 & -0.56 & 0.57 & -0.24 \\ 0.12 & -0.33 & 0.58 & -0.00 & -0.72 \\ 0.47 & -0.51 & 0.36 & -0.00 & 0.61 \\ 0.70 & 0.35 & -0.15 & -0.57 & -0.15 \\ 0.26 & 0.64 & 0.41 & 0.57 & 0.08 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 2.16 & 0 & 0 & 0 & 0 \\ 0 & 1.59 & 0 & 0 & 0 \\ 0 & 0 & 1.27 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 0 & 0.39 \end{pmatrix}$$

$$V^T = \begin{pmatrix} 0.74 & -0.28 & -0.27 & -0.00 & 0.52 & -0.00 \\ 0.27 & -0.52 & 0.74 & 0.00 & -0.28 & 0.00 \\ 0.20 & -0.18 & -0.44 & 0.57 & -0.62 & 0.00 \\ 0.44 & 0.62 & 0.20 & 0.00 & -0.18 & -0.57 \\ 0.32 & 0.21 & -0.12 & -0.57 & -0.40 & 0.57 \\ 0.12 & 0.40 & 0.32 & 0.57 & 0.21 & 0.57 \end{pmatrix}$$



$C = U\Sigma V^T$

$$U = \begin{pmatrix} 0.4403 & -0.2962 & -0.5695 & 0.5774 & -0.2464 \\ 0.1293 & -0.3315 & 0.5870 & -0.0000 & -0.7272 \\ 0.4755 & -0.5111 & 0.3677 & -0.0000 & 0.6144 \\ 0.7030 & 0.3506 & -0.1549 & -0.5774 & -0.1598 \\ 0.2627 & 0.6467 & 0.4146 & 0.5774 & 0.0866 \end{pmatrix}$$

- one row per term
- one column per 'concept'. at most there are min(M,N) of columns. M is the number of terms and N is the number of documents.
- This is an orthonormal matrix:
  - Row vectors have unit length.
  - Any two distinct row vectors are orthogonal to each other.
- Think of the dimensions as "semantic" dimensions that capture distinct topics like politics, sports, economics.
- Each number u_{ij} in the matrix indicates how strongly related term i is to the topic represented by semantic dimension j .

Σ in $C = U\Sigma V^T$

$$\Sigma = \begin{pmatrix} 2.1625 & 0 & 0 & 0 & 0 \\ 0 & 1.5944 & 0 & 0 & 0 \\ 0 & 0 & 1.2753 & 0 & 0 \\ 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0.3939 \end{pmatrix}$$

- This is a square, diagonal matrix of dimensionality $min(M, N) \times min(M, N)$.
- The diagonal consists of the singular values of C.
- The magnitude of the singular value measures the importance of the corresponding semantic dimension.
- We will make use of this by omitting unimportant dimensions.

V^T in $C = U\Sigma V^T$

$$V^T = \begin{pmatrix} 0.7486 & -0.2865 & -0.2797 & -0.0000 & 0.5285 & -0.0000 \\ 0.2797 & -0.5285 & 0.7486 & 0.0000 & -0.2865 & 0.0000 \\ 0.2036 & -0.1858 & -0.4466 & 0.5774 & -0.6255 & 0.0000 \\ 0.4466 & 0.6255 & 0.2036 & 0.0000 & -0.1858 & -0.5774 \\ 0.3251 & 0.2199 & -0.1215 & -0.5774 & -0.4056 & 0.5774 \\ 0.1215 & 0.4056 & 0.3251 & 0.5774 & 0.2199 & 0.5774 \end{pmatrix}$$

- One column per document,
- one row per min(M,N) where M is the number of terms and N is the number of documents.
- This is an orthonormal matrix: (i) Column vectors have unit length. (ii) Any two distinct column vectors are orthogonal to each other.
- These are again the semantic dimensions from the term matrix U that capture distinct topics like politics, sports, economics.
- Each number v_{ij} in the matrix indicates how strongly related document i is to the topic represented by semantic dimension j .

Summary of LSI

- We have decomposed the term-document matrix C into a product of three matrices.
- The term matrix U: consists of one (row) vector for each term
- The document matrix V^T: consists of one (column) vector for each document
- The singular value matrix Σ: diagonal matrix with singular values, reflecting importance of each dimension
- Next: Why are we doing this?

Dimensionality reduction

- Key property: Each singular value tells us how important its dimension is.
- By setting less important dimensions to zero, we keep the important information, but get rid of the 'details'.
- These details may
  - be noise– in that case, reduced LSI is a better representation because it is less noisy
  - make things dissimilar that should be similar–again reduced LSI is a better representation because it represents similarity better.
- Analogy for "fewer details is better"
  - Image of a bright red flower
  - Image of a black and white flower
  - Omitting color makes is easier to see similarity

Reduce the dimensionality to 2

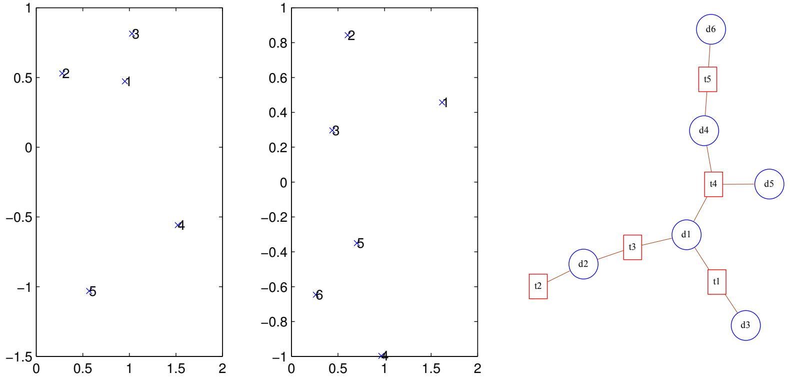
- we only zero out singular values in Σ. This has the effect of setting the corresponding dimensions in U and V T to zero when computing the product $C = U\Sigma V^T$.

U	1	2	3	4	5
ship	−0.44	−0.30	0.00	0.00	0.00
boat	−0.13	−0.33	0.00	0.00	0.00
ocean	−0.48	−0.51	0.00	0.00	0.00
wood	−0.70	0.35	0.00	0.00	0.00
tree	−0.26	0.65	0.00	0.00	0.00

Σ_2	1	2	3	4	5
1	2.16	0.00	0.00	0.00	0.00
2	0.00	1.59	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00

V^T	d_1	d_2	d_3	d_4	d_5	d_6
1	−0.75	−0.28	−0.20	−0.45	−0.33	−0.12
2	−0.29	−0.53	−0.19	0.63	0.22	0.41
3	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00

2D visualization



1	U	0.44	0.30
2		0.13	0.33
3		0.48	0.51
4		0.70	-0.35
5		0.26	-0.65

1	U*S	0.95	0.47
2		0.28	0.53
3		1.03	0.81
4		1.52	-0.56
5		0.57	-1.03

C vs. C2

- We can view C2 as a two-dimensional representation of the matrix.
- We have performed a dimensionality reduction to two dimensions.
- Similarity of d2 and d3 in the original space: 0.
- Similarity of d2 and d3 in the reduced space:

$$0.52 * 0.28 + 0.36 * 0.16 + 0.72 * 0.36 + 0.12 * 0.20 + -0.39 * -0.08 \approx 0.52$$

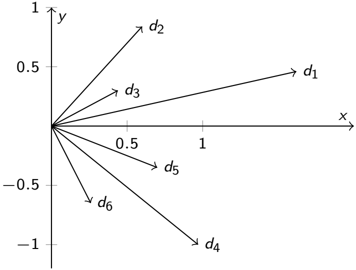
C	d_1	d_2	d_3	d_4	d_5	d_6
ship	1	0	1	0	0	0
boat	0	1	0	0	0	0
ocean	1	1	0	0	0	0
wood	1	0	0	1	1	0
tree	0	0	0	1	0	1

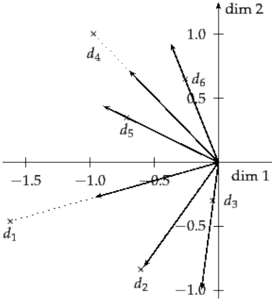
C_2	d_1	d_2	d_3	d_4	d_5	d_6
ship	0.85	0.52	0.28	0.13	0.21	−0.08
boat	0.36	0.36	0.16	−0.20	−0.02	−0.18
ocean	1.01	0.72	0.36	−0.04	0.16	−0.21
wood	0.97	0.12	0.20	1.03	0.62	0.41
tree	0.12	−0.39	−0.08	0.90	0.41	0.49

cosine distance in 2 dimensions

$\Sigma_2 V_2^T =$

$$\begin{matrix} 1.62 & 0.60 & 0.44 & 0.97 & 0.70 & 0.26 \\ 0.46 & 0.84 & 0.30 & -1.00 & -0.35 & -0.65 \end{matrix}$$





plot from IR book

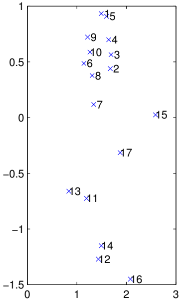
Example for 2D plot

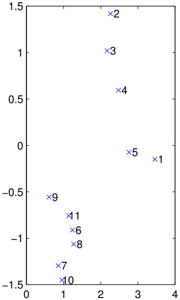
0	1	1	1	1
1	0	1	1	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0
0	0	1	1	1
1	0	0	1	1
1	1	0	0	1
1	1	1	0	0
1	1	0	1	0

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

1	0	0	1	0
1	0	0	0	1
0	0	1	0	0
1	0	0	0	1
1	1	1	1	1
1	0	0	1	1
1	1	1	0	0

0	1	0	0	1	0
0	1	1	0	1	0
0	1	1	0	0	1
1	0	1	0	1	0
1	0	1	0	0	1
1	1	0	1	1	1
1	0	1	1	0	1





$$\lambda_1 = 6.48, \lambda_2 = 3.17, \lambda_3 = 2.83, \ldots$$

Listing 1: Matlab code

```
[u,s,v]=svds(A,2);
dots=(u*s)'
x=dots(1, :);
y=dots(2,:);
m=size(A,2);
n=size(A,1);
subplot(1,2,1);
plot(x,y ,'x')
a=num2str([1:n]');
text(x,y,cellstr(a));
```

Summary

- Standard vector space: Synonyms contribute nothing to document similarity.
- LSI takes documents that are semantically similar, representing them in shorter vectors
- It is proven that word2vec is an implicit matrix factorization.
- Omer Levy and Yoav Goldberg. 2014b. Neural word embedding as implicit matrix factorization. In Advances in neural information processing systems, pages 2177–2185.

SVD and Eigen Decomposition

The singular value decomposition and the eigen decomposition are closely related.

- The left-singular vectors of M are eigenvectors of MM^T.
- The right-singular vectors of M are eigenvectors of $M^T M$.
- The non-zero singular values of M (found on the diagonal entries of Σ) are the square roots of the non-zero eigenvalues of both $M^T M$ and MM^T.

Reading: IIR Chapter 18.