Math Underlying LSA

Promothesh Chatterjee*

Latent Semantic Analysis

We have seen previously that terms are treated as points in high-dimensional space. Spatial closeness between those points represents their semantic association (meaning and relations among terms). LSA creates a vector space representation of the original matrix using singular value decomposition (SVD). SVD is used to identify latent meaning in the documents through dimension reduction based on singular values. Before understanding SVD, we need to understand some related basics.

Eigenvalues and Eigenvectors

The section borrows heavily from Singh 2015 and Savov 2018

Consider a matrix A and a vector u:

$$\mathbf{A} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Multiplying Au, we get

$$\mathbf{Au} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

In other words, we get that $Au = 3u$, more generally $Au = \lambda u$, where A is a square matrix and λ is scalar (vector).

Since A transforms u by scalar multiplication, it only changes the length of the vector (unless $\lambda = \pm 1$). For a non-zero vector u, scalar λ is called eigenvalue and u is called the eigenvector. Think of eigenvalue and eigenvector as tools to understand a matrix. “Eigen” in German refers to “self”. Eigenvalues and eigenvectors help us ‘see’ inside a matrix. Eigenvectors do not change their orientation when multiplied by a matrix, they only shrink or expand.

So the question is how to compute eigenvalue and eigenvector?

$Au = \lambda u$

we can multiply the right hand side of the equation of identity matrix as it makes no difference. Thus, we get

$Au = \lambda I u$

$Au - \lambda I u = O$, where O is zero vector. To solve this equation, we compute determinant of $A - \lambda I$ and equate it to 0 (for an intuitive explanation of determinant see this short video: https://www.youtube.com/watch?v=vvR3JSXO2fo).

For instance, if we were to compute eigenvalue and eigenvector of

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}$$

$A - \lambda I =$

$$\begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} 2-\lambda & 0 \\ 1 & 3-\lambda \end{bmatrix}$$

Computing determinant (we know that that determinant of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is ad-bc)

Thus, det $(A - \lambda I) = 0$

$(2-\lambda)(3-\lambda) = 0$

Thus, we have two values of λ, 2 and 3.

For each eigenvalue, we can compute eigenvector.

For $\lambda = 2$, we can compute u

$(A - \lambda I)u = O$, if

$$\mathbf{u} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\left[\begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\right] u = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

therefore, x + y = 0 or x = -y

$$= \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$$

and similar values are all eigenvectors.

Worth reminding ourselves that eigenvalues and eigenvectors come in pairs, you cannot have one without the other.

Eigen decomposition

Eigen decomposition refers to factorizing a square matrix into three matrices such that the left matrix has columns as eigen-vectors of original matrix, the center matrix is a diagonal matrix that contains eigenvalues of original matrix, and the right matrix is inverse of the left matrix.

Consider

$$\mathbf{A} = \begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix}$$

This square matrix can be eigendecomposed into product of the following three matrices

$$= \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & -1/5 \end{bmatrix}$$

$A = Q \ \ \Lambda \ \ Q^{-1}$ where Q has columns as eigenvectors of A (also called change-of-basis matrix), Λ is a diagonal matrix (a diagonal matrix is a n x n matrix where all the entries to the both sides of the leading diagonal are zero) that contains eigenvalues of A, and Q^{-1} is the inverse of Q matrix.

Please note that if an n x n matrix A is diagonal or triangular (only one side of leading diagonals are zeroes), the values along the leading diagonal are eigenvalues. This is an important property of diagonal matrices (and add the fact that multiplication with a diagonal matrix is a breeze), makes them very important. However it is important to note that an n x n A is diagonalizable if it is similar to a diagonal matrix D.

Therefore matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$$

is diagonalizable because matrix

$$\mathbf{P} = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$$

gives $P^{-1} \ \ A \ \ P = D$ where

$$\mathbf{D} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

is a diagonal matrix (condition for similarity).

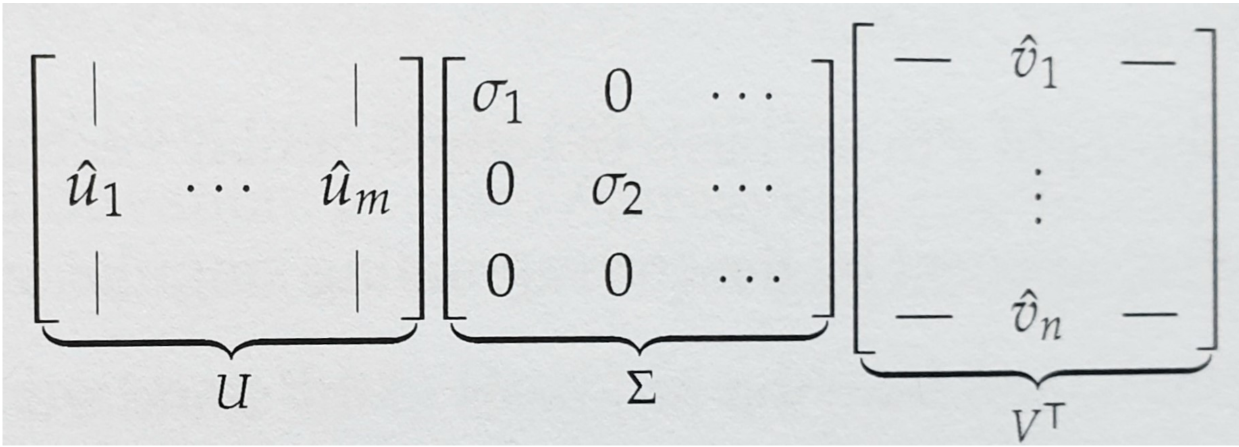
Singular Value Decomposition

The issue with eigen-decomposition is that it needs square matrices. SVD can generalize this to non-square matrices. One trick for turning a non-square matrix (which is the most scenario in data analysis) into square matrix is to multiply the matrix by it's transpose. So if we have $A_{m \times n}$, then $AA^T_{n \times n}$ has same column space as matrix A and $A^T A_{m \times m}$ has same row spaces as the matrix A.

Singular value decomposition breaks a matrix into the product of 3 matrices:

1) U, an $m \times m$ orthogonal matrix (i.e., its inverse is same as its transpose) which consists of left singular vectors.
2) Σ, an $m \times n$ matrix with singular values σ_i on the diagonal,
3) V^T, an $n \times n$ orthogonal matrix of right singular vectors.

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



To find U, Σ, V^T, we perform eigendecomposition on the matrix products AA^T and A^TA.

a) Consider AA^T (a square matrix), so we can perform eigendecomposition giving $U\Lambda U^T$. The eigenvectors of AA^T span the same space as the column space of A. We call these the left singular vectors of A. Therefore, left singular vectors of A (the columns of U) are the eigenvectors of AA^T.

To compute right singular vectors of A (the rows of V^T), we eigendecompose $A^TA = V\Lambda V^T$. We build the orthogonal matrix V^T by stacking the eigenvectors of A^TA as rows.

The eigenvalues of the matrices AA^T and A^TA are same. In both cases, the eigenvalues λ_i correspond to the squares of the singular values of matrix A. $\sigma_i = \sqrt{\lambda_i}$. Thus, $\Sigma_{m\times n}$ contains on its diagonals, the singular values σ_i, which are positive square roots of the eigenvalues λ_i of matrices AA^T or A^TA.

To understand how to solve for SVD, let's take an example.

$$\mathbf{A} = \begin{pmatrix} 4 & 0 & 1 \\ 3 & -5 & 2 \end{pmatrix}$$

First let's compute AA^T

$$\mathbf{A^T} = \begin{pmatrix} 4 & 3 \\ 0 & -5 \\ 1 & 2 \end{pmatrix}$$

$$\mathbf{A.A^T} = \begin{pmatrix} 17 & 14 \\ 14 & 38 \end{pmatrix}$$

First, we find the eigenvector for AA^T

det $(A.A^T - \lambda I) = 0$

$$= \begin{bmatrix} 17-\lambda & 14 \\ 14 & 38-\lambda \end{bmatrix}$$

Solving for λ, we get $\lambda = 10, 45$ (the eigenvalues of AA^T)

Eigenvector for $\lambda = 10$ is:

$$= \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

Eigenvector for $\lambda = 45$ is:

$$= \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$$

For the first eigenvector (0.5,1), Length L = $(\sqrt{0.5^2 + 1^2}) = 1.12$

So, normalizing gives u_1= $(0.5/1.12, 1/1.12) = (0.45, 0.89)$

Similarly, for the second eigenvector (-2, 1), Length L = $(\sqrt{-2^2 + 1^2}) = 2.24$

Again, normalizing gives us $u_2 = (-2/2.24, 1/2.24) = (-0.89, 0.45)$

Next let's compute $A^T A$

$$\mathbf{A^T.A} = \begin{pmatrix} 25 & -15 & 10 \\ -15 & 25 & -10 \\ 10 & -10 & 5 \end{pmatrix}$$

det $(A^T.A - \lambda I) = 0$

$$= \left[\begin{matrix} 25-\lambda & -15 & 10 \\ -15 & 25-\lambda & -10 \\ 10 & -10 & 5-\lambda \end{matrix} \right)$$

Solving for λ, we get $\lambda = 0, 10, 45$ (the eigenvalues of $A^T A$)

Eigenvector for $\lambda = 45$ is:

$$= \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

Eigenvector for $\lambda = 10$ is:

$$= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Eigenvector for $\lambda = 0$ is:

$$= \begin{bmatrix} -0.25 \\ 0.25 \\ 1 \end{bmatrix}$$

For the first eigenvector, (2,-2,1), Length $= (\sqrt{2^2 + -2^2 + 1^2}) = 3$

So, normalizing gives $v_1 = (2/3, -2/3, 1/3) = (0.67, -0.67, 0.33)$

For the second eigenvector, (1,1,0), Length $= (\sqrt{1^2 + 1^2 + 0^2}) = 1.41$

So, normalizing gives $v_2 = (1/1.41, 1/1.41, 0/1.41) = (0.71, 0.71, 0)$

For the third eigenvector, (-0.25,0.25,1), Length $= (\sqrt{-0.25^2 + 0.25^2 + 1^2}) = 1.06$

So, normalizing gives $v_3 = (-0.25/1.06, 0.25/1.06, 1/1.06) = (-0.24, 0.24, 0.94)$

The three matrices Σ, U, V

$$= \begin{pmatrix} \sqrt{45} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} u_1, u_2 \end{pmatrix} = \begin{pmatrix} 0.45 & -0.89 \\ 0.89 & 0.45 \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} v_1, v_2, v_3 \end{pmatrix} = \begin{pmatrix} 0.67 & 0.71 & -0.24 \\ -0.67 & 0.71 & 0.24 \\ 0.33 & 0 & 0.94 \end{pmatrix}$$

You can verify the calculations by multiplying U, Σ, V^T, the result should be original matrix.