Linear algebra: SVD

Statistical Natural Language Processing 1

Çağrı Çöltekin

University of Tübingen
Seminar für Sprachwissenschaft

Winter Semester 2024/2025

version: 796d7ed @2024-11-21

Quick recap

So far we reviewed:

- Vectors, matrices
- Operations on vectors and matrices: scalar multiplication, addition, dot product, matrix multiplication
- Matrices as operators (linear functions / transformations)
- Linearity and linear combinations
- Solving systems of linear equations, elimination
- Finding matrix inverse
- Linear regression
- Eigenvalues and eigenvectors

Today's plan

- Singular value decomposition
- Pseudo inverse

Orthogonal matrices

A short detour

- An orthogonal matrix is a square matrix whose columns (and rows) are orthogonal unit vectors
- Some interesting properties:
  - The product of two orthogonal matrices is another orthogonal matrix
  - Orthogonal matrices are invertible
  - Product of an orthogonal matrix with its transpose is the identity matrix

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$$
$$\Rightarrow \mathbf{Q}^T = \mathbf{Q}^{-1}$$

  - Orthogonal matrices represent length-preserving transformations (rotations and reflections)
  - Determinants of an orthogonal matrix is 1 or -1

Singular Value Decomposition

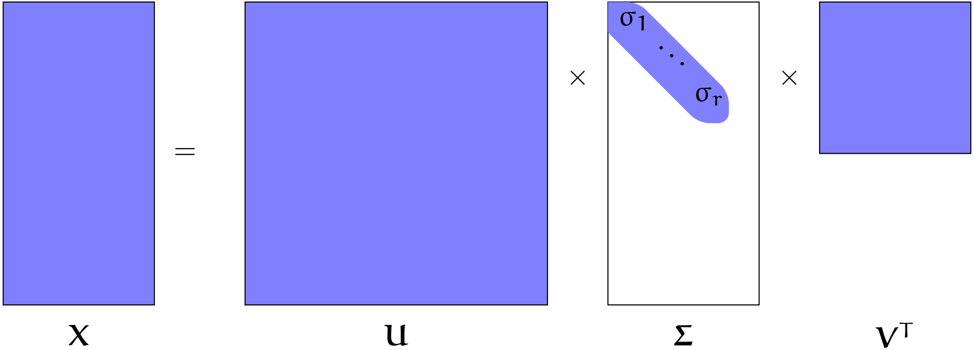
- Singular value decomposition (SVD) of an $n \times m$ matrix $\mathbf{X}$ is

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

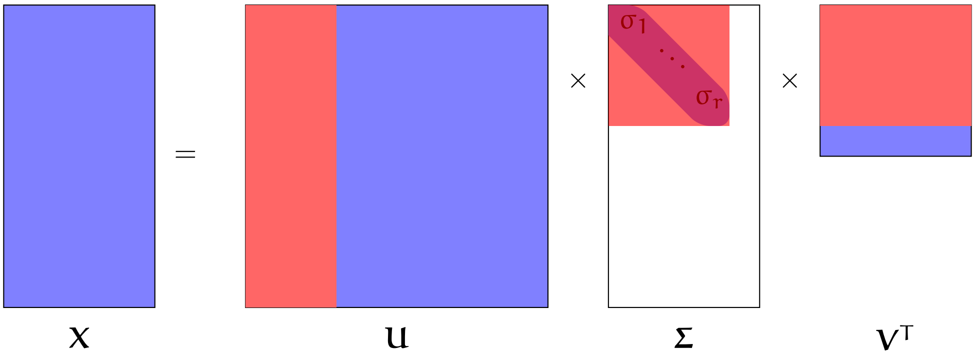
$\mathbf{U}$ is a $n \times n$ orthogonal matrix
$\mathbf{\Sigma}$ is a $n \times m$ diagonal matrix of singular values
$\mathbf{V}^T$ is a $m \times m$ orthogonal matrix.

- Singular vectors in $\mathbf{U}$ are the eigenvalues of $\mathbf{X}\mathbf{X}^T$
- Singular vectors in $\mathbf{V}^T$ are the eigenvalues of $\mathbf{X}^T\mathbf{X}$

Singular Value Decomposition



Singular Value Decomposition



- Since $n - r$ rows and $m - r$ rows of $\mathbf{\Sigma}$ is 0, the decomposition does need the full matrices

Singular value decomposition and $X^T X$

- Assume $X = U \Sigma V^T$

$$
\begin{aligned}
X^T X &= (U \Sigma V^T)^T U \Sigma V^T \\
&= V \Sigma^T U^T U \Sigma V^T \\
&= V \Sigma^T \Sigma V^T \\
&= V \Sigma^2 V^T
\end{aligned}
$$

Singular value decomposition and $X^T X$

- Assume $X = U \Sigma V^T$

$$
\begin{aligned}
X^T X &= (U \Sigma V^T)^T U \Sigma V^T \\
&= V \Sigma^T U^T U \Sigma V^T \\
&= V \Sigma^T \Sigma V^T \\
&= V \Sigma^2 V^T
\end{aligned}
$$

- Columns of V are eigenvectors of $X^T X$

Singular value decomposition and $X^T X$

- Assume $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$

$$
\begin{aligned}
\mathbf{X}^T\mathbf{X} &= (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^T \\
&= \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T
\end{aligned}
$$

- Columns of $\mathbf{V}$ are eigenvectors of $\mathbf{X}^T\mathbf{X}$
- Values in the diagonal matrix $\mathbf{\Sigma}^2$ are the eigenvalues of $\mathbf{X}^T\mathbf{X}$

Singular value decomposition and XX^T

- Assume $X = U\Sigma V^T$

$$
\begin{aligned}
XX^T &= U\Sigma V^T (U\Sigma V^T)^T \\
&= U\Sigma V^T V \Sigma^T U^T \\
&= U\Sigma\Sigma^T U^T \\
&= U\Sigma^2 U^T
\end{aligned}
$$

- Columns of U are eigenvectors of XX^T
- Values in the diagonal matrix Σ^2 are the eigenvalues of XX^T
- X^TX and X^TX share the eigenvalues

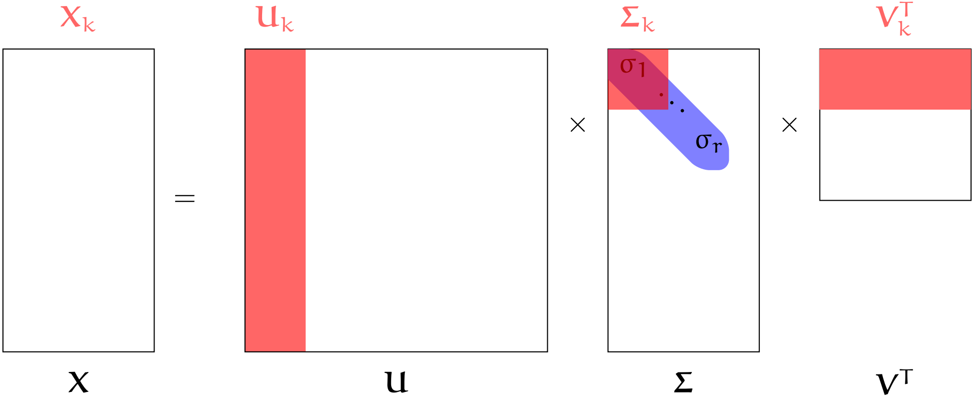
Computing SVD

- Find the eigenvalues and eigenvectors of $\mathbf{X}^T\mathbf{X}$
  - $\mathbf{X}^T\mathbf{X}$ is symmetric (semi) definite, the eigenvectors can be chosen to be orthogonal unit vectors, the eigenvalues are positive
  - $\mathbf{V}$ is the collection of the eigenvectors (of $\mathbf{X}^T\mathbf{X}$)
  - $\sigma_i = \sqrt{\lambda_i}$
- Knowing $\mathbf{V}$ and $\boldsymbol{\Sigma}$,

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$
$$\mathbf{X}\mathbf{V} = \mathbf{U}\boldsymbol{\Sigma}$$
$$\mathbf{X}\mathbf{V}\boldsymbol{\Sigma}^{-1} = \mathbf{U}$$

- In practice there are more efficient ways to compute SVD

Low rank estimation of a matrix



$X_k = U_k \Sigma_k V_k^T$ is the best rank k estimation of matrix X

SVD: properties and applications

- Singular values are related to matrix norms
- SVD has a wide range of applications from image compression to document indexing to semantics of the words
- It is also a method for dimensionality reduction for visualizations
- A large number of statistical methods also rely on SVD (e.g., PCA, we will discuss later)
- The *condition number* of a matrix, an indication of numerical stability, depends on singular values
- SVD can be computed with good numerical accuracy, as a result it is also used for computing other quantities (e.g., matrix inverse)

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ acts as a *left inverse*

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ acts as a *left inverse*
- Similarly we can define *right inverse* as $\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}$ acts as a *left inverse*
- Similarly we can define *right inverse* as $\mathbf{X}^{\mathsf{T}}(\mathbf{X}\mathbf{X}^{\mathsf{T}})^{-1}$
- Remember, however, the existence of $(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}$ requires columns of $\mathbf{X}$ to be independent

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ acts as a *left inverse*
- Similarly we can define *right inverse* as $\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$
- Remember, however, the existence of $(\mathbf{X}^T\mathbf{X})^{-1}$ requires columns of $\mathbf{X}$ to be independent
- A more general solution, *pseudo inverse*, $\mathbf{X}^+$ falls out of SVD

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}$ acts as a *left inverse*

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ acts as a *left inverse*
- Similarly we can define *right inverse* as $\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ acts as a *left inverse*
- Similarly we can define *right inverse* as $\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$
- Remember, however, the existence of $(\mathbf{X}^T\mathbf{X})^{-1}$ requires columns of $\mathbf{X}$ to be independent

Left and right inverses

- For a non-square matrix, or a square matrix with rank lower than n, the inverse is not defined
- From linear regression, we know that $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ acts as a *left inverse*
- Similarly we can define *right inverse* as $\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$
- Remember, however, the existence of $(\mathbf{X}^T\mathbf{X})^{-1}$ requires columns of $\mathbf{X}$ to be independent
- A more general solution falls out of SVD

Computing pseudo inverse

- We want matrix multiplication to get as close to $\mathbf{I}$ as possible. Consider the 3×4 diagonal matrix:

$$\times \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} =$$

Computing pseudo inverse

- We want matrix multiplication to get as close to $\mathbf{I}$ as possible. Consider the 3×4 diagonal matrix:

$$\times \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Computing pseudo inverse

- We want matrix multiplication to get as close to $\mathbf{I}$ as possible. Consider the 3×4 diagonal matrix:

$$\begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Computing pseudo inverse

- We want matrix multiplication to get as close to $\mathbf{I}$ as possible. Consider the 3×4 diagonal matrix:

$$\begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- For an $n \times n$ diagonal matrix $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{+} = \boldsymbol{\Sigma}^{-1}$

Computing pseudo inverse

- We want matrix multiplication to get as close to $\mathbf{I}$ as possible. Consider the 3×4 diagonal matrix:

$$\begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- For an $n \times n$ diagonal matrix $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{+} = \mathbf{\Sigma}^{-1}$
- For any invertible $n \times n$ matrix $\mathbf{X}$, $\mathbf{X}^{+} = \mathbf{X}^{-1}$

Computing pseudo inverse

- We want matrix multiplication to get as close to $\mathbf{I}$ as possible. Consider the 3×4 diagonal matrix:

$$\begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- For an $n \times n$ diagonal matrix $\mathbf{\Sigma}$, $\mathbf{\Sigma}^{+} = \mathbf{\Sigma}^{-1}$
- For any invertible $n \times n$ matrix $\mathbf{X}$, $\mathbf{X}^{+} = \mathbf{X}^{-1}$
- In general, if we use singular value decomposition $\mathbf{X}^{+} = \mathbf{V}\mathbf{\Sigma}^{+}\mathbf{U}^{T}$

Summary / next

- We reviewed SVD and pseudo inverse
- SVD is a very important method. We will return to it multiple times during the course

Summary / next

- We reviewed SVD and pseudo inverse
- SVD is a very important method. We will return to it multiple times during the course

Next:

- A very short introduction to calculus

the SVD song

Further reading

Any of the linear algebra references provided earlier.