SI231b Review: SVD related topic

Mao Song

ShanghaiTech University

maosong@shanghaitech.edu.cn

December 10, 2020

Overview

Definition

Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the Singular Value Decomposition (SVD) of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{1}$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ are orthonormal matrices. $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ is a diagonal matrix, and $\Sigma_{ii} = \sigma_i(\mathbf{A})$ with $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_{\min\{m,n\}}(\mathbf{A}) \geq 0$. This type of SVD is also called Full SVD.

Relatively, we have Thin SVD or Truncated SVD of $\mathbf{A}$ (Suppose that $\mathrm{rank}(\mathbf{A}) = r$), which is given by (more economical!)

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

where $\mathbf{U} \in \mathbb{R}^{m\times r}$ and $\mathbf{V} \in \mathbb{R}^{n\times r}$ are semi-orthogonal matrices ($\mathbf{U}^T\mathbf{U} = \mathbf{I}_r$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_r$). $\mathbf{\Sigma} \in \mathbb{R}^{r\times r}$ is a diagonal matrix, and $\Sigma_{ii} = \sigma_i(\mathbf{A})$ with $\sigma_1(\mathbf{A}) \geq \cdots \geq \sigma_r(\mathbf{A}) > 0$.

Basic Properties

- The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called the **left-singular vectors** and **right-singular vectors** of $\mathbf{A}$, respectively.
- From the SVD point of view, $x \to \mathbf{A}x$ can be seen as : rotating x ($\mathbf{V}$), scaling($\mathbf{\Sigma}$), and rotating again($\mathbf{U}$). *Thus, the singular values can be interpreted as the magnitude of the semiaxes of an ellipse.*
- SVD of a matrix always exists.
- The outer product form: $\mathbf{A} = \sum_{i=1}^{r} \sigma_i(\mathbf{A})\mathbf{u}_i\mathbf{v}_i^T$.
- The SVD of a matrix is not unique. ($\mathbf{\Sigma}$ is unique if Σ_{ii} sorted in descending order, but $\mathbf{U}$ and $\mathbf{V}$ are not unique).
- When we talk about SVD, we often assume its singular values are non-negative and sorted in descending order!

Basic Idea

Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathrm{rank}(\mathbf{A}) = r$.
The eigen-decompositions of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ are given by

$$\mathbf{U}^T\mathbf{A}\mathbf{A}^T\mathbf{U} = \mathbf{diag}(\sigma_1^2(\mathbf{A}), \ldots, \sigma_r^2(\mathbf{A}), \underbrace{0, \ldots, 0}_{m-r})$$

$$\mathbf{V}^T\mathbf{A}^T\mathbf{A}\mathbf{V} = \mathbf{diag}(\sigma_1^2(\mathbf{A}), \ldots, \sigma_r^2(\mathbf{A}), \underbrace{0, \ldots, 0}_{n-r})$$

where $\mathbf{U}$ and $\mathbf{V}$ are eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ respectively.

How to compute SVD by hand

To compute the SVD of $\mathbf{A} \in \mathbb{R}^{m\times n}$ by hand, excute the following steps:

1. Calculate $\mathbf{A}\mathbf{A}^T$.
2. Compute the eigen-decomposition of $\mathbf{A}\mathbf{A}^T$

$$\begin{bmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{bmatrix} \mathbf{A}\mathbf{A}^T \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{\Delta}_r^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{3}$$

   where $\mathbf{U}_1 \in \mathbb{R}^{m\times r}$ is the matrix of eigenvectors corresponding to $\mathbf{\Delta}_r^2$.
3. Let $\mathbf{V}_1 = \mathbf{A}^T\mathbf{U}_1\mathbf{\Delta}_r^{-1}$, and find $\mathbf{V}_2 \in \mathbb{R}^{n\times(n-r)}$ such that $[\mathbf{V}_1, \mathbf{V}_2]$ is an orthonormal matrix. (Eigenvectors of $\mathbf{A}^T\mathbf{A}$ corresponding to eigenvalue 0)
4. Extend $\mathbf{\Delta}_r$ to $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$. Finally, the Full SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \in \mathbb{R}^{m\times m}, \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \in \mathbb{R}^{n\times n}, \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Delta}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{m\times n} \tag{4}$$

Example

Find the SVD of $\mathbf{A}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} \tag{5}$$

Solution Step 1

$\mathbf{A}\mathbf{A}^T$ is given by

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 2 \end{bmatrix} \tag{6}$$

Example

Solution Step 2

$\det(\lambda \mathbf{I} - \mathbf{A}\mathbf{A}^T) = \lambda^2(\lambda - 4)$. thus the eigenvalues of $\mathbf{A}\mathbf{A}^T$ are $\lambda_1 = 4$, $\lambda_2 = \lambda_3 = 0$, and $\sigma_1(\mathbf{A}) = \sqrt{4} = 2$, $\mathbf{\Delta}_1 = [2]$ $(r = 1)$.
The unit eigenvector of $\mathbf{A}\mathbf{A}^T$ corresponding to $\lambda_1 = 4$ is given by $\mathbf{u}_1 = \left[1/\sqrt{2}, 0, 1/\sqrt{2}\right]^T$, $\mathbf{U}_1 = \mathbf{u}_1 = \left[1/\sqrt{2}, 0, 1/\sqrt{2}\right]^T$.

Solution Step 3

$\mathbf{V}_1 = \mathbf{A}^T \mathbf{U}_1 \mathbf{\Delta}_1^{-1} = \left[\begin{smallmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{smallmatrix}\right]$.
Thus the Thin SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}_1 \mathbf{\Delta}_1 \mathbf{V}_1^T$

Example

Solution step 3

Solve the system $\mathbf{A}\mathbf{A}^T\mathbf{v} = 0\mathbf{v}$ to obtain the eigenvectors $\mathbf{U}_2 \in \mathbb{R}^{3\times 2}$ corresponding to 0. Then

1. use the Gram-Schmidt process to orthogonalize $\mathbf{U}_2$.
2. normalize the column vectors of $\mathbf{U}_2$.

we have

$$\mathbf{U}_2 = \begin{bmatrix} -1/\sqrt{2} & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 0 \end{bmatrix} \tag{7}$$

then we solve the system $\mathbf{A}^T\mathbf{A}\mathbf{v} = 0\mathbf{v}$ to obtain the eigenvectors $\mathbf{V}_2 \in \mathbb{R}^{2\times 1}$ corresponding to 0. we have $\mathbf{V}_2 = [\,-1/\sqrt{2},\, 1/\sqrt{2}\,]^T$

Example

Solution Step 4

Thus, the Full SVD of $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \mathbf{\Sigma} \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}^T \tag{8}$$

$$= \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^T \tag{9}$$

Note

1. Do not forget the normalization.
2. Corresponds the eigenvector and the eigenvalue.

Pseudo inverse

Definition

Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the **pseudo inverse** of $\mathbf{A}$, $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ is a matrix satisfying the *Penrose-Moore equations*:

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}, \qquad \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$$

$$(\mathbf{A}\mathbf{A}^\dagger)^T = \mathbf{A}\mathbf{A}^\dagger, \qquad (\mathbf{A}^\dagger\mathbf{A})^T = \mathbf{A}^\dagger\mathbf{A}$$

Pseudo inverse & SVD

If SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, then

$$\mathbf{A}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^T \tag{10}$$

where $\mathbf{\Sigma}^\dagger \in \mathbb{R}^{n\times m}$ is diagonal matrix such that

$$\Sigma^\dagger_{ii} = \begin{cases} \frac{1}{\Sigma_{ii}}, & \text{if } \Sigma_{ii} > 0 \\ 0,, & \text{else} \end{cases} \tag{11}$$

Pseudo inverse & linear system

Solution via pseudo inverse

For a linear system $\mathbf{Ax} = \mathbf{b}$, $\mathbf{x} = \mathbf{A}^\dagger \mathbf{b}$ is always a solution regardless of $\mathbf{A}$ is full rank or not.
When $\mathbf{A}$ is invertible, we have $\mathbf{A}^\dagger = \mathbf{A}^{-1}$.

Theorem

If $\mathrm{rank}[\mathbf{A}, \mathbf{b}] \neq \mathrm{rank}(\mathbf{A})$, *that is,* $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, *then* $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{b}$ *is the solution to the least square problem*

$$\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2 \tag{12}$$

moreover, $\mathbf{x}^*$ *is of minimum 2-norm among all solutions, i.e., if* $\mathbf{u}$ *is also a solution to problem (12), then*

$$\|\mathbf{u}\|_2 \geq \|\mathbf{x}^*\|_2 \tag{13}$$

See HW3, Problem 6 for proof.

Frobenius norm

Frobenius norm is an extension of vector 2-norm. If we vectorize a matrix into a vector (along column or row), the result will become clear.
Thus Frobenius norm is often used to measure the "distance" between two matrices.

$$\|\mathbf{A}\|_F = \sum_{i=1}^{n}\sum_{i=1}^{m} A_{ij}^2 = \sqrt{\mathrm{Trace}(\mathbf{A}^T\mathbf{A})} \tag{14}$$

Property

$$\|\mathbf{A}\|_F = \sqrt{\sum_i \sigma_i^2(\mathbf{A})} \tag{15}$$

The proof is clear if we use the SVD of **A**, and note that **U** and **V** are both orthonormal.

Matrix 2 norm

The matrix 2-norm of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is given by

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x}\neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 \tag{16}$$

Block 3

Suppose SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, then $\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A})$

Proof.

$$\max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{Q}^T\mathbf{D}\mathbf{Q}\mathbf{x} \tag{17}$$

$$= \max_{\|\mathbf{y}\|_2=1} \|\mathbf{D}\mathbf{y}\|_2^2 \leq \lambda_{\max}(\mathbf{A}^T\mathbf{A}) \tag{18}$$

where $\mathbf{Q}^T\mathbf{D}\mathbf{Q}$ is the eigen-decomposition of $\mathbf{A}^T\mathbf{A}$ with $\mathbf{Q}$ orthonormal. By SVD of $\mathbf{A}$, we have $\lambda_{\max}(\mathbf{A}^T\mathbf{A}) = \sigma_1^2(\mathbf{A})$. On the other hand, take $\mathbf{x} = \mathbf{v}_1$, we have $\max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 \geq \sigma_1(\mathbf{A})$. □

Condition number

Recall the definition of condition number:

$$\kappa = \|\mathbf{A}\|\|\mathbf{A}^\dagger\| \tag{19}$$

If we use matrix 2-norm, then the condition number is given by

$$\kappa_2 = \frac{\sigma_1(\mathbf{A})}{\sigma_r(\mathbf{A})} \tag{20}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(\mathbf{A}) = r$.

Projection onto the four subspaces

Suppose that SVD of $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} \tag{21}$$

then we have

- $\mathbf{V}_1\mathbf{V}_1^T$ is the orthogonal projection onto $\mathcal{R}(\mathbf{A}^T)$.
- $\mathbf{V}_2\mathbf{V}_2^T$ is the orthogonal projection onto $\mathcal{N}(\mathbf{A})$.
- $\mathbf{U}_1\mathbf{U}_1^T$ is the orthogonal projection onto $\mathcal{R}(\mathbf{A})$.
- $\mathbf{U}_2\mathbf{U}_2^T$ is the orthogonal projection onto $\mathcal{N}(\mathbf{A}^T)$.

Remark

1. Do not misunderstand the relation of four subspaces: $\mathcal{R}(\mathbf{A}^T) \perp \mathcal{N}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}^T)$.
2. We can also write the above result in the form of pseudo inverse, see Pseudo Projection.

Numerical rank

Though the rank function has a closed form, it is not easy to calculate in the presence of errors in the matrix elements.
Rounding errors and fuzzy data make the rank determination a nontrivial exercise. That is why numerical rank comes out. The numerical rank is useful, especially when the matrix is ill-conditioned. See Example.
The numerical rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as:

$$\text{numerical rank}(\mathbf{A}) = \#\{\sigma_i(A) |\ \sigma_i(A) \geq \epsilon\} \tag{22}$$

where ϵ often depends on the machine precision.

Low Rank Matrix Completion

Low Rank Matrix Completion is a popular topic in recent years, aiming to find a lower rank estimation of a given matrix with partially observed entries. Mathematically, given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\Omega \subseteq [m] \times [n]$, we aim to solve the problem

$$\begin{aligned} \min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \quad & \frac{1}{2} \|\mathcal{P}_\Omega(\mathbf{X}) - \mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \operatorname{rank}(\mathbf{X}) \leq d \end{aligned}$$

where $\mathcal{P}_\Omega$ is the projection such that $\mathcal{P}_\Omega(\mathbf{X}_{ij}) = \mathbf{X}_{ij}$ if $(i,j) \in \Omega$ and $\mathcal{P}_\Omega(\mathbf{X}_{ij}) = 0$ if $(i,j) \notin \Omega$.
A popular method to solve the LRMC problem is to use the Nuclear norm to replace the lower rank constraints, the nuclear norm is defined as $\|\mathbf{X}\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(\mathbf{X})$. You may refer to the famous paper [2] written by Candes for more details.

Spectral functions*

The spectral function reveals the relation of a special class of functions $f : \mathbb{R}^{m\times n} \to (-\infty, \infty]$ to the class of functions $g : \mathbb{R}^{\min\{m,n\}} \to (-\infty, \infty]$ via singular values. For example,

$$\|\mathbf{A}\|_F = \|\sigma(\mathbf{A})\|_2, \|\mathbf{A}\|_* = \|\sigma(\mathbf{A})\|_1, \|\mathbf{A}\|_2 = \|\sigma(\mathbf{A})\|_\infty \tag{23}$$

Thus we can reduce the complexity of the problem.
If you are working with optimization problem on matrices, you may refer to [1] Chapter 7 for more details.

Principal Component Analysis

SVD is also a popular tool for PCA, which is powerful for Dimension Reduction in machine learning and De-noising in signal processing, many packages in machine learning have implemented the PCA model via SVD. You may refer to [3] for more details or select the course SI232 Subspace Learning to investigate the theory behind the SVD.

Reference

Amir Beck.
First Oder Methods in Optimization.
2010.

Emmanuel J Candes and Benjamin Recht.
Exact low-rank matrix completion via convex optimization.
In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 806–812. IEEE, 2008.

René Vidal and Yi Ma.
Generalized Principal Component Analysis.
2019.

The End