Singular Value Decomposition
SVD

JrPhy

Introduction

- SVD always appears in the last part of a Linear Algebra book, the reason is SVD includes the key ideas in Linear Algebra.
- Not only in teaching, but in many applications, like computer vision, statistic, etc.
- I’ll prove this theorem, but it bases on many theorem, those won’t be proved in this ppt.

- Definition:
- Let $\boldsymbol{A}$ in $M_{m \times n}(\mathbb{C})$. The unique scalars σ_1 , σ_2 ,… , σ_n in SVD are called the singular values of A.

- Singular Value Decomposition Theorem:
- Let $\boldsymbol{A}$ is $M_{m \times n}(\mathbb{C})$ and rank($\boldsymbol{A}$) = r with positive scalar singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. $\boldsymbol{A}$ factorization
   - $\boldsymbol{A} = \boldsymbol{U \Sigma V^*}$
- is called a singular value decomposition of $\boldsymbol{A}$. Where $\boldsymbol{U}$ is $M_{m \times m}(\mathbb{C})$ and V is $M_{n \times n}(\mathbb{C})$ are unitary matrices and $\boldsymbol{\Sigma}$ is $M_{m \times n}(\mathbb{C})$ and $\boldsymbol{\Sigma}_{ij} = \sigma_i \delta_{ij}$

Prove the theorem of SVD

- $\because$ ***U***, ***V*** are unitary, $\therefore$ ***AV*** = ***USV*V*** = ***UΣ***

  So that there exists 2 orthonormal basis $\beta = \{v_1, v_2, \dots v_n\}$ and $\gamma = \{u_1, u_2, \dots u_m\}$ in finite-dimension inner product space ***V*** and ***U*** and a linear transform T: $F^n \rightarrow F^m$ of rank r, and $\boldsymbol{\Sigma}$ can be written as

  $$\boldsymbol{\Sigma}_{ij} = \sigma_i \delta_{ij},\ (1 \le i \le r),\ \Sigma_{ij} = 0, \text{ otherwise}$$

  Then $T(v_i) = \sigma_i u_i,\ 1 \le i \le r,\ T(v_i) = 0,\ i > r$

Prove the theorem of SVD

- So that we have

$$\langle u_i, u_j \rangle = \left\langle \frac{1}{\sigma_i} T(v_i), \frac{1}{\sigma_j} T(v_j) \right\rangle = \frac{1}{\sigma_i \sigma_j} \langle T^*T(v_i), v_j \rangle$$

$$= \frac{\sigma_i^{\ 2}}{\sigma_i \sigma_j} \langle v_i, v_j \rangle = \delta_{ij}$$

- So σ_{ij}^2 is the eigenvalue of T^*T and v_i is the corresponding eigenvector.

Procedure for calculating SVD

1. Compute $\boldsymbol{A}^*\boldsymbol{A}$
2. Compute eigenvalue of $\boldsymbol{A}^*\boldsymbol{A}$ its corresponding normalized eigenvector. The square root of eigenvalues are σ_i It forms $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$ matrix
3. Compute $\boldsymbol{AV}/\sigma_i$ to get $\boldsymbol{U}$ matrix

Eigenvalues of ***AA****

- By SVD theorem, $\boldsymbol{A} = \boldsymbol{U\Sigma V}^*$,

$\boldsymbol{A}^* = (\boldsymbol{U\Sigma V}^*)^* = \boldsymbol{V\Sigma^* U^*} = \boldsymbol{V\Sigma U}^* \rightarrow \boldsymbol{A^* U} = \boldsymbol{V\Sigma}$

- Form the viewpoint of linear transformation,

$$\langle v_i, v_j \rangle = \left\langle \frac{1}{\sigma_i} T^*(u_i), \frac{1}{\sigma_j} T^*(u_j) \right\rangle = \frac{1}{\sigma_i \sigma_j} \langle T^*T(u_i), u_j \rangle$$

$$= \frac{\sigma_i^{\ 2}}{\sigma_i \sigma_j} \langle u_i, u_j \rangle = \delta_{ij}$$

- The eigenvalues of ***AA**** and ***A*A*** are the same!

Example

- Give a system of eq: $\begin{cases} x_1 + x_2 - x_3 = 1 \\ x_1 + x_2 - x_3 = 2 \end{cases} \rightarrow \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$AA^T = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \rightarrow \left\| \begin{bmatrix} 3-\lambda & 3 \\ 3 & 3-\lambda \end{bmatrix} \right\| = \lambda(6-\lambda) = 0$$

$$A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{bmatrix}$$

$$\rightarrow \left\| \begin{bmatrix} 2-\lambda & 2 & -2 \\ 2 & 2-\lambda & -2 \\ -2 & -2 & 2-\lambda \end{bmatrix} \right\| = \lambda^2(6-\lambda) = 0$$

AA^T and ***A^TA*** has the same eigenvalue!

Example

- $\because \boldsymbol{A} = \boldsymbol{U\Sigma V^T} \rightarrow \boldsymbol{A^T} = (\boldsymbol{U\Sigma V^T})^T = \boldsymbol{V\Sigma^T U^T}$
- $\boldsymbol{AA^T} = \boldsymbol{U\Sigma V^T V\Sigma^T U^T} = \boldsymbol{U\Sigma^2 U^T}$
- $\boldsymbol{A^T A} = \boldsymbol{V\Sigma^T U^T U\Sigma V^T} = \boldsymbol{V^T\Sigma^2 V}$
- But the corresponding eigenvector are $\boldsymbol{U}$ and $\boldsymbol{V}$ respectively, $\boldsymbol{AA^T}$ to get $\boldsymbol{U}$, $\boldsymbol{A^T A}$ to get $\boldsymbol{V}$, so

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad V = \frac{1}{\sqrt{6}}\begin{pmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ -\sqrt{2} & 0 & 2 \end{pmatrix}$$

Example

- It's easier that we use ***AV = ΣU*** or ***UA = ΣV^T*** to get its SVD. In this example,

$$AV = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ -\sqrt{2} & 0 & 2 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ \sqrt{3} & 0 & 0 \end{bmatrix}$$

- ***U*** is $M_{2\times 2}(\mathbb{R})$, so $u_1 = \frac{1}{\sigma_1} Av_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{3} \\ \sqrt{3} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

- And $(\boldsymbol{AV})_{21}$ is $\frac{1}{\sqrt{2}}$ so $u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \rightarrow U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

The inverse matrix

- Matrix $\boldsymbol{A}$ is invertible $\leftarrow\rightarrow$ $\boldsymbol{A}$ is square

  det($\boldsymbol{A}$) $\neq 0$
- The definition of the a invertible is:
- If A is in $M_{m \times m}(\mathbb{R})$ invertible, then there exists a matrix $\boldsymbol{B}$ in $M_{m \times m}(\mathbb{R})$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}_m$
- $\boldsymbol{AI} = \boldsymbol{A}(\boldsymbol{AB}) = \boldsymbol{A}(\boldsymbol{BA}) = \boldsymbol{A}$
- And it uses to solve least-square problem.

The Pseudo-inverse

- The Pseudo-inverse matrix is a generalized of a inverse matrix,
- If ***A*** is in $M_{m \times n}(\mathbb{R})$, A is not invertible, but its Pseudo-inverse matrix exists.
- The Pseudo-inverse is useful for solving least-square problem, even if the matrix is not full rank.

Definition of Pseudo-inverse

- Let $\boldsymbol{A}$ is in $M_{m \times n}(\mathbb{R})$, then the solution $\boldsymbol{X}$ is the Pseudo-inverse matrix of $\boldsymbol{A}$ satisfies the following eqs:
- $\boldsymbol{AXA} = \boldsymbol{A}$, $\boldsymbol{XAX} = \boldsymbol{X}$, $(\boldsymbol{AX})^* = \boldsymbol{XA}$, $(\boldsymbol{XA})^* = \boldsymbol{AX}$
- The Pseudo-inverse can be used to solve the least-square problem.
- If $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{A}$ is full rank $\leftarrow\rightarrow$ $\boldsymbol{x_0} = (\boldsymbol{A^T A})^{-1}\boldsymbol{A^T b}$
- If $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{A}$ is rank-deficient $\leftarrow\rightarrow$ $\boldsymbol{x}^+ = \boldsymbol{A}^+\boldsymbol{b}$

Pseudo-inverse by SVD

Theorem:

- $\boldsymbol{A}$ is $M_{m \times n}(\mathbb{C})$ and $\boldsymbol{A} = \boldsymbol{U\Sigma V^*}$, then $A^+ = \boldsymbol{V\Sigma}^+\boldsymbol{U^*}$ (Pseudo-inverse of A), $\boldsymbol{\Sigma}^+$ is the pseudo-inverse of $\boldsymbol{\Sigma}$.
- $\boldsymbol{\Sigma}_{ij} = \sigma_i\delta_{ij}$, $(\boldsymbol{\Sigma}_{ij})^+ = \boldsymbol{\Sigma}_{ji} = \sigma_j^{-1}\delta_{ji}$

Note that $\boldsymbol{AA}^+$ and $\boldsymbol{A}^+\boldsymbol{A}$ are not equal to identity matrix necessarily

Least-square solution

Theorem:

a. If $\boldsymbol{Ax} = \boldsymbol{b}$ is consistent, then z is the unique solution to the system having minimum norm.

b. If $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent, then z is the unique best approximation to a solution having minimum norm.

Part a has been proved in another ppt, so here we prove part b.

Least-square solution

Proof :

Suppose $\boldsymbol{x_0}$ is the best-approximation solution of $\boldsymbol{Ax} = \boldsymbol{b}$, then $\| \boldsymbol{Ax} - \boldsymbol{b} \| \geq \| \boldsymbol{Ax_0} - \boldsymbol{b} \|$

$$\boldsymbol{Ax} - \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{A}^{+})\boldsymbol{b} - (\boldsymbol{I} - \boldsymbol{AA}+)(-\boldsymbol{b})$$

Since those vector forms a right triangle by Pythagorean theorem,

$$\begin{aligned} \|\boldsymbol{Ax} - \boldsymbol{b}\|^2 &= \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{AA}^{+})\boldsymbol{b}\|^2 - \|(\boldsymbol{I} - \boldsymbol{AA}+)(-\boldsymbol{b})\|^2 \\ &= \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{AA}^{+})\boldsymbol{b}\|^2 - \|(\boldsymbol{I} - \boldsymbol{AA}+)(-\boldsymbol{b})\|^2 \\ &\geq \|\boldsymbol{Ax_0} - \boldsymbol{b}\|^2 \end{aligned}$$

Example

- Give a system of eq: $\begin{cases} x_1 + x_2 - x_3 = 1 \\ x_1 + x_2 - x_3 = 2 \end{cases} \rightarrow \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix} \quad A^{+} = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \rightarrow z = A^{+} b = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

- We know the system of eq has infinity solutions, so z is the "best-approximate" solutions.