Understanding Singular Value Decomposition

Lin ZHANG, PhD
School of Software Engineering
Tongji University
July, 2020

Outline

- Some prerequisite results
- Singular value decomposition theorem
- Relationship between SVD and EVD
- Economy-sized form and outer product form of SVD
- Matrix norms and singular values
- SVD and linear least squares

Some Prerequisite Results

Result 1: Suppose that $rank(A_{m\times n}) = r$, then $r(A^T A) = r(AA^T) = r$

Proof: Let's prove if $r(A_{m\times n}) = r$, then $r(A^T A) = r(A) = r$

Given an $\mathbf{x} \in \mathbb{R}^{n\times 1}$, if $A\mathbf{x} = \mathbf{0}$ ⟹ $A^T A\mathbf{x} = \mathbf{0}$

Given an $\mathbf{x} \in \mathbb{R}^{n\times 1}$, if $A^T A\mathbf{x} = \mathbf{0}$

⇓

$\mathbf{x}^T A^T A\mathbf{x} = 0$

⇓

$(A\mathbf{x})^T A\mathbf{x} = 0$ ⟹ $A\mathbf{x} = \mathbf{0}$

⟹ $\begin{cases} A^T A\mathbf{x} = \mathbf{0} \\ A\mathbf{x} = \mathbf{0} \end{cases}$ have the same solutions

⇓

$r(A^T A) = r(A) = r$

Following a similar way, we can prove $r(AA^T) = r(A) = r$

Result 2: If A is an $n \times n$ real symmetric matrix and $r(A)=r$, then
A has and only has r non-zero eigen-values

Proof: Since A is an $n \times n$ real symmetric matrix, A can be diagonally decomposed as

$$A = P\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}P^{-1}$$

, where P is an invertible matrix and $\lambda_i s$ are A's n eigen-values

Since multiplied by an invertible matrix, a matrix's rank remains

$$r\left(\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}\right) = r(A) = r \Rightarrow \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

has r non-zeros diagonal entries

Some Prerequisite Results

Result 3: Suppose A is an $m \times n$ matrix, then

$\left(A^T A\right)_{n\times n}$ and $\left(AA^T\right)_{m\times m}$ are both positive semi-definite matrices

Proof:

$$\forall \mathbf{x} \neq \mathbf{0}, 0 \leq \left(A\mathbf{x}\right)^T \left(A\mathbf{x}\right) = \mathbf{x}^T A^T A\mathbf{x}$$

⬇

$\left(A^T A\right)_{n\times n}$ is positive semi-definite

Following a similar way, we can prove $\left(AA^T\right)_{m\times m}$ is also positive semi-definite

Some Prerequisite Results

Result 4: Suppose A is an $m \times n$ matrix, then

$(A^T A)_{n\times n}$ and $(AA^T)_{m\times m}$ have the same nonzero eigen-values

Proof:

Suppose that λ is the eigen-value of $(A^T A)_{n\times n}$ and $\mathbf{x}$ is the associated eigen-vector, then

$$A^T A\mathbf{x} = \lambda \mathbf{x}$$

Multiply A on both sides,

$$AA^T(A\mathbf{x}) = \lambda (A\mathbf{x})$$

It can be seen that, λ and $A\mathbf{x}$ are eigen-value and the associated eigen-vector of AA^T

Outline

- Some prerequisite results
- Singular value decomposition theorem
- Relationship between SVD and EVD
- Economy-sized form and outer product form of SVD
- Matrix norms and singular values
- SVD and linear least squares

Singular Value Decomposition's General Form

SVD decomposition theorem: Any matrix $A_{m\times n}$ can be decomposed as the following form,

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$

where U and V are two orthogonal matrices, $r(A) = r$,

$$\Sigma_{m\times n} = \begin{bmatrix} \Sigma_r & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n} = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n}$$

$\sigma_1, \sigma_2, \ldots, \sigma_r > 0$ are called the **singular values** of A

In general case, $\Sigma_{m\times n}$ is not unique. However, if $\{\sigma_i\}_{i=1}^r$ are arranged in order, $\Sigma_{m\times n}$ is uniquely determined by A. In the following, we require that $\sigma_1 \geq \sigma_2 \geq, \ldots, \geq \sigma_r > 0$

Outline

- Some prerequisite results
- Singular value decomposition theorem
- Relationship between SVD and EVD
- Economy-sized form and outer product form of SVD
- Matrix norms and singular values
- SVD and linear least squares

Relationship between SVD and EVD

Result 5: Suppose that $rank(A_{m\times n}) = r$, then $(A^T A)_{n\times n}$ has and only has *r* positive eigen-values and all its other eigen-values are zeros

Proof:

$r(A_{m\times n}) = r$ —result 1→ $r(A^T A) = r$

$A^T A$ is a real symmetric matrix

—result 2→ $A^T A$ has and only has *r* non-zero eigen-values

Using result 3, $A^T A$ is positive semi-definite → eigen-values of $A^T A$ are all non-negative

→ $A^T A$ has and only has *r* positive eigen-values and all its other eigen-values are zeros

Relationship between SVD and EVD

If $A \in \mathbb{R}^{m\times n}$, $r(A_{m\times n}) = r$, using result 5, $A^T A$ can be orthogonally diagonalized as,

(Note: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$ and V' is an orthogonal matrix)

$$(A^T A)_{n\times n} = V' \begin{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(n-r)} \end{bmatrix}_{n\times n} V'^T$$

On the other hand, based on SVD, we know that $A = U\Sigma V^T$, then

$$(A^T A)_{n\times n} = (U\Sigma_{m\times n} V^T)^T (U\Sigma_{m\times n} V^T)$$

$$= V\Sigma^T U^T U\Sigma V^T = V(\Sigma^T\Sigma)_{n\times n} V^T$$

$$= V \begin{bmatrix} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_r^2 \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(n-r)} \end{bmatrix}_{n\times n} V^T$$

$$\sigma_i = \sqrt{\lambda_i}, 1 \leq i \leq r$$

Relationship between SVD and EVD

- We can have the following conclusions about the SVD of A and the EVD of $A^T A$
  - A has r singular values $\{\sigma_i\}_{i=1}^r$ (positive) and A^TA has r positive eigen-values $\{\lambda_i\}_{i=1}^r$ (all its other eigen-values are zeros) , and $\sigma_i = \sqrt{\lambda_i}$
  - A's right singular matrix V is actually the orthogonal matrix V' obtained when performing orthogonally diagonalization to A^TA; but it needs to be noted that V (V') is not unique
- Using similar derivations, we can have the following conclusions about the SVD of A and the EVD of AA^T
  - AA^T has the same r positive eigen-values $\{\lambda_i\}_{i=1}^r$ (result 4) , and $\sigma_i = \sqrt{\lambda_i}$
  - A's left singular matrix U is actually the orthogonal matrix obtained when performing orthogonally diagonalization to AA^T; but it needs to be noted that both of them are not unique

Relationship between SVD and EVD

- Consider the following special cases
  - If A is an $n \times n$ matrix, how about its eigen-values and singular values?
  - If A is an $n \times n$ real symmetric matrix, how about its eigen-values and singular values?
  - If A is an $n \times n$ real symmetric and positive semi-definite matrix, how about its eigen-values and singular values?

Relationship between SVD and EVD

Result 6: Suppose that $A_{n\times n}$ is a positive semi-definite (real symmetric) matrix. Then, its eigen-values and singular values are the same.

Proof:

Suppose that $rank(A) = r$

$A_{n\times n}$ is positive semi-definite (and real symmetric)

result 5 ⇒ A has and only has r positive eigen-values $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_r > 0$ and all its other eigen-values are zeros

⇓

$$A = U\begin{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_r \end{bmatrix} & O \\ O & O \end{bmatrix}_{n\times n} U^T \Rightarrow A^T A = U\begin{bmatrix} \begin{bmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_r^2 \end{bmatrix} & O \\ O & O \end{bmatrix}_{n\times n} U^T \Rightarrow$$

A's singular values are

$$\sigma_i = \sqrt{\lambda_i^2} = \lambda_i \; (1 \le i \le r)$$

Outline

- Some prerequisite results
- Singular value decomposition theorem
- Relationship between SVD and EVD
- Economy-sized form and outer product form of SVD
- Matrix norms and singular values
- SVD and linear least squares

Economy-sized form of SVD

- For a matrix $A \in \mathbb{R}^{m\times n}$, if $r(A)$ is much smaller then m and n, it will be more memory efficient to store the economy-sized SVD

$$A_{m\times n} = U_{m\times m} \Sigma_{m\times n} V_{n\times n}^T = [\mathbf{u}_1,\ldots,\mathbf{u}_r \mid \mathbf{u}_{r+1},\ldots,\mathbf{u}_m] \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \\ \text{-----} \\ \mathbf{v}_{r+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= [\mathbf{u}_1,\ldots,\mathbf{u}_r] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = U_{m\times r} \Sigma_{r\times r} V_{n\times r}^T$$

Outer product form of SVD

Result 7: **Outer product form of a matrix product**.

In general, if X is an $m \times k$ matrix and Y is a $k \times n$ matrix, the matrix product can be expressed as,

$$XY = \sum_{i=1}^{k} [col(X)_i]_{m\times 1} [row(Y)_i]_{1\times n}$$

Note: each submatrix $[col(X)_i][row(Y)_i]$ is of rank 1

Outer product form of SVD

Let's consider the economy-sized SVD of A, $A_{m\times n}=[\mathbf{u}_1,\ldots,\mathbf{u}_r]\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}\begin{bmatrix}\mathbf{v}_1^T\\ \vdots \\ \mathbf{v}_r^T\end{bmatrix}$

Let $X=[\mathbf{u}_1,\ldots,\mathbf{u}_r]\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}=(\sigma_1\mathbf{u}_1,\sigma_2\mathbf{u}_2,\ldots,\sigma_r\mathbf{u}_r), Y=\begin{bmatrix}\mathbf{v}_1^T\\ \mathbf{v}_2^T\\ \vdots \\ \mathbf{v}_r^T\end{bmatrix}$

$$A=XY=(\sigma_1\mathbf{u}_1,\sigma_2\mathbf{u}_2,\ldots,\sigma_r\mathbf{u}_r)\begin{bmatrix}\mathbf{v}_1^T\\ \mathbf{v}_2^T\\ \vdots \\ \mathbf{v}_r^T\end{bmatrix} \xrightarrow{\text{result 6}} =\sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T$$

Outer product form of SVD

Outline

- Some prerequisite results
- Singular value decomposition theorem
- Relationship between SVD and EVD
- Economy-sized form and outer product form of SVD
- Matrix norms and singular values
- SVD and linear least squares

Matrix norms and singular values

Definition 1: The **spectral norm** of a matrix A is the largest singular value of A i.e. the square root of the largest eigenvalue of the positive semidefinite matrix A^TA (or AA^T):

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} = \sigma_{\max}(A)$$

Definition 2: The **nuclear norm** is the sum of all the singular values of A,

$$\|A\|_* = \sum_{i=1}^{rank(A)} \sigma_i$$

Definition 3: The **Frobenius norm** of a matrix $A_{m\times n}$ is defined as,

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

Matrix norms and singular values

Result 8: Suppose that $rank(A_{m\times n}) = r$ and $\sigma_1, \sigma_2, \ldots, \sigma_r$ are A's singular values, then

$$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

Proof:

$$\|A\|_F = \sqrt{trace\left(A^T A\right)} = \sqrt{trace\left(AA^T\right)} = \sqrt{\sum_{i=1}^{r} \lambda_i} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are the positive eigen-values of A^TA (or AA^T)

Note: From result 5, we can know that A^TA (or AA^T) has and only has r positive eigen-values and all the other eigen-values are zeros.

Outline

- Some prerequisite results
- Singular value decomposition theorem
- Relationship between SVD and EVD
- Economy-sized form and outer product form of SVD
- Matrix norms and singular values
- SVD and linear least squares

SVD and linear least squares

Linear least squares is a general idea for solving linear equations,

$$A_{m\times n}\mathbf{x}_{n\times 1}=\mathbf{b}_{m\times 1} \qquad (1)$$

Using the idea of least squares, Eq. 1 is equivalent to the following problem,

$$\mathbf{x}^*=\arg\min_{\mathbf{x}}\left\|A_{m\times n}\mathbf{x}_{n\times 1}-\mathbf{b}_{m\times 1}\right\|_2^2 \qquad (2)$$

Eq. 2 can be solved by finding the stationary point $\mathbf{x}^*$ of $\left\|A_{m\times n}\mathbf{x}_{n\times 1}-\mathbf{b}_{m\times 1}\right\|_2^2$, i.e. $\mathbf{x}^*$ should satisfy,

$$A^T A\mathbf{x}^*=A^T\mathbf{b} \qquad (3)$$

In Eq. 3, when $rank(A)=n$ (the columns of A are linearly independent) ,

$rank(A^T A)=n$ ➡ $A^T A$ is invertible ➡ $\mathbf{x}^*$ is uniquely determined as $\mathbf{x}^*=(A^T A)^{-1}A^T\mathbf{b}$

How about when rank(A)<n?

SVD and linear least squares

- For solving the linear least squares numerically with a computer, usually we do not use the form of Eq. (3) (though it is elegant) for two reasons
  - When rank(A)<n, $\mathbf{x}^*$ can not be determined
  - Even though A^TA is invertible, the formation of A^TA can dramatically degrade the accuracy of the computation
- Instead, we can use the technique of SVD

SVD and linear least squares

Suppose the SVD form of A is,

$$A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^{T}$$

$$A\mathbf{x}-\mathbf{b} = U\Sigma V^{T}\mathbf{x}-\mathbf{b} = U\left(\Sigma V^{T}\mathbf{x}\right) - U\left(U^{T}\mathbf{b}\right) \triangleq U\left(\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right)$$

where $\mathbf{y}_{n\times 1} = V^{T}\mathbf{x}, \mathbf{c}_{m\times 1} = U^{T}\mathbf{b}$

Since U is an orthogonal matrix,

$$\left\|A\mathbf{x}-\mathbf{b}\right\| = \left\|U\left(\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right)\right\| = \left\|\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right\|$$

Then, our objective is to identify $\mathbf{y}$ that can make $\left\|\Sigma\mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1}\right\|$ have minimum length

SVD and linear least squares

$$\Sigma \mathbf{y}_{n\times 1} = \begin{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} & \boldsymbol{O}_{r\times(n-r)} \\ \boldsymbol{O}_{(m-r)\times r} & \boldsymbol{O}_{(m-r)\times(n-r)} \end{bmatrix}_{m\times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sigma_1 y_1 \\ \sigma_2 y_2 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m\times 1} \Rightarrow \Sigma \mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1} = \begin{bmatrix} \sigma_1 y_1 - c_1 \\ \sigma_2 y_2 - c_2 \\ \vdots \\ \sigma_r y_r - c_r \\ -c_{r+1} \\ \vdots \\ -c_m \end{bmatrix}_{m\times 1}$$

Then, we simply let $y_i = \frac{c_i}{\sigma_i}, 1 \le i \le r$; then, $\left\| \Sigma \mathbf{y}_{n\times 1} - \mathbf{c}_{m\times 1} \right\|$ can get the minimum length $\sqrt{\sum_{i=r+1}^{m} c_i^2}$

Note that $y_{r+1} \sim y_n$ can be arbitrary

SVD and linear least squares

The operation $y_i = \dfrac{c_i}{\sigma_i}, 1 \le i \le r$ can be simply completed by a matrix multiplication,

$$\mathbf{y} = \begin{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_r} \end{bmatrix} & \boldsymbol{O}_{r\times(m-r)} \\ \boldsymbol{O}_{(n-r)\times r} & \boldsymbol{O}_{(n-r)\times(m-r)} \end{bmatrix}_{n\times m} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}_{m\times 1} = \begin{bmatrix} c_1/\sigma_1 \\ c_2/\sigma_2 \\ \vdots \\ c_r/\sigma_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} \triangleq \Sigma^+ \mathbf{c}_{m\times 1}$$

where Σ^+ means transposing Σ and inverting all non-zero diagonal entries

Finally,

$$\mathbf{x} = V\mathbf{y}_{n\times 1} = V\Sigma^+\mathbf{c}_{m\times 1} = V\Sigma^+U^T\mathbf{b}$$

Moore-Penrose inverse

SVD and linear least squares

- Some notes about the generalized inverse used in linear least squares
  - It does not have requirements for the rank of A
  - It can guarantee that the obtained solution can make $\|A\mathbf{x}-\mathbf{b}\|$ having the minimum length; but **the solution may be not unique**


Q+A