MA60053 - Computational Linear Algebra Singular Value Decomposition(SVD) (To be updated)

M. Rajesh Kannan

Department of Mathematics,
Indian Institute of Technology Kharagpur,
email: rajeshkannan1.m@gmail.com, rajeshkannan@maths.iitkgp.ac.in

March 5, 2020

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \leq m$, then an SVD of A is

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \leq m$, then an SVD of A is

$$A = U(\Sigma \; 0) V^T,$$

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \leq m$, then an SVD of A is

$$A = U(\Sigma\ 0)V^T, \text{ where } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}, \sigma_1 \geq \ldots \geq \sigma_n \geq 0,$$

and $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ are orthogonal.

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \leq m$, then an SVD of A is

$$A = U(\Sigma\ 0)V^T, \text{ where } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix}, \sigma_1 \geq \ldots \geq \sigma_n \geq 0,$$

and $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ are orthogonal.

The matrix U is called a left singular vector matrix, V is called a right singular vector matrix, and the scalars σ_j are called singular values.

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$.

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \geq m$, then a singular value decomposition(SVD) of A is a decomposition

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \geq m$, then a singular value decomposition(SVD) of A is a decomposition

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T,$$

Singular Value Decomposition(SVD)

Let $A \in \mathbb{R}^{n \times m}$. If $n \geq m$, then a singular value decomposition(SVD) of A is a decomposition

$$A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T, \text{ where } \Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \end{pmatrix}, \sigma_1 \geq \ldots \geq \sigma_m \geq 0,$$

and $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{m \times m}$ are orthogonal.

Condensed SVD

Let $A \in \mathbb{R}^{n \times m}$ be a nonzero matrix of rank r. Then, there exist $\widehat{U} \in \mathbb{R}^{n \times r}$, $\widehat{\Sigma} \in \mathbb{R}^{r \times r}$ and $\widehat{V} \in \mathbb{R}^{m \times r}$ such that $\widehat{U}^T \widehat{U} = \widehat{V}^T \widehat{V} = I_r$, $\widehat{\Sigma}$ is a diagonal matrix with main diagonal entries $\sigma_1 \geq \cdots \geq \sigma_r > 0$, and $A = \widehat{U} \widehat{\Sigma} \widehat{V}^T$.

Existence of SVD

Theorem

Every matrix has an singular value decomposition.

Proof.

- Let A be an $n \times m$ matrix with rank r and $n \leq m$.

Existence of SVD

Theorem

Every matrix has an singular value decomposition.

Proof.

- Let A be an $n \times m$ matrix with rank r and $n \leq m$.
- Then the matrix AA^T is a symmetric matrix with rank r.

Existence of SVD

Theorem

Every matrix has an singular value decomposition.

Proof.

- Let A be an $n \times m$ matrix with rank r and $n \leq m$.
- Then the matrix AA^T is a symmetric matrix with rank r.
- By Spectral theorem, we have $AA^T = U \Lambda U^T$, where $\Lambda = \mathrm{diag}(d_1, \ldots, d_r, 0, \ldots, 0)$ and $U^T U = I$.

Existence of SVD

Theorem

Every matrix has an singular value decomposition.

Proof.

- Let A be an $n \times m$ matrix with rank r and $n \leq m$.
- Then the matrix AA^T is a symmetric matrix with rank r.
- By Spectral theorem, we have $AA^T = U \Lambda U^T$, where $\Lambda = \mathrm{diag}(d_1, \ldots, d_r, 0, \ldots, 0)$ and $U^T U = I$.
- Take $B = A^T U$, then $B^T B = \Lambda$.

Existence of SVD

Theorem

Every matrix has an singular value decomposition.

Proof.

- Let A be an $n \times m$ matrix with rank r and $n \leq m$.
- Then the matrix AA^T is a symmetric matrix with rank r.
- By Spectral theorem, we have $AA^T = U\Lambda U^T$, where $\Lambda = \text{diag}(d_1, \ldots, d_r, 0, \ldots, 0)$ and $U^T U = I$.
- Take $B = A^T U$, then $B^T B = \Lambda$.
- Define $V = BG$ where $G = \text{diag}(\frac{1}{\sqrt{d_1}}, \ldots, \frac{1}{\sqrt{d_r}}, 0, \ldots, 0(m\text{-times}))$.

Existence of SVD

Theorem

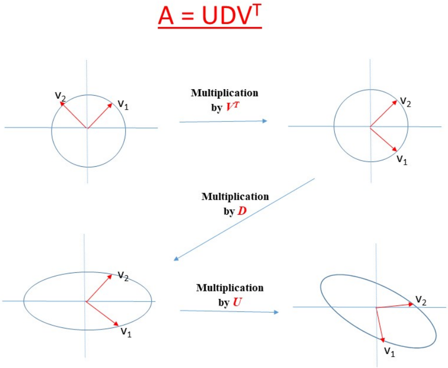
Every matrix has an singular value decomposition.

Proof.

- Let A be an $n \times m$ matrix with rank r and $n \leq m$.
- Then the matrix AA^T is a symmetric matrix with rank r.
- By Spectral theorem, we have $AA^T = U\Lambda U^T$, where $\Lambda = \mathrm{diag}(d_1, \ldots, d_r, 0, \ldots, 0)$ and $U^T U = I$.
- Take $B = A^T U$, then $B^T B = \Lambda$.
- Define $V = BG$ where $G = \mathrm{diag}(\frac{1}{\sqrt{d_1}}, \ldots, \frac{1}{\sqrt{d_r}}, 0, \ldots, 0(m\text{-times}))$.
- Let $\Sigma = (\sqrt{\Lambda}\ 0)$ Verify $U\Sigma V^T$ is a singular value decomposition for A.

$\square$

SVD geometry



Computing SVD

Example

Let us compute SVD for the following 2×3 matrix,

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}.$$

Computing SVD

Example

Let us compute SVD for the following 2×3 matrix,

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}.$$

In order to find U, we have to start with AA^T.

$$AA^T = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}.$$

Computing SVD

Example

Let us compute SVD for the following 2×3 matrix,

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}.$$

In order to find U, we have to start with AA^T.

$$AA^T = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 1 \\ 1 & 11 \end{pmatrix}.$$

Next, we have to find the eigenvalues and corresponding eigenvectors of AA^T. After calculating, we get the following eigenvalues and their corresponding eigenvectors.

$$\lambda = 10; \quad u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\lambda = 12; \quad u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

$$\lambda = 10; \quad u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$\lambda = 12; \quad u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus the matrix A has singular values $\sigma_1 = \sqrt{12}$ and $\sigma_2 = \sqrt{10}$. Now after normalizing u_1 and u_2, we put $U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$.

$$\lambda = 10; \quad u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$
$$\lambda = 12; \quad u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus the matrix A has singular values $\sigma_1 = \sqrt{12}$ and $\sigma_2 = \sqrt{10}$. Now after normalizing u_1 and u_2, we put $U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$.

The calculation of V is similar. V is based on $A^T A$, so we have

$$A^T A = \begin{pmatrix} 3 & -1 \\ 1 & 3 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix}.$$

Eigenvalues and their corresponding eigenvectors are as follows

$$\textit{for } \lambda = 12; \quad v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\textit{for } \lambda = 10; \quad v_2 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$$

$$\textit{for } \lambda = 0; \quad v_3 = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}.$$

Eigenvalues and their corresponding eigenvectors are as follows

$$\textit{for } \lambda = 12; \quad v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\textit{for } \lambda = 10; \quad v_2 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$$

$$\textit{for } \lambda = 0; \quad v_3 = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}.$$

After normalization, we get $V = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} \\ \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{5}} & \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{-5}{\sqrt{30}} \end{pmatrix}$ i.e.,

$$V^T = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}$$

SVD of A is

$$A = U\Sigma V^T, \text{ where } \Sigma = \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix}.$$

SVD of A is

$$A = U\Sigma V^T, \text{ where } \Sigma = \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix}.$$

That is,

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

Geometric form of SVD

Let $A \in \mathbb{R}^{n \times m}$ with $n \leq m$. Then, $\mathbb{R}^n$ has an orthonormal basis $\{u_1, \dots, u_n\}$, $\mathbb{R}^m$ has an orthonormal basis $\{v_1, \dots, v_m\}$ and there exists $\sigma_1 \geq \sigma_2 \geq \dots, \geq \sigma_r > 0$ such that

$$Av_i = \begin{cases} \sigma_i u_i, & \text{if } i = 1, \dots, r, \\ 0 & \text{if } i \geq r+1, \end{cases}$$

Geometric form of SVD

Let $A \in \mathbb{R}^{n \times m}$ with $n \leq m$. Then, $\mathbb{R}^n$ has an orthonormal basis $\{u_1, \ldots, u_n\}$, $\mathbb{R}^m$ has an orthonormal basis $\{v_1, \ldots, v_m\}$ and there exists $\sigma_1 \geq \sigma_2 \geq \ldots, \geq \sigma_r > 0$ such that

$$Av_i = \begin{cases} \sigma_i u_i, & \text{if } i = 1, \ldots, r, \\ 0 & \text{if } i \geq r+1, \end{cases}$$

and

$$A^T u_i = \begin{cases} \sigma_i v_i, & \text{if } i = 1, \ldots, r, \\ 0 & \text{if } i \geq r+1. \end{cases}$$

Proof.

$A = U\Sigma V^T$ implies $AV = U\Sigma$, and $A^T U = V\Sigma$. □

Four fundamental subspaces

For an $n \times m$ matrix A, the following subspaces are called fundamental subspaces.

- ***Range space of** A**:*** $R(A) = \{x \in \mathbb{R}^n : x = Ay \text{ for some } y \in \mathbb{R}^m\}$. (span of columns of A)

Four fundamental subspaces

For an $n \times m$ matrix A, the following subspaces are called fundamental subspaces.

- ***Range space of A:*** $R(A) = \{x \in \mathbb{R}^n : x = Ay \text{ for some } y \in \mathbb{R}^m\}$. (span of columns of A)
- ***Null space of A:*** $N(A) = \{x \in \mathbb{R}^m : Ax = 0\}$.

Four fundamental subspaces

For an $n \times m$ matrix A, the following subspaces are called fundamental subspaces.

- ***Range space of*** A***:*** $R(A) = \{x \in \mathbb{R}^n : x = Ay$ for some $y \in \mathbb{R}^m\}$. (span of columns of A)
- ***Null space of*** A***:*** $N(A) = \{x \in \mathbb{R}^m : Ax = 0\}$.
- ***Range space of*** A^T***:*** $R(A^T) = \{x \in \mathbb{R}^m : x = A^T y$ for some $y \in \mathbb{R}^n\}$.

Four fundamental subspaces

For an $n \times m$ matrix A, the following subspaces are called fundamental subspaces.

- ***Range space of** A**:*** $R(A) = \{x \in \mathbb{R}^n : x = Ay \text{ for some } y \in \mathbb{R}^m\}$. (span of columns of A)
- ***Null space of** A**:*** $N(A) = \{x \in \mathbb{R}^m : Ax = 0\}$.
- ***Range space of** A^T**:*** $R(A^T) = \{x \in \mathbb{R}^m : x = A^T y \text{ for some } y \in \mathbb{R}^n\}$.
- ***Null space of** A^T**:*** $N(A^T) = \{x \in \mathbb{R}^n : A^T x = 0\}$.

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n \times m}$ is a matrix of rank r, and $A = U\Sigma V^T$ is the SVD of A, then

- $R(A) = \text{span}\{u_1, \ldots, u_r\}$,

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n \times m}$ is a matrix of rank r, and $A = U\Sigma V^T$ is the SVD of A, then

- $R(A) = \text{span}\{u_1, \ldots, u_r\}$,
- $N(A) = \text{span}\{v_{r+1}, \ldots, v_m\}$,

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n \times m}$ is a matrix of rank r, and $A = U \Sigma V^T$ is the SVD of A, then

- $R(A) = \mathrm{span}\{u_1, \ldots, u_r\}$,
- $N(A) = \mathrm{span}\{v_{r+1}, \ldots, v_m\}$,
- $R(A^T) = \mathrm{span}\{v_1, \ldots, v_r\}$,

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n \times m}$ is a matrix of rank r, and $A = U\Sigma V^T$ is the SVD of A, then

- $R(A) = \text{span}\{u_1, \ldots, u_r\}$,
- $N(A) = \text{span}\{v_{r+1}, \ldots, v_m\}$,
- $R(A^T) = \text{span}\{v_1, \ldots, v_r\}$,
- $N(A^T) = \text{span}\{u_{r+1}, \ldots, u_n\}$.

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n\times m}$ is a matrix of rank r, and $A = U\Sigma V^T$ is the SVD of A, then

- $R(A) = \text{span}\{u_1, \ldots, u_r\}$,
- $N(A) = \text{span}\{v_{r+1}, \ldots, v_m\}$,
- $R(A^T) = \text{span}\{v_1, \ldots, v_r\}$,
- $N(A^T) = \text{span}\{u_{r+1}, \ldots, u_n\}$.

By products

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n\times m}$ is a matrix of rank r, and $A = U\Sigma V^T$ is the SVD of A, then

- $R(A) = \text{span}\{u_1, \dots, u_r\}$,
- $N(A) = \text{span}\{v_{r+1}, \dots, v_m\}$,
- $R(A^T) = \text{span}\{v_1, \dots, v_r\}$,
- $N(A^T) = \text{span}\{u_{r+1}, \dots, u_n\}$.

By products

- $R(A)^\perp = N(A^T)$ and $N(A)^\perp = R(A^T)$,

Basis for fundamental subspaces

If $A \in \mathbb{R}^{n \times m}$ is a matrix of rank r, and $A = U\Sigma V^T$ is the SVD of A, then

- $R(A) = \text{span}\{u_1, \ldots, u_r\}$,
- $N(A) = \text{span}\{v_{r+1}, \ldots, v_m\}$,
- $R(A^T) = \text{span}\{v_1, \ldots, v_r\}$,
- $N(A^T) = \text{span}\{u_{r+1}, \ldots, u_n\}$.

By products

- $R(A)^\perp = N(A^T)$ and $N(A)^\perp = R(A^T)$,
- If $A \in \mathbb{R}^{n \times m}$, then $\dim(R(A)) + \dim(N(A)) = m$.

Illustration

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

Illustration

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

- $R(A) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\},$

Illustration

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

- $R(A) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\}$,
- $N(A) = span \{0\}$,

Illustration

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

- $R(A) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\},$
- $N(A) = span \{0\},$
- $R(A^T) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} \end{pmatrix} \right\},$

Illustration

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{12} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} & 0 \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} & \frac{-5}{\sqrt{30}} \end{pmatrix}.$$

- $R(A) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\},$
- $N(A) = span\,\{0\},$
- $R(A^T) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{30}} \end{pmatrix} \right\},$
- $N(A^T) = span \left\{ \begin{pmatrix} \frac{1}{\sqrt{6}} \\ 0 \\ \frac{-5}{\sqrt{30}} \end{pmatrix} \right\}.$

SVD - equivalent (and useful) form

Theorem

Let $A \in \mathbb{R}^{m \times n}$, and let $\sigma_1, \ldots, \sigma_r$ be the nonzero singular values of A, with associated right and left singular vectors $v_1, \ldots, v_r$ and $u_1, \ldots, u_r$, respectively. Then

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T.$$

SVD - equivalent (and useful) form

Theorem

Let $A \in \mathbb{R}^{m\times n}$, and let $\sigma_1, \ldots, \sigma_r$ be the nonzero singular values of A, with associated right and left singular vectors $v_1, \ldots, v_r$ and $u_1, \ldots, u_r$, respectively. Then

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^T.$$

If $A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix}$, then

$$A = (\sqrt{12}) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{30}} \right) + (\sqrt{10}) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix} \left(\frac{2}{\sqrt{6}}, \frac{-1}{\sqrt{5}}, \frac{2}{\sqrt{30}} \right).$$

Properties

1. An $n \times m$ matrix has $\min\{m, n\}$ singular values.

Properties

1. An $n \times m$ matrix has $\min\{m, n\}$ singular values.
2. The singular values are unique, but the singular vector matrices are not. Although an SVD is not unique, one often says "the SVD" instead of "a SVD."

Properties

1. An $n \times m$ matrix has $\min\{m, n\}$ singular values.
2. The singular values are unique, but the singular vector matrices are not. Although an SVD is not unique, one often says "the SVD" instead of "a SVD."
3. Let $A \in \mathbb{R}^{n \times m}$ with $n \geq m$. If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ is an SVD of A, then $A^T = V(\Sigma\ 0)U^T$ is an SVD of A^T. Therefore, A and A^T have the same singular values.

Properties

1. An $n \times m$ matrix has $\min\{m, n\}$ singular values.
2. The singular values are unique, but the singular vector matrices are not. Although an SVD is not unique, one often says "the SVD" instead of "a SVD."
3. Let $A \in \mathbb{R}^{n\times m}$ with $n \geq m$. If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ is an SVD of A, then $A^T = V(\Sigma\ 0)U^T$ is an SVD of A^T. Therefore, A and A^T have the same singular values.
4. $A \in \mathbb{R}^{n\times n}$ is nonsingular if and only if all singular values are nonzero, i.e., $\sigma_j > 0, 1 \leq j \leq n$.

Properties

1. An $n \times m$ matrix has $\min\{m, n\}$ singular values.
2. The singular values are unique, but the singular vector matrices are not. Although an SVD is not unique, one often says "the SVD" instead of "a SVD."
3. Let $A \in \mathbb{R}^{n \times m}$ with $n \geq m$. If $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ is an SVD of A, then $A^T = V(\Sigma \; 0)U^T$ is an SVD of A^T. Therefore, A and A^T have the same singular values.
4. $A \in \mathbb{R}^{n \times n}$ is nonsingular if and only if all singular values are nonzero, i.e., $\sigma_j > 0, 1 \leq j \leq n$.
5. If $A = U \Sigma V^T$ is an SVD of A, then $A^{-1} = V \Sigma^{-1} U^T$ is an SVD of A^{-1}.

Theorem

If $A \in \mathbb{R}^{n \times m}$ *has singular values* $\sigma_1 \geq \ldots \geq \sigma_p$, *where* $p = min\{m, n\}$, *then* $\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1$, *and* $\|A\|_2 = \|A^T\|_2$.

Theorem

If $A \in \mathbb{R}^{n\times m}$ *has singular values* $\sigma_1 \geq \ldots \geq \sigma_p$, *where* $p = min\{m, n\}$, *then* $\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1$, *and* $\|A\|_2 = \|A^T\|_2$.

Theorem

$\|A\|_F = (\sigma_1^2 + \cdots + \sigma_r^2)^{\frac{1}{2}}$.

Properties

Theorem (Condition number)

If $A \in \mathbb{R}^{n\times n}$ *is nonsingular, then* $\kappa_2(A) = \frac{\sigma_1}{\sigma_n} = \frac{\text{maxmag}(A)}{\text{minmag}(A)}$.

Properties

Theorem (Condition number)

If $A \in \mathbb{R}^{n \times n}$ *is nonsingular, then* $\kappa_2(A) = \frac{\sigma_1}{\sigma_n} = \frac{\mathrm{maxmag}(A)}{\mathrm{minmag}(A)}$.

Theorem (Magnification)

If $A \in \mathbb{R}^{m \times n}$ *has singular values* $\sigma_1 \geq \ldots \geq \sigma_p$, *where* $p = min\{m, n\}$, *then*

$$\max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1 = \mathrm{maxmag}(A),$$

Properties

Theorem (Condition number)

If $A \in \mathbb{R}^{n\times n}$ *is nonsingular, then* $\kappa_2(A) = \frac{\sigma_1}{\sigma_n} = \frac{\text{maxmag}(A)}{\text{minmag}(A)}$.

Theorem (Magnification)

If $A \in \mathbb{R}^{m\times n}$ *has singular values* $\sigma_1 \geq \ldots \geq \sigma_p$*, where* $p = \min\{m, n\}$*, then*

$$\max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1 = \text{maxmag}(A),$$
$$\min_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_p = \text{minmag}(A).$$

Properties

Theorem (Condition number)

If $A \in \mathbb{R}^{n\times n}$ *is nonsingular, then* $\kappa_2(A) = \frac{\sigma_1}{\sigma_n} = \frac{\text{maxmag}(A)}{\text{minmag}(A)}$.

Theorem (Magnification)

If $A \in \mathbb{R}^{m\times n}$ *has singular values* $\sigma_1 \geq \ldots \geq \sigma_p$, *where* $p = \min\{m, n\}$, *then*

$$\max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1 = \text{maxmag}(A),$$
$$\min_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_p = \text{minmag}(A).$$

Theorem

Let $A \in \mathbb{R}^{n\times m}$. *Then* $\|A^T A\|_2 = \|A\|_2^2$, *and* $\kappa_2(A^T A) = \kappa(A)^2$.

Theorem

Let $A \in \mathbb{R}^{n \times m}$, $n \geq m$, $\mathrm{rank}(A) = m$, *with singular values* $\sigma_1 \geq \ldots \sigma_m > 0$. *Then,*

1. $\|(A^T A)^{-1}\|_2 = \frac{1}{\sigma_m^2}$,
2. $\|(A^T A)^{-1} A^T\|_2 = \frac{1}{\sigma_m}$,
3. $\|A^T (A^T A)^{-1}\|_2 = \frac{1}{\sigma_m}$, *and*
4. $\|A^T (A^T A)^{-1} A^T\|_2 = 1$.

Full rank matrices are dense

Theorem

Let $A \in \mathbb{R}^{n \times m}$ with rank r such that $r < \min\{n, m\}$. Then for every $\epsilon > 0$, there exists a full rank matrix $A_\epsilon \in \mathbb{R}^{n \times m}$ such that $\|A - A_\epsilon\|_2 = \epsilon$.

Full rank matrices are dense

Theorem

Let $A \in \mathbb{R}^{n \times m}$ *with rank* r *such that* $r < \min\{n, m\}$*. Then for every* $\epsilon > 0$*, there exists a full rank matrix* $A_\epsilon \in \mathbb{R}^{n \times m}$ *such that* $\|A - A_\epsilon\|_2 = \epsilon$*.*

Corollary

Full column rank matrices are dense on $\mathbb{R}^{n \times m}$*, for* $n \leq m$*.*

Low rank approximation using SVD

Theorem (Eckart and Young (1936))

Let $A \in \mathbb{R}^{n\times m}$ have a SVD as in previous definition. If $k < \mathrm{rank}(A)$, then the absolute distance of A to the set of rank k matrices is

$$\sigma_{k+1} = \min_{B\in\mathbb{R}^{n\times m}, \mathrm{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2,$$

where $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$.

Low rank approximation using SVD

Theorem (Eckart and Young (1936))

Let $A \in \mathbb{R}^{n\times m}$ have a SVD as in previous definition. If $k < \operatorname{rank}(A)$, then the absolute distance of A to the set of rank k matrices is

$$\sigma_{k+1} = \min_{B\in\mathbb{R}^{n\times m}, \operatorname{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2,$$

where $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$.

Corollary

Let $A \in \mathbb{R}^{n\times m}$ has full rank. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, where $r = \min\{n, m\}$. If $B \in \mathbb{R}^{n\times m}$ and $\|A - B\|_2 < \sigma_r$. Then B has full rank.

Low rank approximation using SVD

Theorem (Eckart and Young (1936))

Let $A \in \mathbb{R}^{n\times m}$ have a SVD as in previous definition. If $k < \mathrm{rank}(A)$, then the absolute distance of A to the set of rank k matrices is

$$\sigma_{k+1} = \min_{B\in\mathbb{R}^{n\times m}, \mathrm{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2,$$

where $A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^T$.

Corollary

Let $A \in \mathbb{R}^{n\times m}$ has full rank. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, where $r = \min\{n, m\}$. If $B \in \mathbb{R}^{n\times m}$ and $\|A - B\|_2 < \sigma_r$. Then B has full rank. Full rank matrices are open subset of $\mathbb{R}^{n\times m}$.

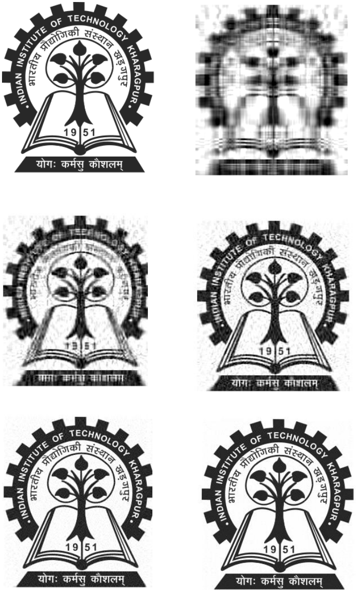
Relative distance to singular matrices

Theorem

Let $A \in \mathbb{R}^{n \times n}$ be a nonsingular matrix. Let A_s be the singular matrix closest to A in the sense that $\|A - A_s\|_2$ is as small as possible. Then, $\|A - A_s\|_2 = \sigma_n$ and

$$\frac{\|A - A_s\|_2}{\|A\|_2} = \frac{\sigma_n}{\|A\|_2} = \frac{1}{\|A\|_2 \|A^{-1}\|_2} = \frac{1}{\kappa_2(A)}.$$

Applications of SVD - I - Image compression



Applications of SVD - II - Least squares problems

Applications of SVD - III - Handwritten digit classification

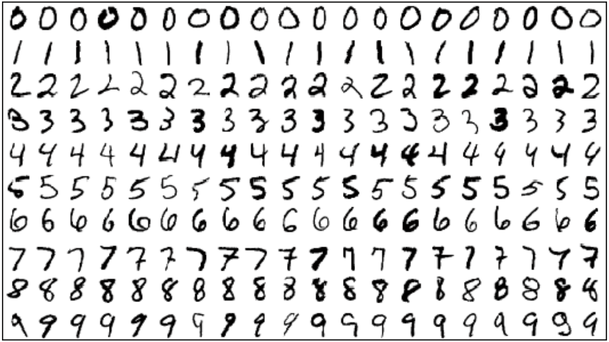
Problem: How to classify unknown digit?

Applications of SVD - III - Handwritten digit classification

Problem: How to classify unknown digit? Precisely, given a set of manually classified digits (the training set), classify a set of unknown digits (the test set).

Applications of SVD - III - Handwritten digit classification

Problem: How to classify unknown digit? Precisely, given a set of manually classified digits (the training set), classify a set of unknown digits (the test set).



A simple algorithm: Distance to the known digits

- Measure the distance between the unknown digit to the known digits using the Euclidean distance.

A simple algorithm: Distance to the known digits

- Measure the distance between the unknown digit to the known digits using the Euclidean distance.
- stack the columns of the image in a vector and identify each digit as a vector in $\mathbb{R}^{256}$.

A simple algorithm: Distance to the known digits

- Measure the distance between the unknown digit to the known digits using the Euclidean distance.
- stack the columns of the image in a vector and identify each digit as a vector in $\mathbb{R}^{256}$.
- Then define the distance function

$$d(x, y) = ||x - y||_2 = \sqrt{(x_1 - y_1)^2 + \cdots + (x_{256} - y_{256})^2}.$$

A simple algorithm: Distance to the known digits

- Measure the distance between the unknown digit to the known digits using the Euclidean distance.
- stack the columns of the image in a vector and identify each digit as a vector in $\mathbb{R}^{256}$.
- Then define the distance function

$$d(x, y) = ||x - y||_2 = \sqrt{(x_1 - y_1)^2 + \cdots + (x_{256} - y_{256})^2}.$$

- All the digits of one kind in the training set form a cluster of points in the Euclidean space $\mathbb{R}^{256}$. (Assumption)

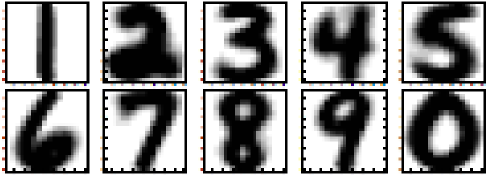
A simple algorithm: Distance to the known digits

- Measure the distance between the unknown digit to the known digits using the Euclidean distance.
- stack the columns of the image in a vector and identify each digit as a vector in $\mathbb{R}^{256}$.
- Then define the distance function

$$d(x, y) = ||x - y||_2 = \sqrt{(x_1 - y_1)^2 + \cdots + (x_{256} - y_{256})^2}.$$

- All the digits of one kind in the training set form a cluster of points in the Euclidean space $\mathbb{R}^{256}$. (Assumption)
- Ideally the clusters are well separated, and the separation between the clusters depends on how well written the training digits are.

The means (“averages“) of all digits in the training set.



A simple algorithm: Distance to the known digits

Algorithm

- Given the manually classified training set, compute the means m_i, $i = 0, 1, 2, \ldots, 9$, of all the 10 digits.

A simple algorithm: Distance to the known digits

Algorithm

- Given the manually classified training set, compute the means m_i, $i = 0, 1, 2, \ldots, 9$, of all the 10 digits.
- For each digit in the test set, classify it as k if m_k is the closest mean.

A simple algorithm: Distance to the known digits

Algorithm

- Given the manually classified training set, compute the means m_i, $i = 0, 1, 2, \ldots, 9$, of all the 10 digits.
- For each digit in the test set, classify it as k if m_k is the closest mean.

For some test set, the success rate of this algorithm is around 75%.

A simple algorithm: Distance to the known digits

Algorithm

- Given the manually classified training set, compute the means m_i, $i = 0, 1, 2, \ldots, 9$, of all the 10 digits.
- For each digit in the test set, classify it as k if m_k is the closest mean.

For some test set, the success rate of this algorithm is around 75%. The

reason for the relatively bad performance is that the algorithm does not use any information about the variation within each class of digits.

A simple algorithm: Distance to the known digits

Algorithm

- Given the manually classified training set, compute the means m_i, $i = 0, 1, 2, \ldots, 9$, of all the 10 digits.
- For each digit in the test set, classify it as k if m_k is the closest mean.

For some test set, the success rate of this algorithm is around 75%. The

reason for the relatively bad performance is that the algorithm does not use any information about the variation within each class of digits.

Using singular value decomposition(SVD), we will see a classification algorithm, for which the success rate is around 93%.

Algo. for classification of handwritten digits using SVD

- Let us consider the 16×16 matrix representation of the image as vector in $\mathbb{R}^{256\times 1}$, by stacking all the columns of the image above each other.

Algo. for classification of handwritten digits using SVD

- Let us consider the 16×16 matrix representation of the image as vector in $\mathbb{R}^{256\times 1}$, by stacking all the columns of the image above each other.
- The matrix consisting of all the training digits of one kind. the $3's$ say, is an element of the space $A \in \mathbb{R}^{256\times n}$.

Algo. for classification of handwritten digits using SVD

- Let us consider the 16×16 matrix representation of the image as vector in $\mathbb{R}^{256\times 1}$, by stacking all the columns of the image above each other.
- The matrix consisting of all the training digits of one kind. the $3's$ say, is an element of the space $A \in \mathbb{R}^{256\times n}$.
- Each column of A represents an image of a digit 3.

Algo. for classification of handwritten digits using SVD

- Let us consider the 16×16 matrix representation of the image as vector in $\mathbb{R}^{256\times 1}$, by stacking all the columns of the image above each other.
- The matrix consisting of all the training digits of one kind. the $3's$ say, is an element of the space $A \in \mathbb{R}^{256\times n}$.
- Each column of A represents an image of a digit 3. If $A = \sum_{i=1}^{m} \sigma_i u_i v_i^T$, then the left singular vectors u_i forms an orthonormal basis for the range space of A. i.e., "the image of $3's$".

Algo. for classification of handwritten digits using SVD

- Let us consider the 16×16 matrix representation of the image as vector in $\mathbb{R}^{256\times 1}$, by stacking all the columns of the image above each other.
- The matrix consisting of all the training digits of one kind. the $3's$ say, is an element of the space $A \in \mathbb{R}^{256\times n}$.
- Each column of A represents an image of a digit 3. If $A = \sum_{i=1}^{m} \sigma_i u_i v_i^T$, then the left singular vectors u_i forms an orthonormal basis for the range space of A. i.e., "the image of $3's$".
- Each digit is well characterized by a few of the first left singular values of its own kind.

Algo. for classification of handwritten digits using SVD

- Let us consider the 16×16 matrix representation of the image as vector in $\mathbb{R}^{256\times 1}$, by stacking all the columns of the image above each other.
- The matrix consisting of all the training digits of one kind. the $3's$ say, is an element of the space $A \in \mathbb{R}^{256\times n}$.
- Each column of A represents an image of a digit 3. If $A = \sum_{i=1}^{m} \sigma_i u_i v_i^T$, then the left singular vectors u_i forms an orthonormal basis for the range space of A. i.e., "the image of $3's$".
- Each digit is well characterized by a few of the first left singular values of its own kind.

Algorithm

Training: For the training set of known digits, compute the SVD of each set of digits of one kind.
Classification: For a given test digit, compute its relative residual in all 10 bases. If one residual is significantly smaller than all the others, classify as that. Otherwise give up.