충북대학교 전자정보대학
소프트웨어학과

특이값 분해
Singular Value Decomposition

Keon M. Lee

- ❖ 특이값
- ❖ 특이값 분해 (SVD)
- ❖ 특이값 분해 계산 방법
- ❖ 특이값 분해 활용분야

특이값 (singular value)

❖ mxn 행렬 A의 특이값(singular value)

- A^TA의 eigenvalue에 제곱근값

$$\sigma_i(A) = \sqrt{\lambda_i(A^\top A)}$$

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

$$A^\top A = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$ 대칭 행렬

$$\lambda = \{36, 9, 0\}$$

$$\sigma_1 = 6,\ \sigma_2 = 3,\ \sigma_3 = 0$$

특이값 분해

❖ 특이값 분해 (Singular Value Decomposition)

- 임의의 m×n 형렬 A를 $\mathbf{A} = U\Sigma V^{\mathrm{T}}$ 로 분해하는 것

  U 는 m×m 직교 행렬

  Σ 는 m×n 행렬로 처음 r개의 값이 특이값 $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$인 대각행렬

  V는 n×n 직교 행렬

$$\underset{A}{\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix}} = \underset{U}{\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}} \underset{V^{\top}}{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \sqrt{0.2} & 0 & 0 & 0 & \sqrt{0.8} \\ 0 & 0 & 0 & 1 & 0 \\ -\sqrt{0.8} & 0 & 0 & 0 & \sqrt{0.2} \end{bmatrix}}$$

특이값 분해

❖ **선형변환** $L(\mathbf{x}) = A\mathbf{x} : \mathcal{R}^n \to \mathcal{R}^m$

- A^TA의 **정규직교기저**(orthonormal basis) 존재 $\{\mathbf{v}_1 \cdots \mathbf{v}_n\} \in \mathcal{R}^n$
  - A^TA가 nxn 공간의 **대칭행렬**이므로, **대각화가능**(diagonalizable)하여 **고유값 분해**를 통해 정규직교기저 추출 가능
- $A\mathbf{v}_1, \cdots, A\mathbf{v}_n$ 은 서로 직교(orthogonal)
- $||A\mathbf{v}_i|| = \sigma_i$

From $\sigma_i = \sqrt{\lambda_i}$ for $A^\top A$, we have $\sigma_i^2 = \lambda_i$.

$$<A\mathbf{v}_i, A\mathbf{v}_j> = (A\mathbf{v}_i)^\top (A\mathbf{v}_j) = \mathbf{v}_i^\top A^\top A\mathbf{v}_j$$

$$= \mathbf{v}_i^\top \lambda_j \mathbf{v}_j = \lambda_j \mathbf{v}_i^\top \mathbf{v}_j = \sigma_j^2(\mathbf{v}_i^\top \mathbf{v}_j) = \begin{cases} 0 & \text{if } i \neq j \\ \sigma_j^2 & \text{if } i = j \end{cases}$$

$$||A\mathbf{v}_i|| = \sqrt{\langle A\mathbf{v}_i, A\mathbf{v}_i \rangle} = \sqrt{(A\mathbf{v}_i)^\top (A\mathbf{v}_i)}$$

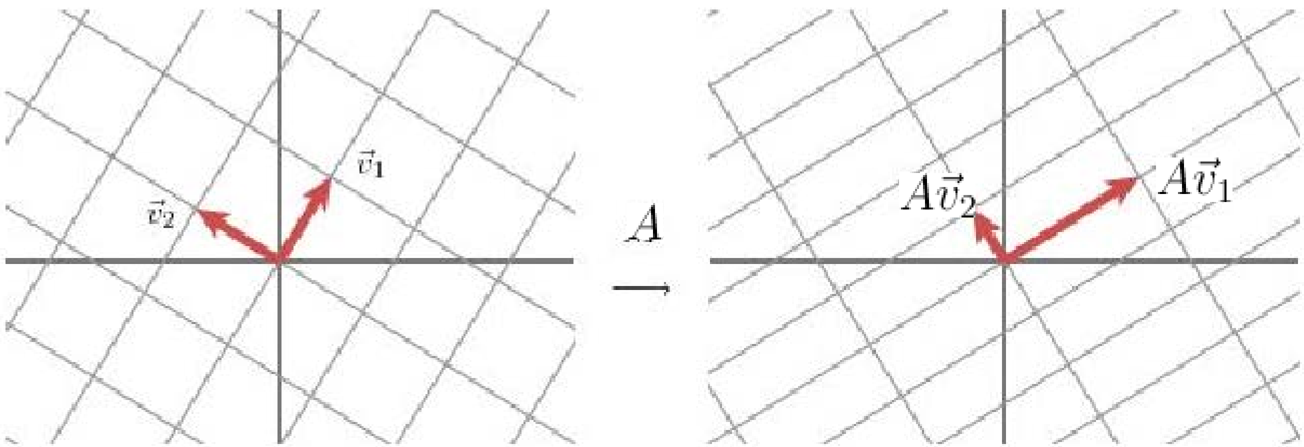
$$= \sqrt{\mathbf{v}_i^\top A^\top A\mathbf{v}_i} = \sqrt{\mathbf{v}_i^\top \sigma_i^2 \mathbf{v}_i} = \sqrt{\sigma_i^2} = \sigma_i$$

특이값 분해

❖ 선형변환 $L(\mathbf{x}) = A\mathbf{x} : \mathcal{R}^n \rightarrow \mathcal{R}^m$

$$\{\mathbf{v}_1 \cdots \mathbf{v}_n\} \in \mathcal{R}^n \quad \Rightarrow \quad \{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\} \in \mathcal{R}^m$$

$A^{\mathrm{T}}A$의 **정규직교기저**



$$||A\mathbf{v}_i|| = \sigma_i$$

특이값 분해

❖ **A^TA**의 eigenvector를 사용한 **정규 직교기저 $\mathbf{v}_1$, …, $\mathbf{v}_n$**

$$A^\top A\mathbf{v}_1 = \sigma_1^2\mathbf{v}_1,\ ||\mathbf{v}_1|| = 1,\ \cdots,\ A^\top A\mathbf{v}_r = \sigma_r^2\mathbf{v}_r,\ ||\mathbf{v}_r|| = 1$$

$$A^\top A\mathbf{v}_{r+1} = 0\,\mathbf{v}_{r+1} = \mathbf{0},\ ||\mathbf{v}_{r+1}|| = 1,\ \cdots,\ A^\top A\mathbf{v}_n = \mathbf{0},\ ||\mathbf{v}_n|| = 1$$

$$\mathbf{u}_1 = A\mathbf{v}_1/\sigma_1, \cdots, \mathbf{u}_r = A\mathbf{v}_r/\sigma_r$$

$$\begin{aligned}\langle\mathbf{u}_i, \mathbf{u}_j\rangle &= \langle A\mathbf{v}_i/\sigma_i, A\mathbf{v}_j/\sigma_j\rangle = \mathbf{v}_i^\top A^\top A\mathbf{v}_j/(\sigma_i\sigma_j) \\ &= \sigma_j^2\mathbf{v}_i^\top\mathbf{v}_j/(\sigma_i\sigma_j) = \begin{cases}0 & \text{if } i \neq j \\ 1 & \text{if } i = j\end{cases}\end{aligned}$$

$$\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r\} \Rightarrow \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r, \cdots, \mathbf{u}_m\}$$

$$\begin{aligned}A\mathbf{v}_i &= \sigma_i\mathbf{u}_i \text{ for } i = 1, \cdots, r, \\ A\mathbf{v}_i &= \mathbf{0} \text{ for } i = r+1, \cdots, m\end{aligned}$$

특이값 분해

❖ SVD 존재에 대한 증명

(mxn)(nxn) (mxn)

$$A[\mathbf{v}_1 \cdots \mathbf{v}_r \mathbf{v}_{r+1} \cdots \mathbf{v}_n] = [\sigma_1 \mathbf{u}_1 \cdots \sigma_r \mathbf{u}_r \ \mathbf{0} \cdots \mathbf{0}]$$

$$= [\mathbf{u}_1 \cdots \ \mathbf{u}_r \ \mathbf{0} \cdots \mathbf{0}] \begin{bmatrix} \sigma_1 & & & \mathbf{0} \\ & \ddots & & \vdots \\ & & \sigma_r & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix} = [\mathbf{u}_1 \cdots \ \mathbf{u}_r \ \mathbf{u}_{r+1} \cdots \mathbf{u}_m] \begin{bmatrix} \sigma_1 & & & \mathbf{0} \\ & \ddots & & \vdots \\ & & \sigma_r & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

(mxm) (mxn) (mxm) (mxn)

$$= U\Sigma$$

$$AV = U\Sigma$$

$$A = U\Sigma V^{-1} = U\Sigma V^{\top}$$ because V is an orthogonal basis.

$$= \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^{\top}$$

SVD 방법

❖ SVD 계산 방법

$$A = U\Sigma V^\top$$

$$A^\top A = (U\Sigma V^\top)^\top U\Sigma V^\top = V\Sigma^\top U^\top U\Sigma V^\top = V\Sigma^\top \Sigma V = V\Sigma' V$$

$$A^\top AV = V\Sigma' V^\top V$$

$$(A^\top A)V = V\Sigma'$$

$$AA^\top = U\Sigma V^\top (U\Sigma V^\top)^\top = U\Sigma V^\top V\Sigma^\top U^\top = U\Sigma\Sigma^\top U^\top = U\Sigma'' U^\top$$

$$AA^\top U = U\Sigma'' U^\top U$$

$$(AA^\top)U = U\Sigma''$$

SVD 예

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

$$A^\top A = \begin{bmatrix} 9 & -9 \\ -9 & 1 \end{bmatrix} \qquad \lambda_1 = 18, \mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \qquad \lambda_2 = 0, \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$V = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad \sigma_1 = \sqrt{18}, \quad \sigma_2 = 0$$

$$\Sigma = \begin{bmatrix} \color{red}{\sqrt{18}} & 0 \\ 0 & \color{red}{0} \\ 0 & 0 \end{bmatrix}$$

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix} \qquad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{u}_1 = A\mathbf{v}_1/\sigma_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

2 orthogonal vectors s.t. $\mathbf{u}_1^\top \mathbf{x} = 0$

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

By Gram-Schmidt procedure

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix} \qquad \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

SVD 예

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \qquad A = U\Sigma V^{\top}$$

$$A = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} \sqrt{18} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Octave / Matlab

```
>> A= [1 -1; -2 2; 2 -2]

A =

     1    -1
    -2     2
     2    -2
```

```
>> [U, S, V] = svd(A)

U =

    0.3333   -0.9428    0.0000
   -0.6667   -0.2357   -0.7071
    0.6667    0.2357   -0.7071


S =

    4.2426         0
         0    0.0000
         0         0


V =

    0.7071    0.7071
   -0.7071    0.7071
```

SVD와 rank

❖ 행렬의 rank

- SVD에서 0이 아닌 **특이값**(singular value)의 **개수**에 해당

$$N(A^\top A) = N(A)$$

$\Rightarrow$

Let $\mathbf{v} \in N(A^\top A)$. It implies $\mathbf{v}^\top A^\top A\mathbf{v} = \mathbf{v}^\top \mathbf{0} = 0$.

$$(A\mathbf{v})^\top (A\mathbf{v}) = (A\mathbf{v})^\top (A\mathbf{v}) = ||A\mathbf{v}||^2 = 0$$

$$A\mathbf{v} = \mathbf{0}$$

$\therefore$ If $\mathbf{v} \in N(A^\top A)$ then $\mathbf{v} \in N(A)$.

$\Leftarrow$

Let $\mathbf{v} \in N(A)$.

$$A\mathbf{v} = \mathbf{0}$$

$$A^\top A\mathbf{v} = A^\top \mathbf{0} = \mathbf{0}$$

$\therefore$ If $\mathbf{v} \in N(A)$ then $\mathbf{v} \in N(A^\top A)$.

rank-nullity theorem

$$nullity(A) + rank(A) = m$$

$$nullity(A^\top A) = nullity(A)$$

$$rank(A^\top A) = rank(A)$$

특이값은 A^TA에 대한 eigenvalue에 의해 결정되고,
eigenvalue가 0인 A**x** = 0를 만족하는 것이 nullity에 대응하므로,
0이 아닌 eigenvalue의 개수(중복도 고려)가 rank가 됨

SVD를 이용한 Low-rank 근사

❖ **Low-rank 근사 (low-rank approximation)**

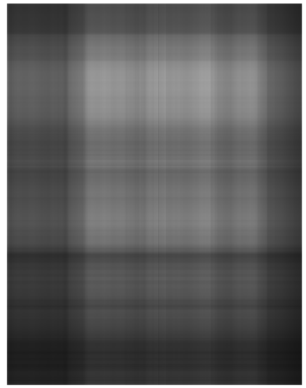
$$\underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \bullet & & \\ & \bullet & \\ & & \bullet \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{V^T}$$

$$A = U\Lambda V^\top = \sum_{i=1}^{r} \lambda_i \mathbf{u}_i \mathbf{v}_i^\top \qquad \tilde{A} = \sum_{i=1}^{k} \lambda_i \mathbf{u}_i \mathbf{v}_i^\top \qquad (k < r)$$

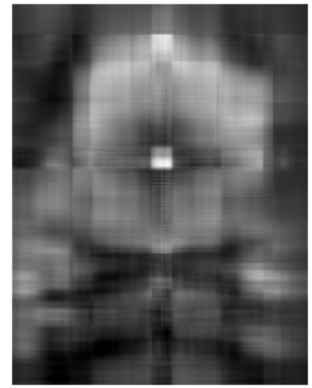
$$\begin{aligned} ||A - \tilde{A}||_F^2 &= tr[(U\Lambda V^\top - U\Lambda_k V^\top)^\top (U\Lambda V^\top - U\Lambda_k V^\top)] \\ &= tr[V(\Lambda - \Lambda_k)^\top U^\top U(\Lambda - \Lambda_k)V^\top] \\ &= tr[V^\top V(\Lambda - \Lambda_k)^\top (\Lambda - \Lambda_k)] \\ &= tr[(\Lambda - \Lambda_k)^\top (\Lambda - \Lambda_k)] \\ &= \sum_{i=k+1}^{\min(m,n)} \lambda_i^2 \end{aligned}$$

SVD를 이용한 Low-rank 근사

600×465 pixels



Rank-1 근사



Rank- 5 근사

Rank-20 근사

Rank-30 근사

Rank-50 근사

Rank-100 근사

Rank-150 근사

SVD와 역행렬

❖ nxn 행렬 A의 역행렬

- 모든 특이값(singular value)이 0이 아닐 때 가역행렬

$$N(A^\top A) = N(A)$$

- A의 SVD

$$A = U\Sigma V^\top$$

- A의 역행렬

$$A^{-1} = V\Sigma^{-1}U^\top \qquad \Sigma^{-1} = diag(1/\sigma_1, \ldots, 1/\sigma_n)$$

$$AA^{-1} = (U\Sigma V^\top)(V\Sigma^{-1}U^\top) = U\Sigma(V^\top V)\Sigma^{-1}U^\top$$

$$= U\Sigma\Sigma^{-1}U^\top = UU^\top = I$$

SVD와 최소제곱문제의 해

❖ 최소제곱문제(least-squares problem)

- A**x** = **b** (변수보다 데이터가 많고, 정확한 해가 없는 문제)
- 최소제곱해

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||A\mathbf{x} - \mathbf{b}||^2$$

$$\hat{\mathbf{x}} = (A^\top A)^{-1} A^\top \mathbf{b} = A_g \mathbf{b}$$

$$A_g = V \Lambda_g U^\top = \sum_{i=1}^{p} \lambda_i^* \mathbf{v}_i \mathbf{u}_i^\top \quad \text{where} \quad \lambda_i^* = \begin{cases} \lambda_i^{-1} & \text{if } \lambda_i \neq 0 \\ 0 & \text{if } \lambda_i = 0 \end{cases}$$

SVD 적용분야

- ❖ 데이터 축소 (Data Reduction)
- ❖ 최소제곱해 문제(Solving linear least square Problems)
- ❖ 영상처리 및 압축(Image Processing and Compression)
- ❖ 데이터 군집화에서 군집개수 결정(k-Selection for k-means clustering)
- ❖ 이상치 감지(Multivariate Outliers Detection)
- ❖ 잡음제거(Noise Filtering)
- ❖ 경향 분석(Trend detection in the observations and the variables)
- ❖ 행렬의 rank 결정

Summary

- 행렬 A의 **특이값**(singular value)은 Gram 행렬 A^TA의 eigenvalue에 제곱근값이다.
- **특이값 분해**(SVD)는 임의의 m×n 형렬 A를 $A = U\Sigma V^T$ 로 분해하는 것으로, U와 V는 직교행렬이고, Σ 는 대각성분에 특이값을 갖는 사각행렬이다.
- **행렬의 rank**는 SVD에서 0이 아닌 **특이값**(singular value)의 **개수**와 같다.
- SVD는 행렬에 대한 **low-rank 근사**를 통해 데이터를 압축하는데 사용될 수 있다.
- SVD는 **역행렬, pseudo-inverse** 등의 계산을 쉽게 할 수 있게 한다.
- SVD는 **데이터 압축, 영상 및 신호 처리, 행렬의 효과적인 연산** 등 다양한 분야에서 활용되고 있다.