# 8강. 자료의 처리

**[ 연 습 문 제 ]**

1. 다음 물음에 답하시오.
   (1) 세 점 $(0,-1),(1,2),(-1,0)$을 지나는 이차함수를 구하시오.
   (2) 네 점 $(-1,-2),(0,-4),(1,0),(2,16)$을 지나는 삼차함수를 구하시오.
2. 최소제곱법을 이용하여 다음 물음에 답하시오.
   (1) 세 점 $(0,0),(2,-1),(3,4)$에 근사한 일차함수를 구하시오.
   (2) 네 점 $(1,6),(2,1),(-1,5),(-2,2)$에 근사한 이차함수를 구하시오.
3. 제약조건 $x^2+y^2=1$ 하에서 $4x^2+2y^2$ 의 최댓값과 최솟값을 구하고, 그 때의 $x$와 $y$의 값을 각각 구하시오.
4. 제약조건 $x^2+y^2+z^2=1$ 하에서 $x^2-3y^2+8z^2$ 의 최댓값과 최솟값을 구하고, 그때의 $x$와 $y$의 값을 각각 구하시오.

1.

(1) $\begin{pmatrix} 1 & 0 & 0 & -1 \\ 1 & 1 & 1 & 2 \\ 1 & -1 & 1 & 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 3 \\ 0 & -1 & 1 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 2 & 4 \end{pmatrix}$

$\Rightarrow \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$ $\therefore f(x) = -1 + x + 2x^2$

(2) $\begin{pmatrix} 1 & -1 & 1 & -1 & -2 \\ 1 & 0 & 0 & 0 & -4 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 4 & 8 & 16 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & -4 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ $\therefore f(x) = -4 + 3x^2 + x^3$

2. (1) $A = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$, $A^T = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \end{pmatrix}$

$\rightarrow A^TA = \begin{pmatrix} 3 & 5 \\ 5 & 13 \end{pmatrix}$ & $B = \begin{pmatrix} 0 \\ -1 \\ 4 \end{pmatrix}$

$\rightarrow (A^TA)^{-1} = \frac{1}{14}\begin{pmatrix} 13 & -5 \\ -5 & 3 \end{pmatrix}$

$\therefore \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = (A^TA)^{-1}A^TB = \frac{1}{14}\begin{pmatrix} -11 \\ 15 \end{pmatrix}$

$\therefore f(x) = -\frac{11}{14} + \frac{15}{14}x$

3. $4x^2+2y^2 = (x\ y)\begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$

$\det(\lambda I_2 - A) = \det\begin{pmatrix} \lambda-4 & 0 \\ 0 & \lambda-2 \end{pmatrix} = (\lambda-4)(\lambda-2) = 0$

$\therefore \begin{cases} \lambda_1 = 4 \\ \lambda_2 = 2 \end{cases}$

① $\lambda_1 = 4$일 때 $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$ $\therefore V_1 = (1,0)$ ⇒ 최댓값

② $\lambda_2 = 2$ 일때 $\begin{pmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\therefore V_2 = (0,1)$ ⇒ 최솟값

— Index —



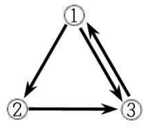
# 1. 우선순위 평가

## (1) 인접행렬

### ① 개념

요소간의 연결 관계를 나타내는 정사각 행렬

ex>



$$\Rightarrow \quad \begin{array}{c} \boxed{①②③} \\ \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array} \boxed{\begin{array}{c} ① \\ ② \\ ③ \end{array}}$$

연관된

연관한

— Index —

1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례

2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

Authority

### ② 권위벡터와 허브벡터

$n \times n$ 인접행렬 $A = (a_{ij})$ 에 대하여

$$\begin{pmatrix} \sum_{i=1}^{n} a_{i1} \\ \sum_{i=1}^{n} a_{i2} \\ \vdots \\ \sum_{i=1}^{n} a_{in} \end{pmatrix} \text{와} \begin{pmatrix} \sum_{j=1}^{n} a_{1j} \\ \sum_{j=1}^{n} a_{2j} \\ \vdots \\ \sum_{j=1}^{n} a_{nj} \end{pmatrix}$$ 을 각각 $A$의

권위벡터와 허브벡터라 하며, 각 벡터의 성분을 권위가중치와 허브가중치라 한다.

예

$$\begin{array}{c|cccc} & ① & ② & ③ & ④ \\ \hline ① & 1 & 1 & 0 & 1 \\ ② & 1 & 0 & 0 & 0 \\ ③ & 1 & 1 & 1 & 1 \\ ④ & 1 & 0 & 0 & 1 \end{array} \Rightarrow v_0 = \begin{pmatrix} 3 \\ 1 \\ 4 \\ 2 \end{pmatrix}$$

: Hub Vector

$u_0 = (4, 2, 1, 3)$ : Authority Vector

→ 얼마나 연관 됐는지

연쇄적 반응 (서로 영향을 주며 상호작용)

— Index —

1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례

2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

## (2) 순위평가 원리

인접행렬 $A$와 초기권위벡터 $u_0$ 와 초기허브벡터 $v_0$ 에 대하여

$$u_k = \begin{cases} u_0 & , k = 0 \\ \dfrac{A^T v_k}{\|A^T v_k\|} & , k > 0 \end{cases} \qquad v_k = \begin{cases} v_0 & , k = 0 \\ \dfrac{A u_{k-1}}{\|A u_{k-1}\|} & , k > 0 \end{cases}$$

와 같이 새로운 정규화된 권위벡터 $u_k$와 허브벡터 $v_k$를 정의한다. ($k$는 정수)

이때 $u_k$, $v_k$ 를 연립하면 다음과 같이

① 정규화 이유?

→ 너무 커지거나 작아지거나

②

$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 1 \\ 3 \end{pmatrix} = 4\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 9 \\ 4 \\ 10 \\ 7 \end{pmatrix} = v_1$$

$$\frac{v_1}{\|v_1\|} = \frac{Au_0}{\|Au_0\|}$$

정규화된 $u_k$와 $v_k$의 점화식을 얻을 수 있다.

$$u_k = \frac{A^T v_k}{\|A^T v_k\|} = \frac{A^T\left(\frac{Au_{k-1}}{\|Au_{k-1}\|}\right)}{\left\|A^T\left(\frac{Au_{k-1}}{\|Au_{k-1}\|}\right)\right\|}$$
$$= \frac{(A^T A)u_{k-1}}{\|(A^T A)u_{k-1}\|}$$

마찬가지로 $v_k = \frac{(AA^T)v_{k-1}}{\|(AA^T)v_{k-1}\|}$

이 벡터들이 안정화가 되었다고 판단되는 상태로부터 각각 최종 중요도를 판별한다.

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

# (3) 사례

* 검색 결과의 중요도를 판단할 때
Authority vs Hub
→ 참조된 가중치

10개의 인터넷 페이지(㉠~㉩)들 간의 인접행렬 $A$가 다음과 같다고 하자.

$$\begin{array}{cc} \begin{matrix} ㉠ & ㉡ & ㉢ & ㉣ & ㉤ & ㉥ & ㉦ & ㉧ & ㉨ & ㉩ \end{matrix} & \\ \begin{pmatrix} 0&1&0&0&1&0&0&1&0&0\\ 0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&1&1&0&0&0\\ 0&0&0&0&0&0&0&1&0&0\\ 0&1&1&1&1&0&0&1&0&1\\ 0&0&0&0&0&0&0&0&0&1\\ 0&0&0&0&1&0&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0\\ 0&0&0&0&0&1&0&0&0&0 \end{pmatrix} & \begin{matrix} ㉠\\㉡\\㉢\\㉣\\㉤\\㉥\\㉦\\㉧\\㉨\\㉩ \end{matrix} \end{array}$$

Authority + 정규화

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

(2)에서 소개된 절차에 따라 $A$의 정규화된 권위벡터가 안정화 될 때까지 반복계산한 결과는 다음과 같다.

변화가 크지 않음 (안정화)

| | $u_0$ | $u_1$ | $u_2$ | | $u_9$ | $u_{10}$ | |
|---|---|---|---|---|---|---|---|
| ㄱ | 0 | 0 | 0 | | 0 | 0 | |
| ㄴ | 0.2722 | 0.4006 | 0.4165 | | 0.4199 | 0.4199 | ③ |
| ㄷ | 0.1361 | 0.2337 | 0.2492 | | 0.2534 | 0.2534 | ⑤ |
| ㄹ | 0.1361 | 0.2337 | 0.2492 | | 0.2534 | 0.2534 | ⑤ |
| ㅁ | 0.6804 | 0.6509 | 0.6341 | | 0.6260 | 0.6260 | ① |
| ㅂ | 0.4083 | 0.1669 | 0.0632 | ⋯ | 0.0001 | 0.0001 | |
| ㅅ | 0.1361 | 0.0668 | 0.0260 | | 0.0000 | 0.0000 | |
| ㅇ | 0.4083 | 0.4506 | 0.4667 | | 0.4717 | 0.4717 | ② |
| ㅈ | 0 | 0 | 0 | | 0 | 0 | |
| ㅊ | 0.2722 | 0.2670 | 0.2789 | | 0.2846 | 0.2846 | ④ |

권위와 허브가 엮이면서 순서가 확 바뀌는 경우도 존재

다음의 우선순위로 검색결과 노출

따라서 $u_{10}$ 권위가중치로부터 페이지 ㉠, ㉥, ㉦, ㉨ 는 관련이 적고, 그 외의 페이지는 중요도가 높은 것부터 ㉤ > ㉧ > ㉡ > ㉩ > ㉢=㉣ 순서대로 검색엔진에서 노출되어야 함을 알 수 있다.

Google 검색엔진 원리

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

# 2. 자료압축

## (1) 특잇값 분해 Singular Value Decomposition

### ① 분해

한 행렬을 여러 행렬들의 곱으로 표현하는 것.

ex> QR분해, LU분해, LDU분해, 고윳값분해,
헤센버그 분해, 슈르 분해, 특잇값 분해 등

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

항상 정사각행렬이기 때문에 A가 정사각행렬이 아니어도 특잇값을 구할 수 있다.

### ② 특잇값

$m \times n$ 행렬 $A$에 대하여 $\lambda_1, \lambda_2, \cdots, \lambda_n$이 $A^TA$의 고윳값일 때

$$\sigma_1 = \sqrt{\lambda_1},\ \sigma_2 = \sqrt{\lambda_2},\ \cdots,\ \sigma_n = \sqrt{\lambda_n}$$

을 $A$의 특잇값이라 한다.

ex> 행렬 $A=\begin{pmatrix}1&1\\0&1\\1&0\end{pmatrix}$ 에 대하여

$A^TA=\begin{pmatrix}1&0&1\\1&1&0\end{pmatrix}\begin{pmatrix}1&1\\0&1\\1&0\end{pmatrix}=\begin{pmatrix}2&1\\1&2\end{pmatrix}$ 이므로

$A^TA$의 고유방정식은

$\lambda^2-4\lambda+3=(\lambda-1)(\lambda-3)=0$ 이다.

따라서 $A$의 두 특잇값은 각각

$\sqrt{3}$, $1$ 이다.

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

## ③ 특잇값 분해

영행렬이 아닌 임의의 $m\times n$행렬 $A$는 다음과 같이 나타낼 수 있다.

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

이때 $U$, $V$는 직교행렬이며, $\Sigma$는 주대각성분이 $A$의 특잇값이고 나머지 성분들은 0인 $m\times n$행렬이다.

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

ex> 행렬 $A=\begin{pmatrix}1&1\\0&1\\1&0\end{pmatrix}$ 는 다음과 같이

특잇값 분해된다.

기저

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

$\sigma_1$ $\sigma_2$

$v_1$ $v_2$ 고유값에 대응하는 고유벡터

$\frac{Av_1}{a_1}$ $\frac{Av_2}{a_2}$

앞의 둘과 정규직교인 벡터

## (2) 축소된 특잇값 분해 (Reduced SVD)

특잇값 분해에서 0인 성분들로만 이루어진, 대수적으로 무의미한 행 또는 열을 제거한 형태를 축소된 특잇값 분해라 한다. 즉,

$$A = U_1 \Sigma_1 V_1^T$$

$$= (u_1\, u_2\, \cdots\, u_k) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{pmatrix} \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{pmatrix}$$

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

또한 축소된 특잇값 분해를 이용하여 행렬 $A$를 다음과 같이 전개한 것을 $A$의 축소된 특잇값 전개라 한다.

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$$

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리

ex>

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{array}{c} \quad u_1 \qquad u_2 \qquad u_3 \\ \begin{pmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{pmatrix} \end{array} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{array}{c} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix} \end{array} \begin{array}{l} \leftarrow v_1^T \\ \leftarrow v_2^T \end{array}$$

$$= \sqrt{3}\, u_1 v_1^T + u_2 v_2^T$$

## (3) 자료압축 원리

압축되지 않은 $m \times n$ 행렬 $A$를 위한 필요 저장 공간은 $mn$이다.

$A$를 축소된 특잇값 분해한 결과가 $A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T$ 라면, 이제 필요한 저장 공간은 $k + km + kn = k(1+m+n)$이다. $(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k)$

충분히 작다고 판단되는 $\sigma_{r+1}, \cdots, \sigma_k$에 대응하는 항들을 추가로 제거하면, 이때 필요한 저장 공간은 $r(1+m+n)$ 뿐이다.

$k$ : 특잇값 개수 = $\Sigma$의 (행 개수, 열 개수)

$m$ : $U$의 행 개수 = $u_i$의 성분개수

$n$ : $V$의 열 개수 = $v_i^T$의 성분 개수

$$\begin{array}{cc} & n \\ m & \begin{pmatrix} mn \end{pmatrix} \end{array} = \begin{array}{cc} & k \\ m & \begin{pmatrix} km \end{pmatrix} \end{array} \begin{array}{cc} & k \\ k & \begin{pmatrix} k \end{pmatrix} \end{array} \begin{array}{cc} & n \\ k & \begin{pmatrix} kn \end{pmatrix} \end{array}$$

대각성분이라 $k^2$이 아닌 $k$개 필요

— Index —
1. 우선순위 평가
(1) 인접행렬
(2) 순위평가 원리
(3) 사례
2. 자료압축
(1) 특잇값 분해
(2) 축소된 특잇값 분해
(3) 자료압축 원리