

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.

— Index —

1. 우선순위 평가

(1) 인접행렬

(2) 순위평가 원리

(3) 사례

2. 자료압축

(1) 특잇값 분해

(2) 축소된 특잇값 분해

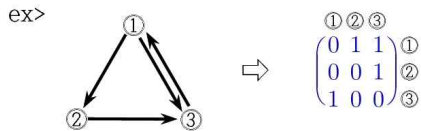
(3) 자료압축 원리

1. 우선순위 평가

(1) 인접행렬

① 개념

요소간의 연결 관계를 나타내는 정사각
행렬



— Index —

1. 우선순위 평가

(1) 인접행렬

(2) 순위평가 원리

(3) 사례

2. 자료압축

(1) 특잇값 분해

(2) 축소된 특잇값 분해

(3) 자료압축 원리

② 권위벡터와 허브벡터

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

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

권위벡터와 허브벡터라 하며, 각 벡터의
성분을 권위가중치와 허브가중치라 한다.

— 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

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

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

— Index —

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

정규화된 u_k 와 v_k 의 점화식을 얻을 수 있다.

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

$$= \frac{(A^T A) u_{k-1}}{\|(A^T A) u_{k-1}\|}$$

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

이 벡터들이 안정화가 되었다고 판단되는 상태에서부터 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

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

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

— 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.

$$\begin{matrix} u_0 & u_1 & u_2 & u_9 & u_{10} \\ \begin{pmatrix} 0 \\ 0.2722 \\ 0.1361 \\ 0.1361 \\ 0.6804 \\ 0.4083 \\ 0.1361 \\ 0.4083 \\ 0 \\ 0.2722 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.4006 \\ 0.2337 \\ 0.2337 \\ 0.6509 \\ 0.1669 \\ 0.0668 \\ 0.4506 \\ 0 \\ 0.2670 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.4165 \\ 0.2492 \\ 0.2492 \\ 0.6341 \\ 0.0632 \\ 0.0260 \\ 0.4667 \\ 0 \\ 0.2789 \end{pmatrix} & \cdots & \begin{pmatrix} 0 \\ 0.4199 \\ 0.2534 \\ 0.2534 \\ 0.6260 \\ 0.0001 \\ 0.0000 \\ 0.4717 \\ 0 \\ 0.2846 \end{pmatrix} & \begin{pmatrix} 0 \\ 0.4199 \\ 0.2534 \\ 0.2534 \\ 0.6260 \\ 0.0001 \\ 0.0000 \\ 0.4717 \\ 0 \\ 0.2846 \end{pmatrix} \end{matrix}$$

— Index —

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

따라서 u_{10} 권위가중치로부터 페이지 ㉟,
 ㉟, ㉟, ㉟ 는 관련이 적고, 그 외의
 페이지는 중요도가 높은 것부터 ㉟ > ㉟
 > ㉟ > ㉟ > ㉟=㉟ 순서대로
 검색엔진에서 노출되어야 함을 알 수
 있다.

— 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

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

② 특잇값

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

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

을 A 의 특잇값이라 한다.

— 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 $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$ 에 대하여

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

$A^T A$ 의 고유방정식은

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

따라서 A 의 두 특잇값은 각각

$$\sqrt{3}, 1 \text{ 이다.}$$

— 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, Σ 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.

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

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

— Index —

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

(2) 축소된 특잇값 분해

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

$$\begin{aligned}
 A &= U_1 \Sigma_1 V_1^T \\
 &= (u_1 \ u_2 \ \cdots \ u_k) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{pmatrix} \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_k^T \end{pmatrix}
 \end{aligned}$$

— 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 전개**라 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

$$\begin{aligned}
 \text{EX>} \quad & \begin{matrix} u_1 & u_2 & u_3 \end{matrix} \\
 \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} &= \begin{pmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{pmatrix} \begin{matrix} \leftarrow v_1^T \\ \leftarrow v_2^T \end{matrix} \\
 &= \sqrt{3} u_1 v_1^T + u_2 v_2^T
 \end{aligned}$$

— Index —

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

(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.

— Index —

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

