STA0067: Regression Analysis II **Spring 2024**

Lecture 01: Matrix Theory

Lecturer: Kipoong Kim

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

1.1 행렬 이론

1.2 기본 이론

1.2.1 행렬의 종류

(1) (i,j)번째 성분이 a_{ij}인 $m \times n$ 행렬 $\mathbf{A}$는 다음과 같이 주어집니다

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

그리고 $\mathbf{A} = (a_{ij})$ 또는 $\{a_{ij}\}$로 표기됩니다.

(2) $\mathbf{A}$의 j번째 열벡터:

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

$\mathbf{A}$의 i번째 행벡터: $\mathbf{A}_i = (a_{i1}, a_{i2}, \cdots, a_{in})$ 또는 열벡터와의 혼동을 피하기 위해 $\mathbf{a}_i$로 표기하기도 함.

(3) 정방 (square) 행렬 : $m = n$인 경우

(4) 대각 (diagonal) 행렬 : 비대각원소 $a_{ij} = 0$인 $(i \neq j)$ 정방행렬, $\mathbf{A} = \mathrm{diag}\,(a_{11}, \ldots, a_{nn})$로 표기됨

(5) 단위행렬 또는 항등 (identity) 행렬 : $a_{ii} = 1, i = 1, \cdots, n$인 대각행렬, $\mathbf{I}$ 또는 $\boldsymbol{I}_n$으로 표기됨

(6) 전치 (transpose) 행렬: $m \times n$ 행렬 $\mathbf{A} = (a_{ij})$의 전치 (transpose) 행렬은 $n \times m$ 행렬 $\mathbf{A}' = (a_{ji})$이며, $\mathbf{A}^T$ 또는 $\mathbf{A}'$로도 표기됩니다. 또한, $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$임이 명확합니다.

(7) 멱등 (idempotent) 행렬: 정방행렬 $\mathbf{A}$가 $\mathbf{A}^2 = \mathbf{A}$를 만족하면 멱등 (idempotent) 행렬이라고 부릅니다.

1.2.2 대각합 (trace)

(1) 정방행렬 $\mathbf{A}$의 대각합은 $\mathbf{A}$의 대각 원소들의 합이며, $\mathrm{tr}(\mathbf{A})$로 표기됩니다. 따라서, $\mathrm{tr}(\mathbf{A}) = \sum a_{ii}$입니다.

(2) 대각합의 성질

$$
\begin{aligned}
&\mathrm{tr}(\mathbf{A} \pm \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \pm \mathrm{tr}(\mathbf{B}) \\
&\mathrm{tr}(k\mathbf{A}) = k\,\mathrm{tr}(\mathbf{A}), \quad k : \text{상수} \\
&\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}),
\end{aligned}
$$

여기서 $\mathbf{AB}$와 $\mathbf{BA}$는 정의되어야 합니다.

1.2.3 행렬식 (determinant)

(1) $n \times n$ 정방행렬 $\mathbf{A}$의 행렬식은 $\det(A)$ 또는 $|\mathbf{A}|$로 표기되며, 다음과 같이 정의됩니다:

$$\det(\mathbf{A}) = \sum_{j=1}^{n} a_{ij}c_{ij}, \forall i = 1, \cdots, n,$$

여기서 $c_{ij} = (-1)^{i+j}d_{ij}$는 a_{ij}의 여인수라 불리며, d_{ij}는 $\mathbf{A}$에서 i번째 행과 j번째 열을 제거한 $(n-1) \times (n-1)$ 행렬 $\mathbf{A}_{(i,j)}$의 행렬식입니다.

(2) 예시

$n = 1$인 경우 : $|\mathbf{A}| = a_{11}$, 즉, 값 자체

$n = 2$인 경우 :

$$
\begin{aligned}
|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} &= a_{11}a_{22} - a_{12}a_{21}, \quad (i = 1 \text{ 경우 }) \\
&= -a_{21}a_{12} + a_{22}a_{11}, \quad (i = 2 \text{ 경우 })
\end{aligned}
$$

$n = 3$인 경우 :

$$
\begin{aligned}
|\mathbf{A}| = &\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
=&a_{11}c_{11} + a_{12}c_{12} + a_{13}c_{13} \\
=&a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
=&a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} \\
&+ a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22}
\end{aligned}
$$

(3) 행렬식의 성질

$$
\begin{aligned}
&|\mathbf{AB}| = |\mathbf{BA}| = |\mathbf{A}||\mathbf{B}|, \text{ 만약 } \mathbf{A} \text{ 와 } \mathbf{B} \text{ 가 } n \times n \text{ 정방행렬이면} \\
&|\mathbf{A}| = |\mathbf{A}'| \\
&|k\mathbf{A}| = k^n|\mathbf{A}|, k \text{ 는 상수}
\end{aligned}
$$

(4) 정방행렬 $\mathbf{A}$가 $|\mathbf{A}| \neq 0$이면 비특이(non-singular)라고 하고, $|\mathbf{A}| = 0$이면 특이(singular)라고 합니다.

1.3 역행렬 (inverse matrix)

1.3.1 선형 독립과 종속 (linearly independent & dependent)

(1) 선형 종속 (linearly dependent)

n차원 벡터 $\mathbf{v}_1, \cdots, \mathbf{v}_k$의 선형 결합이 $\mathbf{0}$일 때, 즉

$$c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$$

에서 적어도 하나의 $\{c_i\}$가 0이 아니라면, $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k$를 선형 종속이라고 한다.

(2) 선형 독립 (linearly independent)
반대로, 모든 c_i가 0이라면, $\mathbf{v}_1, \cdots, \mathbf{v}_k$를 선형 독립이라고 한다.

(3) **예제**: 두 벡터 $(1, 1)$와 $(-3, 2)$는 선형 독립이다. 왜냐하면

$$c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

는 $c_1 - 3c_2 = 0$, $c_1 + 2c_2 = 0$을 주고, 우리는 반드시 $c_1 = 0, c_2 = 0$을 가져야 하기 때문이다.

(4) **주의**: $\mathbf{v}_1, \cdots, \mathbf{v}_k$가 선형 종속이라고 가정하자, $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$는 0이 아닌 c_i를 가진다. 그러면, $c_j \neq 0$에 대해,

$$\mathbf{v}_j = -\frac{1}{c_j}\left(c_1\mathbf{v}_1 + \cdots + c_{j-1}\mathbf{v}_{j-1} + c_{j+1}\mathbf{v}_{j+1} + \cdots + c_k\mathbf{v}_k\right);$$

즉, $\mathbf{v}_1, \cdots, \mathbf{v}_k$가 선형 종속이면, 하나의 벡터는 다른 벡터들의 선형 결합으로 표현될 수 있다.

1.3.2 행렬의 계수 (rank)

(1) 정의 : $\mathbf{A}$를 m개의 행 벡터와 n개의 열 벡터를 가진 $m \times n$ 행렬이라 하자. m^*를 m개의 행 벡터 중 선형 독립인 벡터의 최대 개수라 하고, n^*를 n개의 열 벡터 중 선형 독립인 벡터의 최대 개수라 하자. 그러면, 우리는 반드시 $m^* = n^*$를 가지며, 이를 $\mathbf{A}$의 계수라 하고, $r(\mathbf{A})$로 표기한다. 따라서, $r(\mathbf{A}) \leq \min(m, n)$이다.

(2) 예제 : 5×4 행렬 $\mathbf{A}$의 계수를 계산하라.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 3 \\ 3 & 0 & 6 \\ 2 & 1 & 3 \end{bmatrix}$$

먼저, 우리는 $r(\mathbf{A}) \leq 4$를 가져야 한다.
첫 번째와 두 번째 행 벡터가 선형 독립임을 주목하고, 3–4 번째 행 벡터는 첫 번째와 두 번째 행 벡터의 선형 결합으로 표현될 수 있다. 예를 들어, 세 번째 행 벡터는 $(1, 2, 0, 1) + 2(1, -1, 3, 2)$로 표현될 수 있다. 따라서, $r(\mathbf{A}) = 2$이다.

(3) 계수의 성질

(a) $r(\mathbf{AB}) \leq \min(r(\mathbf{A}), r(\mathbf{B}))$

(b) $n \times n$ 행렬 $\mathbf{A}$가 비특이행렬이면, $r(\mathbf{A}) = n$이고, 특이행렬이면, $r(\mathbf{A}) < n$이다.

(c) $\mathbf{A}$의 계수는 비특이행렬을 곱해도 변하지 않는다.

(d) $\mathbf{AGA} = \mathbf{A}$이면, $r(\mathbf{A}) = r(\mathbf{GA})$이다.

(e) $r(\mathbf{A} : \mathbf{B}) \leq r(\mathbf{A}) + r(\mathbf{B})$, 여기서 $\mathbf{A}$는 $m \times n_1$, $\mathbf{B}$는 $m \times n_2$이고, $\mathbf{A} : \mathbf{B}$는 $m \times (n_1 + n_2)$ 행렬로, 확장 또는 연결 행렬이라 한다. 예를 들어,

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \quad B = \begin{pmatrix} 7 & 9 \\ 8 & 10 \end{pmatrix}$$

$$A : B = \begin{pmatrix} 1 & 3 & 5 & 7 & 9 \\ 2 & 4 & 6 & 8 & 10 \end{pmatrix}$$

(f) $r(\mathbf{A}+\mathbf{B}) \leq r(\mathbf{A}:\mathbf{B}) \leq r(\mathbf{A}) + r(\mathbf{B})$

(g) $\mathbf{A}$가 $n \times n$ 행렬이면, $r(\boldsymbol{AB}) \geq r(\mathbf{A}) + r(\mathbf{B}) - n$이다.

(h) $\mathbf{A}$가 $n \times n$ 멱등행렬이면, $r(\boldsymbol{I} - \mathbf{A}) = n - r(\mathbf{A})$이다.

(i) $r(\mathbf{A}) = r(\mathbf{A}') = r(\mathbf{A}'\mathbf{A}) = r(\mathbf{A}\mathbf{A}')$

1.3.3 역행렬 (inverse matrix)

(1) 동기 : 방정식 $ax = b$의 해 x를 계산하기 위해, a의 역수, 즉 $1/a$를 양변에 곱합니다. 즉,

$$ax = b \Rightarrow \frac{1}{a}ax = \frac{1}{a}b \Rightarrow x = \frac{b}{a}$$

여기서 $a \neq 0$이어야 합니다. a의 역수가 존재하려면 $a \neq 0$일 때만 가능하기 때문입니다. 이제 n개의 방정식과 n개의 미지수를 고려해봅시다.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad & \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

이제 $\mathbf{A} = (a_{ij}), i = 1, \cdots, n; j = 1, \cdots, n$를 $n \times n$ 정방행렬, $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}', \mathbf{B} = \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix}'$라 하면, 위의 방정식들은 다음과 같이 표현될 수 있습니다:

$$\mathbf{Ax} = \mathbf{b}$$

만약 $\mathbf{A}^{-1}$가

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I},$$

를 만족한다면, 이를 $\mathbf{A}$의 역행렬이라고 부릅니다. 스칼라 a가 역수를 가지려면 $a \neq 0$이어야 합니다. 마찬가지로, 정방행렬 $\mathbf{A}$가 역행렬을 가지려면 $|\mathbf{A}| \neq 0$, 즉 비특이행렬이어야 합니다. 따라서, $\mathbf{A}$가 비특이행렬이면 $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \;\Rightarrow\; \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$입니다. 실제로, $\mathbf{A}^{-1}$는 다음과 같이 주어집니다:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ \vdots & \vdots & & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix}$$

여기서 c_{ij}는 a_{ij}의 여인수입니다.

(2) 성질

(a) $\mathbf{A}^{-1}$는 유일합니다.

(b) $\mathbf{A}^{-1}$는 비특이행렬입니다.

(c) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$

(d) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

(3) 직교행렬 (orthogonal matrix)

정방행렬 $\mathbf{A}$가 $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \boldsymbol{I}$를 만족하면, 이를 직교행렬이라고 부르며, 다음과 같은 성질을 가집니다:

(a) $\operatorname{tr}(\mathbf{A}'\mathbf{B}\mathbf{A}) = \operatorname{tr}(\mathbf{B})$

(b) $|\mathbf{A}'\mathbf{B}\mathbf{A}| = |\mathbf{B}|$

(c) $|\mathbf{A}| = \pm 1$

1.3.4 특수 행렬의 역행렬

(1) $\mathbf{J}$를 모든 성분이 1인 $n \times n$ 정방행렬이라 하고, $a \neq 0, a+nb \neq 0$이면, 다음이 성립합니다:

$$(a\mathbf{I}+b\mathbf{J})^{-1} = \frac{1}{a}\left(\boldsymbol{I} - \frac{b}{a+nb}\boldsymbol{J}\right)$$

(2) $\{\mathrm{diag}\,(a_1,\cdots,a_n)\}^{-1} = \mathrm{diag}\,(1/a_1,\cdots,1/a_n)\,,\quad a_i \neq 0,\ i=1,\cdots,n$

(3) $\boldsymbol{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{n-1} = (\mathbf{A}^n - \mathbf{I})\,(\mathbf{A}-\mathbf{I})^{-1}$

(4) $\left(\mathbf{I}+\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}(\mathbf{A}+\mathbf{I})^{-1}$

(5) $(\mathbf{I}+\mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I}+\mathbf{BA})^{-1}\mathbf{B}$

(6) $(\mathbf{A}+\mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBV}\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{UBV}\right)^{-1}\mathbf{A}^{-1}$

1.3.5 일반화 역행렬 (generalized inverse)

여기서는 $\mathbf{A}$가 비특이행렬일 때의 $\mathbf{A}$의 역행렬을 정의하고, $\mathbf{A}$가 정방행렬이 아닐 때의 $\mathbf{A}$의 역행렬도 정의합니다.

1. Moore–Penrose 역행렬 $\mathbf{A}$를 $p \times q$ 행렬이라 하면, $\mathbf{A}$의 Moore–Penrose 역행렬은 다음 3가지 조건을 만족하는 $q \times p$ 행렬 $\mathbf{M}$입니다;

(1) $\mathbf{AMA} = \mathbf{A}$

(2) $\mathbf{MAM} = \mathbf{M}$

(3) $\mathbf{AM}$와 $\mathbf{MA}$는 대칭행렬이다

Moore–Penrose 역행렬 $\mathbf{M}$은 유일하다는 점에 주목하세요. 예를 들어, 3×4 행렬

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{bmatrix}$$

의 Moore–Penrose 역행렬은

$$\mathbf{M} = \frac{1}{18}\begin{bmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{bmatrix}$$

입니다. 또한, $\mathbf{A}$가 정방행렬이면 $\mathbf{M}$ 역시 정방행렬이고 $\mathbf{M} = \mathbf{A}^{-1}$임에 주목하세요.

2. 일반화 역행렬 (generalized inverse) 위의 3가지 조건 중, $\mathbf{G}$가 첫 번째 조건만 만족한다면, 즉 $\mathbf{AGA} = \mathbf{A}$이면, $\mathbf{G}$를 $\mathbf{A}$의 일반화 역행렬이라 하고, $\mathbf{A}^-$로 표기합니다. 사실, $\mathbf{G}$는 <u>유일하지 않습니다</u>. 여기서는 $\mathbf{G}$를 계산하는 한 가지 대중적인 방법을 소개합니다.

먼저, $\mathbf{A}_{p\times q}$의 정방이고 비특이인 부분행렬 $\mathbf{A}_{11}$을 찾고, 다른 모든 성분을 $\mathbf{0}$로 둡니다. 마지막으로, $\mathbf{A}_{11}^{-1}$을 찾습니다. 즉,

$$\mathbf{A}_{p\times q} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \Rightarrow \mathbf{G}_{q\times p} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

예를 들어, 다음을 고려해봅시다

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 & 3 \\ 3 & -1 & 2 & -2 \\ 5 & -4 & 0 & -7 \end{bmatrix}$$

그리고

$$\mathbf{A}_{11} = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}, \mathbf{A}_{12} = \begin{bmatrix} 4 & 3 \\ 2 & -2 \end{bmatrix}, \mathbf{A}_{21} = \begin{bmatrix} 5 & -4 \end{bmatrix}, \mathbf{A}_{22} = \begin{bmatrix} 0 & -7 \end{bmatrix},$$

라 하면

$$\mathbf{G} = \frac{1}{7}\begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

입니다.

또 다른 예로, 일원분류를 고려해봅시다

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, 2, j = 1, 2, 3$$

즉,

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{bmatrix}$$

행렬 표기법으로,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

그러면 최소제곱법의 정규방정식은

$$(\mathbf{X}'\mathbf{X})\,\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}.$$

가 됩니다. 하지만,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix}$$

는 특이행렬이므로, $(\mathbf{X}'\mathbf{X})^{-1}$은 존재하지 않습니다. 따라서, $\boldsymbol{\beta}$의 추정량으로 일반화 역행렬 $(\mathbf{X}'\mathbf{X})^{-}$를 사용하고, 추정량은 $(\mathbf{X}'\mathbf{X})^{-}\,\mathbf{x}'\mathbf{y}$가 됩니다. 가능한 한 방법은 다음과 같습니다:

$$(\mathbf{X}'\mathbf{X})^{-} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

3. 역행렬 계산의 문제점 다음을 주목하세요:

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix}$$

은 $|\mathbf{A}| = 0$이므로 특이행렬이고, 따라서 역행렬이 존재하지 않습니다. 이제,

$$\mathbf{B} = \begin{bmatrix} 1.9998 & 0.9999 \\ 5.9994 & 3.0009 \end{bmatrix}$$

는 행렬식이 0.0024이므로 비특이행렬입니다. $\mathbf{B}$가 수학적으로는 비특이행렬이지만, 행렬식이 0에 매우 가깝기 때문에 거의 특이행렬에 가깝다는 점에 주목하세요. 이제 다음을 고려해봅시다:

$$\begin{bmatrix} 2.0 & 2.5 \\ 2.5 & 3.0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix},$$

이때 해는 $(0.1, 0.1)'$입니다. 그러나

$$\begin{bmatrix} 2.04 & 2.49 \\ 2.49 & 3.04 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.55 \end{bmatrix}$$

의 해는 $(-1, 1)'$이고, 행렬식은 0에 매우 가깝습니다(실제로 0.0015입니다). 이 경우, 성분의 작은 변화가 해의 큰 변화를 초래할 수 있습니다. 행렬식의 값이 작은 행렬을 나쁜 조건의 행렬(ill-conditioned matrix)이라고 부르며, 이는 $n \times p$ 행렬 A의 가장 큰 특이값과 가장 작은 특이값의 비율로 정의되는 조건수로 측정할 수 있습니다. 조건수가 크면 행렬이 나쁜 조건을 가질 수 있습니다.

1.4 분할 행렬 (partitioned matrix)

행렬 $\boldsymbol{P}$를 분할하는 것을 고려해봅시다.

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

그러면, 다음 항등식에 의해,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

우리는 다음을 얻습니다:

$$\begin{aligned} \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} &= |\mathbf{A}| \left|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right|, \ \text{만약 } \mathbf{A}^{-1} \text{ 가 존재한다면} \\ &= |\mathbf{D}| \left|\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right|, \ \text{만약 } \mathbf{D}^{-1} \text{ 가 존재한다면} \end{aligned}$$

또한, $\boldsymbol{P}$의 역행렬은 다음과 같습니다:

$$\begin{aligned} &\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} \\ =&\begin{bmatrix} \boldsymbol{I} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1} & \left(\mathbf{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix}, \ \text{만약 } \mathbf{A}^{-1} \text{ 가 존재한다면} \\ =&\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{I} \\ -\mathbf{D}^{-1}\mathbf{C} \end{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} \left[\boldsymbol{I} - \mathbf{B}\mathbf{D}^{-1}\right], \ \text{만약 } \mathbf{D}^{-1} \text{ 가 존재한다면} \end{aligned}$$

예를 들어, 만약

$$\boldsymbol{P} = \begin{bmatrix} 2 & 0 & 0 & 1 & -1 \\ 0 & 2 & 0 & 1 & -1 \\ 0 & 0 & 2 & 1 & -1 \\ 2 & 2 & 2 & 1 & 0 \\ -1 & -1 & -1 & 0 & 1 \end{bmatrix}$$

라면 다음과 같이 둡니다:

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix},\ \mathbf{B} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix},$$
$$\mathbf{C} = \begin{bmatrix} 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix},\ \mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

그리고 우리는 다음을 얻습니다:

$$\begin{aligned} \mathbf{A}^{-1} &= \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix} \cdot \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 3 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \end{aligned}$$

따라서

$$|\boldsymbol{P}| = 8\left(1 - \frac{9}{2}\right) = -28$$
$$\boldsymbol{P}^{-1} = \frac{1}{14}\begin{bmatrix} 4 & -3 & -3 & 2 & -2 \\ -3 & 4 & -3 & 2 & -2 \\ -3 & -3 & 4 & 2 & -2 \\ 4 & 4 & 4 & 2 & 12 \\ -2 & -2 & -2 & 6 & 8 \end{bmatrix}$$

1.5 고유값과 고유벡터 (eigenvalues and eigenvectors)

1.5.1 고유값과 고유벡터

정방행렬 $\mathbf{A}$, 벡터 $\mathbf{u}$, 그리고 상수 λ에 대해 다음을 고려해봅시다:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$$

이는 다음과 동등합니다:

$$(\mathbf{A} - \lambda\boldsymbol{I})\mathbf{u} = \mathbf{0}$$

만약 $\mathbf{A} - \lambda\boldsymbol{I}$가 비특이행렬이면, $\mathbf{u}$에 대한 해는 $\mathbf{u} = \mathbf{0}$입니다. 하지만, $\mathbf{A} - \lambda\boldsymbol{I}$가 특이행렬이면, u와 상수 λ에 대한 영이 아닌 해가 존재합니다. 즉,

$$|\mathbf{A} - \lambda\mathbf{I}| = \mathbf{0}$$

는 $\mathbf{u}$와 상수 λ에 대한 영이 아닌 해를 제공합니다. 위 방정식을 $\mathbf{A}$의 특성방정식이라고 하며, $\mathbf{A}$가 $n \times n$이므로 이는 λ에 대한 n차 다항식입니다. n개의 해를 $\lambda_1, \cdots, \lambda_n$이라고 하면, 이들을 고유값, 특성근, 또는 잠재근이라고 부릅니다. 더 나아가, 각 λ_i에 대해 다음을 만족하는 벡터 $\mathbf{u}_i$를

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i, \quad i = 1, \cdots, n$$

λ_i에 대응하는 고유벡터라고 부릅니다.

예를 들어, 만약

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

이라면 특성방정식 (characteristic equation)은

$$\begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0$$

이고 고유값은 $\lambda = -5$ 또는 $\lambda = 7$입니다. 또한, 우리는 다음을 얻습니다:

$$\mathbf{A}\begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5\begin{bmatrix} 2 \\ -3 \end{bmatrix}$$
$$\mathbf{A}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7\begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

따라서, $\lambda = -5$와 $\lambda = 7$에 대응하는 고유벡터는 각각

$$\begin{bmatrix} 2 \\ -3 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

입니다.

1.5.2 고유값의 성질

(1) λ가 $\mathbf{A}$의 고유값이면,

  (a) $\mathbf{A}^k$의 고유값은 λ^k이다.
  (b) $\mathbf{A}^{-1}$의 고유값은 $1/\lambda$이다.
  (c) $c\mathbf{A}$의 고유값은 $c\lambda$이다.
  (d) $\mathbf{A} + c\boldsymbol{I}$의 고유값은 $\lambda + c$이다.
  (e) $(\mathbf{A} + c\boldsymbol{I})^{-1}$의 고유값은 $1/(\lambda + c)$이다.

(2) $\mathrm{tr}(\mathbf{A}) = \sum \lambda_i, \quad |\mathbf{A}| = \prod \lambda_i$

(3) $\mathbf{A}$가 대칭이고 그 성분들이 실수값이면,

  (a) 모든 고유값은 실수이다.
  (b) 고유벡터들은 직교한다.
  (c) $\mathbf{A}$의 계수는 0이 아닌 고유값의 개수이다.

(4) 멱등행렬의 고유값은 0 또는 1이지만, 그 역은 성립하지 않는다.

1.6 이차형식 (quadratic form)과 양의 정부호 행렬 (positive definite matrix)

1.6.1 이차형식 (quadratic form)

벡터 $\mathbf{x}$와 행렬 $\mathbf{A}$에 대해, $\mathbf{x}'\mathbf{A}\mathbf{x}$를 $\mathbf{x}$의 이차형식이라고 한다. 예를 들어, $\mathbf{x} = (x_1, x_2, x_3)'$이고

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix}$$

이면

$$\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + 7x_2^2 + 5x_3^2 + (4+2)x_1x_2 + (2+3)x_1x_3 + (-2+6)x_2x_3$$

이는 x_1, x_2, x_3의 이차함수이다. 일반적으로, $\mathbf{x} = (x_1, \cdots, x_n)'$이고 $\mathbf{A}_{n\times n} = (a_{ij})$이면,

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \sum a_{ii}x_i^2 + \sum_{i\neq j}\sum a_{ij}x_ix_j \\ &= \sum a_{ii}x_i^2 + \sum_{i<j}\sum (a_{ij}+a_{ji})\, x_ix_j \end{aligned}$$

$i \neq j$일 때 x_ix_j의 계수가 a_{ij}와 a_{ji}의 합이므로, 대응하는 행렬 $\mathbf{A}$는 유일하지 않다. 즉,

$$\begin{aligned} & x_1^2 + 7x_2^2 + 5x_3^2 + (4+2)x_1x_2 + (2+3)x_1x_3 + (-2+6)x_2x_3 \\ =& x_1^2 + 7x_2^2 + 5x_3^2 + (3+3)x_1x_2 + (3+2)x_1x_3 + (0+4)x_2x_3 \end{aligned}$$

따라서, 만약 우리가

$$\mathbf{B} = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 7 & 0 \\ 2 & 4 & 5 \end{bmatrix}$$

를 취하면

$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$이다. 그러므로, 이차형식을 정의할 때 우리는 종종 $\mathbf{A}$가 대칭이라고 가정한다. 그러면 특정 이차형식에 대응하는 행렬은 유일해진다. 위의 이차형식에 대응하는 대칭행렬은

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 5/2 \\ 3 & 7 & 2 \\ 5/2 & 2 & 5 \end{bmatrix}$$

이고 이는 유일하다.

1.6.2 양의 정부호 행렬 (positive definite matrix)

(1) 정의:
스칼라는 양수 또는 음수임이 명확하지만, 행렬이 양수 또는 음수라고 정의하는 것은 불가능하다. 따라서 행렬의 부호는 이차형식을 사용하여 정의된다. 행렬

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

를 고려하고 대응하는 이차형식은

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3 \\ &= (x_1 + 2x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2\end{aligned}$$

이는 모든 x_1, x_2, x_3가 0인 경우를 제외하고는 항상 양수이다. 이 경우 우리는 A를 양의 정부호라고 부른다. 일반적으로, $\mathbf{x}'\mathbf{A}\mathbf{x} > 0, \forall x \neq 0$이면 A를 양의 정부호(p.d.)라고 한다.

반면에,

$$\mathbf{A} = \begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}$$

를 고려하면 그 이차형식은

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_1x_3 - 6x_2x_3 \\ &= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2\end{aligned}$$

이고 $\mathbf{x} = (213)'$이 $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$을 주기 때문에 $\mathbf{x} \neq \mathbf{0}$임에도 0이 될 수 있다.

이 행렬을 양의 준정부호(p.s.d.) 행렬이라고 한다. 일반적으로, $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0, \forall \mathbf{x}$이고 어떤 $\mathbf{x}$에 대해 $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$이면 A를 양의 준정부호(p.s.d.)라고 한다.

양의 정부호와 양의 준정부호 행렬은 종종 비음정부호(n.n.d.)라고 불린다. 이런 의미에서 음의 정부호(n.d.), 음의 준정부호(n.s.d.), 비양정부호(n.p.d.)를 정의할 수 있다. 또한, $\mathbf{A}$가 어떤 종류의 정부호성으로도 분류될 수 없으면 부정부호라고 한다.

(2) 성질

(a) $\mathbf{A}_{n\times n} = (a_{ij})$가 p.d.이면,
  (i) $r(\mathbf{A}) = n$
  (ii) $a_{ii} > 0, \quad i = 1, \cdots, n$
  (iii) 임의의 $n \times n$ 정방행렬 $\mathbf{P}$에 대해 $\mathbf{P}'\mathbf{A}\mathbf{P}$는 p.d.이다.

(b) $\mathbf{A}_{n\times n} = (a_{ij})$가 p.s.d.이면,
  (i) $r(\mathbf{A}) < n$
  (ii) $a_{ii} \geq 0, \quad i = 1, \cdots, n$
  (iii) 임의의 $n \times n$ 정방행렬 $\boldsymbol{P}$에 대해 $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$는 n.n.d.이다.

(c) 대칭행렬 $\mathbf{A}_{n\times n}$가 p.d.가 되기 위한 필요충분조건은
  (i) $\mathbf{B}'\mathbf{B} = \mathbf{A}$를 만족하는 풀랭크 행렬 $\mathbf{B}_{n\times n}$이 존재한다
  (ii) $\mathbf{A}$의 모든 고유값이 양수이다
  (iii) $a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \cdots, |\mathbf{A}| > 0$

(d) 대칭행렬 $\mathbf{A}_{n\times n}$가 p.d.가 아닌 p.s.d.가 되기 위한 필요충분조건은
  (i) $r(\mathbf{A}) \leq n$인 $\mathbf{B}'\mathbf{B} = \mathbf{A}$를 만족하는 행렬 $\mathbf{B}_{m\times n}$, $(m \leq n)$이 존재한다
  (ii) $\mathbf{A}$의 모든 고유값이 0보다 크거나 같고, 적어도 하나의 고유값은 0이어야 한다

(e) $\mathbf{A}_{m\times n}$의 계수가 m $(m < n)$이면,
  (i) $\mathbf{A}'\mathbf{A}$와 $\mathbf{A}\mathbf{A}'$는 p.s.d.이다

(f) $\mathbf{A}_{m\times n}$의 계수가 r $(r < m, r < n)$이면,
  (i) $\mathbf{A}'\mathbf{A}$와 $\mathbf{A}\mathbf{A}'$는 p.s.d.이다

1.7 행렬의 투영과 분해 (projection and decomposition)

1.7.1 투영 (projection)

(1) 벡터에 대한 투영

$\mathbf{y}$를 벡터 $\mathbf{x}$에 투영하면, 그 결과는 $c\mathbf{x}$가 되며, 여기서 $c = \dfrac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}$입니다.

(2) 열 공간에 대한 투영

$\mathbf{x} = (\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_{p-1})$를 $n \times p$ 행렬이라 하고, 여기서 $\mathbf{x}_i$, $i = 1, \ldots, p-1$, 는 n-벡터입니다. 그러면 $\mathbf{x}$의 열 공간은 다음과 같이 정의됩니다:

$$
\begin{aligned}
\mathcal{C}_{\mathbf{x}} &\equiv \operatorname{span}\{\mathbf{1}, \mathbf{x}_1, \cdots, \mathbf{x}_{p-1}\} \\
&= \{\beta_0 \mathbf{1} + \beta_1 x_1 + \cdots, +\beta_{p-1} x_{p-1} \mid \beta_0, \cdots, \beta_{p-1} \in R\} \\
&= \{\mathbf{x}\boldsymbol{\beta} \mid \boldsymbol{\beta} \in R^p\}
\end{aligned}
$$

벡터 $\mathbf{y}$를 열 공간 $\mathcal{C}_{\mathbf{x}}$에 투영하면, 그 결과는 $\boldsymbol{H}\mathbf{y}$가 되며, 여기서 $\boldsymbol{H} = \mathbf{x}\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{y}$이고, 이를 투영 행렬이라고 합니다.

(3) 그램-슈미트 직교화 (Gram-Schmidt orthogonalization)

$\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{p-1})$를 $\mathbf{x}_0 = \mathbf{1}$인 $n \times p$ 행렬이라 합시다. 또한, $\Pi(\mathbf{x} \mid \mathbf{z})$를 벡터 $\mathbf{x}$를 벡터 $\mathbf{z}$에 투영한 것이라 하면, 즉

$$\Pi(\mathbf{x} \mid \mathbf{z}) = (\mathbf{z}'\mathbf{x}/\mathbf{z}'\mathbf{z})\,\mathbf{z}$$

$\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{p-1})$를 $\mathbf{Z} = (\mathbf{z}_0, \mathbf{z}_1, \cdots, \mathbf{z}_{p-1})$로 다음과 같이 변환합니다:

$$
\begin{aligned}
\mathbf{z}_0 &= \mathbf{x}_0 \\
\mathbf{z}_1 &= \mathbf{x}_1 - \Pi\left(\mathbf{x}_1 \mid \mathbf{z}_0\right) \\
\mathbf{z}_2 &= \mathbf{x}_2 - \Pi\left(\mathbf{x}_2 \mid \mathbf{z}_0\right) - \Pi\left(\mathbf{x}_2 \mid \mathbf{z}_1\right) \\
&\vdots \\
\mathbf{z}_{p-1} &= \mathbf{x}_{p-1} - \Pi\left(\mathbf{x}_{p-1} \mid \mathbf{z}_0\right) - \Pi\left(\mathbf{x}_{p-1} \mid \mathbf{z}_1\right) - \cdots - \Pi\left(\mathbf{x}_{p-1} \mid \mathbf{z}_{p-2}\right)
\end{aligned}
$$

그러면 $\mathcal{C}_{\mathbf{x}} = \mathcal{C}_{\mathbf{Z}}$이고, $\mathbf{z}_0, \mathbf{z}_1, \cdots, \mathbf{z}_{p-1}$는 서로 직교한다는 점에 주목하세요. 이 변환을 그램-슈미트 직교화라고 합니다.

1.7.2 행렬의 분해 (decomposition)

(1) QR 분해

정의. $n \times p$ 행렬 $\mathbf{x}$는 $\mathbf{x} = \mathbf{QR}$ 형태로 쓸 수 있습니다. 여기서 $\mathbf{Q}$는 $n \times p$ 직교행렬이고 $\mathbf{R}$은 $p \times p$ 상삼각행렬입니다. 이를 $\mathbf{x}$의 $\mathbf{QR}$ 분해라고 합니다.

$\mathbf{QR}$ 분해를 계산하는 방법에는 3가지가 있습니다: (i) 그램-슈미트 과정, (ii) 하우스홀더 변환, (iii) 기븐스 회전. 여기서는 그램-슈미트 과정을 소개합니다. $\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{p-1})$가 그램-슈미트 직교화에 의해 $\mathbf{Z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_{p-1})$로 변환되었다고 가정하고, 다음을 정의합니다:

$$
\gamma_{ij} = \begin{cases} \mathbf{x}_j'\mathbf{Z}_i/\mathbf{Z}_i'\mathbf{Z}_i & , \quad i < j \\ 1 & , \quad i = j \\ 0 & , \quad i > j \end{cases}
$$

이제 $\mathbf{\Gamma} = (\gamma_{ij})$라 하면, 이는 $p \times p$ 상삼각행렬이고, $\mathbf{X} = \mathbf{Z\Gamma}$를 얻습니다. 따라서 $\mathbf{X} = \mathbf{QR}$이며, 여기서 $\mathbf{Q} = \mathbf{Z}$이고 $\mathbf{R} = \mathbf{\Gamma}$입니다. QR 분해는 $\mathbf{X}'\mathbf{X}$의 역행렬을 계산할 때 매우 유용합니다. 왜냐하면 $\mathbf{X}'\mathbf{X} = \mathbf{R}'\mathbf{Q}'\mathbf{QR} = \mathbf{R}'\mathbf{R}$이고 $\mathbf{R}'\mathbf{R}$의 역행렬을 계산하는 것이 매우 쉽기 때문입니다.

(2) 촐레스키 분해 (Cholesky decomposition)

대칭이고 양정치인 행렬 $\mathbf{A}$에 대해, $\mathbf{A} = \mathbf{R}'\mathbf{R}$을 만족하는 상삼각행렬 $\mathbf{R}$이 존재하며, 이 결과를 촐레스키 분해라고 합니다.

주어진 $\mathbf{A}$에 대한 $\mathbf{R}$의 계산은 다음 알고리즘을 기반으로 합니다.

단계 1. $r_{11} = a_{11}^{1/2}, \;\; r_{ij} = a_{1j}/r_{11}, \quad j = 2, \cdots, p$

단계 2. $2 \le i \le p$에 대해,

$$
\begin{aligned}
r_{ii} &= \left(a_{ii} - \sum_{k=1}^{i-1} r_{ki}^2 \right)^{1/2}, \\
r_{ij} &= \left(a_{ij} - \sum_{k=1}^{i-1} r_{ki} r_{kj} \right) / r_{ii}, \quad i+1 \le j \le p
\end{aligned}
$$

(3) 스펙트럴 분해 (spectral decomposition)

모든 $n \times n$ 대칭행렬 $\mathbf{A}$는 다음과 같이 쓸 수 있습니다:

$$\mathbf{A} = \mathbf{\Gamma D \Gamma}'$$

여기서 $\mathbf{\Gamma} = (\mathbf{u}_1, \cdots, \mathbf{u}_n)$는 $n \times n$ 직교행렬이고 $\mathbf{D} = \mathrm{diag}\,(\lambda_1, \cdots, \lambda_n)$는 대각행렬입니다. 여기서 λ_i, $i = 1, \cdots, n$는 A의 고유값이고, $\mathbf{u}_i$는 λ_i에 대응하는 고유벡터입니다. 이 분해를 스펙트럴 분해(또는 고유값 분해)라고 합니다. 또한, 다음과 같이 쓸 수 있음에 주목하세요:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i'$$

(4) 특이값 분해 (singular value decomposition)

모든 $n \times p$ $(p < n)$ 행렬 $\mathbf{X}$는 다음과 같이 쓸 수 있습니다:

$$\mathbf{X} = \mathbf{USV}'$$

여기서

(a) $\mathbf{U}$는 $\mathbf{XX}'$의 n개 고유값 중 가장 큰 p개의 고유값에 대응하는 고유벡터로 구성된 $n \times p$ 직교행렬입니다.

(b) $\mathbf{S} = \mathrm{diag}\,(s_1, \cdots, s_p)\,, s_1 \ge \cdots \ge s_p \ge 0$는 $p \times p$ 대각행렬이며, 여기서 s_i, $i = 1, \cdots, p$를 $\mathbf{X}$의 특이값이라고 합니다. $\mathbf{X}$의 특이값은 $\mathbf{X}'\mathbf{X}$의 고유값의 양의 제곱근임을 상기하세요. 왜냐하면

$$\mathbf{X}'\mathbf{X} = \mathbf{VSU}'\mathbf{USV}' = \mathbf{VS}^2\mathbf{V}'$$

이고, 스펙트럴 분해에 의해 $\mathbf{X}'\mathbf{X}$의 고유값이 $\mathbf{S}^2$의 대각 원소임을 알 수 있습니다.

(c) $\mathbf{V}$는 $\mathbf{X}'\mathbf{X}$의 고유벡터로 구성된 $p \times p$ 직교행렬입니다.
이 분해를 $n \times p$ $(p < n)$ 행렬 $\mathbf{X}$의 특이값 분해(SVD)라고 합니다.

1.8 행렬의 기타 사항 (Miscellaneous)

1.8.1 합산 벡터 (Summing Vector)와 중심화 행렬 (Centering Matrix)

$\mathbf{1}_n = (1, 1, \cdots, 1)'$를 합산 벡터라고 부릅니다. n차원 벡터 $\mathbf{x} = (x_1, \cdots, x_n)'$에 대해 $\mathbf{1}'\mathbf{x} = \sum x_i$이기 때문입니다. 또한, $\mathbf{1}_r\mathbf{1}_s'$는 모든 성분이 1인 $r \times s$ 행렬이며, 종종 $\boldsymbol{J}_{r\times s}$로 표기됩니다. 더 나아가, $\boldsymbol{J}_{n\times n}$을 $\boldsymbol{J}_n$으로 표기하며, $\boldsymbol{J}_n^2 = n\boldsymbol{J}_n$임을 쉽게 보일 수 있습니다. 특히,

$$\mathbf{C} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{J}_n$$

를 중심화 행렬이라고 부르며, 다음을 만족합니다:

$$\mathbf{C} = \mathbf{C}' = \mathbf{C}^2, \quad \mathbf{C1} = \mathbf{0}, \quad \mathbf{CJ} = \mathbf{JC} = \mathbf{0}$$

중심화 행렬의 예로, 다음과 같은 $\mathbf{x} = (x_1, \cdots, x_n)'$에 대한 표본 분산 s^2를 생각할 수 있다.

$$\begin{aligned}(n-1)s^2 = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 &= \mathbf{x}'\mathbf{x} - n\left(\frac{1}{n}\mathbf{1}'\mathbf{x}\right)^2 \\ &= \mathbf{x}'\mathbf{x} - \frac{1}{n}\mathbf{x}'\mathbf{1}\mathbf{1}'\mathbf{x} = \mathbf{x}'\mathbf{x} - \frac{1}{n}\mathbf{x}'\boldsymbol{J}\mathbf{x} = \mathbf{x}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\mathbf{x} = \mathbf{x}'\mathbf{C}\mathbf{x}\end{aligned}$$

1.8.2 행렬의 미분 (differentiation)

(i) $\mathbf{x}, \mathbf{y}$가 n차원 벡터라면,

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{y}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{y}'\mathbf{x}) = \mathbf{y}$$

(ii) $\mathbf{x}$가 n차원 벡터이고 $\mathbf{A}$가 $n \times n$ 행렬이라면,

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}) &= \mathbf{A} \\ \frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) &= \mathbf{A}'\end{aligned}$$

(iii) n차원 벡터 $\mathbf{x}$의 이차형식의 미분은 다음과 같다

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) &= \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x} \\ &= 2\mathbf{A}\mathbf{x}, \quad \text{만약 } \mathbf{A} \text{ 가 대칭이면}\end{aligned}$$

(iv) $f(\mathbf{x}) = f(x_1, \cdots, x_n)$의 x_i와 x_j에 대한 2차 미분은

$$\boldsymbol{H} = \left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\} = \frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}'}$$

이는 $n \times n$ 행렬이며 헤시안 행렬이라고 불린다.

1.8.3 크로네커 곱 (Kronecker Product)

(1) 정의

$\mathbf{A}_{p\times q}$와 $\mathbf{B}_{m\times n}$의 크로네커 곱 또는 직접 곱은 다음과 같이 정의된다

$$\mathbf{A}_{p\times q} \otimes \mathbf{B}_{m\times n} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix}$$

이는 $pm \times qn$ 행렬이다.

(2) 성질

(a) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$

(b) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{x} \otimes \mathbf{y}) = \mathbf{Ax} \otimes \mathbf{By}$

(c) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, 여기서 $\mathbf{A}$와 $\mathbf{B}$는 정방행렬이다

(d) $r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B})$

(e) $\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})$

(f) $|\mathbf{A}_{p\times p} \otimes \mathbf{B}_{m\times m}| = |\mathbf{A}|^m |\mathbf{B}|^p$

(g) $\mathbf{A} \otimes \mathbf{B}$의 고유값은 $\mathbf{A}$의 고유값과 $\mathbf{B}$의 고유값의 곱이다

1.8.4 벡터화 (vectorization)

(1) 정의

행렬 $\mathbf{A}_{m\times n}$을 $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_n]$로 쓸 때, $\mathbf{A}_i$는 m차원 i번째 열벡터이다. 그러면 $\mathrm{vec}(\mathbf{A})$는 다음과 같이 정의된다

$$\mathrm{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

즉, $\mathrm{vec}(\mathbf{A})$는 mn차원 벡터이다. 예를 들어,

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}, \quad \mathrm{vec}(\mathbf{A}) = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

(2) 성질

(a) $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\,\mathrm{vec}\,\mathbf{B}$

(b) $\mathrm{tr}(\mathbf{AB}) = (\mathrm{vec}\,\mathbf{A}')'\ \mathrm{vec}\ \mathbf{B}$

(c) $\mathrm{tr}\,(\mathbf{AZ}'\mathbf{BZC}) = (\mathrm{vec}\,\mathbf{Z}')'\,(\mathbf{CA} \otimes \mathbf{B}')\,\mathrm{vec}\,\mathbf{Z}$