Part.02

회귀분석

회귀분석을 위한 통계 수학적 개념이해 – 벡터, 행렬 연산 및 미분

FASTCAMPUS
ONLINE
머신러닝과 데이터분석 A-Z

강사. 김강진

I Matrix 표기법

- Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & . & . & a_{1n} \\ a_{22} & a_{22} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{m1} & . & . & . & a_{mn} \end{bmatrix} = [a_{ij}]$$

- Vector
  - 행 또는 열의 수가 1인 경우. 전자는 row vector, 후자는 column vector.
- Transpose and symmetric

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \Rightarrow A' = A^T = A^t = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

  - Symmetric 정의: $A = A^T$일 때 A 는 Symmetric이다.

I Matrix

- Scalar: 1 by 1 matrix.
- Identity matrix

$$I = \begin{bmatrix} 1 & 0 & . & . & 0 \\ 0 & 1 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & 1 \end{bmatrix} \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Diagonal matrix
  - $diag(a_1, \ldots, a_n) = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_n \end{bmatrix}$

I Matrix

- Equality
  - 정의: 모든 i, j에 대하여 $a_{ij} = b_{ij}$ 이면 $A = B$

- 합, 차의 성질

$$A \pm B = B \pm A$$
$$(A \pm B) \pm C = A \pm (B + C)$$
$$(A \pm B)^T = A^T \pm B^T$$

I Matrix

- 곱
  - 상수배
    - $B = kA \quad \Leftrightarrow \quad b_{ij} = ka_{ij}$

- 행렬곱
  - $A = (m \times p), \mathrm{B} = (p \times n),$
  - $C = AB, C = (m \times n)$

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

  - 일반적으로, $AB \neq BA$

Matrix

- 내적
  - Row vector와 Column vector의 곱
    - 예시: $a = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, b = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$
    - $a^T \cdot b = 0$

- 행렬 곱의 성질
  - $A(B + C) = AB + AC$
  - $A(BC) = (AB)C$
  - $(AB)^T = B^T A^T$
  - $AB = 0$ 이더라도 A, B 어느것도 0이 아닐 수 있음.

$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

I Matrix

- Trace

$$\mathrm{tr}(A) = a_{11} + a_{22} + \ldots. + a_{\mathrm{nn}}$$

- Trace의 성질
  - $\mathrm{tr}(A^{\mathrm{T}}) = \mathrm{tr}(A)$
  - $\mathrm{tr}(\mathrm{k}A) = \mathrm{ktr}(A)$
  - $\mathrm{tr}(I_n) = \mathrm{n}$
- A, B가 정사각 행렬인 경우,
  - $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$
  - $\mathrm{tr}(AB) = \mathrm{tr}(BA)$
  - $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$

I Matrix

- 행렬식 (determinant) $|A|$ 구하기
- 2 by 2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \Rightarrow |A| = a_{11}a_{22} - a_{12}a_{21}$$

- In general,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

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

$$= a_{11}\left(a_{22}a_{33} - a_{23}a_{32}\right) - a_{12}\left(a_{21}a_{33} - a_{23}a_{31}\right) + a_{13}\left(a_{21}a_{32} - a_{22}a_{31}\right)$$

I Matrix

- 역행렬 (Inverse)
  - 아래를 만족할 때, B는 A의 역행렬이고 A는 B의 역행렬이다.
  - $AB = BA = I_n$
- 역행렬 성질
  - $I^{-1} = I$
  - $(A^{-1})^{-1} = A$
  - $AB = I, BA = I$
- Idempotent
  - $AA = A$

I Matrix 미분

- 미분 표기법의 종류
    - Matrix 미분을 헷갈리게 하는 이유: 서로 다른 표기법이 존재.
    - 대부분의 책에서 명시하지 않고 사용하기에 두 표기법의 개념을 모두 알아두는 것이 좋음.
  - Numerator layout
    - 미분 당하는 변수 (혹은 함수)를 기준으로 결과의 형태를 표기
  - Denumerator layout
    - 미분을 하는 변수 (혹은 함수)를 기준으로 결과의 형태를 표기

- 이 강의록에서 표현한 모든 vector는 칼럼 vector로 통일.
- 핵심은, 의도한 미분을 수행했을 때 결과값의 차원.

I Matrix 미분

- Scalar를 vector로 미분

Numerator layout	Denumerator layout
$\frac{\delta y}{\delta \mathbf{x}} = \begin{bmatrix} \frac{\delta y}{\delta x_1} & \cdots & \frac{\delta y}{\delta x_p} \end{bmatrix}$	$\frac{\delta y}{\delta \mathbf{x}} = \begin{bmatrix} \frac{\delta y}{\delta x_1} \\ \vdots \\ \frac{\delta y}{\delta x_p} \end{bmatrix}$
$\frac{\delta y}{\delta \mathbf{x}^T} = \begin{bmatrix} \frac{\delta y}{\delta x_1} \\ \vdots \\ \frac{\delta y}{\delta x_p} \end{bmatrix}$	$\frac{\delta y}{\delta \mathbf{x}^T} = \begin{bmatrix} \frac{\delta y}{\delta x_1} & \cdots & \frac{\delta y}{\delta x_p} \end{bmatrix}$

I Matrix 미분

- Vector를 scalar로 미분

Numerator layout	Denumerator layout
$\frac{\delta \mathbf{x}}{\delta y} = \begin{bmatrix} \frac{\delta x_1}{\delta y} \\ \vdots \\ \frac{\delta x_p}{\delta y} \end{bmatrix}$	$\frac{\delta \mathbf{x}}{\delta y} = \begin{bmatrix} \frac{\delta x_1}{\delta y} & \cdots & \frac{\delta x_p}{\delta y} \end{bmatrix}$

I Matrix 미분

- Scalar를 matrix로 미분

Numerator layout	Denumerator layout
$\frac{\delta y}{\delta \mathbf{M}} = \begin{bmatrix} \frac{\delta y}{\delta x_{11}} & \cdots & \frac{\delta y}{\delta x_{\mathrm{m}1}} \\ \vdots & \ddots & \vdots \\ \frac{\delta y}{\delta x_{1n}} & \cdots & \frac{\delta y}{\delta x_{mn}} \end{bmatrix}$	$\frac{\delta y}{\delta \mathbf{M}} = \begin{bmatrix} \frac{\delta y}{\delta x_{11}} & \cdots & \frac{\delta y}{\delta x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\delta y}{\delta x_{m1}} & \cdots & \frac{\delta y}{\delta x_{mn}} \end{bmatrix}$

- Matrix를 scalar로 미분

Numerator layout	Denumerator layout
$\frac{\delta \mathbf{M}}{\delta y} = \begin{bmatrix} \frac{\delta x_{11}}{\delta y} & \cdots & \frac{\delta x_{1n}}{\delta y} \\ \vdots & \ddots & \vdots \\ \frac{\delta x_{m1}}{\delta y} & \cdots & \frac{\delta x_{mn}}{\delta y} \end{bmatrix}$	$\frac{\delta \mathbf{M}}{\delta y} = \begin{bmatrix} \frac{\delta x_{11}}{\delta y} & \cdots & \frac{\delta x_{\mathrm{m}1}}{\delta y} \\ \vdots & \ddots & \vdots \\ \frac{\delta x_{1n}}{\delta y} & \cdots & \frac{\delta x_{mn}}{\delta y} \end{bmatrix}$

I Matrix 미분

- Vector를 vector로 미분

Numerator layout	Denumerator layout
$\frac{\delta \mathbf{y}}{\delta \mathbf{x}} = \begin{bmatrix} \frac{\delta y_1}{\delta x_1} & \cdots & \frac{\delta y_1}{\delta x_{\mathrm{q}}} \\ \vdots & \ddots & \vdots \\ \frac{\delta y_{\mathrm{p}}}{\delta x_1} & \cdots & \frac{\delta y_p}{\delta x_q} \end{bmatrix}$	$\frac{\delta \mathbf{y}}{\delta \mathbf{x}} = \begin{bmatrix} \frac{\delta y_1}{\delta x_1} & \cdots & \frac{\delta y_p}{\delta x_1} \\ \vdots & \ddots & \vdots \\ \frac{\delta y_1}{\delta x_q} & \cdots & \frac{\delta y_p}{\delta x_q} \end{bmatrix}$

I Matrix 미분

- (예시) $\mathbf{a}^T\mathbf{x}$의 미분 (Numerator layout)
  - $\mathbf{a}^T\mathbf{x}$ 는 스칼라이므로, column벡터로 미분한 결과값은 row 벡터일 것.
  - $\frac{\delta\mathbf{a}^T\mathbf{x}}{\delta\mathbf{x}} = \begin{bmatrix}\frac{\delta\mathbf{a}^T\mathbf{x}}{\delta x_1} & \cdots & \frac{\delta\mathbf{a}^T\mathbf{x}}{\delta x_p}\end{bmatrix} = \begin{bmatrix}\frac{\delta(a_1x_1+\ldots+a_px_p)}{\delta x_1} & \cdots & \frac{\delta(a_1x_1+\ldots+a_px_p)}{\delta x_p}\end{bmatrix} = \begin{bmatrix}a_1 & \cdots & a_n\end{bmatrix} = \mathbf{a}^T$
  - $\frac{\delta\mathbf{x}^T\mathbf{a}}{\delta\mathbf{x}} = \begin{bmatrix}a_1 & \cdots & a_n\end{bmatrix} = \mathbf{a}^T$
  - $\frac{\delta\mathbf{a}^T\mathbf{x}}{\delta\mathbf{x}^T} = \begin{bmatrix}\frac{\delta\mathbf{x}^T\mathbf{a}}{\delta x_1} \\ \vdots \\ \frac{\delta\mathbf{x}^T\mathbf{a}}{\delta x_p}\end{bmatrix} = \begin{bmatrix}\frac{\delta(a_1x_1+\ldots+a_px_p)}{\delta x_1} \\ \vdots \\ \frac{\delta(a_1x_1+\ldots+a_px_p)}{\delta x_p}\end{bmatrix} = \begin{bmatrix}a_1 \\ \vdots \\ a_n\end{bmatrix} = \mathbf{a}$
  - $\frac{\delta\mathbf{x}^T\mathbf{a}}{\delta\mathbf{x}^T} = \begin{bmatrix}a_1 \\ \vdots \\ a_n\end{bmatrix} = \mathbf{a}$

I Matrix 미분

- (예시) $\mathbf{Ax}$의 미분 ($\mathbf{A}$: n×p, $\mathbf{x}$: p×1, Numerator layout)
  - $\mathbf{Ax}$ (n× 1)
  - Numerator layout이므로, 미분한 결과의 행의 수는 n개여야 한다. 결과값은 n×p일 것.
  - $$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \mathbf{Ax} = \begin{bmatrix} \sum_{j=1}^{p} a_{1j}\, x_j \\ \vdots \\ \sum_{j=1}^{p} a_{nj}\, x_j \end{bmatrix}$$
  - $$\frac{\delta \mathbf{Ax}}{\delta \mathbf{x}} = \begin{bmatrix} \frac{\delta \sum_{j=1}^{p} a_{1j}x_j}{\delta x_1} & \cdots & \frac{\delta \sum_{j=1}^{p} a_{1j}x_j}{\delta x_p} \\ \vdots & \ddots & \vdots \\ \frac{\delta \sum_{j=1}^{p} a_{nj}x_j}{\delta x_1} & \cdots & \frac{\delta \sum_{j=1}^{p} a_{nj}x_j}{\delta x_p} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{np} \end{bmatrix} = \boldsymbol{A}$$

I Matrix 미분

- 주요 미분 결과 정리 (Numerator layout)
  - $\frac{\delta \mathbf{a}^T \mathbf{x}}{\delta \mathbf{x}} = \frac{\delta \mathbf{x}^T \mathbf{a}}{\delta \mathbf{x}} = \mathbf{a}^T$
  - $\frac{\delta \mathbf{a}^T \mathbf{x}}{\delta \mathbf{x}^T} = \frac{\delta \mathbf{x}^T \mathbf{a}}{\delta \mathbf{x}^T} = \mathbf{a}$
  - $\frac{\delta \mathbf{A}\mathbf{x}}{\delta \mathbf{x}} (= \frac{\delta \mathbf{A}\mathbf{x}}{\delta \mathbf{x}^T}) = \mathbf{A}$
  - $\frac{\delta \mathbf{x}^T \mathbf{A}}{\delta \mathbf{x}} (= \frac{\delta \mathbf{x}^T \mathbf{A}}{\delta \mathbf{x}^T}) = \mathbf{A}^{T}$
  - $\frac{\delta \mathbf{x}^T \mathbf{A}\mathbf{x}}{\delta \mathbf{x}} = \mathbf{x}^{\mathbf{T}}(\mathbf{A} + \mathbf{A}^{\mathbf{T}})$, $\mathbf{x}$: $\mathrm{p} \times 1$, $\mathbf{A}$: $\mathrm{p} \times \mathrm{p}$
  - $\frac{\delta \mathbf{x}^T \mathbf{A}\mathbf{x}}{\delta \mathbf{x}^T} = (\mathbf{A} + \mathbf{A}^{\mathbf{T}})\mathbf{x}$, x: $\mathrm{p} \times 1$, A: $\mathrm{p} \times \mathrm{p}$

I Matrix 미분

- 주요 미분 결과 간편 활용 방법
  - 내적 형태
    - $\frac{\delta \mathbf{a}^T \mathbf{x}}{\delta \mathbf{x}} = \mathbf{a}^T$
      - 미분하면 $\mathbf{a}$ 꼴이 나올 것. 스칼라 – 벡터의 미분 -> 분모 차원의 반대. $1 \times \mathrm{p}$ 만들어주기.
  - Matrix – vector 곱 형태 (Linear form)
    - $\frac{\delta \mathbf{A}\mathbf{x}}{\delta \mathbf{x}} = \mathbf{A}$
      - 미분하면 $\mathbf{A}$ 꼴이 나올 것. 벡터 – 벡터의 미분 -> 분자 차원 유지. A의 행차원과 같이 만들어주기.
  - 제곱 형태 (Quadratic form)
    - $\frac{\delta \mathbf{x}^T \mathbf{A}\mathbf{x}}{\delta \mathbf{x}} = \mathbf{x}^{\mathbf{T}}(\mathbf{A} + \mathbf{A}^{\mathbf{T}})$, $\mathbf{x}\text{:}\ \mathrm{p} \times 1, \mathbf{A}\text{:}\ \mathrm{p} \times \mathrm{p}$
      - 미분하면 $(\mathbf{A} + \mathbf{A}^{\mathbf{T}})\mathrm{x}$ 꼴이 나올 것. 스칼라 – 벡터의 미분 -> 분모 차원의 반대. $1 \times \mathrm{p}$ 만들어 주기.

I Matrix 미분

- 회귀분석에 적용
  - 모델: $\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$
  - 오차제곱합: $R(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X\beta})^T(\mathbf{y} - \mathbf{X\beta})$
  - 오차제곱합을 최소로하는 $\widehat{\boldsymbol{\beta}}$을 구하고자 함.
  - $\boldsymbol{\beta}$는 column vector.
  - 미분 진행
    - $\frac{\delta R(\boldsymbol{\beta})}{\delta \boldsymbol{\beta}^T} = \frac{\delta(\mathbf{y}-\mathbf{X\beta})^T(\mathbf{y}-\mathbf{X\beta})}{\delta \boldsymbol{\beta}^T} = \frac{\delta\left(\mathbf{y}^T\mathbf{y}-\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y}-\mathbf{y}^T\mathbf{X\beta}+\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X\beta}\right)}{\delta \boldsymbol{\beta}^T} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X\beta}$
  - 미분한 함수를 0이되게 하는 해를 찾음
    - $-2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\widehat{\boldsymbol{\beta}} = 0$
    - $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$