예제 1

(1) $\frac{\partial}{\partial x} x^T y = y$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

풀이 : $x^T y = [x_1 \dots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + \cdots + x_n y_n$

그런데 $\frac{\partial}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}$ 이므로,

$$\frac{\partial}{\partial x}(x^T y) = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} (x_1 y_1 + \cdots + x_n y_n) = \begin{bmatrix} \frac{\partial}{\partial x_1}(x_1 y_1 + \cdots + x_n y_n) \\ \vdots \\ \frac{\partial}{\partial x_n}(x_1 y_1 + \cdots + x_n y_n) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} x_1 y_1 + \frac{\partial}{\partial x_1} x_2 y_2 + \cdots + \frac{\partial}{\partial x_1} x_n y_n \\ \vdots \\ \frac{\partial}{\partial x_n} x_1 y_1 + \frac{\partial}{\partial x_n} x_2 y_2 + \cdots + \frac{\partial}{\partial x_n} x_n y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

다른풀이 :

$$\frac{\partial}{\partial x}(x^T y) = \underbrace{\left(\frac{\partial}{\partial x} x^T\right)} y = I y = y.$$

$$\frac{\partial}{\partial x} x^T = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} [x_1 \dots x_n] = \begin{bmatrix} \frac{\partial}{\partial x_1} x_1 & \frac{\partial}{\partial x_1} x_2 & \cdots & \frac{\partial}{\partial x_1} x_n \\ \frac{\partial}{\partial x_2} x_1 & \frac{\partial}{\partial x_2} x_2 & \cdots & \frac{\partial}{\partial x_2} x_n \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_n} x_1 & \frac{\partial}{\partial x_n} x_2 & \cdots & \frac{\partial}{\partial x_n} x_n \end{bmatrix} = I$$

(2) $\frac{\partial}{\partial x} y^T x = y$.

풀이: $y^T x = x^T y$ 이므로 $\frac{\partial}{\partial x} y^T x = \frac{\partial}{\partial x} x^T y = \underbrace{\left(\cancel{\frac{\partial}{\partial x} x^T}\right)}^{I} y$

(3) $\frac{\partial}{\partial \beta} y^T X \beta = X^T y$.

풀이: $y^T X \beta$ 가 스칼라 이므로,

$$\frac{\partial}{\partial \beta} y^T X \beta = \frac{\partial}{\partial \beta} \beta^T X^T y = \underbrace{\left(\frac{\partial}{\partial \beta} \beta^T\right)}_{I} X^T y = X^T y.$$

다른풀이: $\frac{\partial}{\partial \beta}(y^T X \beta) = X^T y$ 임을 보이는 다른풀이.

$$X\beta = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} x_{11}\beta_1 + \cdots + x_{1p}\beta_p \\ x_{21}\beta_1 + \cdots + x_{2p}\beta_p \\ \vdots \\ x_{n1}\beta_1 + \cdots + x_{np}\beta_p \end{bmatrix}$$

$$y^T X \beta = [y_1 \ y_2 \ \cdots \ y_n] \begin{bmatrix} x_{11}\beta_1 + \cdots + x_{1p}\beta_p \\ x_{21}\beta_1 + \cdots + x_{2p}\beta_p \\ \vdots \\ x_{n1}\beta_1 + \cdots + x_{np}\beta_p \end{bmatrix}$$

$$= y_1(x_{11}\beta_1 + \cdots + x_{1p}\beta_p) + y_2(x_{21}\beta_1 + \cdots + x_{2p}\beta_p) + \cdots$$

$$+ y_n(x_{n1}\beta_1 + \cdots + x_{np}\beta_p)$$

$= A_1 + A_2 + \cdots + A_n$. , $A_1, A_2, \ldots, A_n$은 모두 스칼라.

$$\cdot \ \frac{\partial}{\partial \beta}(y^T X\beta) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} \\ \vdots \\ \frac{\partial}{\partial \beta_p} \end{bmatrix} (A_1 + A_2 + \cdots + A_n)$$

$$= \begin{bmatrix} \frac{\partial}{\partial \beta_1} A_1 \\ \vdots \\ \frac{\partial}{\partial \beta_p} A_1 \end{bmatrix} + \begin{bmatrix} \frac{\partial}{\partial \beta_1} A_2 \\ \vdots \\ \frac{\partial}{\partial \beta_p} A_2 \end{bmatrix} + \cdots + \begin{bmatrix} \frac{\partial}{\partial \beta_1} A_n \\ \vdots \\ \frac{\partial}{\partial \beta_p} A_n \end{bmatrix}$$

$$\frac{\partial}{\partial \beta_1} A_1 = \frac{\partial}{\partial \beta_1} \left\{ y_1 (x_{11}\beta_1 + x_{21}\beta_2 + \cdots + x_{1p}\beta_p) \right\} = y_1 x_{11} + 0 + \cdots + 0.$$

$$\frac{\partial}{\partial \beta_2} A_1 = \frac{\partial}{\partial \beta_2} \left\{ y_2 (x_{11}\beta_1 + x_{21}\beta_2 + \cdots + x_{1p}\beta_p) \right\} = 0 + y_2 x_{21} + \cdots + 0.$$

$$\vdots$$

$$= \begin{bmatrix} y_1 x_{11} \\ y_2 x_{21} \\ \vdots \\ y_n x_{n1} \end{bmatrix} + \begin{bmatrix} y_2 x_{21} \\ y_2 x_{22} \\ \vdots \\ y_n x_{2n} \end{bmatrix} + \cdots + \begin{bmatrix} y_n x_{n1} \\ y_n x_{n2} \\ \vdots \\ y_n x_{np} \end{bmatrix}$$

$$= \begin{bmatrix} y_1 x_{11} + y_2 x_{21} + \cdots + y_n x_{n1} \\ y_2 x_{21} + y_2 x_{22} + \cdots + y_n x_{n2} \\ \vdots \\ y_n x_{n1} + y_n x_{2n} + \cdots + y_n x_{np} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{21} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = X^T y.$$

또다른풀이

$$y^T X\beta = y^T \begin{bmatrix} X_1 & X_2 & \cdots & X_p \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = y^T (X_1\beta_1 + \cdots + X_p\beta_p)$$

$$= y^T x_1 \beta_1 + y^T X_2 \beta_2 + \cdots + y^T X_p \beta_p$$

$$\frac{\partial}{\partial \beta}(y^T X \beta) = \begin{bmatrix} \frac{\partial}{\partial \beta_1} \\ \vdots \\ \frac{\partial}{\partial \beta_p} \end{bmatrix} (y^T X_1 \beta_1 + y^T X_2 \beta_2 + \cdots + y^T X_p \beta_p)$$

$$= \begin{bmatrix} y^T x_1 \\ y^T X_2 \\ \vdots \\ y^T X_p \end{bmatrix} = \begin{bmatrix} X_1^T y \\ X_2^T y \\ \vdots \\ X_p^T y \end{bmatrix} = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_p^T \end{bmatrix} y = X^T y$$

<u>예제2</u>

(1) $\frac{\partial}{\partial y} y^T y = 2y$.

<u>풀이</u>: $y^T y = [y_1 \ y_2 \cdots y_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y_1^2 + y_2^2 + \cdots y_n^2$.

따라서 $\frac{\partial}{\partial y}(y^T y) = \begin{bmatrix} \frac{\partial}{\partial y_1} \\ \vdots \\ \frac{\partial}{\partial y_n} \end{bmatrix} (y_1^2 + \cdots + y_n^2) = \begin{bmatrix} 2y_1 \\ 2y_2 \\ \vdots \\ 2y_n \end{bmatrix}$

<u>틀린풀이</u>: $\frac{\partial}{\partial y}(y^T y) \underset{\text{틀림.}}{=} \Big(\underbrace{\frac{\partial}{\partial y} y^T}_{I}\Big) y = Iy = y.$

틀린풀이의 스칼라 버전: $\frac{d}{dy}(y^2) = \left(\frac{d}{dy}y\right)y = y$

올바른풀이 스칼라 버전:

$$\frac{d}{dy}y^2 = \frac{d}{dy}y\cdot y = \frac{d}{dy}f(y)g(y) = \left(\frac{d}{dy}f(y)\right)g(y) + f(y)\left(\frac{d}{dy}g'(y)\right)$$

$$= 1\cdot y + y\cdot 1 = 2y.$$

※ 곱의미분: 함수 $f(x), g(x)$가 x에 대하여 미분가능하면,

$$\{f(x)\cdot g(x)\}' = f'(x)g(x) + f(x)g'(x).$$

다시 벡터로 돌아오면

올바른 풀이: $\left(\frac{\partial}{\partial y}\right)(y^T y) = A + B = Iy + Iy = 2y$

A: 빨간 y만 변수로 보고 미분

B: 파란 y만 변수로 보고 미분.

$$A = \left(\frac{\partial}{\partial y}\right)(y^T y) = \left(\frac{\partial}{\partial y}y^T\right)y = Iy$$

$$B = \left(\frac{\partial}{\partial y}\right)(y^T y) = \left(\frac{\partial}{\partial y}\right)(y^T y) = \left(\frac{\partial}{\partial y}y^T\right)y = Iy$$

(2) $\frac{\partial}{\partial \beta}\beta^T X^T X\beta = 2X^T X\beta$

풀이: $\frac{\partial}{\partial \beta}(\beta^T X^T X \beta) = A + B = X^T X \beta + X^T X \beta = 2X^T X \beta.$

$A = \left(\frac{\partial}{\partial \beta}\beta^T\right) X^T X \beta = X^T X \beta.$

$B = \frac{\partial}{\partial \beta}(\beta^T X^T X \beta) = \frac{\partial}{\partial \beta}(\beta^T X^T X \beta) = \left(\frac{\partial}{\partial \beta}\beta^T\right) X^T X \beta = X^T X \beta.$

예제 3: $loss = (y - X\beta)^T (y - X\beta)$ 를 최소화하는 β를 구하라.

풀이: loss 는 convex-function 이므로 $\frac{\partial}{\partial \beta} loss = 0$ 을 풀면 된다.

$$loss = (y - X\beta)^T (y - X\beta) = (y^T - \beta^T X^T)(y - X\beta)$$

$$= y^T y - y^T X\beta - \beta^T X^T y + \underline{\beta^T X^T X \beta}.$$

$$\frac{\partial}{\partial \beta} loss = 0 - X^T y - X^T y + 2X^T X \beta = 0.$$

$$2X^T X \beta = 2X^T y.$$

$\therefore \hat{\beta} = (X^T X)^{-1} X y$ 에서 loss 값이 최소가 된다.

예제4 : $loss = (y - X\beta)^T(y - X\beta) + \lambda\beta^T\beta$ 를 최소로

하는 β를 구해보자.

풀이 $\frac{\partial}{\partial\beta} loss = \underbrace{-2X^Ty + 2X^TX\beta}_{\text{예제3에서 함.}} + \frac{\partial}{\partial\beta}\lambda\beta^T\beta.$

$= -2X^Ty + 2X^TX\beta + 2\lambda\beta.$

$= -2X^Ty + 2(X^TX + \lambda I)\beta. = 0 \quad \cdots\cdots (*)$

$(*) \Leftrightarrow X^Ty = (X^TX + \lambda I)\beta.$

$\therefore \hat{\beta} = (X^TX + \lambda I)^{-1}X^Ty$ 에서 최소가 된다.

Note: 이때, $(X^TX + \lambda I)$의 역행렬은 항상 존재함. ($\because \lambda > 0$)

예제5 : 예제3과 동일한 loss를 아래와 같이 재표현하자.

- $u = X\beta$
- $v = y - u$
- $loss = v^Tv$

$\rightarrow \frac{\partial}{\partial\beta} loss = \frac{\partial}{\partial\beta}u^T \frac{\partial}{\partial u}v^T \frac{\partial}{\partial v} loss.$

① $= \frac{\partial}{\partial\beta}u^T = \frac{\partial}{\partial\beta}\beta^TX^T = X^T$

② $= \frac{\partial}{\partial u}v^T = \frac{\partial}{\partial u}(y^T - u^T) = \frac{\partial}{\partial u}(-u^T) = -I.$

③ $= \frac{\partial}{\partial v} loss = \frac{\partial}{\partial v} v^T v = 2v = 2(y-u) = 2(y-XB)$

$\therefore \frac{\partial}{\partial B} u^T \frac{\partial}{\partial u} v^T \frac{\partial}{\partial v} loss =$ ① $\times$ ② $\times$ ③ $= X^T(-I)2(y-XB)$

$= -X^T(2y-2XB) = -2X^Ty + 2X^TXB. = \frac{\partial}{\partial B} loss$

$\therefore$ 예제 3 과 예제 5는 동일한계산이다. 이러한 결과가 성립하는

이유는 체인룰 때문이다.