벡터미분 / 매트릭스 미분.

① 정의 1: 벡터로 미분하는 경우.

$$\frac{\partial}{\partial y} := \begin{bmatrix} \frac{\partial}{\partial y_1} \\ \vdots \\ \frac{\partial}{\partial y_n} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

② 정의 2: 매트릭스로 미분하는 경우.

$$\frac{\partial}{\partial X} := \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \cdots & \frac{\partial}{\partial x_{1p}} \\ \vdots & & \\ \frac{\partial}{\partial x_{n1}} & \cdots & \frac{\partial}{\partial x_{np}} \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

---

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

pf: $x^T y = [x_1 \cdots x_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$

$\cdot \ \frac{\partial}{\partial x} := \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}$

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

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

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

$$= \begin{bmatrix} y_1 + 0 + \cdots + 0 \\ 0 + y_2 + \cdots + 0 \\ \vdots \\ 0 + 0 + \cdots + y_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y.$$

$\blacksquare$

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

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

· $$\frac{\partial}{\partial x} x^T = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} [x_1 \cdots x_n]$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} x_1 & \frac{\partial}{\partial x_1} x_2 & \cdots & \frac{\partial}{\partial x_1} x_n \\ \frac{\partial}{\partial x_2} x_1 & \frac{\partial}{\partial x_2} x_2 & \cdots & \frac{\partial}{\partial x_2} x_n \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_n} x_1 & \frac{\partial}{\partial x_n} x_2 & \cdots & \frac{\partial}{\partial x_n} x_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I \quad ,$$

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

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

③ $\frac{\partial}{\partial w}(y^T X w) = X^T y$. 단 w: $p\times 1$ vector.
X: $n\times p$ matrix
y: $n\times 1$ vector.

pf. $y^T X w$ 는 스칼라이므로, $y^T X w = w^T X^T y$.

$\therefore \frac{\partial}{\partial w}(y^T X w) = \frac{\partial}{\partial w}(w^T X^T y) = \left(\frac{\partial}{\partial w} w^T\right) X^T y$.

$= I X^T y = X^T y$.

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

$$\cdot\ Xw = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_p \end{bmatrix} = \begin{bmatrix} x_{11}w_1 + \cdots + x_{1p}w_p \\ x_{21}w_1 + \cdots + x_{2p}w_p \\ \vdots \\ x_{n1}w_1 + \cdots + x_{np}w_p \end{bmatrix}$$

$$\cdot\ y^T X w = [y_1 \cdots y_n] \begin{bmatrix} x_{11}w_1 + \cdots + x_{1p}w_p \\ x_{21}w_1 + \cdots + x_{2p}w_p \\ \vdots \\ x_{n1}w_1 + \cdots + x_{np}w_p \end{bmatrix}$$

$$= y_1(x_{11}w_1 + \cdots + x_{1p}w_p) + y_2(x_{21}w_1 + \cdots + x_{2p}w_p)$$

$$+ \cdots + y_n(x_{n1}w_1 + \cdots + x_{np}w_p) := A_1 + A_2 + \cdots + A_n$$

- note: $A_1, A_2, \dots, A_n$은 모두 스칼라.

- $$\frac{\partial}{\partial w}(y^T X w) = \begin{bmatrix} \frac{\partial}{\partial w_1} \\ \vdots \\ \frac{\partial}{\partial w_p} \end{bmatrix} (A_1 + A_2 + \cdots + A_n)$$

$$= \begin{bmatrix} \frac{\partial}{\partial w_1} A_1 \\ \vdots \\ \frac{\partial}{\partial w_p} A_1 \end{bmatrix} + \cdots + \begin{bmatrix} \frac{\partial}{\partial w_1} A_n \\ \vdots \\ \frac{\partial}{\partial w_p} A_n \end{bmatrix}$$

- $$\frac{\partial}{\partial w_1} A_1 = \frac{\partial}{\partial w_1} \left(y_1 (x_{11} w_1 + \cdots + x_{1p} w_p) \right)$$

$$= \frac{\partial}{\partial w_1} y_1 x_{11} w_1 + \frac{\partial}{\partial w_1} y_1 x_{12} w_2 + \cdots + \frac{\partial}{\partial w_1} y_1 x_{1p} w_p.$$

$$= y_1 x_{11}$$

- $$\frac{\partial}{\partial w_p} A_1 = \frac{\partial}{\partial w_p} \left(y_1 (x_{11} w_1 + \cdots + x_{1p} w_p) \right)$$

$$= y_1 x_{1p}$$

$$\therefore \begin{bmatrix} \frac{\partial}{\partial w_1} A_1 \\ \vdots \\ \frac{\partial}{\partial w_p} A_1 \end{bmatrix} = \begin{bmatrix} y_1 x_{11} \\ y_1 x_{12} \\ \vdots \\ y_1 x_{1p} \end{bmatrix}$$

또한 $\begin{bmatrix} \frac{\partial}{\partial w_1} A_2 \\ \vdots \\ \frac{\partial}{\partial w_p} A_2 \end{bmatrix} = \begin{bmatrix} y_2 x_{21} \\ y_2 x_{22} \\ \vdots \\ y_2 x_{2p} \end{bmatrix}, \cdots, \begin{bmatrix} \frac{\partial}{\partial w_1} A_n \\ \vdots \\ \frac{\partial}{\partial w_p} A_n \end{bmatrix} = \begin{bmatrix} y_n x_{n1} \\ y_n x_{n2} \\ \vdots \\ y_n x_{np} \end{bmatrix}$

이므로, $\frac{\partial}{\partial w}(y^T X w) = \begin{bmatrix} y_1 x_{11} \\ y_1 x_{12} \\ \vdots \\ y_1 x_{1p} \end{bmatrix} + \begin{bmatrix} y_2 x_{21} \\ y_2 x_{22} \\ \vdots \\ y_2 x_{2p} \end{bmatrix} + \cdots + \begin{bmatrix} y_n x_{n1} \\ y_n x_{n2} \\ \vdots \\ y_n x_{np} \end{bmatrix}$

$$= \begin{bmatrix} y_1 x_{11} + y_2 x_{21} + \cdots + y_n x_{n1} \\ y_1 x_{12} + y_2 x_{22} + \cdots + y_n x_{n2} \\ \vdots \\ y_1 x_{1p} + y_2 x_{2p} + \cdots + y_n x_{np} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & & & \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$= X^T y$. ■.

④ $\frac{\partial}{\partial y} y^T y = 2y$.

(pf) $y^T y = [y_1 \cdots y_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = y_1^2 + y_2^2 + \cdots + y_n^2$.

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

$$= \begin{bmatrix} \frac{\partial}{\partial y_1}(y_1^2 + \cdots + y_n^2) \\ \vdots \\ \frac{\partial}{\partial y_n}(y_1^2 + \cdots + y_n^2) \end{bmatrix} = \begin{bmatrix} 2y_1 + 0 + \cdots + 0 \\ 0 + 2y_2 + \cdots + 0 \\ \vdots \\ 0 + 0 + \cdots + 2y_n \end{bmatrix} = 2y.$$

(틀린풀이) $\frac{\partial}{\partial y}(y^T y) = \left(\frac{\partial}{\partial y} y^T\right) y.$

그런데 $\frac{\partial}{\partial y} y^T = I.$ 이므로 $\frac{\partial}{\partial y}(y^T y) = Iy = y.$

틀린풀이의 비유 $\frac{d}{dy}(y \cdot y) = \left(\frac{d}{dy} y\right) \cdot y = y.$

올바른풀이 : $\frac{d}{dy}(y\,y) = \left(\frac{d}{dy} y\right) \cdot y + y \cdot \left(\frac{d}{dy} y\right) = y + y$

곱의미분.

Note: 곱의미분 : 함수 $f(x), g(x)$ 가 x에 대하여 미분가능할때

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

예제 $f(x) = x$, $g(x) = x$ 를 대입하면

$$\frac{d}{dx} x\,x = \left(\frac{d}{dx} x\right) x + x \left(\frac{d}{dx} x\right)$$

이제 벡터비 전으로 돌아오자!

(올바른풀이) $\left(\frac{\partial}{\partial y}\right)(y^T y) = A + B = Iy + Iy = 2y.$

A: 빨간 y를 변수로 보고 미분.

B: 파란 y를 변수로 보고 미분.

$y^T y = y^T y$

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

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

$= Iy$

⑤ $\frac{\partial}{\partial w} w^T X^T X w = 2X^T X w.$

(pf) $\frac{\partial}{\partial w}(w^T X^T X w) = \left(\frac{\partial}{\partial w} w^T X^T\right) X w + \left(\frac{\partial}{\partial w} w^T X^T\right) X w$

등호가 좀 이상함.

$= \left(\frac{\partial}{\partial w} w^T\right) X^T X w + \left(\frac{\partial}{\partial w} w^T\right) X^T X w$

$= I X^T X w + I X^T X w = 2X^T X w.$

예제: $(y - Xw)^T (y - Xw) = \text{loss}$ 라고하자.

$\frac{\partial}{\partial w} \text{loss} = \frac{\partial}{\partial w}\left(y^T y - y^T X w - w^T X^T y + w^T X^T X w\right)$

$$= \frac{\partial}{\partial w}\left(-y^T X w\right) + \frac{\partial}{\partial w}\left(-w^T X^T y\right) + \frac{\partial}{\partial w}\left(w^T X^T X w\right)$$

$$= \frac{\partial}{\partial w}\left(-w^T X^T y\right) + \frac{\partial}{\partial w}\left(-w^T X^T y\right) + 2X^T X w$$

$$= -2\left(\frac{\partial}{\partial w} w^T\right) X^T y + 2X^T X w$$

$$= -2 I X^T y + 2X^T X w$$

$$= 2X^T X w - 2X^T y$$