Basic knowledge ②

matrix Derivation

For variable is a matrix.

In this case, we denote:

$$f(X),\quad X_{m\times n} = (x_{ij})_{i=1,j=1}^{m,n}$$

$$\nabla_X f(X) = \frac{\partial f(X)}{\partial X_{m\times n}} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & & & \\ & & & \frac{\partial f}{\partial x_{mn}} \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \cdots & \end{bmatrix}_{m\times n}$$

1、Constant.

$$\frac{\partial c}{\partial X} = 0_{m\times n},\quad \text{here } c \text{ is a constant.}$$

prove:

$$\frac{\partial c}{\partial X} = \begin{bmatrix} \frac{\partial c}{\partial x_{11}} & \frac{\partial c}{\partial x_{12}} & \cdots & \frac{\partial c}{\partial x_{1n}} \\ \vdots & & & \\ \frac{\partial c}{\partial x_{m1}} & \cdots & \cdots & \frac{\partial c}{\partial x_{mn}} \end{bmatrix}_{m\times n} = \begin{bmatrix} 0 & 0 & 0 \\ & \ddots & \\ & & \ddots \end{bmatrix}_{m\times n}$$

$$= 0_{m\times n}$$

2、Linear Rule.

$$\frac{\partial[c_1 f(X) + c_2 g(X)]}{\partial X} = c_1\frac{\partial f(X)}{\partial X} + c_2\frac{\partial g(X)}{\partial X},\quad \text{here } c_1, c_2 \text{ are constant.}$$

prove:

$$\frac{\partial[c_1 f(X)+c_2 g(X)]}{\partial X} = \begin{bmatrix} \frac{\partial[c_1 f(X)+c_2 g(X)]}{\partial x_{11}}, & \frac{\partial[c_1 f(X)+c_2 g(X)]}{\partial x_{12}} & \cdots & \frac{\partial[c_1 f(X)+c_2 g(X)]}{\partial x_{1n}} \\ \vdots & & & \\ \frac{\partial[c_1 f(X)+c_2 g(X)]}{\partial x_{m1}}, & \cdots & \cdots & \frac{\partial[c_1 f(X)+c_2 g(X)]}{\partial x_{mn}} \end{bmatrix}$$

$$= \begin{bmatrix} c_1\frac{\partial f}{\partial x_{11}} + c_2\frac{\partial g}{\partial x_{11}}, & c_1\frac{\partial f}{\partial x_{12}} + c_2\frac{\partial g}{\partial x_{12}} & \cdots & c_1\frac{\partial f}{\partial x_{1n}} + c_2\frac{\partial g}{\partial x_{1n}} \\ \vdots & & & \\ c_1\frac{\partial f}{\partial x_{m1}} + c_2\frac{\partial g}{\partial x_{m1}} & \cdots & \cdots & c_1\frac{\partial f}{\partial x_{mn}} + c_2\frac{\partial g}{\partial x_{mn}} \end{bmatrix}$$

$$= c_1\begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & & \\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix} + c_2\begin{bmatrix} \frac{\partial g}{\partial x_{11}} & \cdots & \frac{\partial g}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial g}{\partial x_{m1}} & & \frac{\partial g}{\partial x_{mn}} \end{bmatrix} = c_1\frac{\partial f(X)}{\partial X} + c_2\frac{\partial g(X)}{\partial X}$$

3. multiply Rule.

$$\frac{d[f(x)\,g(x)]}{dX} = \frac{df(x)}{dX}\,g(x) + f(x)\,\frac{dg(x)}{dX}$$

prove: $\frac{d[f(x)g(x)]}{dX} = \begin{bmatrix} \frac{d(fg)}{dx_{11}} & \cdots & \frac{d(fg)}{dx_{1n}} \\ \vdots & & \\ \frac{d(fg)}{dx_{m1}} & \cdots & \frac{d(fg)}{dx_{mn}} \end{bmatrix}$

$$= \begin{bmatrix} \frac{df}{dx_{11}}\cdot g + f\cdot\frac{dg}{dx_{11}}, & \cdots & \frac{df}{dx_{1n}}\cdot g + f\cdot\frac{dg}{dx_{1n}} \\ \vdots & & \\ \frac{df}{dx_{m1}}\cdot g + f\cdot\frac{dg}{dx_{m1}} & \cdots & \frac{df}{dmn}\cdot g + f\cdot\frac{dg}{dmn} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{df}{dx_{11}}\cdot g, & \cdots & \frac{df}{dx_{1n}}\cdot g \\ \frac{df}{dx_{m1}}\cdot g, & \cdots & \frac{df}{dx_{mn}}\cdot g \end{bmatrix} + \begin{bmatrix} f\cdot\frac{dg}{dx_{11}}, & \cdots & f\cdot\frac{dg}{dx_{1n}} \\ \vdots & & \\ f\cdot\frac{dg}{dx_{m1}}, & \cdots & f\cdot\frac{dg}{dx_{mn}} \end{bmatrix}$$

$$= \frac{df(x)}{dX}\cdot g(x) + f(x)\cdot\frac{dg(x)}{dX}$$

//

5. Division Rule.

$$\frac{d\left[\frac{f(x)}{g(x)}\right]}{dX} = \frac{1}{g^2(x)}\left[\frac{df(x)}{dX}\,g(x) - f(x)\,\frac{dg(x)}{dX}\right], \quad g(x) \neq 0$$

prove: $\frac{d\left[\frac{f(x)}{g(x)}\right]}{dX} = \begin{bmatrix} \frac{d}{dx_{11}}\left(\frac{f}{g}\right) & \cdots & \frac{d}{dx_{1n}}\left(\frac{f}{g}\right) \\ \frac{d}{dx_{m1}}\left(\frac{f}{g}\right) & \cdots & \frac{d}{dx_{mn}}\left(\frac{f}{g}\right) \end{bmatrix} = \begin{bmatrix} \frac{1}{g^2}\left(\frac{df}{dx_{11}}g - f\frac{dg}{dx_{11}}\right), \cdots \\ \\ \end{bmatrix}$

$$= \frac{1}{g^2}\left(\begin{bmatrix} \frac{df}{dx_{11}}, & \cdots & \frac{df}{dx_{1n}} \\ \vdots & & \\ & \cdots & \end{bmatrix} g - f\cdot\begin{bmatrix} \frac{dg}{dx_{11}} & \cdots & \frac{dg}{dx_{1n}} \\ \vdots & \cdots & \vdots \end{bmatrix}\right)$$

$$= \frac{1}{g^2(x)}\left[\frac{df(x)}{dX}\,g(x) - f(x)\,\frac{dg(x)}{dX}\right]$$

Some equations:

① $\frac{d(a^T X b)}{dX} = ab^T$, here a, b are constant vectors,
$a = (a_1 \cdots a_m)^T$, $b = (b_1 \cdots b_n)^T$

prove:

$$\underset{1\times m}{a^T} \underset{m\times n}{X} \underset{n\times 1}{b} = [a_1 X_{11} + a_2 X_{21} + \cdots a_m X_{m1}, a_1 X_{12} + a_2 X_{22} + \cdots a_m X_{m2}, a_1 X_{1n} + \cdots a_m X_{mn}] \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$$= b_1 (a_1 X_{11} + a_2 X_{21} + \cdots a_m X_{m1}) + b_2 (a_1 X_{12} + a_2 X_{22} + \cdots a_m X_{m2}) + \cdots + b_n (a_1 X_{1n} + \cdots a_m X_{mn})$$

$$\frac{d(a^T X b)}{dX} = \begin{bmatrix} a_1 b_1, & a_1 b_2, & a_1 b_3 & \cdots & a_1 b_n \\ a_2 b_1, & a_2 b_2, & a_2 b_3 & \cdots & a_2 b_n \\ \vdots \\ a_m b_1, & a_m b_2, & a_m b_3, & \cdots & a_m b_n \end{bmatrix}_{m\times n}$$

$$= [b_1 \cdot a, b_2 \cdot a, b_3 \cdot a, \cdots b_n \cdot a]_{m\times n}$$

$$= b^T a = ab^T$$

② $\frac{d(a^T X^T b)}{dX} = ba^T$, Here $a_{m\times1}, b_{m\times1}$ are constant vectors. $a=(a_1, a_2 \cdots a_n)^T$, $(b_1, b_2, \cdots b_m)^T$.

Prove:

$$\frac{d(a^T X^T b)}{dX} = \frac{d(b^T X a)}{dX} \stackrel{②}{=} ba^T$$

③ $\frac{d(a^T X X^T b)}{dX} = ab^T X + ba^T X$. Here $a_{m\times1}, b_{m\times1}$ are constant vectors. $a=(a_1 \cdots a_m)^T, b=(b_1 \cdots b_m)^T$

Prove:

$$a^T X X^T b = \underset{1\times m}{(a_1, a_2 \cdots a_m)} \underset{m\times m}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & & & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}} \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & & & \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

$$= (a_1 \cdots a_m) \underset{m\times m}{\begin{pmatrix} x_{11}^2 + x_{12}^2 + \cdots x_{1m}^2, & x_{11}x_{21} + x_{12}x_{22} + x_{13}x_{23} + \cdots x_{1m}x_{2m}, & x_{11}x_{31} + \cdots x_{1m}x_{3m}, \cdots, & x_{11}x_{m1} + \cdots x_{1m}x_{mm} \\ x_{21}x_{11} & & & \\ & \cdot & & \end{pmatrix}} \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

$$= (a_1 b_1)(x_{11}x_{11} + x_{12}x_{12} + \cdots x_{1n}x_{1n}) + (a_1 b_2)(x_{11}x_{21} + x_{12}x_{22} + \cdots + x_{1n}x_{2n})$$
$$+ \cdots + (a_1 b_m)(x_{11}x_{m1} + x_{12}x_{m2} + \cdots + x_{1m}x_{mn})$$
$$+ (a_2 b_1)(x_{21}x_{11} + x_{22}x_{12} + \cdots x_{2n}x_{1n}) + (a_2 b_2)(x_{21}x_{21} + x_{22}x_{22} + \cdots + x_{2n}x_{2n})$$
$$+ \cdots + (a_2 b_m)(x_{21}x_{m1} + x_{22}x_{m2} + \cdots + x_{2n}x_{mn})$$
$$+ \cdots + (a_m b_1)(x_{m1}x_{11} + x_{m2}x_{12} + \cdots + x_{mn}x_{1n}) +$$
$$(a_m b_2)(x_{m1}x_{21} + x_{m2}x_{22} + \cdots + x_{mn}x_{2n}) + \cdots + (a_m b_m)(x_{m1}x_{m1} + \cdots x_{mn}x_{mn})$$

so..

$$\frac{d\, a^T x x^T b}{dx} = \begin{bmatrix} (a_1 b_1 x_{11} + a_1 b_2 x_{21} + \cdots + a_1 b_m x_{m1}) + (b_1 a_1 x_{11} + b_1 a_2 x_{21} + \cdots + b_1 a_m x_{m1}), & \cdots \\ (a_2 b_1 x_{11} + a_2 b_2 x_{21} + \cdots + a_2 b_m x_{m1}) + (b_2 a_1 x_{11} + b_2 a_2 x_{21} + \cdots + b_2 a_m x_{m1}), & \cdots \\ \vdots & \\ (a_m b_1 x_{11} + a_m b_2 x_{21} + \cdots + a_m b_m x_{m1}) + (b_m a_1 x_{11} + \cdots b_m a_2 x_{21} + b_m a_m x_{m1}) + & \end{bmatrix}$$

$$= A + B$$

$$= \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_m \\ \vdots & & & \\ a_m b_1 & a_m b_2 & \cdots & a_m b_m \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & & & x_{2n} \\ \vdots & & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} + \begin{bmatrix} b_1 a_1 & \cdots & b_1 a_m \\ b_2 a_1 & & \vdots \\ \vdots & & \vdots \\ b_m a_1 & \cdots & b_m a_m \end{bmatrix} X$$

$$= a b^T X + b a^T X$$

④ $\dfrac{d(a^T X^T X b)}{dX} = Xba^T + Xab^T$, $a_{n\times1}, b_{n\times1}$ are constant vectors.
$a = (a_1 \cdots a_n)^T$
$b = (b_1 \cdots b_n)^T$

prove:

we know:

$$D_x f(x) = \frac{df(X)}{dX^T_{m\times n}} = \begin{bmatrix} \frac{df}{dx_{11}} & \frac{df}{dx_{21}} & \cdots & \frac{df}{dx_{m1}} \\ \vdots & & & \\ \frac{df}{dx_{1n}} & \frac{df}{dx_{2n}} & \cdots & \frac{df}{dx_{mn}} \end{bmatrix}_{n\times m}$$

$$\nabla_x f(x) = \frac{df(x)}{dX_{m\times n}} = \begin{bmatrix} \frac{df}{dx_{11}} & \frac{df}{dx_{12}} & \cdots & \frac{df}{dx_{1n}} \\ \vdots & & & \\ \frac{df}{dx_{m1}} & \frac{df}{dx_{m2}} & \cdots & \frac{df}{dx_{mn}} \end{bmatrix}_{m\times n}$$

So,

$$\frac{df(x)}{dX^T_{m\times n}} = \left(\frac{df(x)}{dX_{m\times n}}\right)^T$$

From ③: $\dfrac{d(a^T XX^T b)}{dX^T} = \left(\dfrac{d(a^T XX^T b)}{dX}\right)^T = X^T b a^T + X^T a b^T$

And: $\dfrac{d(a^T X^T X b)}{dX} = \dfrac{d(a^T (X^T)(X^T)^T b)}{d(X^T)^T} = (X^T)^T b a^T + (X^T)^T a b^T$

$= Xba^T + Xab^T$

proved //