矩阵求导法则 $\begin{cases}\text{简洁}\\ \text{加速计算机计算速度}\end{cases}$

$f: R \to R^{2\times 2}$ $f(x) = \begin{bmatrix} x & x^2 \\ x^3 & x^4 \end{bmatrix}$

$f: R^2 \to R^{2\times 2}$ $f(x_1, x_2) = \begin{bmatrix} x_1+x_2 & x_1^2+x_2^2 \\ x_1^3+x_2^3 & x_1^4+x_2^4 \end{bmatrix}$

输入为向量函数 $\longrightarrow$ 输出为向量函数

矩阵求导:

$\frac{dA}{dB}$: 矩阵A中的每个元素对矩阵B中每一个元素求导

举例说明: $A \Rightarrow 1\times 1$, $B \Rightarrow 1\times 1$, 则 $\frac{dA}{dB} \Rightarrow 1\times 1$

$A \Rightarrow 1\times p$, $B \Rightarrow 1\times n$, 则 $\frac{dA}{dB} \Rightarrow p\times n$

$A \Rightarrow q\times p$, $B \Rightarrow m\times n$, 则 $\frac{dA}{dB} \Rightarrow p\times q \times m\times n$.

求导秘术: YX拉伸 $\begin{cases}\text{① 标量不变, 向量拉伸}\\ \text{② 前面横向拉, 后面纵向拉.}\end{cases}$

例1: $\frac{df(x)}{dx}$ 若 $f(x)$ 是标量函数 $x$ 是向量

即 $f(x) = f(x_1, x_2, \cdots x_n)$

则 $\frac{df(x)}{dx} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$

例2: $\frac{df(x)}{dx}$ 若 $f(x)$ 是向量 $x$ 是标量

即 $f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix}$

则 $\frac{df(x)}{dx} = \left[\frac{\partial f_1}{\partial x} \quad \frac{\partial f_2}{\partial x} \quad \cdots\cdots \quad \frac{\partial f_n}{\partial x}\right]$

例3: $\frac{df(x)}{dx}$  若 $f(x)$ 是向量，$x$ 是向量

$$f(x)=\begin{bmatrix} f_1(x)\\ f_2(x)\\ \cdots \\ f_m(x)\end{bmatrix} \qquad x=\begin{bmatrix}x_1\\ x_2\\ \cdots\\ x_n\end{bmatrix}$$

$$\text{则}\ \frac{\partial f}{\partial x}=\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \cdots & \cdots & & \cdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

的转置

常见矩阵求导公式:

① $f(x)=A^TX$  $A: n\times 1$  $X: n\times 1$

则 $f'(x)=A$  同理若 $f(x)=X^TA$ 则 $\frac{df(x)}{dx}=A$.

② $f(x)=X^TAX$  $X: n\times 1$  $A: n\times n$

则 $\frac{df(x)}{dx}=(A+A^T)X$

最小二乘为例

$$Y=\begin{bmatrix}y_1\\ y_2\\ \cdots\\ y_n\end{bmatrix} \qquad X=\begin{bmatrix}x_1^T\\ x_2^T\\ \cdots\\ x_n^T\end{bmatrix} \qquad b=\begin{bmatrix}b_1\\ b_2\\ \cdots\\ b_p\end{bmatrix}$$

$X: n\times p$
$b: p\times 1$
$Y: n\times 1$

$$J(b)=\sum_{k=1}^{n}(y_i-x_i^Tb)^2$$

向量化形式为 $J(b)=(Y-Xb)^T(Y-Xb)$

$$= (Y^T - b^T X^T)(Y - Xb)$$

$$= Y^T Y - Y^T X b - b^T X^T Y + b^T X^T X b$$

$$= Y^T Y - 2Y^T X b + b^T X^T X b$$

则 $\frac{\partial J(b)}{\partial b} = -2X^T Y + (X^T X + X^T X) b$

$$= -2X^T Y + 2X^T X b$$

$$= 2X^T (X - Y) b$$

向量对向量求导:

a是与x无关的函数 $\frac{\partial a}{\partial x} = 0$

$f(x) = x$ $\frac{\partial f}{\partial x} = 1$

$\frac{\partial Ax}{\partial x} = A^T$

$\frac{\partial x^T A}{\partial x} = A$

a是与x无关函数. $\frac{\partial a u}{\partial x} = a \frac{\partial u}{\partial x}$

$a = a(x)$ $u = u(x)$ $\frac{\partial a u}{\partial x} = a \frac{\partial u}{\partial x} + \frac{\partial a}{\partial x} u^T$

A是矩阵,与x无关 $\frac{\partial A u}{\partial x} = \frac{\partial u}{\partial x} A^T$