# 矩阵求导笔记

## Notes 01：

标量函数：输出为标量的函数

eg： 1. $f(x)=x^2$  $R \to R$

2. $f(x_1,x_2)=x_1^2+x_2^2$  $R^2 \to R$

向量函数：输出为向量的函数

eg： 1. $f(x)=\begin{bmatrix} f_1(x)=x \\ f_2(x)=x^2 \end{bmatrix}$  $R \to R^2$  $x \to \begin{bmatrix} x \\ x^2 \end{bmatrix}$

2. $f(x)=\begin{bmatrix} f_{11}(x)=x_1+x_2 & f_{12}(x)=x_1^2+x_2^2 \\ f_{21}(x)=x_1^3+x_2^3 & f_{22}(x)=x_1^4+x_2^4 \end{bmatrix}$  $R^2 \to R^{2\times 2}$  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to \begin{bmatrix} x_1+x_2 & x_1^2+x_2^2 \\ x_1^3+x_2^3 & x_1^4+x_2^4 \end{bmatrix}$

矩阵求导的本质：

设 $A,B$ 分别为两个矩阵，则 $\frac{dA}{dB}$ = 矩阵A中的每一个元素对矩阵B中的每一个元素求导

| | $A$ | $B$ | $\frac{dA}{dB}$ |
|---|---|---|---|
| 形状： | $1\times1$ | $1\times1$ | $1\times1$ |
| 形状： | $1\times p$ | $1\times n$ | $p\times n$ |
| 形状： | $q\times p$ | $m\times n$ | $p\times q\times m\times n$ |

## Notes 02：

两大法则 { 标量不变，向量拉伸
Y/f(x) 横向拉，X 纵向拉. (分母布局)

eg 1：设 $f(x)$ 是标量函数，$x$ 是向量 即 $f(x)=f(x_1,x_2,x_3\cdots x_n)$, $X=[x_1,x_2,\cdots x_n]^T$，求 $\frac{df(x)}{dx}$

此时 $f(x)$ 是标量函数，不需拉伸，$x$ 纵向拉指将 $x$ 表示成列向量的形式，实际是将多元函数的偏导写成列向量

$$\frac{df(x)}{dx}=\begin{bmatrix} \frac{df(x)}{dx_1} \\ \frac{df(x)}{dx_2} \\ \vdots \\ \frac{df(x)}{dx_n} \end{bmatrix}$$

$X.shape=(n,)$, $f(x).shape=(1)$ $\longrightarrow$ $\frac{df(x)}{dx}.shape=(n,)$

eg 2：设 $f(x)$ 是向量函数，$x$ 为标量，即 $f(x)=[f_1(x),f_2(x),f_3(x),\cdots,f_n(x)]^T$

此时 $f(x)$ 是向量，需横向拉伸，写成行向量的形式，$x$ 为标量不变

$$\frac{df(x)}{dx}=\left[\frac{df_1(x)}{dx},\frac{df_2(x)}{dx},\cdots,\frac{df_n(x)}{dx}\right]$$

eg 3：设 $f(x)$，$x$ 均为向量，即 $f(x)=[f_1(x),f_2(x),\cdots f_n(x)]^T$，$x=[x_1,x_2,\cdots,x_n]^T$

① 第一步，$f(x)$ 为向量，则先进行横向拉伸

$f(x)$ 对 $x$ 求导即 $f(x)$ 中每个元素分别对 $x$ 求导，即：

$$\frac{df(x)}{dx}=\left[\frac{\partial f_1(x)}{\partial x},\frac{\partial f_2(x)}{\partial x},\frac{\partial f_3(x)}{\partial x},\cdots,\frac{\partial f_n(x)}{\partial x}\right]$$

② $f(x)$ 中的元素 $f_i(x)$ 为标量，第一步先用 $f_i(x)$ 对 $x$ 求导

即

$$\frac{df_1(x)}{dx}=\begin{pmatrix}\frac{\partial f_1(x)}{\partial x_1}\\ \frac{\partial f_1(x)}{\partial x_2}\\ \vdots\\ \frac{\partial f_1(x)}{\partial x_n}\end{pmatrix}$$

类似 eg 1，$f_1(x)$ 为标量，$x$ 为向量，$x$ 纵向拉伸为列向量

则

$$\frac{df(x)}{dx}=\begin{bmatrix}\frac{\partial f_1(x)}{\partial x_1}&\frac{\partial f_2(x)}{\partial x_1}&\frac{\partial f_3(x)}{\partial x_1}&\cdots&\frac{\partial f_n(x)}{\partial x_1}\\ \frac{\partial f_1(x)}{\partial x_2}&\frac{\partial f_2(x)}{\partial x_2}&\frac{\partial f_3(x)}{\partial x_2}&\cdots&\frac{\partial f_n(x)}{\partial x_2}\\ \vdots&\vdots&\vdots&\cdots&\vdots\\ \frac{\partial f_1(x)}{\partial x_n}&\frac{\partial f_2(x)}{\partial x_n}&\frac{\partial f_3(x)}{\partial x_n}&\cdots&\frac{\partial f_n(x)}{\partial x_n}\end{bmatrix}$$

求导完为 $n\times n$ 的形状。

## Notes 03：

eg 1：$f(x)=A^T\cdot x=\sum_{i=1}^{n}a_ix_i$，$A=[a_1,a_2,a_3\cdots a_n]^T\in R^{n\times1}$，$x=[x_1,x_2\cdots x_n]^T\in R^{n\times1}$

求 $\frac{df(x)}{dx}$，即标量函数对向量求导

$$\frac{df(x)}{dx}=\begin{bmatrix}\frac{\partial f(x)}{\partial x_1}\\ \frac{\partial f(x)}{\partial x_2}\\ \vdots\\ \frac{\partial f(x)}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\frac{\partial(a_1x_1+a_2x_2+\cdots+a_nx_n)}{\partial x_1}\\ \vdots\\ \frac{\partial(a_1x_1+a_2x_2+\cdots+a_nx_n)}{\partial x_n}\end{bmatrix}=\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_{n-1}\\ a_n\end{bmatrix}=A$$

公式总结1：$\frac{dA^T\cdot X}{dX}=\frac{dX^T\cdot A}{dX}=A$

eg2：$f(x)=X^T\cdot A X$（二次型形式），其中 $X=[x_1,x_2,\cdots x_n]^T$，$A=\begin{bmatrix} a_{11},a_{12},\cdots a_{1n}\\ a_{21},a_{22},\cdots a_{2n}\\ \vdots \\ a_{n1},a_{n2},\cdots a_{nn}\end{bmatrix}$

求 $\frac{df(x)}{dX}$

分析：$f(x)=\underset{(1\times n)}{X^T}\cdot\underset{(n\times n)}{A}\cdot\underset{(n\times 1)}{X}$，其中 X 为 $(n\times 1)$，A为 $(n\times n)$，即 $f(x)$ 为 $1\times 1$ 的标量函数

对应

$$f(x)=[x_1,x_2,\cdots x_n]\begin{bmatrix} a_{11},a_{12},\cdots a_{1n}\\ a_{21},a_{22},\cdots a_{2n}\\ \vdots \\ a_{n1},a_{n2},\cdots a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j$$

对应

$$\frac{df(x)}{dX}=\begin{bmatrix}\frac{\partial f(x)}{\partial x_1}\\\frac{\partial f(x)}{\partial x_2}\\\vdots\\\frac{\partial f(x)}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\frac{\partial(\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j)}{\partial x_1}\\\vdots\\\frac{\partial(\sum_{i=1}^n\sum_{j=1}^n a_{ij}x_ix_j)}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^n a_{1j}x_j+\sum_{i=1}^n a_{i1}x_i\\\sum_{j=1}^n a_{2j}x_j+\sum_{i=1}^n a_{i2}x_i\\\vdots\\\sum_{j=1}^n a_{nj}x_j+\sum_{i=1}^n a_{in}x_i\end{bmatrix}=\begin{bmatrix}\sum_{j=1}^n a_{1j}x_j\\\sum_{j=1}^n a_{2j}x_j\\\vdots\\\sum_{j=1}^n a_{nj}x_j\end{bmatrix}+\begin{bmatrix}\sum_{i=1}^n a_{i1}x_i\\\sum_{i=2}^n a_{i2}x_j\\\vdots\\\sum_{j=1}^n a_{in}x_j\end{bmatrix}$$

$$=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&&&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}+\begin{bmatrix}a_{11}&a_{21}&\cdots&a_{n1}\\a_{12}&a_{22}&\cdots&a_{n2}\\\vdots&&&\vdots\\a_{1n}&a_{2n}&\cdots&a_{nn}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=AX+A^TX$$

公式总结2：$\frac{dX^TAX}{dX}=(A+A^T)X$

Notes 04：

1. 两种布局 $\begin{cases}\text{分母布局}\quad(\text{Y 横向拉, X 纵向拉})\\ \text{分子布局}\quad(\text{X 横向拉, Y 纵向拉})\end{cases}$ ⇒ 求导后元素的排列方式不同

(分母布局$)^T$ = 分子布局

eg1：$f(x)=X^T\cdot X$，其中 $X=[x_1,x_2,\cdots x_n]^T$

X的形状为 $n\times 1$ 列向量，即 $\underset{(1\times n)}{X^T}\cdot\underset{(n\times 1)}{X}=1\times 1$，$f(x)$ 为标量。

分子布局：X 横向拉。 $\frac{df(x)}{dX}=\left[\frac{\partial f(x)}{\partial x_1}\;\frac{\partial f(x)}{\partial x_2}\cdots\frac{\partial f(x)}{\partial x_n}\right]=[2x_1,2x_2,\cdots 2x_n]=2X^T$

分母布局： X 纵向排.

$$\frac{df(x)}{dx}=\begin{bmatrix}\frac{\partial f(x)}{\partial x_1}\\ \frac{\partial f(x)}{\partial x_2}\\ \vdots\\ \frac{\partial f(x)}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\frac{\partial \sum_{i=1}^{n} x_i^2}{\partial x_1}\\ 2x_2\\ \vdots\\ 2x_n\end{bmatrix}=\begin{bmatrix}2x_1\\ 2x_2\\ \vdots\\ 2x_n\end{bmatrix}=2X$$

eg 2：设 $U=[u_1(x),\ u_2(x),\ \cdots\ u_n(x)]^T_{n\times1}$ , $V=[v_1(x),\ v_2(x),\ v_3(x)\cdots v_n(x)]^T_{n\times1}$

$X=[x_1,x_2,x_3\cdots x_n]^T_{n\times1}$

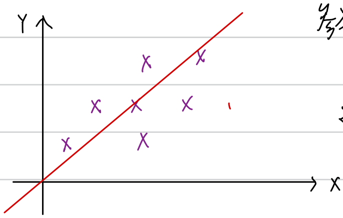
① $\frac{du^Tv}{dx}=\frac{\partial u}{\partial x}\cdot v+\frac{\partial v}{\partial x}u$ 矩阵求导的乘法公式

② $\frac{d(v+u)}{dx}=\frac{dv}{dx}+\frac{du}{dx}$ 矩阵求导的加法公式

Notes 05:

最小二乘法（以分母布局推导） 样本：$X=[x_1,x_2,\cdots x_n]^T$ , $Y=[y_1,y_2\cdots y_n]^T$

参数：$W=[w_1,w_2,\cdots w_n]^T$




损失函数 $L=\sum_{i=1}^{n}(y_i-x_i^Tw)^2 \xrightarrow{\text{向量化}} (Y-XW)^T\cdot(Y-XW)$

优化目标：使损失函数最小.

$$\begin{aligned}L&=(Y-XW)^T\cdot(Y-XW)\\&=(Y^T-W^TX^T)(Y-XW)\\&=Y^TY-Y^TXW-W^TX^TY-W^TX^T\cdot XW\\&=Y^TY-2Y^TXW-W^TX^TXW\end{aligned}$$

由于 $(Y^TXW)^T=W^TX^TY$ 且 $(Y^TXW)$ 为一个标量 即 $Y^TXW=W^TX^TY$

$$\frac{dL(x)}{dw}=\frac{dY^TY}{dw}-\frac{2d\overset{\to A}{Y^TX}W}{dw}-\frac{dW^T\overset{\to A}{X^TX}W}{dw}$$

$$=\begin{bmatrix}0\\0\\0\\ \vdots\\0\end{bmatrix}_{n\times1}-2(Y^TX)^T-(X^TX+X^TX)W$$

套公式① $\frac{dA^Tx}{dx}=\frac{dx^TA}{dx}=A$

套公式② $\frac{dx^TAx}{dx}=(A+A^T)x$

令 $\frac{dL(W)}{dW} = -2X^TY + 2X^TXW = 0$ 即 $X^TY = X^TXW \Rightarrow \hat{W} = (X^TX)^{-1}X^TY$