矩阵微分法

主要内容：

一. 相对于<u>数量变量</u>的微分法

二. 相对于<u>向量（变量）</u>的微分法

- 1.数量函数对向量的导数
- 2.向量函数对向量的导数
- 3.矩阵函数对向量的导数

一. 相对于数量变量的微分

定义1：

$$a(t)=\begin{bmatrix}a_1(t) & a_2(t) & \cdots & a_n(t)\end{bmatrix}^{\mathrm{T}}$$

$$\Rightarrow \quad \frac{da(t)}{dt}=\begin{bmatrix}\frac{da_1(t)}{dt} & \frac{da_2(t)}{dt} & \cdots & \frac{da_n(t)}{dt}\end{bmatrix}^{\mathrm{T}}$$

一. 相对于数量变量的微分

定义2:

$$A=[a_{ij}(t)]_{m\times n}$$

$$\Longrightarrow \quad \frac{\mathrm{d}A}{\mathrm{d}t}=\left[\frac{\mathrm{d}a_{ij}(t)}{\mathrm{d}t}\right]_{m\times n}$$

一. 相对于数量变量的微分

运算法则:

$$\frac{d(A \pm B)}{dt} = \frac{dA}{dt} \pm \frac{dB}{dt}$$

$$\frac{d(\lambda A)}{dt} = \frac{d\lambda}{dt} A + \lambda \frac{dA}{dt}$$

$$\frac{d}{dt}(a^{\mathrm{T}} b) = \frac{da^{\mathrm{T}}}{dt} b + a^{\mathrm{T}} \frac{db}{dt}$$

$$\frac{d}{dt}(AB) = \frac{dA}{dt} B + A \frac{dB}{dt}$$

什么是向量的函数？

$$y = f(x_1, x_2) = 5x_1 + 3x_2 = \begin{bmatrix} 5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

定义列向量：

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathrm{T}}$$

y可以表示为： $y = f(x) \qquad x \in R^2$

说明：若无特殊声明，小写、加黑且无下标的字母一般都表示列向量（λ 除外）。

二. 相对于向量的微分

1. 数量函数相对于向量的微分

定义3:

$$f(x)=f(x_1,x_2,\cdots x_n) \qquad x=[x_1,x_2\cdots x_n]^{\mathrm{T}}$$

$$\begin{cases} \dfrac{df(x)}{dx}=\begin{bmatrix}\dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n}\end{bmatrix}^{\mathrm{T}} \\ \dfrac{df(x)}{dx^{\mathrm{T}}}=\begin{bmatrix}\dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n}\end{bmatrix} \end{cases}$$

数学中梯度的定义，表示为

$grad[f(x)]$

或$\nabla f(x)$

二. 相对于向量的微分

1. 数量函数相对于向量的微分

运算法则： $f(x), g(x)$

$$\frac{d(f \pm g)}{dx} = \frac{df}{dx} \pm \frac{dg}{dx}$$

$$d(fg) = \frac{df}{dx} g + f \frac{dg}{dx}$$

二. 相对于向量的微分

2. 向量函数相对于向量的微分

设：

$$a(x)=\begin{bmatrix} a_1(x) \\ a_2(x) \\ \vdots \\ a_m(x) \end{bmatrix} \qquad x=\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}}$$

其中：$a_i(x)$ ——数量函数

二. 相对于向量的微分

2. 向量函数相对于向量的微分

定义4:

$$\frac{\mathrm{d}a(x)}{\mathrm{d}x^{\mathrm{T}}}=\begin{bmatrix}\frac{\partial a_1}{\partial x_1} & \frac{\partial a_1}{\partial x_2} & \cdots & \frac{\partial a_1}{\partial x_n}\\ \frac{\partial a_2}{\partial x_1} & \frac{\partial a_2}{\partial x_2} & \cdots & \frac{\partial a_2}{\partial x_n}\\ \vdots & \vdots & \vdots & \vdots\\ \frac{\partial a_m}{\partial x_1} & \frac{\partial a_m}{\partial x_2} & \cdots & \frac{\partial a_m}{\partial x_n}\end{bmatrix}=\left[\frac{\partial a_i}{\partial x_j}\right]_{m\times n}$$

二. 相对于向量的微分

2. 向量函数相对于向量的微分

定义4:

$$\frac{\mathrm{d}a^{\mathrm{T}}(x)}{\mathrm{d}x}=\begin{bmatrix}\frac{\partial a_1}{\partial x_1} & \frac{\partial a_2}{\partial x_1} & \cdots & \frac{\partial a_m}{\partial x_1}\\ \frac{\partial a_1}{\partial x_2} & \frac{\partial a_2}{\partial x_2} & \cdots & \frac{\partial a_m}{\partial x_2}\\ \vdots & \vdots & \vdots & \vdots\\ \frac{\partial a_1}{\partial x_n} & \frac{\partial a_2}{\partial x_n} & \cdots & \frac{\partial a_m}{\partial x_n}\end{bmatrix}=\left[\frac{\partial a_j}{\partial x_i}\right]_{n\times m}$$

二. 相对于向量的微分

2. 向量函数相对于向量的微分

运算法则：

$$\frac{d}{dx}(a^{\mathrm{T}} \pm b^{\mathrm{T}}) = \frac{d(a^{\mathrm{T}})}{dx} \pm \frac{d(b^{\mathrm{T}})}{dx}$$

$$\frac{d}{dx}(\lambda a^{\mathrm{T}}) = \frac{d\lambda}{dx} a^{\mathrm{T}} + \lambda \frac{da^{\mathrm{T}}}{dx}$$

$$\frac{d}{dx}(a^{\mathrm{T}} b) = \frac{d(a^{\mathrm{T}})}{dx} b + \frac{d(b^{\mathrm{T}})}{dx} a$$

二. 相对于向量的微分

2. 向量函数相对于向量的微分

两个有用的等式:

$$\frac{dx}{dx^{\mathrm{T}}} = I$$

$$\frac{dx^{\mathrm{T}}}{dx} = I$$

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

设：

$$A(x)=\begin{bmatrix} a_{11}(x) & a_{12}(x) & \cdots & a_{1l}(x) \\ a_{21}(x) & a_{22}(x) & \cdots & a_{2l}(x) \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}(x) & a_{m2}(x) & \cdots & a_{ml}(x) \end{bmatrix}_{m\times l}$$

其中： $a_{ij}(x)$为数量函数

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

定义5：

$$\frac{dA(x)}{dx}=\begin{bmatrix}\frac{\partial A(x)}{\partial x_1}\\ \frac{\partial A(x)}{\partial x_2}\\ \vdots\\ \frac{\partial A(x)}{\partial x_n}\end{bmatrix}_{nm\times l}$$

$m\times l$

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

定义5：

$$\frac{dA(x)}{dx^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial A(x)}{\partial x_1} & \frac{\partial A(x)}{\partial x_2} & \cdots & \frac{\partial A(x)}{\partial x_n} \end{bmatrix}_{m\times \mathrm{l}n}$$

$m\times l$

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

定义5：

$$\frac{dA(x)}{dx^{\mathrm{T}}}=\begin{bmatrix}\dfrac{\partial A(x)}{\partial x_1} & \dfrac{\partial A(x)}{\partial x_2} & \cdots & \dfrac{\partial A(x)}{\partial x_n}\end{bmatrix}_{m\times ln}$$

$m\times l$

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

其中：

$$\frac{\partial A(x)}{\partial x_i} = \begin{bmatrix} \frac{\partial a_{11}(x)}{\partial x_i} & \frac{\partial a_{12}(x)}{\partial x_i} & \cdots & \frac{\partial a_{1l}(x)}{\partial x_i} \\ \frac{\partial a_{21}(x)}{\partial x_i} & \frac{\partial a_{22}(x)}{\partial x_i} & \cdots & \frac{\partial a_{2l}(x)}{\partial x_i} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial a_{m1}(x)}{\partial x_i} & \frac{\partial a_{m2}(x)}{\partial x_i} & \cdots & \frac{\partial a_{ml}(x)}{\partial x_i} \end{bmatrix}_{m\times l}$$

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

运算法则（加法）：

$$\frac{d(A \pm C)}{dx} = \frac{dA}{dx} \pm \frac{dC}{dx}$$

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

运算法则（数乘）：

$$\frac{d(\lambda A)}{dx} = \left[\frac{d\lambda}{dx}\right] A + \lambda \frac{dA}{dx}$$

$$\left[\frac{d\lambda}{dx}\right] A \stackrel{\Delta}{=\!=} \begin{bmatrix} \dfrac{\partial \lambda}{\partial x_1} A \\ \dfrac{\partial \lambda}{\partial x_2} A \\ \vdots \\ \dfrac{\partial \lambda}{\partial x_n} A \end{bmatrix}$$

说明：$\left[\dfrac{d\lambda}{dx}\right] A$ 仅为一种表示方式。

二. 相对于向量的微分

3. 矩阵函数相对于向量的微分

运算法则（乘法）：

$$\frac{d}{dx}(AB)=\frac{dA}{dx}B+A\left[\frac{dB}{dx}\right]$$

$$A\left[\frac{dB}{dx}\right]\xrightarrow{\triangleq}\begin{bmatrix}A\frac{\partial B}{\partial x_1}\\ A\frac{\partial B}{\partial x_2}\\ \vdots\\ A\frac{\partial B}{\partial x_n}\end{bmatrix}$$