§1 回顾变量的微分学

在学习向量和矩阵的求导之前, 最好回顾关于一元函数、多元函数的微分学.

一元函数的微分学 微分算子 $\mathscr{D}$ 是一个线性算子. 这就意味着对于映射 f, g、常数 k, l, 有 $\mathscr{D}(kf+lg) = k\mathscr{D}(f)+l\mathscr{D}(g)$. 此外, 我们还有链式法则, 即 $f(g(x)) = f'_{g(x)}g'_x$.

多元函数的微分学 最重要的是考虑多元函数的复合函数.

§2 向量求导: 两种格式

一句话概括: 向量求导无非就是将多个变量排成了一个比较好的格式. 并且在这个比较好的格式上面求微分. 再继续之前, 我们先来回顾几个重要的技巧:

1. **作为点积的求和**. 对于形如 $\sum_i u_i v_i$ 的式子, 可以使用两个列向量 $\mathbf{u}, \mathbf{v}$ 并且写作点积的形式: 即 $\sum u_i v_i = \mathbf{u}\cdot\mathbf{v} = \mathbf{u}'\mathbf{v}$.
2. **作为矩阵乘法的点积和**. 对于一个矩阵 $\mathbf{X}$, 其中如果第 i 行第 j 列的元素是 $\mathbf{u}_i\cdot\mathbf{v}_j$, 可以定义两个矩阵 $\mathbf{U} = \begin{bmatrix} | & | & & | & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_{n-1} & \mathbf{u}_n \\ | & | & & | & | \end{bmatrix}$、$\mathbf{V} = \begin{bmatrix} | & | & & | & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{n-1} & \mathbf{v}_n \\ | & | & & | & | \end{bmatrix}$, 使得 $\mathbf{X} = \mathbf{U}'\mathbf{V}$.

2.1. 基本定义.

定义 1. 向量 $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_m \end{bmatrix}'$ 对于标量 x 的求导记作

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} := \begin{bmatrix} \frac{\mathrm{d}y_1}{\mathrm{d}x} & \frac{\mathrm{d}y_2}{\mathrm{d}x} & \frac{\mathrm{d}y_3}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_m}{\mathrm{d}x} \end{bmatrix}'.$$

注. (1) 这通常称为向量 $\mathbf{y}$ 的切向量. 实际上这里映射 $\mathbf{y}: \mathbb{R}^1 \to \mathbb{R}^m$.

(2) 这里我们采用竖着排版的形式. 这样的情形只是一个习惯性操作. 倘若定义 $\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} := \begin{bmatrix} \frac{\mathrm{d}y_1}{\mathrm{d}x} & \frac{\mathrm{d}y_2}{\mathrm{d}x} & \frac{\mathrm{d}y_3}{\mathrm{d}x} & \cdots & \frac{\mathrm{d}y_m}{\mathrm{d}x} \end{bmatrix}$, 那么后续的内容除了符号上有更改以外, 没有太多区别.

定义 2. 标量 y 对于向量 $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_m \end{bmatrix}'$ 的**分子形式**记作

$$\frac{\partial y}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\mathrm{d}y}{\mathrm{d}x_1} & \frac{\mathrm{d}y}{\mathrm{d}x_2} & \frac{\mathrm{d}y}{\mathrm{d}x_3} & \cdots & \frac{\mathrm{d}y}{\mathrm{d}x_m} \end{bmatrix}.$$

其**分母形式**可以类似记作

$$\frac{\partial y}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\mathrm{d}y}{\mathrm{d}x_1} & \frac{\mathrm{d}y}{\mathrm{d}x_2} & \frac{\mathrm{d}y}{\mathrm{d}x_3} & \cdots & \frac{\mathrm{d}y}{\mathrm{d}x_m} \end{bmatrix}'.$$