矩阵计算（矩阵求导）

标量导数

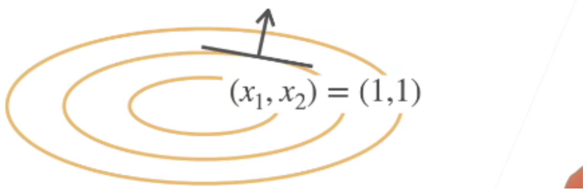
- 导数是切线的斜率

y	a	x^n	$\exp(x)$	$\log(x)$	$\sin(x)$
$\frac{dy}{dx}$	0	nx^{n-1}	$\exp(x)$	$\frac{1}{x}$	$\cos(x)$

- 指向值变化最大的方向

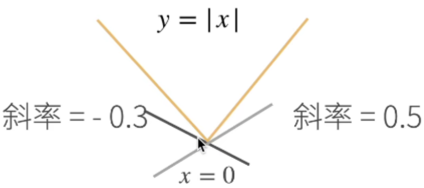
$$\frac{\partial}{\partial \mathbf{x}} x_1^2 + 2x_2^2 = [2x_1, 4x_2]$$

方向 (2, 4) 跟等高线正交



亚导数

- 将导数拓展到不可微的函数



$$\frac{\partial |x|}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ a & \text{if } x = 0, \quad a \in [-1,1] \end{cases}$$

函数与标量，向量，矩阵

1、f 为是一个标量

1.1 input是一个标量

$$f(x) = x + 2$$

1.2 input是一个向量

$$\boldsymbol{x} = [x_1, x_2, x_3]^T$$

$$f(\boldsymbol{x}) = a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2 + a_4 x_1 x_2$$

1.3 input是一个矩阵

$$\boldsymbol{X}_{3\times 2} = (x_{ij})_{i=1,j=1}^{3,2}$$

$$f(\boldsymbol{X}) = a_1 x_{11}^2 + a_2 x_{12}^2 + a_3 x_{21}^2 + a_4 x_{22}^2 + a_5 x_{31}^2 + a_6 x_{32}^2$$

2、f是一个向量

- **f**是由若干个f(标量)组成的向量

2.1 input是一个标量

$$\boldsymbol{f}_{3\times 1}(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix} = \begin{bmatrix} x+1 \\ 2x+1 \\ 3x^2+1 \end{bmatrix}$$

2.2 input是一个标量

$$\boldsymbol{x} = [x_1, x_2, x_3]^T$$

$$\boldsymbol{f}_{3\times 1}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ f_3(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_1^2 + 2x_2 + 2x_3 \\ x_1 x_2 + x_2 + x_3 \end{bmatrix}$$

2.3 input是一个矩阵

$$\boldsymbol{X}_{3\times 2} = (x_{ij})_{i=1,j=1}^{3,2}$$

$$\boldsymbol{f}_{3\times 1}(\boldsymbol{X}) = \begin{bmatrix} f_1(\boldsymbol{X}) \\ f_2(\boldsymbol{X}) \\ f_3(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} x_{11} + x_{12} + x_{21} + x_{22} + x_{31} + x_{32} \\ x_{11} + x_{12} + x_{21} + x_{22} + x_{31} + x_{32} + x_{11}x_{12} \\ 2x_{11} + x_{12} + x_{21} + x_{22} + x_{31} + x_{32} + x_{11}x_{12} \end{bmatrix}$$

3、F是一个矩阵

- **F**是一个由若干**f**组成的一个矩阵

3.1 input是一个标量

$$\boldsymbol{F}_{3\times 2}(x) = \begin{bmatrix} f_{11}(x) & f_{12}(x) \\ f_{21}(x) & f_{22}(x) \\ f_{31}(x) & f_{32}(x) \end{bmatrix} = \begin{bmatrix} x+1 & 2x+2 \\ x^2+1 & 2x^2+1 \\ x^3+1 & 2x^3+1 \end{bmatrix}$$

3.2 input是一个向量

$$\boldsymbol{x} = [x_1, x_2, x_3]^T$$

$$\boldsymbol{F}_{3\times 2}(\boldsymbol{x}) = \begin{bmatrix} f_{11}(\boldsymbol{x}) & f_{12}(\boldsymbol{x}) \\ f_{21}(\boldsymbol{x}) & f_{22}(\boldsymbol{x}) \\ f_{31}(\boldsymbol{x}) & f_{32}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 + x_3 & 2x_1 + 2x_2 + x_3 \\ 2x_1 + 2x_2 + x_3 & x_1 + 2x_2 + x_3 \\ 2x_1 + x_2 + 2x_3 & x_1 + 2x_2 + 2x_3 \end{bmatrix}$$

3.3 input是一个向量

$$\boldsymbol{X}_{3\times 2} = (x_{ij})_{i=1,j=1}^{3,2}$$

$$\boldsymbol{F}_{3\times 2}(\boldsymbol{X}) = \begin{bmatrix} f_{11}(\boldsymbol{X}) & f_{12}(\boldsymbol{X}) \\ f_{21}(\boldsymbol{X}) & f_{22}(\boldsymbol{X}) \\ f_{31}(\boldsymbol{X}) & f_{32}(\boldsymbol{X}) \end{bmatrix}$$

$$= \begin{bmatrix} x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} & 2x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \\ 3x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} & 4x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \\ 5x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} & 6x_{11}+x_{12}+x_{21}+x_{22}+x_{31}+x_{32} \end{bmatrix}$$

求导的本质

$$f(x_1, x_2, x_3) = x_1^2 + x_1 x_2 + x_2 x_3$$

可以将f对x1，x2，x3的偏导分别求出来，即

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 2x_1 + x_2 \\[2ex] \dfrac{\partial f}{\partial x_2} = x_1 + x_3 \\[2ex] \dfrac{\partial f}{\partial x_3} = x_2 \end{cases}$$

- 矩阵求导也是一样的，**本质就是** function 中的**每个** f **分别对变元中的每个元素逐个求偏导，只不过写成了向量、矩阵形式而已。**

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}_{3\times 1}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 + x_3 \\ x_2 \end{bmatrix}$$

（课上是按行向量展开的）

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_n}\right]$$

X为矩阵时，先把矩阵变元 $\boldsymbol{X}$ 进行**转置**，再对**转置后**的**每个位置**的元素逐个求偏导，结果布局和**转置布局一样。**（课上讲的是这种展开方式）

$$
\mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^T_{m\times n}}
$$

$$
= \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{21}} & \cdots & \frac{\partial f}{\partial x_{m1}} \\ \frac{\partial f}{\partial x_{12}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{m2}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial x_{1n}} & \frac{\partial f}{\partial x_{2n}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}_{n\times m}
$$

- 所以，如果 function 中有 m 个 f (标量)，变元中有 n 个元素，那么，每个 f 对变元中的每个元素逐个求偏导后，我们就会产生 $m \times n$ 个结果。

矩阵求导的布局

1.**分子布局**，就是分子是**列向量**形式，分母是**行向量**形式 （课上讲的）

$$
\frac{\partial \boldsymbol{f}_{2\times 1}(\boldsymbol{x})}{\partial \boldsymbol{x}^T_{3\times 1}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} \end{bmatrix}_{2\times 3}
$$

2.**分母布局**，就是分母是**列向量**形式，分子是**行向量**形式

$$
\frac{\partial \boldsymbol{f}^T_{2\times 1}(\boldsymbol{x})}{\partial \boldsymbol{x}_{3\times 1}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} \\ \frac{\partial f_1}{\partial x_3} & \frac{\partial f_2}{\partial x_3} \end{bmatrix}_{3\times 2}
$$