矩阵导数

1. 矩阵 Y=F(*x*)对标量 *x* 求导

相当于每个元素求导数

$$\frac{d\mathbf{Y}}{dx}=\begin{bmatrix}\frac{df_{11}(x)}{dx} & \frac{df_{12}(x)}{dx} & \cdots & \frac{df_{1n}(x)}{dx}\\ \frac{df_{21}(x)}{dx} & \frac{df_{22}(x)}{dx} & \cdots & \frac{df_{2n}(x)}{dx}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{df_{m1}(x)}{dx} & \frac{df_{m2}(x)}{dx} & \cdots & \frac{df_{mn}(x)}{dx}\end{bmatrix}$$

2. 标量 *y* 对列向量 x 求导

注意与上面不同，这次括号内是求偏导，对 m×1 向量求导后还是 m×1 向量

$$y=f(\mathbf{x})\rightarrow\frac{dy}{d\mathbf{x}}=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \frac{\partial f}{\partial x_2}\\ \vdots\\ \frac{\partial f}{\partial x_m}\end{bmatrix}$$

3. 行向量 $\mathbf{y}^{\mathrm{T}}$ 对列向量 x 求导

注意 1×n 向量对 m×1 向量求导后是 m×n 矩阵。

将 **y** 的每一列对 **x** 求偏导，将各列构成一个矩阵。

$$\frac{d\mathbf{y}^T}{d\mathbf{x}}=\begin{bmatrix}\frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_1} & \cdots & \frac{\partial f_n(x)}{\partial x_1}\\ \frac{\partial f_1(x)}{\partial x_2} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_n(x)}{\partial x_2}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_1(x)}{\partial x_m} & \frac{\partial f_2(x)}{\partial x_m} & \cdots & \frac{\partial f_n(x)}{\partial x_m}\end{bmatrix}$$

重要结论：

$$\frac{d\mathbf{x}^T}{d\mathbf{x}}=\mathbf{I}$$

$$\frac{d(\mathbf{A}\mathbf{x})^T}{d\mathbf{x}}=\mathbf{A}^T$$

4. 列向量 y 对行向量 $\mathbf{x}^T$ 求导

转化为行向量 $\mathbf{y}^T$ 对列向量 **x** 的导数，然后转置。

注意 m×1 向量对 1×n 向量求导结果为 m×n 矩阵。

$$\frac{d\mathbf{y}}{d\mathbf{x}^T} = \left(\frac{d\mathbf{y}^T}{d\mathbf{x}}\right)^T = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_1} \\ \frac{\partial f_1(x)}{\partial x_2} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(x)}{\partial x_n} & \frac{\partial f_2(x)}{\partial x_n} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}^T$$

重要结论：

$$\frac{d\mathbf{x}}{d\mathbf{x}^T} = \mathbf{I}$$

$$\frac{d(\mathbf{A}\mathbf{x})}{d\mathbf{x}^T} = \mathbf{A}$$

5. 向量积对列向量 x 求导运算法则

注意与标量求导有点不同。

$$\frac{d(\mathbf{u}^T\mathbf{v})}{d\mathbf{x}} = \frac{d(\mathbf{u}^T)}{d\mathbf{x}} \cdot \mathbf{v} + \frac{d(\mathbf{v}^T)}{d\mathbf{x}} \cdot \mathbf{u}$$

重要结论：

$$\frac{d(\mathbf{x}^T\mathbf{x})}{d\mathbf{x}} = \frac{d(\mathbf{x}^T)}{d\mathbf{x}} \cdot \mathbf{x} + \frac{d(\mathbf{x}^T)}{d\mathbf{x}} \cdot \mathbf{x} = 2\mathbf{x}$$

$$\frac{d(\mathbf{x}^T\mathbf{A}\mathbf{x})}{d\mathbf{x}} = \frac{d(\mathbf{x}^T)}{d\mathbf{x}} \cdot \mathbf{A}\mathbf{x} + \frac{d(\mathbf{x}^T\mathbf{A}^T)}{d\mathbf{x}} \cdot \mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$$

6. 矩阵 Y 对列向量 x 求导

将 Y 对 **x** 的每一个分量求偏导，构成一个超向量。
注意该向量的每一个元素都是一个矩阵。

$$\mathbf{Y}=\mathbf{F}(\mathbf{x}) \to \frac{d\mathbf{Y}}{d\mathbf{x}}=\begin{bmatrix} \frac{\partial \mathbf{F}}{\partial x_1} \\ \frac{\partial \mathbf{F}}{\partial x_2} \\ \vdots \\ \frac{\partial \mathbf{F}}{\partial x_m} \end{bmatrix}$$

7. 标量 y 对矩阵 X 的导数

类似标量 y 对列向量 **x** 的导数，把 y 对每个 X 的元素求偏导，不用转置。

$$\frac{dy}{d\mathbf{X}}=\begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

重要结论：

$$\frac{d(\mathbf{u}^T\mathbf{X}\mathbf{v})}{d\mathbf{X}}=\mathbf{u}\cdot\mathbf{v}^T$$

$$\frac{d(\mathbf{u}^T\mathbf{X}^T\mathbf{X}\mathbf{u})}{d\mathbf{X}}=2\mathbf{X}\mathbf{u}\cdot\mathbf{u}^T$$

$$\frac{d[(\mathbf{X}\mathbf{u}-\mathbf{v})^T(\mathbf{X}\mathbf{u}-\mathbf{v})]}{d\mathbf{X}}=2(\mathbf{X}\mathbf{u}-\mathbf{v})\mathbf{u}^T$$

8. 矩阵 Y 对矩阵 X 的导数

将 **Y** 的每个元素对 **X** 求导，然后排在一起形成超级矩阵。