向量对向量求导

对于向量$\boldsymbol{x} \in \mathbb{R}^{n\times 1}$和函数$\boldsymbol{y} = f(\boldsymbol{x}) \in \mathbb{R}^{m\times 1}$，则向量$\boldsymbol{y}$关于向量$\boldsymbol{x}$的导数为

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{n\times m}$$

常用示例

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{I} \qquad \frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^T \qquad \frac{\partial \boldsymbol{x}^T\boldsymbol{A}}{\partial \boldsymbol{x}} = \boldsymbol{A}$$

链式求导法则

若$x \in \mathbb{R}$，$\boldsymbol{y} = g(x) \in \mathbb{R}^{m\times 1}$，$\boldsymbol{z} = f(\boldsymbol{y}) \in \mathbb{R}^{n\times 1}$，则 $\dfrac{\partial \boldsymbol{z}}{\partial x} = \dfrac{\partial \boldsymbol{y}}{\partial x}\dfrac{\partial \boldsymbol{z}}{\partial \boldsymbol{y}} \in \mathbb{R}^{1\times n}$

若$\boldsymbol{x} \in \mathbb{R}^{m\times 1}$，$\boldsymbol{y} = g(\boldsymbol{x}) \in \mathbb{R}^{k\times 1}$，$\boldsymbol{z} = f(\boldsymbol{y}) \in \mathbb{R}^{n\times 1}$，则 $\dfrac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = \dfrac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\dfrac{\partial \boldsymbol{z}}{\partial \boldsymbol{y}} \in \mathbb{R}^{m\times n}$

若$\boldsymbol{X} \in \mathbb{R}^{m\times n}$，$\boldsymbol{y} = g(\boldsymbol{X}) \in \mathbb{R}^{k\times 1}$，$z = f(\boldsymbol{y}) \in \mathbb{R}$，则

$$\frac{\partial z}{\partial x_{ij}} = \frac{\partial \boldsymbol{y}}{\partial x_{ij}}\frac{\partial z}{\partial \boldsymbol{y}} \in \mathbb{R}$$

若$\boldsymbol{X} \in \mathbb{R}^{m\times n}, \boldsymbol{a} \in \mathbb{R}^{n}, \boldsymbol{b} \in \mathbb{R}^{m}$，$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a} + \boldsymbol{b} \in \mathbb{R}^{m}$，$z = f(\boldsymbol{y}) \in \mathbb{R}$，则

$$\frac{\partial z}{\partial \boldsymbol{X}} = \frac{\partial z}{\partial \boldsymbol{y}}\boldsymbol{a}^T \in \mathbb{R}^{m\times n} \qquad \frac{\partial z}{\partial \boldsymbol{X}^T} = \left[\frac{\partial z}{\partial \boldsymbol{y}}\boldsymbol{a}^T\right]^T = \boldsymbol{a}\left[\frac{\partial z}{\partial \boldsymbol{y}}\right]^T \in \mathbb{R}^{n\times m}$$