1. Vector derivatives

1.1 Vector - by - Scaler

$y = [y_1, y_2, \cdots, y_m]^T \in \mathbb{R}^m \qquad x \in \mathbb{R}$

$$\frac{dy}{dx} = \begin{bmatrix} \frac{dy_1}{dx} \\ \frac{dy_2}{dx} \\ \vdots \\ \frac{dy_m}{dx} \end{bmatrix} \in \mathbb{R}^m$$

1.2 Scaler - by - Vector

$y \in \mathbb{R} \qquad x = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$

$$\frac{dy}{dx} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \cdots, \frac{\partial y}{\partial x_n} \right] \in \mathbb{R}^{1\times n} \qquad D_x(y)$$

$$\nabla y = \left(\frac{dy}{dx} \right)^T \qquad \nabla_x(y)$$

Note: Numerator-layout ($\to D$) v.s. Denominator-layout ($\to \nabla$)

分子式布局 v.s. 分母式布局

Note: 结果行数与分子/母相同，二者就是转置关系

1.3 Vector - by - Vector

$y = [y_1, y_2, \cdots, y_m]^T \in \mathbb{R}^m \qquad x = [x_1, x_2, \cdots, x_n] \in \mathbb{R}^n$

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

2. Matrix derivatives

2.1 Matrix-by-Scaler

$Y = [y_{ij}(x)]_{m\times n} \in \mathbb{R}^{m\times n}$ $\quad x \in \mathbb{R}$

$$\frac{\partial Y}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix} \in \mathbb{R}^{m\times n}$$

2.2 Scaler-by-Matrix

$y = f(X) \in \mathbb{R}$ $\quad X = [x_{ij}]_{p\times q} \in \mathbb{R}^{p\times q}$

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix} \in \mathbb{R}^{p\times q} \qquad \nabla_X(y)$$

3. Common Formula （分子布局）

(a) $A \in \mathbb{R}^{m\times n}$ $\quad x \in \mathbb{R}^n$

$$\boxed{\frac{d\,Ax}{d\,x^T} = A}$$

(b). $A \in \mathbb{R}^{m\times n}$ $\quad y \in \mathbb{R}^m$ $\quad x \in \mathbb{R}^n$

$$\boxed{\frac{\partial (y^T A x)}{\partial x^T} = x^T A^T \left(\frac{\partial y}{\partial x^T}\right) + y^T A}$$

(c). $A \in \mathbb{R}^{m\times n}$ $x \in \mathbb{R}^{n}$

$$\boxed{\frac{\partial (x^T A x)}{\partial x^T} = x^T(A+A^T)}$$

(d). $A \in \mathbb{R}^{m\times n}$ $y \in \mathbb{R}^{m}$ $x \in \mathbb{R}^{n}$

，y, x 是 z 的函数，$z \in \mathbb{R}^{q}$

$$\boxed{\frac{\partial (y^T A x)}{\partial z^T} = x^T A^T \left(\frac{\partial y}{\partial z^T}\right) + y^T A \left(\frac{\partial x}{\partial z^T}\right)}$$

(e). A 可逆， $A \in \mathbb{R}^{m\times m}$ 且是 α 的函数， $\alpha \in \mathbb{R}$

$$\boxed{\frac{\partial A^{-1}}{\partial \alpha} = -A^{-1} \cdot \frac{\partial A}{\partial \alpha} \cdot A^{-1}}$$