# 矩阵微积分

- 矩阵微积分

定义1：若$f:R\rightarrow R^m$，即，$f(x)=\begin{bmatrix}f_1(x)\\ f_2(x)\\ \vdots \\ f_m(x)\end{bmatrix}$，则$\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f_1}{\partial x}\\ \frac{\partial f_2}{\partial x}\\ \vdots \\ \frac{\partial f_m}{\partial x}\end{bmatrix}$

定义2：若$f:R^n\rightarrow R$，即，$f(x)=f(x_1,x_2,\ldots,x_n)$，则$\frac{\partial f}{\partial x}=\begin{bmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \ldots & \frac{\partial f}{\partial x_n}\end{bmatrix}$

定义3：若$f:R^n\rightarrow R^m$，即，$f(x)=\begin{bmatrix}f_1(x_1,x_2,\ldots,x_n)\\ f_2(x_1,x_2,\ldots,x_n)\\ \vdots \\ f_m(x_1,x_2,\ldots,x_n)\end{bmatrix}$，则$\frac{\partial f}{\partial x}=$
$\begin{bmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \ldots & \frac{\partial f_1}{\partial x_n}\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \ldots & \frac{\partial f_2}{\partial x_n}\\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \ldots & \frac{\partial f_m}{\partial x_n}\end{bmatrix}$

一些常见的例子：
例1：若$z=Wx,W\in R^{m\times n}$，则：

$$\begin{aligned}\frac{\partial z}{\partial x}&=\begin{bmatrix}\frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \ldots & \frac{\partial z_1}{\partial x_n}\\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \ldots & \frac{\partial z_2}{\partial x_n}\\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_m}{\partial x_1} & \frac{\partial z_m}{\partial x_2} & \ldots & \frac{\partial z_m}{\partial x_n}\end{bmatrix} && \text{(根据定义3)}\\ &=W && \left(\frac{\partial z_i}{\partial x_j}=W_{ij}\right)\end{aligned}$$

例2：若$z=f(x),x\in R^n$，其中$z_i=f(x_i)$，则：

$$\frac{\partial z}{\partial x}=\begin{bmatrix}f'(x_1) & 0 & \ldots & 0\\ 0 & f'(x_2) & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & f'(x_n)\end{bmatrix} \qquad \text{(根据定义3)}$$

例3：若$z=g(y), y=f(x), x\in R^n, y\in R^p, z\in R^m$，则：

$$
\begin{aligned}
\frac{\partial z}{\partial x} &= \begin{bmatrix} \frac{\partial z_1}{\partial y}\frac{\partial y}{\partial x_1} & \frac{\partial z_1}{\partial y}\frac{\partial y}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial y}\frac{\partial y}{\partial x_n} \\ \frac{\partial z_2}{\partial y}\frac{\partial y}{\partial x_1} & \frac{\partial z_2}{\partial y}\frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_2}{\partial y}\frac{\partial y}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial z_m}{\partial y}\frac{\partial y}{\partial x_1} & \frac{\partial z_m}{\partial y}\frac{\partial y}{\partial x_2} & \cdots & \frac{\partial z_m}{\partial y}\frac{\partial y}{\partial x_n} \end{bmatrix} \qquad \text{(根据定义3)} \\
&= \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}
\end{aligned}
$$

例4：若$\frac{\partial J}{\partial z}=\delta, z=Wx, W\in R^{m\times n}, J\in R$，则：

$$
\frac{\partial J}{\partial W} = \begin{bmatrix} \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{11}} & \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{12}} & \cdots & \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{1n}} \\ \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{21}} & \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{22}} & \cdots & \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{m1}} & \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{m2}} & \cdots & \frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{mn}} \end{bmatrix}
$$

其中，

$$
\frac{\partial z_k}{\partial W_{ij}} = \begin{cases} x_j\text{，若}k=i \\ 0\text{，其它} \end{cases}
$$

于是，

$$
\frac{\partial J}{\partial z}\frac{\partial z}{\partial W_{ij}} = \delta_i x_j
$$

综上，$\frac{\partial J}{\partial W}=\delta^T x^T$

例5，人工神经网络：

$$
\begin{aligned}
&x \\
&z = Wx + b_1 \\
&h = ReLU(z) \\
&r = Uh + b_2 \\
&\hat{y} = Softmax(r) \\
&J = CE(y - \hat{y})
\end{aligned}
$$

已知$\frac{\partial J}{\partial r}=\delta_1$，则：

$$
\begin{aligned}
\frac{\partial J}{\partial U} &= \delta_1^T h^T \\
\frac{\partial J}{\partial h} &= \frac{\partial J}{\partial r}\frac{\partial r}{\partial h} = \delta_1 U \\
\frac{\partial J}{\partial b_2} &= \frac{\partial J}{\partial r}\frac{\partial r}{\partial b_2} = \delta_1 \\
\frac{\partial J}{\partial z} &= \frac{\partial J}{\partial h}\frac{\partial h}{\partial z} = \delta_1 U \circ ReLU'(z) = \delta_2 \\
\frac{\partial J}{\partial W} &= \delta_2^T x^T \\
\frac{\partial J}{\partial b_1} &= \frac{\partial J}{\partial z}\frac{\partial z}{\partial b_1} = \delta_2
\end{aligned}
\tag{根据例4}
$$