Jacobian 矩阵：

$f: \mathbb{R}^n \to \mathbb{R}^m$

$g: \mathbb{R}^m \to \mathbb{R}^l$

$g \circ f: \mathbb{R}^n \to \mathbb{R}^m \to \mathbb{R}^l$

$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \quad \begin{pmatrix} z_1 \\ \vdots \\ z_l \end{pmatrix}$

$$\frac{\partial z_i}{\partial x_j} = \sum_{k=1}^{m} \frac{\partial z_i}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_j}$$

$$\begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_l}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial x_n} & \cdots & \frac{\partial z_l}{\partial x_n} \end{pmatrix}_{n\times l} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}_{n\times m} \begin{pmatrix} \frac{\partial z_1}{\partial y_1} & \cdots & \frac{\partial z_l}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_1}{\partial y_m} & \cdots & \frac{\partial z_l}{\partial y_m} \end{pmatrix}_{m\times l}$$

$$\Rightarrow \underset{n\times l}{J(g\circ f)} = \underset{n\times m}{J(f)} \cdot \underset{m\times l}{J(g)}$$

$f = L(\phi(w)) : \mathbb{R}^p \to \mathbb{R}$

$\phi : \mathbb{R}^p \to \mathbb{R}^o$ （输出 z^L）

$L : \mathbb{R}^o \to \mathbb{R}$

$$\underset{p\times 1}{J(f)} = \underset{p\times o}{J_\phi} \cdot \underset{o\times 1}{J_L}$$

$$J_i = \frac{\partial f}{\partial w_i} = \sum_k \underbrace{\frac{\partial f}{\partial z_k^L}}_{J_L} \cdot \underbrace{\frac{\partial z_k^L}{\partial w_i}}_{J_\phi}$$

$$\underset{(p\times p)}{H_{ij}} = \frac{\partial^2 f}{\partial w_i \partial w_j} = \frac{\partial}{\partial w_j} \sum_k \frac{\partial f}{\partial z_k^L} \cdot \frac{\partial z_k^L}{\partial w_i}$$

$$= \sum_k \frac{\partial f}{\partial z_k^L} \cdot \frac{\partial^2 z_k^L}{\partial w_i \partial w_j} + \sum_k \frac{\partial}{\partial w_j}\left(\frac{\partial f}{\partial z_k^L}\right) \cdot \frac{\partial z_k^L}{\partial w_i}$$

$$= \sum_k \frac{\partial f}{\partial z_k^L} \cdot \frac{\partial^2 z_k^L}{\partial w_i \partial w_j} + \sum_k \left(\sum_m \frac{\partial^2 f}{\partial z_k^L \partial z_m^L} \cdot \frac{\partial z_m^L}{\partial w_j}\right) \cdot \frac{\partial z_k^L}{\partial w_i}$$

$$= \sum_k \frac{\partial f}{\partial z_k^L} \cdot \frac{\partial^2 z_k^L}{\partial w_i \partial w_j} + \sum_{k,m} \frac{\partial z_m^L}{\partial w_j} \cdot \frac{\partial^2 f}{\partial z_k^L \partial z_m^L} \cdot \frac{\partial z_k^L}{\partial w_i}$$

$$\approx \sum_{k,m} \frac{\partial z_m^L}{\partial w_j} \cdot \frac{\partial^2 f}{\partial z_k^L \cdot \partial z_m^L} \cdot \frac{\partial z_k^L}{\partial w_i}$$

$$\underset{p\times o}{J_\phi} = \begin{pmatrix} \frac{\partial z_1^L}{\partial w_1} & \cdots & \frac{\partial z_o^L}{\partial w_1} \\ \vdots & & \\ \frac{\partial z_1^L}{\partial w_p} & \cdots & \frac{\partial z_o^L}{\partial w_p} \end{pmatrix}_{p\times o}$$

$$\therefore H_{p\times p} = \underset{p\times o}{J_\phi} \cdot \underset{o\times o}{H_L} \cdot \underset{o\times p}{J_\phi^T}$$

$$\therefore \underset{p\times p}{H}\, \underset{p\times 1}{v} = J_\phi\left(H_L\left(J_\phi^T v\right)\right) = J_\phi\left(H_L\left(v^T J_\phi\right)^T\right)$$

定义: $R_v\{f(w)\} = \frac{\partial}{\partial r} f(w+rv)\Big|_{r=0} = v^T J_f$

证明:

$f: \mathbb{R}^p \to \mathbb{R}^o \qquad w' \to f(w') \qquad J_f \in \mathbb{R}^{p\times o}$

$g: \mathbb{R} \to \mathbb{R}^p \qquad r \to w+rv \qquad J_g \in \mathbb{R}^{1\times p}$

$f\circ g: \mathbb{R} \to \mathbb{R}^p \to \mathbb{R}^o \qquad J_{f\circ g} = J_g \cdot J_f$ ☆

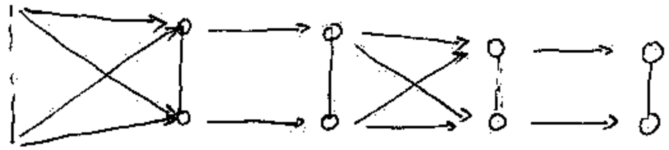
$\underset{p\times 1}{dg} = d(w+rv) = \underset{p\times 1}{v}\cdot \underset{1\times 1}{dr}$

$\therefore J_g = v^T$ 代入 ☆

$\therefore \underset{1\times o}{\frac{\partial f(w+rv)}{\partial r}} = \underset{1\times p}{v^T}\cdot \underset{p\times o}{J_f}$

设有 Net:

$$x \xrightarrow[0]{W(1)} z(1) \xrightarrow[1]{f} a(1) \xrightarrow[2]{W(2)} z(2) \xrightarrow[3]{s} a(2) \cdots y$$



正向:

$z_i^l = \sum_j w_{ij}^l a_j^{l-1} \qquad R\{z_i^l\} = \sum_j v_{ij}^l a_j^{l-1}$

$a_i^l = f_l(z_i^l) \qquad R\{a_i^l\} = f_l'(z_i^l)\cdot R\{z_i^l\}$

$z_i^L = \sum_j w_{ij}^{l+1}\cdot a_j^l \qquad R\{z_i^L\} = \sum_j w_{ij}^{l+1} R\{a_j^l\} + \sum_j v_{ij}^{l+1} a_j^l$

令 $\underset{1\times p}{v^T}\cdot \underset{p\times o}{\frac{\partial z^L}{\partial w}} = \underset{1\times p}{v^T}\underset{p\times o}{J_\phi} = R_v\{z^L\}$

③

设 $\dfrac{\partial z^L}{\partial z^l} = \delta^l \in \mathbb{R}^{n\times o}$

$$\underset{m\times 1}{a^{l-1}} \xrightarrow[n\times m]{W^l} \underset{n\times 1}{z^l} \xrightarrow{f_l} \underset{n\times 1}{a^l} \xrightarrow[o\times m]{W^{l+1}} \underset{o\times 1}{z^{l+1}} = z^L$$

$$\underset{n\times 1}{dz^l} = \underset{n\times m}{dW^l} \cdot \underset{m\times 1}{a^{l-1}}$$

$$\therefore \underset{o\times 1}{\mathrm{vec}(dz^L)} = \delta^{lT}\,\mathrm{vec}(dz^l)$$

$$= \delta^{lT}\,\mathrm{vec}(I_n \cdot \underset{n\times m}{dW^l} \cdot \underset{m\times 1}{a^{l-1}})$$

$$= \underset{o\times n}{\delta^{lT}} \cdot \underset{n\times mn}{\underset{1\times m}{a^{(l-1)T}} \otimes \underset{n\times n}{I_n}} \cdot \underset{mn\times 1}{\mathrm{vec}(dW^l)}$$

$$\therefore \boxed{\underset{mn\times o}{\frac{\partial z^L}{\partial W^l}} = \underset{mn\times n}{(\underset{m\times 1}{a^{l-1}} \otimes \underset{n\times n}{I_n})} \cdot \underset{n\times o}{\delta^l}}$$

再设已知：$\dfrac{\partial z^L}{\partial z^{l+1}} = \delta^{l+1}$ $\underline{o\times o}$

则 $\mathrm{vec}(dz^L) = \delta^{(l+1)T}\,\mathrm{vec}(dz^{l+1})$

又 $dz^{l+1} = W^{l+1}\,\mathrm{diag}(f_l'(z^l))\,dz^l$

$$\therefore \mathrm{vec}(dz^{l+1}) = \mathrm{vec}(\underline{W^{l+1}\mathrm{diag}(f_l'(z^l))} \cdot \underline{dz^l} \cdot \underline{I_1})$$

$$= W^{l+1}\,\mathrm{diag}(f_l'(z^l)) \cdot \mathrm{vec}(dz^l)$$

$$\therefore \mathrm{vec}(dz^L) = \delta^{(l+1)T}W^{l+1}\,\mathrm{diag}(f_l'(z^l)) \cdot \mathrm{vec}(dz^l)$$

$$\therefore \boxed{\delta^l = \underset{n\times n}{\mathrm{diag}(f_l'(z^l))} \cdot \underset{n\times o}{W^{(l+1)T}} \cdot \underset{o\times o}{\delta^{l+1}}}$$

$$\boxed{\delta^L = I_{o\times o}}$$

$\therefore \underset{p\times o}{J_\phi}\ \underset{o\times 1}{v}$ = $\underline{\text{反向传播}}$. 说明见下页

正向传播:

$$\underset{mn\times o}{\frac{\partial z^L}{\partial w^l}} = \begin{bmatrix} \cancel{a_1^{l-1}} \\ \vdots \\ \cancel{a_m^{l-1}} \end{bmatrix}_{mn\times n} \begin{bmatrix} \delta^l \end{bmatrix}_{n\times o} = \begin{bmatrix} a_1^{l-1}\delta^l \\ \vdots \\ a_m^{l-1}\underset{n\times o}{\delta^l} \end{bmatrix}_{mn\times o} = \begin{bmatrix} a_1^{l-1}A \\ \vdots \\ a_m^{l-1}\underset{n\times o}{A} \end{bmatrix}_{mn\times o} = \begin{bmatrix} a_1^{l-1}\boxed{A}_{n\times o} \\ \\ a_m^{l-1}\boxed{A}_{n\times o} \end{bmatrix}_{mn\times o}$$

$$\underset{n\times o}{\delta^l} = \underset{n\times n}{diag(f_l'(z^l))} \cdot \underset{n\times o}{W^{(l+1)T}}\, \delta^L_{o\times o} = \underset{n\times o}{A}\ \underset{o\times o}{\delta^L} = A_{n\times o}$$

$$\underset{mn\times o}{\frac{\partial z^L}{\partial w^l}} \cdot \underset{o\times 1}{u} = \begin{bmatrix} a_1^{l-1}A_{n\times o} \\ \\ a_m^{l-1}A_{n\times o} \end{bmatrix}_{mn\times o} \begin{bmatrix} u \end{bmatrix}_{o\times 1} = \begin{bmatrix} a_1^{l-1}\underset{n\times 1}{(A_{n\times o}u)} \\ \vdots \\ a_m^{l-1}A_{n\times o}u_{n\times 1} \end{bmatrix}_{mn\times 1}$$

$$W_{n\times m}$$

$$\begin{array}{c|cccc} & 1 & 2 & \cdots & m \\ \hline n行 & | & | & \cdots & | \end{array}$$

$$w_{ij} \to j列\ i行 \to a_j^{l-1}[A_{n\times o}u]_i \to a_j^{l-1}\sum_k A_{ik}u_k$$

相等

反向传播:

$$\underset{o\times o}{\delta^L} \cdot \underset{o\times 1}{u} = u$$

$$\delta^l \cdot u = \underset{n\times n}{diag(f_l'(z^l))}\ \underset{n\times o}{W^{(l+1)T}} \cdot u_{o\times 1} = \underset{n\times o}{A}\, u_{o\times 1}$$

$$(a^{l-1} \otimes I_n) \cdot \delta^l u = \begin{bmatrix} \cancel{a_1^{l-1}} \\ \\ \cancel{a_m^{l-1}} \end{bmatrix}_{mn\times n} \cdot \underset{n\times o}{Au} = \begin{bmatrix} a_1^{l-1}Au \\ \\ a_m^{l-1}Au \end{bmatrix}_{mn\times 1}$$

$$\Delta_z = J_\phi \left(H_L J_\phi^T z + J_L \right) + \lambda z$$

$$z = \rho z - \beta \Delta z$$

Step 1: forward, get $\underset{=a-y}{J_L}$, $\underset{=a(1-a)}{H_L}$

Step 2: CB forward, get $J_\phi^T z$

Step 3: backword, get $J_\phi \left(H_L J_\phi^T z + J_L \right) + \lambda z$

Step 4: CB forward, get $J_\phi^T \Delta z$ to compute β. ρ

$$\begin{bmatrix} -\beta \\ \rho \end{bmatrix} = - \begin{bmatrix} \Delta_z^T J_\phi H_L J_\phi^T \Delta z & z^T J_\phi H_L J_\phi^T \Delta z \\ z^T J_\phi H_L J_\phi^T \Delta z & z^T J_\phi H_L J_\phi^T z \end{bmatrix} \begin{bmatrix} J_L^T J_\phi^T \Delta z \\ J_L^T J_\phi^T z \end{bmatrix}$$