$\hat{R}$ $\quad \mathbb{R}$

$L(\cdot)$ $\quad \mathbb{R}^m \to \mathbb{R}$

$$\frac{\partial \hat{R}}{\partial \vec{y}_3} = \left(\frac{\partial \hat{R}}{\partial y_3[1]} \ \frac{\partial \hat{R}}{\partial y_3[2]} \ \cdots \ \frac{\partial \hat{R}}{\partial y_3[m]} \right) \quad (1 \times m)$$

$(m \times 1)$ $\quad \nabla_{\vec{y}_3} \hat{R}$

engineering notation

Actually $\dfrac{\partial R \text{ (scalar)}}{\partial \vec{u} \text{ (col-vec)}}$ is a row-vec (Numerator Layout)

sometimes we write col-vec for engineering requirement

$\vec{y}_3$ $\quad \mathbb{R}^m$

$h(\cdot)$

$$\frac{\partial \vec{y}_3}{\partial \vec{z}_3} = \begin{pmatrix} h'(z_3[1]) & 0 & \cdots & 0 \\ 0 & h'(z_3[2]) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h'(z_3[m]) \end{pmatrix} \quad (m \times m)$$

diagonal

$$\underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{z}_3}} = \underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{y}_3}} \ \underset{(m\times m)}{\frac{\partial \vec{y}_3}{\partial \vec{z}_3}} = \begin{pmatrix} h'(z_3[1]) \cdot \partial \hat{R}/\partial y_3[1] \\ h'(z_3[2]) \cdot \partial \hat{R}/\partial y_3[2] \\ \cdots \\ h'(z_3[m]) \cdot \partial \hat{R}/\partial y_3[m] \end{pmatrix}^{T} \quad \text{or } h'(\vec{z}_3) \odot \nabla_{\vec{y}_3} \hat{R}$$

elementwise-product

post-hoc marker because at first we wrote $\frac{\partial R}{\partial y}$ as col-vec

$\vec{z}_3$ $\quad \mathbb{R}^m$

W_3 $(m \times n)$

$$\frac{\partial \vec{z}_3}{\partial \vec{y}_2} = \begin{pmatrix} \partial z_3[1]/\partial y_2[1] & \partial z_3[1]/\partial y_2[2] & \cdots & \partial z_3[1]/\partial y_2[n] \\ \partial z_3[2]/\partial y_2[1] & \partial z_3[2]/\partial y_2[2] & \cdots & \partial z_3[2]/\partial y_2[n] \\ \vdots & \vdots & \ddots & \vdots \\ \partial z_3[m]/\partial y_2[1] & \partial z_3[m]/\partial y_2[2] & \cdots & \partial z_3[m]/\partial y_2[n] \end{pmatrix} = \begin{pmatrix} W_3[1,1] & W_3[1,2] & \cdots & W_3[1,n] \\ W_3[2,1] & W_3[2,2] & \cdots & W_3[2,n] \\ \vdots & \vdots & \ddots & \vdots \\ W_3[m,1] & W_3[m,2] & \cdots & W_3[m,n] \end{pmatrix} = W_3$$

$$\underset{(1\times n)}{\frac{\partial \hat{R}}{\partial \vec{y}_2}} = \underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{z}_3}} \cdot \underset{(m\times n)}{W_3}$$

$\dfrac{\partial \vec{z}_3}{\partial W_3}$ = shit ! it's a 3D tensor!

$$\begin{pmatrix} z_3[1] \\ z_3[2] \\ \vdots \\ z_3[m] \end{pmatrix} = \begin{pmatrix} W_3[1,1] & W_3[1,2] & \cdots & W_3[1,n] \\ W_3[2,1] & W_3[2,2] & \cdots & W_3[2,n] \\ \vdots & \vdots & \ddots & \vdots \\ W_3[m,1] & W_3[m,2] & \cdots & W_3[m,n] \end{pmatrix} \begin{pmatrix} y_2[1] \\ y_2[2] \\ \cdots \\ y_2[n] \end{pmatrix}$$

$$\frac{\partial z_3[1]}{\partial W_3[1,1]} = y_2[1] \qquad \frac{\partial z_3[1]}{\partial W_3[1,2]} = y_2[2] \qquad \cdots \qquad \frac{\partial z_3[1]}{\partial W_3[1,n]} = y_2[n]$$

but we have $\boxed{\dfrac{\partial z_3[m]}{\partial W_3[m,n]} = y_2[n], \quad \dfrac{\partial z_3[m]}{\partial W_3[p,n]} \ (p \neq m) = 0}$

$\vec{y}_2$ $\quad \mathbb{R}^n$

$g(\cdot)$

$$\frac{\partial \vec{y}_2}{\partial \vec{z}_2} = \mathrm{diag}(g'(\vec{z}_2)) \quad (n \times n)$$

$$\frac{\partial \hat{R}}{\partial \vec{z}_2} = \frac{\partial \hat{R}}{\partial \vec{y}_2} \cdot \frac{\partial \vec{y}_2}{\partial \vec{z}_2}$$

$$\frac{\partial \hat{R}}{\partial W_3} = \begin{pmatrix} \partial \hat{R}/\partial W_3[1,1] & \partial \hat{R}/\partial W_3[1,2] & \cdots & \partial \hat{R}/\partial W_3[1,n] \\ \partial \hat{R}/\partial W_3[2,1] & \partial \hat{R}/\partial W_3[2,2] & \cdots & \partial \hat{R}/\partial W_3[2,n] \\ \vdots & \vdots & \ddots & \vdots \\ \partial \hat{R}/\partial W_3[m,1] & \partial \hat{R}/\partial W_3[m,2] & \cdots & \partial \hat{R}/\partial W_3[m,n] \end{pmatrix}^{T}$$

$$= \underset{(n\times 1)}{\vec{y}_2} \cdot \underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{z}_3}}$$

(this is a $n \times m$ matrix, but $\nabla_W R$ is a $m \times n$, for substraction compatibility)

post-hoc marker

$$\frac{\partial \hat{R}}{\partial W_3[m,n]} = \frac{\partial \hat{R}}{\partial z_3[m]} \ \frac{\partial z_3[m]}{\partial W_3[m,n]} = \frac{\partial \hat{R}}{\partial z_3[m]} \cdot y_2[n]$$

$\vec{z}_2$ $\quad \mathbb{R}^n$

W_2 $(n \times p)$

$$\frac{\partial \vec{z}_2}{\partial \vec{y}_1} = W_2$$

$$\underset{(1\times p)}{\frac{\partial \hat{R}}{\partial \vec{y}_1}} = \underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{y}_3}} \cdot \underset{\substack{(m\times m) \\ \partial \vec{y}_3/\partial \vec{z}_3}}{\mathrm{diag}(h'(\vec{z}_3))} \cdot \underset{\substack{(m\times n) \\ \partial \vec{z}_3/\partial \vec{y}_2}}{W_3} \cdot \underset{\substack{(n\times n) \\ \partial \vec{y}_2/\partial \vec{z}_2}}{\mathrm{diag}(g'(\vec{z}_2))} \cdot \underset{\substack{(n\times p) \\ \partial \vec{z}_2/\partial \vec{y}_1}}{W_2}$$

$$\underset{(p\times n)}{\frac{\partial \hat{R}}{\partial W_2}} = \vec{y}_1 \cdot \frac{\partial \hat{R}}{\partial \vec{z}_2}$$

$$= \underset{(p\times 1)}{\vec{y}_1} \cdot \underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{z}_3}} \cdot \underset{(m\times n)}{W_3} \cdot \underset{(n\times n)}{\mathrm{diag}(g'(\vec{z}_2))}$$

$\vec{y}_1$ $\quad \mathbb{R}^p$

$f(\cdot)$ $\quad \mathrm{diag}(f'(\vec{z}_1))$

$\vec{z}_1$ $\quad \mathbb{R}^p \xleftarrow[(p\times q)]{W_1} \ x \quad \mathbb{R}^q$

$$\frac{\partial \vec{z}_1}{\partial x} = W_1$$

$$(1\times n) \leftarrow \frac{\partial \hat{R}}{\partial \vec{y}_L} = \underset{(1\times m)}{\frac{\partial \hat{R}}{\partial \vec{z}_{(L+1)}}} \cdot \underset{(m\times n)}{W_{(L+1)}}$$

$$(1\times n) \leftarrow \frac{\partial \hat{R}}{\partial \vec{z}_L} = \underset{(1\times n)}{\frac{\partial \hat{R}}{\partial \vec{y}_L}} \cdot \underset{(n\times n)}{\mathrm{diag}(f'(\vec{z}_L))}$$

layer $L-1$: p

layer L: n

layer $l+1$: m

$$(p\times n) \leftarrow \frac{\partial \hat{R}}{\partial W_L} = \underset{(p\times 1)}{\vec{y}_{L-1}} \cdot \underset{(1\times n)}{\frac{\partial \hat{R}}{\partial \vec{z}_L}}$$