CODE

For 1 example

For m examples

Almost $\neq$

output $\boxed{\delta^L} = (a^L - y^L)$

hidden layers $\boxed{\delta^l} = (w^{l+1}\delta^{l+1}) \odot \sigma'(z^l)$

$$\frac{\partial L}{\partial b^l} = \delta^l$$

$$\frac{\partial L}{\partial w^l} = a^{l-1}(\delta^l)^T$$

$$\delta^L = a^L - y^L$$

$$\delta^l = (\delta^{l+1} \cdot w^{l+1\,T}) \odot \sigma'(z^l)$$

$$\frac{\partial L}{\partial b^l} = \frac{1}{m} \cdot np.sum(\delta^l, axis=0)$$

$$\frac{\partial L}{\partial w^l} = a^{l-1\,T} \cdot \delta^l$$

output $\delta^L = (a - y)$

hidden layers $\delta^l = (w^{l+1}\delta^{l+1}) \odot \sigma'(z^l)$

$$\begin{cases} \dfrac{\partial L}{\partial b^l} = \delta^l \\ \dfrac{\partial L}{\partial w^l} = a^{l-1}(\delta^l)^T \end{cases}$$

$\delta^l = (\delta^{l+1} \cdot w^{l+1}) \odot \sigma'(z^l)$

$\dfrac{\partial L}{\partial b^l} = \dfrac{1}{m} np.sum(\delta^l, axis=0)$

$\dfrac{\partial L}{\partial w^l} = a^{l-1^T} \cdot \delta^l$

$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}$ $(n,1)$

$$\begin{array}{cc} & n \\ X = m & \begin{bmatrix} \text{---} & x^{(1)} & \text{---} \\ \text{---} & x^{(2)} & \text{---} \\ \text{---} & x^{(3)} & \text{---} \\ & \vdots & \\ \text{---} & x^{(m)} & \text{---} \end{bmatrix} \end{array}$$ $m \times n$

Prob

$$a_L = \begin{bmatrix} \vdots \\ a^{(i)} \\ \vdots \\ K \end{bmatrix} \Big\} \text{Argmax}$$

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{array}{l} \leftarrow \text{cat} \\ \leftarrow \text{dog} \\ \leftarrow \text{horse} \end{array}$$

$$a_L = \begin{bmatrix} — a^{(1)} — \\ — a^{(2)} — \\ — a^{(i)} — \\ \vdots \\ — a^{(m)} — \end{bmatrix} \quad m \times c$$

1

$$y = \begin{bmatrix} 0\,0\,0\,1\,0\,0\,0 \\ \text{———} \\ \text{———} \\ \text{———} \\ \text{———} \end{bmatrix} \quad m \times c$$

$$\begin{bmatrix} \vdots \\ b_3^l \end{bmatrix} \quad \begin{bmatrix} \vdots \\ \delta b_3 \end{bmatrix}$$

$$b^l = b^l - \eta \cdot \delta^l$$

$\delta_1^{l(1)}, \delta^{l(2)} \ldots$

$\delta_2^{l(1)}, \delta^{l(2)} \ldots$

$\delta_3^{l(1)}, \delta^{l(3)} \ldots$

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{array}{l} \leftarrow \text{cat} \\ \leftarrow \text{dog} \\ \leftarrow \text{horse} \end{array}$$

$a^{(i)}$

$a^{(m)}$

$m \times c$

$$y = \begin{bmatrix} 0\ 0\ 0\ 1\ 0\ 0\ 0 \\ \text{———} \\ \text{———} \\ \text{———} \\ \text{———} \end{bmatrix}_{m \times c}$$

$$b^l = \begin{bmatrix} b_1^l \\ b_2^l \\ b_3^l \end{bmatrix} \qquad \delta^l = \begin{bmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{bmatrix}$$

$$b^l = b^l - \eta \cdot \delta^l$$

$$\delta^L = m \times c$$

$$= \begin{bmatrix} \text{—} \ \delta^{l(1)} \ \text{—} \\ \text{—} \ \delta^{l(2)} \ \text{—} \\ \vdots \\ \text{—} \ \delta^{l(m)} \ \text{—} \end{bmatrix}$$

42 %

$$b^l = \begin{bmatrix} b_1^l \\ b_2^l \\ b_3^l \end{bmatrix} \qquad \delta^l = \begin{bmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{bmatrix}$$

$$b^l = b^l - \eta \cdot \delta^l$$

$$\delta^L = m \times c$$

$$= \begin{bmatrix} \text{---} \; \delta^{l(1)} \; \text{---} \\ \text{---} \; \delta^{l(2)} \; \text{---} \\ \vdots \\ \text{---} \; \delta^{l(m)} \; \text{---} \end{bmatrix}$$

b^1 $\delta_1^{l(1)}, \delta^{l(2)} \ldots$

b_2 $\delta^{l(1)}, \delta^{l(2)} \ldots$

$\delta_1^{l(1)}, \delta^{l(2)} \ldots$

$\delta_2^{l(1)}, \delta^{l(2)} \ldots$

$\delta_3^{l(1)}, \delta^{l(3)} \ldots$

$$b_1^l = b_1^l - \eta \cdot \sum_{i=1}^{m} \frac{\partial L}{\partial b_1^{l(i)}}$$

$$b^l = \begin{bmatrix} b_1^l \\ b_2^l \\ b_3^l \end{bmatrix} \qquad \delta^l = \begin{bmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{bmatrix}$$

$$b^l = b^l - \eta \cdot \delta^l$$

$$\delta^l = m \times c$$

$$= \begin{bmatrix} \text{---}\ \delta^{l(1)}\ \text{---} \\ \text{---}\ \delta^{l(2)}\ \text{---} \\ \vdots \\ \text{---}\ \delta^{l(m)}\ \text{---} \end{bmatrix}$$

$$= \Big[\ \delta_1^{l(1)} \quad \delta_2^{l(1)} \quad \delta_3^{l(3)} \ \dots$$

$$\delta_1^{l(2)}$$

$$\delta_1^{l(3)}$$

$$\delta_1^{l(4)}$$

$$\vdots$$

$$\delta_1^{l(m)}$$

b_1: $\delta_1^{l(1)}, \ \delta^{l(2)} \dots$

b_2: $\delta_2^{l(1)}, \ \delta^{l(2)} \dots$

$b^l = b - \eta \cdot \delta$

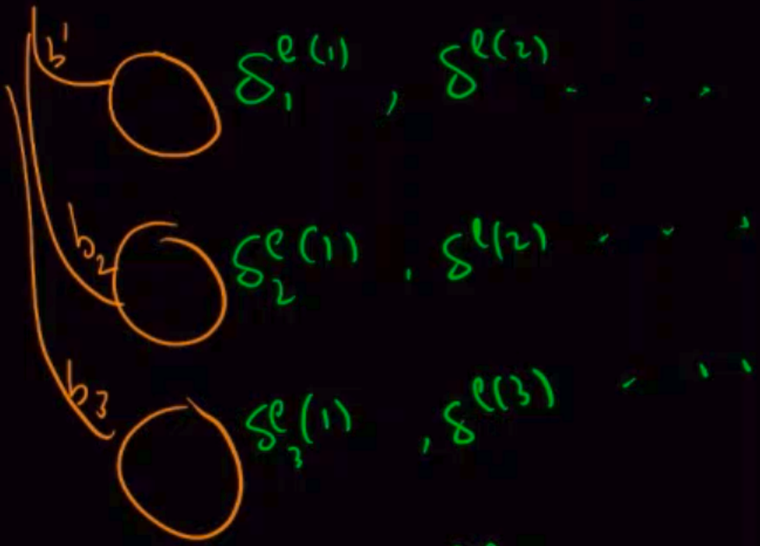
$b_1 = [\ \delta^{l(m)}\ ]$

b_2

$= \begin{bmatrix} \delta_1^{l(1)} & \delta_2^{l(1)} & \delta_3^{l(3)} & \cdots \\ \delta_2^{l(2)} & & & \end{bmatrix}$

$\begin{bmatrix} \delta_1^{l(3)} \\ \delta_1^{l(4)} \\ \vdots \\ \delta_1^{l(m)} \end{bmatrix} \cdots$

b_1



$$b_1^l = b_1^l - \eta \cdot \sum_{i=1}^{m} \frac{\partial L}{\partial b_1^{l(i)}}$$

For 1 example

output $\delta^L = \boxed{(a^L - y^L)}$

hidden layers $\delta^l = (w^{l+1}\delta^{l+1}) \odot \sigma'(z^l)$

$\boxed{\frac{\partial L}{\partial b^l} = \delta^l}$

$\frac{\partial L}{\partial w^l} = a^{l-1}(\delta^l)^T$

Almost =

For m examples

$\delta^L = \boxed{a^L - y^L}$

$\delta^l = (\delta^{l+1} \cdot w^{l+1^T}) \odot \sigma'(z^l)$

$\frac{\partial L}{\partial b^l} = \frac{1}{m} \text{np.sum}(\delta^l, \text{axis}=0)$

$\frac{\partial L}{\partial w^l} = a^{l-1^T} \cdot \delta^l$

$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}$

$X = \begin{bmatrix} \text{—— } x^{(1)} \text{ ——} \\ \text{—— } x^{(2)} \text{ ——} \\ \text{—— } x^{(3)} \text{ ——} \\ \vdots \end{bmatrix}$ m, n output → $\delta^L = \boxed{(a^L - y^L)}$

hidden layers → $\delta^l = (W^{l+1}\delta^{l+1}) \odot \sigma'(z^l)$

$$\boxed{\frac{\partial L}{\partial b^l} = \delta^l}$$

$$\frac{\partial L}{\partial W^l} = a^{l-1}(\delta^l)^T$$

h_{l-1}

$$[\;] = [\;][\;]$$

h_{l-1}, h_l $\quad$ $h_{l-1}\times 1$ $\quad$ $1 \times h_l$

Almost =

$$\delta^L = \boxed{a^L - y^L}$$

$$\delta^l = (\delta^{l+1} \cdot W^{l+1\,T}) \odot \sigma'(z^l)$$

$$\frac{\partial L}{\partial b^l} = \frac{1}{m}\text{np.sum}(\delta^l, \text{axis}=0)$$

$$\frac{\partial L}{\partial W^l} = a^{l-1\,T} \cdot \delta^l$$

n

$$X = \begin{bmatrix} — x^{(1)} — \\ — x^{(2)} — \\ — x^{(3)} — \\ \vdots \\ — x^{(m)} — \end{bmatrix}_{m \times n}$$

m

Prob

output → $\delta^L = \boxed{(a^L - y^L)}$

hidden layers → $\delta^l = (W^{l+1}\delta^{l+1}) \odot \sigma'(z^l)$

$$\boxed{\frac{\partial L}{\partial b^l} = \delta^l}$$

$$\frac{\partial L}{\partial W^l} = a^{l-1} (\delta^l)^T$$

$W = (h_{l-1}, h_l)$

h_{l-1}

h_{l-1} h_l

$[\ ] = [\ ][\ ]$

h_{l-1}, h_l $\quad h_{l-1} \times 1 \quad 1 \times h_l$

Almost =

$$\delta^L = \boxed{a^L - y^L}$$

$$\delta^l = (\delta^{l+1} \cdot W^{l+1\,T}) \odot \sigma'(z^l)$$

$$\frac{\partial L}{\partial b^l} = \frac{1}{m} np.sum(\delta^l, axis=0)$$

$$\frac{\partial L}{\partial W^l} = a^{l-1\,T} \cdot \boxed{\delta^l}$$

Do make dimensions

$$\frac{\partial L}{\partial W^l} = [\ ]\ [\ ]$$

$h^{l-1} \times m \qquad m \times h^l$

$= (h^{l-1}, h^l)$

Prob