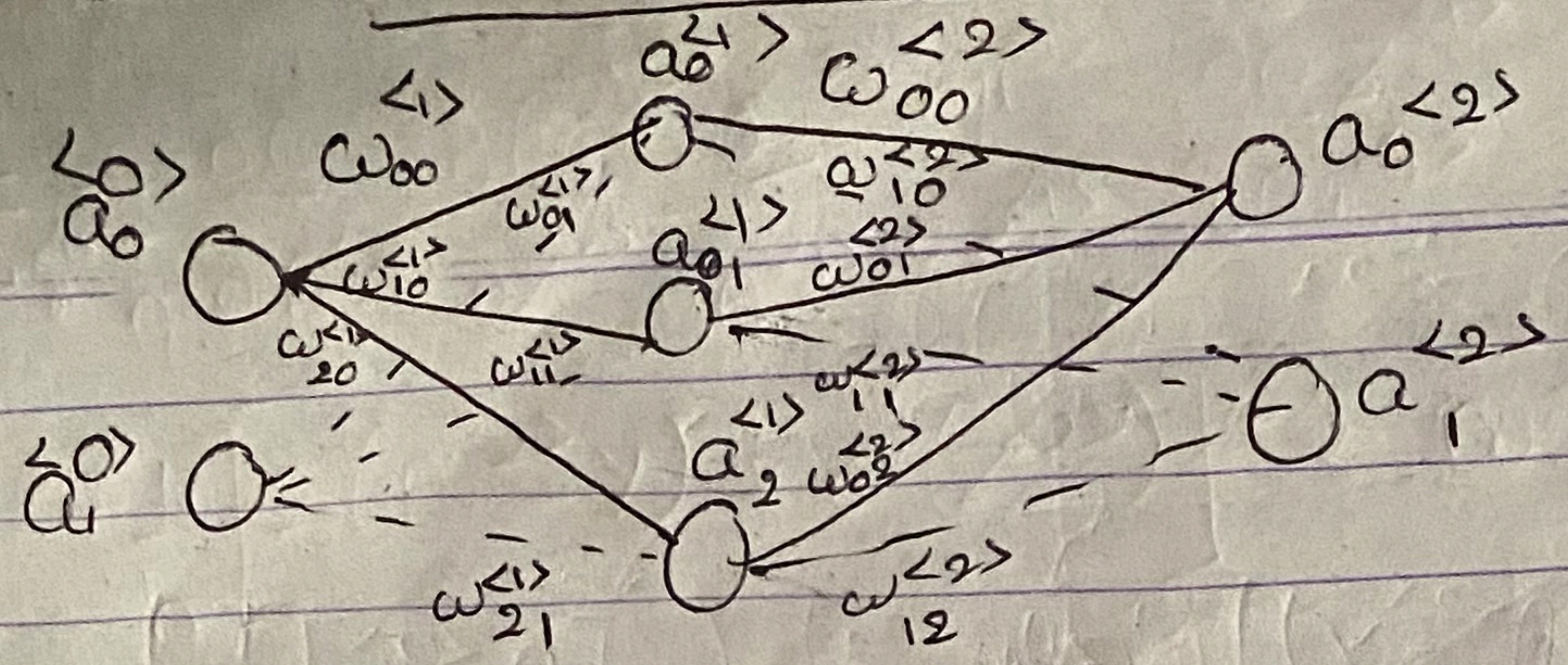
2-3-2 Neural network

$2 = m \rightarrow$ no. of input (features)
$n \rightarrow$ no. of data



Here; $w^{<1>} = \begin{bmatrix} w_{00}^{<1>} & w_{01}^{<1>} \\ w_{10}^{<1>} & w_{11}^{<1>} \\ w_{20}^{<1>} & w_{21}^{<1>} \end{bmatrix}_{3\times2}$, $w^{<2>} = \begin{bmatrix} w_{00}^{<2>} & w_{01}^{<2>} & w_{02}^{<2>} \\ w_{10}^{<2>} & w_{11}^{<2>} & w_{12}^{<2>} \end{bmatrix}_{2\times3}$

Now,

$C = (A^{<2>} - y)^2$

$Z^{<2>} = w^{<2>} A^{<1>} + b^{<2>}$

$A^{<2>} = \sigma(Z^{<2>})$

$Z^{<1>} = w^{<1>} A^{<0>} + b^{<1>}$

$A^{<1>} = \sigma Z^{<1>}$

Matrics	Shape
$A^{<0>}$	$2 \times n$
$A^{<1>}; Z^{<1>}$	$3 \times n$
$A^{<2>}; Z^{<2>}$	$2 \times n$
$B^{<1>}$	3×1
$B^{<2>}$	2×1

Now; (for one set of inputs)

$$\frac{\partial C}{\partial w^{<2>}} = \begin{bmatrix} \partial C/\partial w_{00}^{<2>} & \partial C/\partial w_{01}^{<2>} & \partial C/\partial w_{02}^{<2>} \\ \partial C/\partial w_{10}^{<2>} & \partial C/\partial w_{11}^{<2>} & \partial C/\partial w_{12}^{<2>} \end{bmatrix}_{3\times2} = Jw_2$$

This gives, how $w^{<2>}$ shall change to maximize C
(avg of n inputs)

Note; it is same shape as weight

Similarly $\frac{\partial C}{\partial b^{<2>}} = \begin{bmatrix} \partial C/\partial b_0^{<2>} \\ \partial C/\partial b_1^{<2>} \end{bmatrix}_{2\times1}$

---

also; for m terms

$$\frac{\partial C}{\partial w^{<2>}} = \frac{1}{m} \frac{\partial C}{\partial A^{<2>}} \frac{\partial A^{<2>}}{\partial Z^{<2>}} \frac{\partial Z^{<2>}}{\partial w^{<2>}}$$

(in python code) $= \frac{1}{m}(2(A^{<2>} - y) * \sigma'(Z^{<2>}))$ @ $(A^{<2>})^T$

How?? $\longrightarrow$ ($\omega_{00}^{<2>}$ change changes $z_0^{<2>}$; changes $a_0^{<2>}$ changes C)

$$\rightarrow \frac{\partial C}{\partial \omega_{00}^{2}} = \frac{\partial z_0^{<2>}}{\partial \omega_{00}^{<2>}} \cdot \frac{\partial a_0^{<2>}}{\partial z_0^{<2>}} \cdot \frac{\partial C}{\partial a_0^{<2>}} = a_0^{1} \cdot \sigma'(z_0^{<2>})\, 2(a_0^{<2>} - y)$$

$$\rightarrow \frac{\partial C}{\partial \omega_{01}^{<2>}} = a_1^{<1>} \sigma'(z_0^{<2>})\, 2(a_0^{<2>} - y_0) \Rightarrow \frac{\partial C}{\partial \omega_{02}^{<2>}} = a_2^{<1>} \sigma'(z_0^{<2>})\, 2(a_0^{<2>} - y)$$

$$\frac{\partial C}{\partial \omega_{10}^{<2>}} = \frac{\partial z_1^{<2>}}{\partial \omega_{10}^{<2>}} \cdot \frac{\partial a_1^{<2>}}{\partial z_1^{<2>}} \cdot \frac{\partial C}{\partial a_1^{<2>}} = a_0^{<1>} \sigma'(z_1^{<2>})\, 2(a_1^{<2>} - y_1)$$

$\rightarrow \frac{\partial C}{\partial \omega_{11}^{<2>}}$ & $\frac{\partial C}{\partial \omega_{12}^{<2>}}$ follows above:

Using these six values in $\frac{\partial C}{\partial \omega^{<2>}}$ matrix

$$\frac{\partial C}{\partial \omega^{<2>}} = \begin{bmatrix} a_0^{<1>}\sigma'(z_0^{<2>}).2(a_0^{<2>} - y_0) & a_1^{<1>}\sigma'(z_0^{<2>})2(a_0^{<2>} - y_0) & a_2^{<1>}\sigma'(z_0^{<2>})2(a_0^{<2>} - y_0) \\ a_0^{<1>}\sigma'(z_1^{<2>}).2(a_1^{<2>} - y_1) & a_1^{<1>}\sigma'(z_1^{<2>})2(a_1^{<2>} - y_1) & a_2^{<1>}\sigma'(z_1^{<2>})2(a_1^{<2>} - y_1) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma'(z_0^{<2>}) \\ \sigma'(z_1^{<2>}) \end{bmatrix} * \begin{bmatrix} 2(a_0^{<2>} - y_0) \\ 2(a_1^{<2>} - y_1) \end{bmatrix} @ \begin{bmatrix} a_0^{<1>} & a_1^{<1>} & a_2^{<1>} \end{bmatrix}$$

$$= \sigma'(Z^{<2>}) * (2(A^{<2>} - Y)) @ [A^{<2>}]^T$$

if n inputs:

$$\frac{1}{m}\left(\underbrace{\begin{bmatrix} \sigma'(z_0^{<2>}) \ldots \\ \sigma'(z_1^{<2>}) \ldots \end{bmatrix}}_{2\times n} * \underbrace{\begin{bmatrix} 2(a_0^{<2>} - y_0) & \ldots \\ 2(a_1^{<2>} - y_1) & \ldots \end{bmatrix}}_{2\times n} \right)_{2\times n} @ \underbrace{\begin{bmatrix} a_0^{<1>} & a_1^{<1>} & a_2^{<1>} \\ \vdots & \vdots & \vdots \end{bmatrix}}_{n\times 3}$$

$$\Downarrow$$

$$2\times 3$$

$$\text{Avg } \frac{\partial C}{\partial \omega^{<2>}} = \frac{1}{n}\left(2(A^{<2>} - Y) * \sigma'(Z^{<2>})\right) @ \left((A^{<2>})\right)^T$$

$$\text{Avg } \frac{\partial C}{\partial b^{<2>}} = \frac{1}{n}\left(\underbrace{2(A^{<2>} - Y)}_{2\times n} * \underbrace{\sigma'(Z^{<2>})}_{2\times n}\right)$$

But we need 2×1;

$$\frac{\partial C}{\partial b^{<2>}} = \frac{1}{n} \text{ Sum across Row } (2(A^{<2>} - Y * \sigma'(Z^{<2>}))$$

$$(0+0+0)$$

also,

$$\frac{\partial C}{\partial \omega^{<1>}} = \frac{1}{n}\frac{\partial C}{\partial A^{<2>}} \frac{\partial A^{<2>}}{\partial Z^{<2>}} \frac{\partial Z^{<2>}}{\partial A^{<1>}} \frac{\partial A^{<1>}}{\partial Z^{<1>}} \frac{\partial Z^{<1>}}{\partial \omega^{<1>}}$$

$$= \frac{1}{n}\left[2(A^{<2>} - Y) * \sigma'(Z^{<2>})\right]_{2\times n} \times \omega^{<2>}_{2\times 3} * \sigma'(Z^{<1>})_{3\times n} \, (A^{<0>})_{2\times n}$$

So

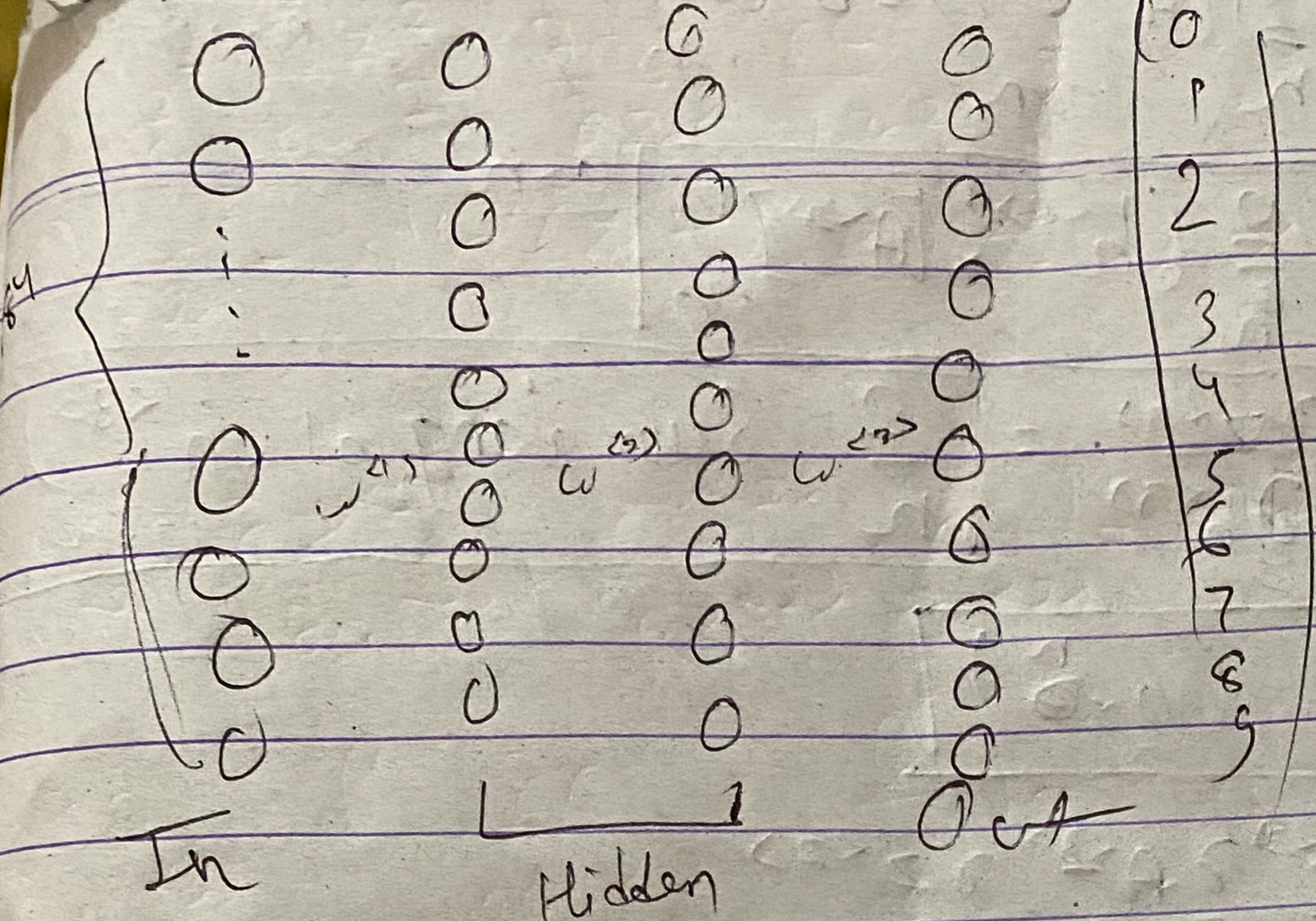
$$= \frac{1}{n}\left[\left[2(A^{<2>} - Y) * \sigma'(Z^{<2>})\right]^T_{n\times 2} @ \, \omega^{<2>}_{2\times 3}\right]^T_{3\times n} * \sigma'(Z^{<1>})_{3\times n} \, @ \left[A^{<0>}\right]^T_{n\times 2}$$

$$\frac{\partial C}{\partial b^{<1>}} = \text{Sum of cols}\left(\frac{1}{n}\left[\underbrace{\left[2(A^{<2>} - Y) * \sigma'(Z^{<2>})\right]^T @ \, \omega^{<2>}}_{3\times 2 \text{ (matrix)}}\right]^T * \sigma'(Z^{<1>})\right)$$

and

$$\omega^{<1>}_{i+1} = \omega^{<1>}_{i} - \alpha \, \partial C/\partial \omega^{<1>}$$
$$\omega^{<2>}_{i+1} = \omega^{<2>}_{i} - \alpha \, \partial C/\partial \omega^{<2>}$$
$$b^{<1>}_{i+1} = b^{<1>}_{i} - \alpha \, \partial C/\partial b^{<1>}$$
$$b^{<2>}_{i+1} = b^{<2>}_{i} - \alpha \, \partial C/\partial b^{<2>}$$

Our Network



$$\sigma = \frac{1}{1+e^{-x}}$$

$$\sigma' = \frac{e^{-x}}{(1+e^{-x})^2} = \sigma(x)(1-\sigma(x))$$

$$\boxed{C = (A^{<3>} - Y)^2}$$

one data set

Input: $X = \begin{bmatrix} \cdots \\ \cdots \\ \cdots \end{bmatrix} \Big\} 784$

n

$784 \times n$

$W^{<1>} = (10 \times 784)$ $\quad W^{<2>} = 10 \times 10$ $\quad W^{<3>} = 10 \times 10$

Forward

$\rightarrow A^{<0>} = X \ (784 \times n)$

$\rightarrow \underset{10\times n}{Z^{<1>}} = \underset{10\times 784}{W^{<1>}} \underset{784\times n}{A^{<0>}} + \underset{10\times n}{b^{<1>}}$

$A^{<1>} = \sigma(Z^{<1>})$

$\rightarrow \underset{10\times n}{Z^{<2>}} = \underset{10\times 10}{W^{<2>}} \underset{10\times n}{A^{<1>}} + \underset{10\times n}{b^{<2>}}$

$A^{<2>} = \sigma(Z^{<2>})$

$\rightarrow \underset{10\times n}{Z^{<3>}} = \underset{10\times 10}{W^{<3>}} \underset{10\times n}{A^{<2>}} + \underset{10\times 16}{b^{<3>}}$

$A^{<3>} = \sigma(Z^{<3>})$

Backward

$$\underbrace{\frac{\partial C}{\partial \omega^{<3>}}}_{10\times 10} = \frac{1}{n}\frac{\partial C}{\partial A^{<3>}}\frac{\partial A^{<3>}}{\partial Z^{<3>}}\frac{\partial Z^{<3>}}{\partial \omega^{<3>}} = \frac{2}{n}\Big(\underbrace{(A^{<3>}-Y)}_{10\times n}\circ \underbrace{\sigma'(Z^{<3>})}_{10\times n}\Big)\underbrace{[A^{<2>}]^T}_{n\times 10}$$

$$\boxed{= \frac{2}{n}\Big(h^{<3>}\times [A^{<2>}]^T\Big)_{10\times 10}}$$

$$\underbrace{\frac{\partial C}{\partial b^{<3>}}}_{10\times 1} = \frac{1}{n}\frac{\partial C}{\partial A^{<3>}}\frac{\partial A^{<3>}}{\partial Z^{<3>}}\frac{\partial Z^{<3>}}{\partial b^{<3>}} = \boxed{\frac{2}{n}\text{ Sum along col }(\underbrace{h^{<3>}}_{10\times n})}$$

$$\boxed{h^{<3>} = (A^{<3>}-Y)\circ\sigma'(Z^{<3>}) \quad (10\times n)}$$

$$\frac{\partial C}{\partial \omega^{<2>}} = \frac{1}{n}\frac{\partial C}{\partial A^{<3>}}\frac{\partial A^{<3>}}{\partial Z^{<3>}}\frac{\partial Z^{<3>}}{\partial A^{<2>}}\frac{\partial A^{<2>}}{\partial Z^{<2>}}\frac{\partial Z^{<2>}}{\partial \omega^{<2>}}$$

$$= \frac{2}{n}\underbrace{\Big[(h^{<3>})^T_{n\times 10}\times \omega^{<3>}_{10\times 10}\Big]^T}_{10\times n}\circ \underbrace{\sigma'(Z^{<2>})}_{10\times n}\times \underbrace{[A^{<1>}]^T}_{n\times 10}$$

let $\boxed{h^{<2>}_{10\times n} = [(h^{<3>})^T\times \omega^{<3>}]^T\circ\sigma'(Z^{<2>}) \quad (10\times n)}$

also,

$$\underset{10\times 1}{\frac{\partial C}{\partial b^{<2>}}} = \boxed{\frac{2}{n}\text{ Sum along col }(h^{<2>})}$$

$$\underset{10\times 10}{\frac{\partial C}{\partial \omega^{<2>}}} = \boxed{\frac{2}{n}h^{<2>}\times[A^{<1>}]^T}$$

Finally,

$$\underset{10\times 784}{\frac{\partial C}{\partial \omega^{<1>}}} = \frac{1}{n}\frac{\partial C}{\partial A^{<3>}}\frac{\partial A^{<3>}}{\partial Z^{<3>}}\frac{\partial Z^{<3>}}{\partial A^{<2>}}\frac{\partial A^{<2>}}{\partial Z^{<2>}}\frac{\partial Z^{<2>}}{\partial A^{<1>}}\frac{\partial A^{<1>}}{\partial Z^{<1>}}\frac{\partial Z^{<1>}}{\partial \omega^{<1>}}$$

$$= \frac{2}{n}\Big(\underbrace{(h^{<2>})^T}_{n\times 10}\times \underbrace{\omega^{<2>}}_{10\times 10}\Big)^T_{10\times n}\circ \underbrace{\sigma'(Z^{<1>})}_{10\times n}\cdot\underbrace{[A^{<0>}]^T}_{n\times 784}$$

let $\boxed{h^{<1>} = [(h^{<2>})^T\times\omega^{<2>}]^T\circ\sigma'(Z^{<1>})}\ (10\times n)$

So,

$$\boxed{\frac{\partial C}{\partial \omega^{<1>}} = \frac{2}{n}\underbrace{h^{<1>}\times[A^{<0>}]^T}_{10\times 784}}$$

&

$$\boxed{\underset{10\times 1}{\frac{\partial C}{\partial b^{<1>}}} = \frac{2}{n}\text{ Sum along col }(h^{<1>})}$$