$\boxed{784}$ one training image (784 pixels)

28×28

{There are m training images}

$$X = \begin{bmatrix} ---- & x^{(1)} & ---- \\ ---- & x^{(2)} & ---- \\ & \vdots & \\ ---- & x^{(m)} & ---- \end{bmatrix}^T$$

FORWARD PROPOGATION:

$$A^{(0)} = X \;(784 \times m)$$

$$Z^{(1)} = W^{(1)} A^{(0)} + b^{(1)}$$

$$A^{(1)} = g(Z^{(1)}) \Rightarrow ReLU(Z^{(1)})$$

$$Z^{(2)} = W^{(2)} A^{(1)} + b^{(2)}$$

$$A^{(2)} = g(Z^{(2)}) \Rightarrow Softmax(Z^{(2)})$$

{Cost funtion or lost = $(A^{(2)} - y)^2$}

$$\frac{\delta L_0}{\delta w^1} = \frac{\delta z^1}{\delta w^1} \cdot \frac{\delta A^1}{\delta z^1} \cdot \frac{\delta z^2}{\delta A^1} \cdot \frac{\delta A^2}{\delta z^2} \cdot \frac{\delta L_0}{\delta z^2}$$

$$= \sum_{i=1}^{m} f'(z1) \cdot A_i^{(0)} \cdot w_i^{(2)} \cdot 2(A_i^{(2)} - y) \cdot \frac{1}{m} \quad \text{--- (3)}$$

$$\frac{\delta L_0}{\delta b^1} = \frac{\delta z^1}{\delta b^1} \cdot \frac{\delta A^1}{\delta z^1} \cdot \frac{\delta z^2}{\delta A^1} \cdot \frac{\delta A^2}{\delta z^2} \cdot \frac{\delta L_0}{\delta z^2}$$

$$= \sum_{i=1}^{m} f'(z1) \cdot w_i^{(2)} \cdot 2(A_i^{(2)} - y) \cdot \frac{1}{m} \quad \text{--- (4)}$$

$dz2 = 2(A^{(2)} - y)/m$ — (A)

$dw2 = dz2 \cdot A^{(1)}$ — (B)

$db2 = \sum_{i=1}^{m} (dz2)$ — (C)

$dz1 = w2 \cdot dz2 \cdot f'(z1)$ — (D)

$dw1 = dz1 \cdot X$ — (E)

$db1 = \sum_{i=1}^{m} (dz1)$ — (F)

Activation Functions

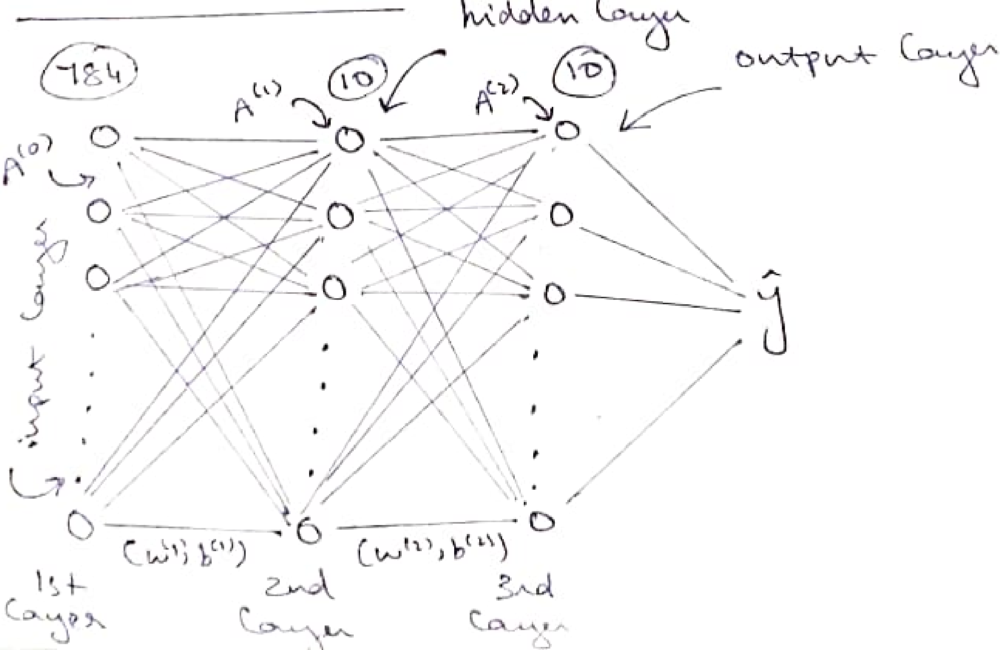
1. **ReLU:**

   $f(x) = \max(0, x)$

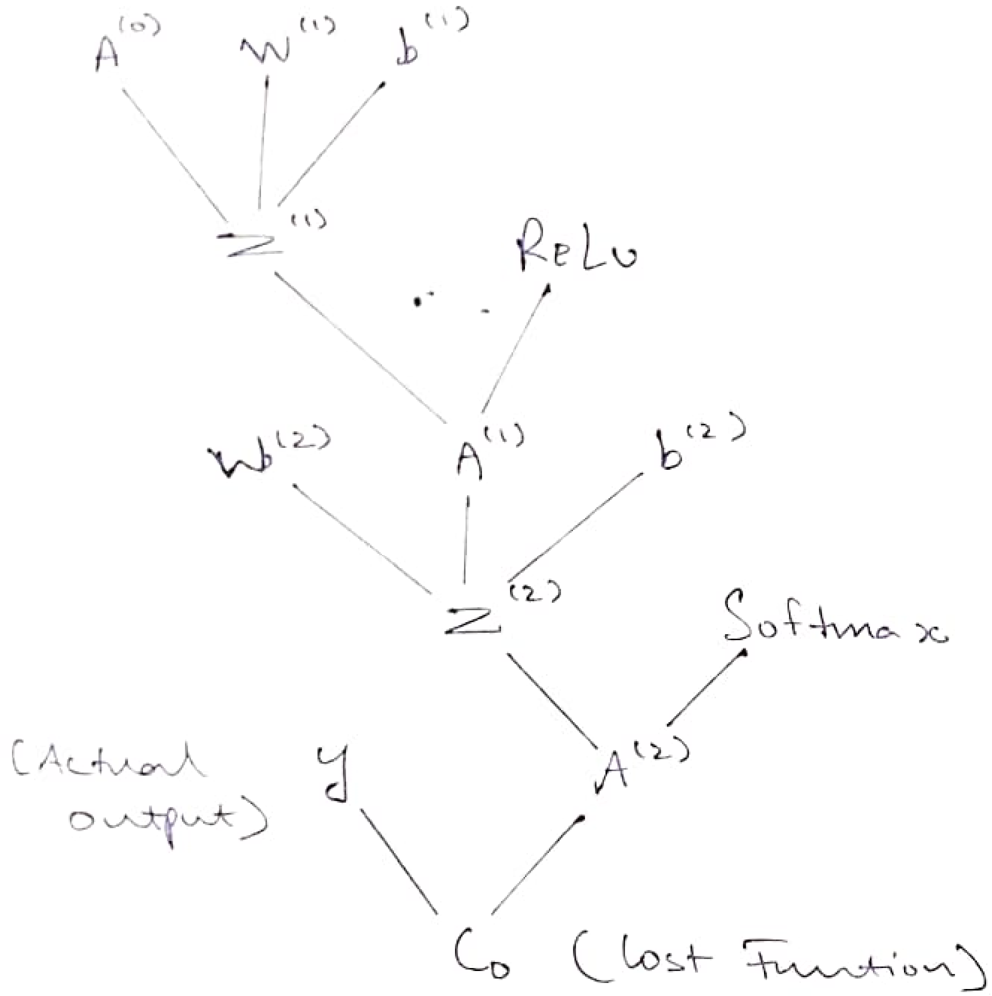
2. **Softmax:**

   $$\frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Neural Network:



Backward Propogation:



$$\frac{\delta C_0}{\delta w^2} = \frac{\delta z^2}{\delta w^2} \cdot \frac{\delta A^2}{\delta z^2} \cdot \frac{\delta C_0}{\delta A^2}$$

$$= A^{(1)} \cdot 2(A^{(2)} - y)$$

$$\Rightarrow \sum_{i=1}^{m} \frac{1}{m} \left[A_i^{(1)} \cdot 2(A_i^{(2)} - y) \right] \quad \text{——} \quad (1)$$

$$\frac{\delta C_0}{\delta b^2} = \frac{\delta z^2}{\delta b^2} \cdot \frac{\delta A^2}{\delta z^2} \cdot \frac{\delta C_0}{\delta A^2}$$

$$= 2(A^{(2)} - y) \Rightarrow \sum_{1}^{m} \frac{1}{m} \left[2(A_i^{(2)} - y) \right] \quad \text{——} \quad (2)$$