Muhammad Yasir

191236

AI LAB Project:-

∴ Computing Number of Neurons in Hidden layer.

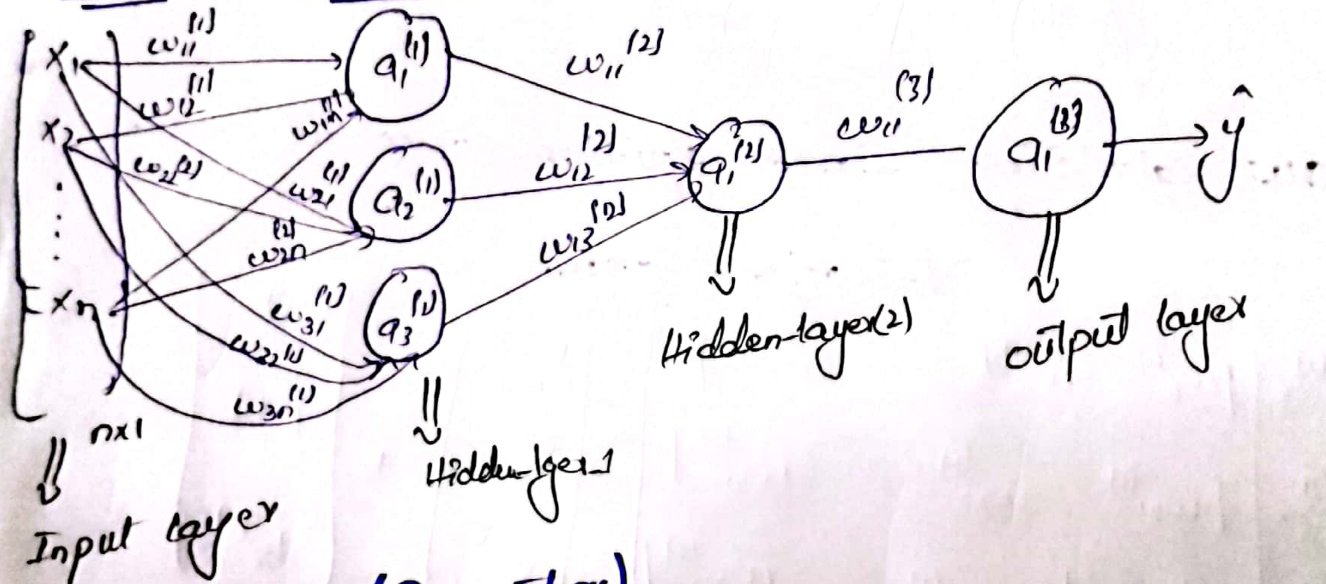
Neurons = (191205, 191216, 191236, 191235, 191869) / 5

= 155/5 = 31

Hidden layer 1 = 3 Neurons:-

" " 2 = 1 "

Network Architecture:-



Calculating no:(Parameters)

Input images = (28×28, 1)

$n = (784, 1)$

1st Hidden layer:-

$= 3n + 3$

$= 3(784) + 3 \Rightarrow (2,355)$

2nd Hidden layer:

$= 1n + 1$

$= 1(3) + 1 = 4$

Output layer:-

$$1n+1 = 1(1)+1 = 2$$

Total para = 9,361

∴ Hidden-layer-1:

- $Z_1^{[1]} = W_1^{[1]T} \cdot X + b_1^{[1]}$

  $a_1^{[1]} = \delta(Z_1^{[1]})$

- $Z_2^{[1]} = W_2^{[1]T} \cdot X + b_2^{[1]}$

  $a_2^{[1]} = \delta(Z_2^{[1]})$

- $Z_3^{[1]} = W_3^{[1]T} \cdot X + b_3^{[1]}$

  $a_3^{[1]} = \delta(Z_3^{[1]})$

Now calculating single equation for whole layer.

Vectorized form.

$$\begin{array}{l} Z_1^{[1]} = \\ Z_2^{[1]} = \\ Z_3^{[1]} = \end{array} \underbrace{\underbrace{\begin{bmatrix} [w_{11}^{[1]} & w_{12}^{[1]} & \cdots & w_{1n}^{[1]}] \\ [w_{21}^{[1]} & w_{22}^{[1]} & \cdots & w_{2n}^{[1]}] \\ [w_{31}^{[1]} & w_{32}^{[1]} & \cdots & w_{3n}^{[1]}] \end{bmatrix}_{3\times n} \cdot \begin{bmatrix} X_{n\times 1} \\ X_{n\times 1} \\ X_{n\times 1} \end{bmatrix}_{n\times 1}}_{3\times 1} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \end{bmatrix}_{3\times 1}}_{3\times 1}$$

$$a_{3\times 1}^{[1]} = \delta(Z_{3\times 1}^{[1]})$$

∴ **For all layers.**

Forward Propogation:-

Calculating Sigmoid:-

i) $\underset{3\times1}{z^{[1]}} = \underbrace{\underbrace{\underset{3\times n}{W^{[1]}} \cdot \underset{n\times1}{X}}_{3\times1} + \underset{3\times1}{b^{[1]}}}_{3\times1}$

$\underset{3\times1}{a^{[1]}} = \sigma(\underset{3\times1}{z^{[1]}})$

ii) $\underset{1\times1}{Z^{[2]}} = \underbrace{\underbrace{\underset{1\times3}{W^{[2]}} \cdot \underset{3\times1}{a^{[1]}}}_{1\times1} + \underset{1\times1}{b^{[2]}}}_{1\times1}$

$\underset{1\times1}{a^{[2]}} = \sigma(\underset{1\times1}{z^{[2]}})$

iii) $\underset{1\times1}{Z^{[3]}} = \underbrace{\underbrace{\underset{1\times1}{W^{[3]}} \cdot \underset{1\times1}{a^{[2]}}}_{1\times1} + \underset{1\times1}{b^{[3]}}}_{1\times1}$

$\hat{y} = \underset{1\times1}{a^{[3]}} = \sigma(\underset{1\times1}{z^{[3]}})$

Likelihood:-

$$\mathcal{L} = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})]$$

Cost function:-

$$J(y,\hat{y}) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}^{(i)}$$

Learning Parameters:-

$w^{[1]}, w^{[2]}, w^{[3]}, b^{[1]}, b^{[2]}, b^{[3]}$

$$\frac{\partial J}{\partial w^{(l)}} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial w^{(l)}} \quad ; \; l = 1, 2, 3$$

$$\frac{\partial J}{\partial b^{(l)}} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial w^{(l)}}$$

~~Bo~~ Back Propogation

Using chain Rule:-

$$\frac{\partial L}{\partial w^{(3)}} = \frac{\partial L}{\partial a^{(3)}} \cdot \frac{\partial (a^{(3)})}{\partial (z^{3})} \cdot \frac{\partial (z^{(3)})}{\partial (w^{(3)})} \Longrightarrow (1)$$

$$\frac{\partial L}{\partial w^{(2)}} = \frac{\partial L}{\partial a^{(3)}} \cdot \frac{\partial (a^{(3)})}{\partial (z^{(3)})} \cdot \frac{\partial (z^{(3)})}{\partial a^{(2)}} \cdot \frac{\partial (a^{(2)})}{\partial (z^{(2)})} \cdot \frac{\partial (z^{(2)})}{\partial (w^{(2)})} \Longrightarrow (2)$$

$$\frac{\partial L}{\partial w^{(1)}} = \frac{\partial L}{\partial a^{(3)}} \cdot \frac{\partial (a^{(3)})}{\partial (z^{(3)})} \cdot \frac{\partial (z^{(3)})}{\partial (a^{(2)})} \cdot \frac{\partial (a^{(2)})}{\partial (z^{(2)})} \cdot \frac{\partial (z^{(2)})}{\partial (a^{(1)})} \cdot \frac{\partial (a^{(1)})}{\partial (z^{(1)})} \cdot \frac{\partial (z^{(1)})}{\partial (w^{(1)})}$$

$$\Downarrow$$

(3)

$$\frac{\partial \mathcal{L}}{\partial(b^{(3)})} = \frac{\partial \mathcal{L}}{\partial(a^{(3)})} \cdot \frac{\partial(a^{(3)})}{\partial(z^{(3)})} \cdot \frac{\partial(z^{(3)})}{\partial(b^{(3)})} \rightarrow (4)$$

$$\frac{\partial \mathcal{L}}{\partial(b^{(2)})} = \frac{\partial \mathcal{L}}{\partial(a^{(3)})} \cdot \frac{\partial(a^{(3)})}{\partial(z^{(3)})} \cdot \frac{\partial(z^{(3)})}{\partial(a^{(2)})} \cdot \frac{\partial(a^{(2)})}{\partial(z^{(2)})} \cdot \frac{\partial(z^{(2)})}{\partial(b^{(2)})} \rightarrow (5)$$

$$\frac{\partial \mathcal{L}}{\partial(b^{(1)})} = \frac{\partial \mathcal{L}}{\partial(a^{(3)})} \cdot \frac{\partial(a^{(3)})}{\partial(z^{(3)})} \cdot \frac{\partial(z^{(3)})}{\partial(a^{(2)})} \cdot \frac{\partial(a^{(2)})}{\partial(z^{(2)})} \cdot \frac{\partial(z^{(2)})}{\partial(a^{(1)})} \cdot \frac{\partial(a^{(1)})}{\partial(z^{(1)})} \cdot \frac{\partial(z^{(1)})}{\partial(b^{(1)})} \Downarrow (6)$$

Solving (1)

- $\frac{\partial \mathcal{L}}{\partial a^{(3)}} = \frac{a^{(3)} - y}{a^{(3)}(1 - a^{(3)})}$
- $\frac{\partial(a^{(3)})}{\partial(z^{(3)})} = a^{(3)}(1 - a^{(3)})$
- $\frac{\partial(z^{(3)})}{\partial(w^{(3)})} = a^{(2).T}$ post

so;

$$\frac{\partial \mathcal{L}}{\partial w^{(3)}} = \frac{a^{(3)} - y}{\cancel{a^{(3)}(1 - a^{(3)})}} \cdot \cancel{a^{(3)}(1 - a^{(3)})} \cdot a^{(2).T}$$

$$\underset{1\times 1}{\frac{\partial \mathcal{L}}{\partial w^{(3)}}} = \underbrace{\underbrace{(\underset{(1\times1)}{a^{(3)}} - \underset{(1\times1)}{y})}_{(1\times1)} \cdot \underset{(1\times1)}{a^{(2).T}}}_{1\times1}$$

$$w^{(1)} = w^{(1)} - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}(a^{(3)} - y) \cdot a^{(2).T}\right]$$

Solving (2)

- $\frac{\partial \mathcal{L}}{\partial a^{(3)}} = \frac{a^{(3)} - y}{a^{(3)}(1 - a^{(3)})}$
- $\frac{\partial(z^{(3)})}{\partial(a^{(2)})} = w^{(3).T}$ pre
- $\frac{\partial a^{(3)}}{\partial(z^{(3)})} = a^{(3)}(1 - a^{(3)})$
- $\frac{\partial(a^{(2)})}{\partial(z^{(2)})} = a^{(2)}(1 - a^{(2)})$
- $\frac{\partial(z^{(2)})}{\partial(w^{(2)})} = a^{(1).T}$ post

$$\frac{\partial \mathcal{L}}{\partial w^{(2)}} = w^{(3).T}\,\frac{a^{(3)}-y}{\cancel{a^{(3)}(1-a^{(3)})}} \cdot \cancel{a^{(3)}(1-a^{(3)})} \cdot a^{(2)}(1-a^{(2)}) . \; a^{(1).T}$$

$$\frac{\partial \mathcal{L}}{\partial w^{(2)}} = w^{(3).T}(a^{(3)}-y).(a^{2}(1-a^{(2)})) \;.\; a^{(1).T}$$

1x1   1x1   1x1   1x1   1x1   1x1   1x1   1x1

1x1   1x1   1x1

1x1   1x1

1x1

1x1   1x1

1x1

$$w^{(2)} = w^{(2)} - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}\left(w^{(3).T}(a^{3}-y)(a^{2}(1-a^{2})).\; a^{(1).T}\right]\right.$$

$$\frac{\partial \mathcal{L}}{\partial w^{(1)}} = ? \quad \text{Solving (3)}$$

- $\dfrac{\partial \mathcal{L}}{\partial a^{(3)}} = \dfrac{a^{3}-y}{a^{3}(1-a^{3})}$
- $\dfrac{\partial a^{(3)}}{\partial (z^{(3)})} = a^{(3)}(1-a^{(3)})$
- $\dfrac{\partial (z^{(3)})}{\partial (a^{(2)})} = w^{(3).T}\,pre$
- $\dfrac{\partial (a^{(2)})}{\partial (z^{(2)})} = a^{(2)}(1-a^{(2)})$
- $\dfrac{\partial (z^{(2)})}{\partial (a^{(1)})} = w^{(2).T}\,pre$
- $\dfrac{\partial (a^{(1)})}{\partial (z^{(1)})} = a^{(1)}(1-a^{(1)})$
- $\dfrac{\partial (z^{(1)})}{\partial (w^{(1)})} = \cancel{a^{(1).T}\,post}\; X^{.T}\,post$

$$\underset{3\times 6131}{\frac{\partial \mathcal{L}}{\partial w^{(1)}}} = (w^{(2).T})(w^{(3).T})\,(a^{(3)}-y)\,(a^{(2)}(1-a^{(2)}))(a^{(1)}(1-a^{(1)})).X^{T}$$

1×1 1×1 1×1 1×1 1×1

3×1 1×1

1×1 1×1 1×1 3×1 1×6131

1×1 1×1

1×6131

1×1

3×1

1×6131

3×6131

Same in way we can find Equations of b.

$$\frac{\partial \mathcal{L}}{\partial (b^{(3)})} = a^{(3)} - y$$

$$b^{(3)} = b^{(3)} - \alpha\left[\frac{1}{m}\sum_{i=1}^{m}(a^{(3)}-y)\right]$$

$$b^{(2)} = b^{(2)} - \alpha\left[\frac{1}{m}\sum_{i=1}^{m} w^{(3).T}(a^{(3)}-y).a^{(2)}(1-a^{(2)})\right]$$

$$b^{(1)} = b^{(1)} - \alpha\left[\frac{1}{m}\sum_{i=1}^{m} w^{(2).T}.w^{(3).T}(a^{(3)}-y)(a^{(2)}(1-a^{(2)}))(a^{(1)}(1-a^{(1)}))\right].$$