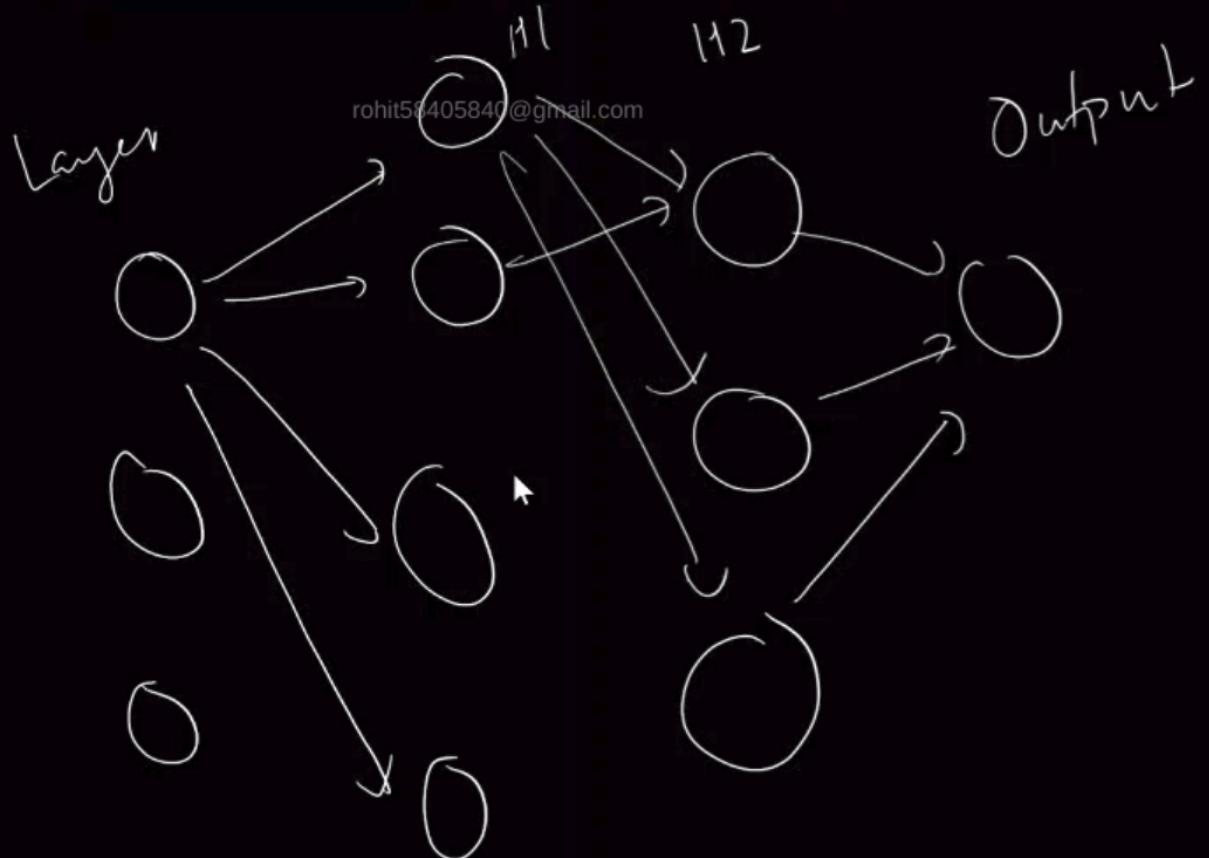




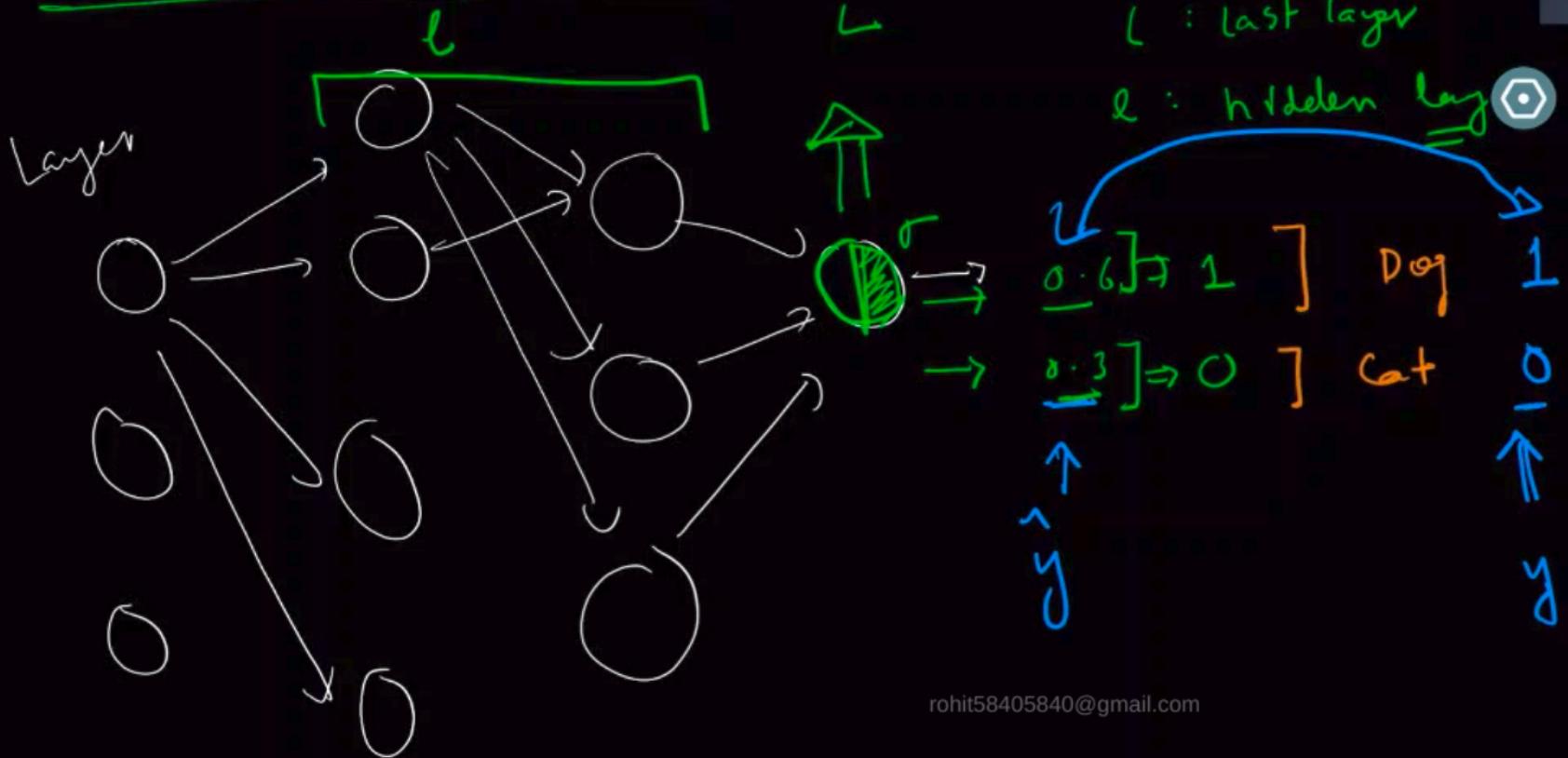
## PART-3

# Backpropagation





# Backpropagation

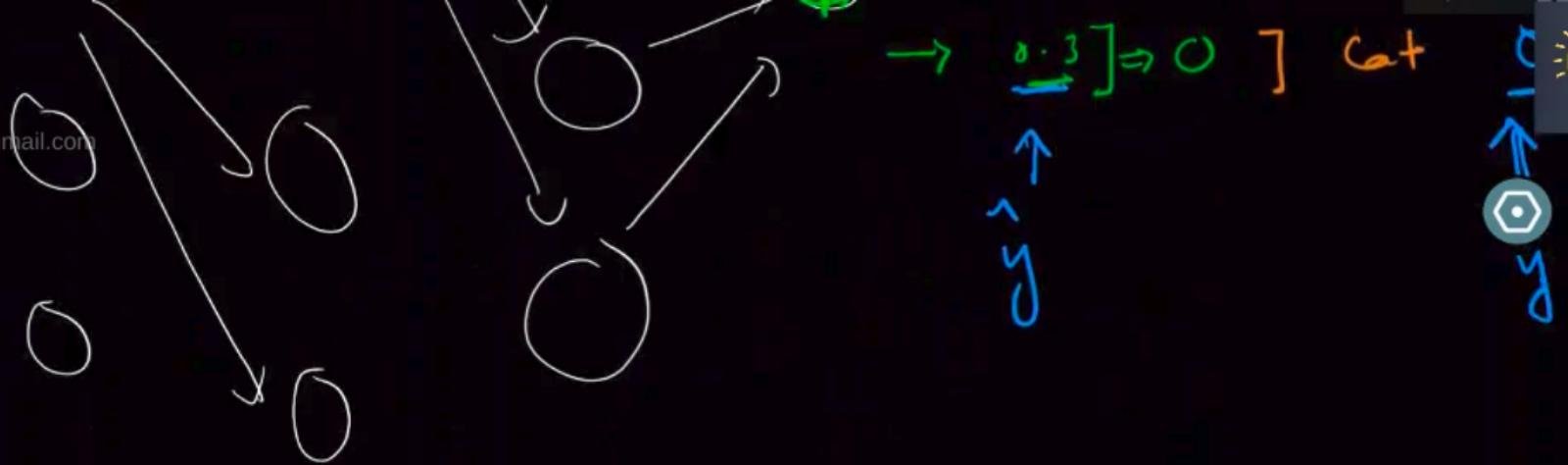


rohit58405840@gmail.com



Coding Blocks | Online

rohit58405840@gmail.com



For simplification -

$$MSE = L(\hat{y}, y) = \frac{1}{2} \sum_{i=1}^m (\hat{y}^{(i)} - y^{(i)})^2$$

$$MSE = L(y, \hat{y}) = \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



case-1

output layer

$$\frac{\partial L}{\partial W_{i,j}^L} = ?$$

rohit58405840@gmail.com

$$\frac{\partial L}{\partial b^L} = ?$$

case-2

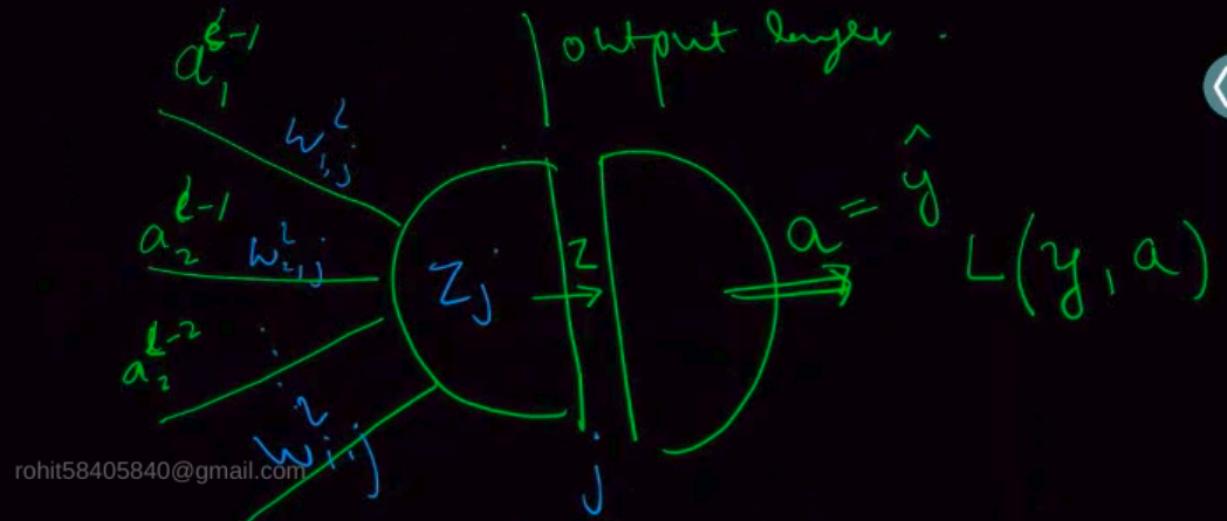
hidden layer

$$\frac{\partial L}{\partial W_{i,j}^l} = ?$$

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

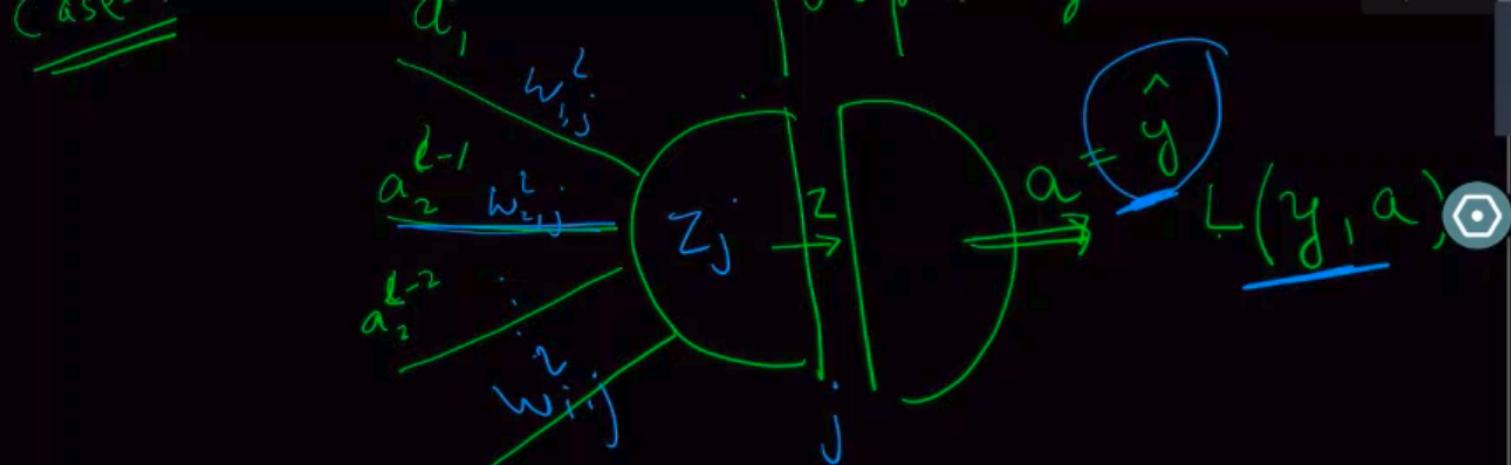


Case-1



rohit58405840@gmail.com

$$z_j = \sum_i^l w_{ij}^l a_i^{l-1} + b_j^l$$



$$w_{i,j} \rightarrow z_j \rightarrow a_j \rightarrow L(a, y)$$
$$z_j = \sum_i^l w_{i,j} a_i^{l-1} + b_j^l$$

$$\frac{\partial L}{\partial w_{i,j}} = \frac{\partial L}{\partial a_j}$$



$w_{ij} \rightarrow z_j \rightarrow a_i^j \rightarrow L(a, y)$

$$z_j = \sum_i^l w_{ij} a_i^{l-1} + b_j$$

rohit58405840@gmail.com

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ij}}$$

Chain Rule

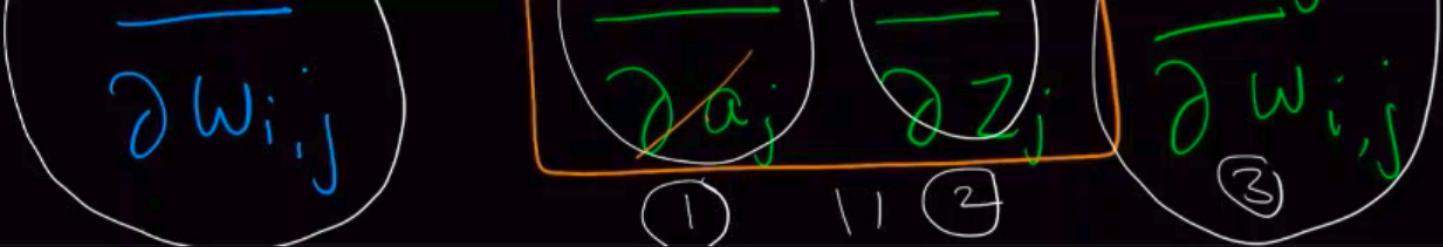
$$\frac{\partial L}{\partial w_{i,j}} = \frac{\partial L}{\partial a_j} \frac{\partial a_j}{\partial z_j} \frac{\partial z_j}{\partial w_{i,j}}$$

(1)      ||      (2)      (3)

Chain  
Rule

$\downarrow$

$$S_j^L \Leftarrow \frac{\partial L}{\partial z_j}$$



↓

$$S_j^L \Leftarrow \frac{\partial L}{\partial z_j}$$

$a_L$	$\hat{a}_j$	$y^L$
0	0.3	0
0	0.5	1
0	0.2	0

$$\textcircled{1} \quad \frac{\partial L}{\partial a_j} = (y_j - a_j)$$

rohit58405840@gmail.com

$$\begin{aligned} L &= \frac{1}{2} \sum (y_i^L - \hat{a}_i^L)^2 \\ \frac{\partial L}{\partial a_j} &= \frac{1}{2} \times (y_j - a_j) \end{aligned}$$

$$S_j \Leftarrow \frac{\partial L}{\partial z_j}$$

;  $\sigma(a_i)$   
 

$$\textcircled{1} \quad \frac{\partial L}{\partial a_j} = (y_j - a_j)$$

$$\textcircled{2} \quad \frac{\partial a_j}{\partial z_j} = \sigma'(z_j)$$

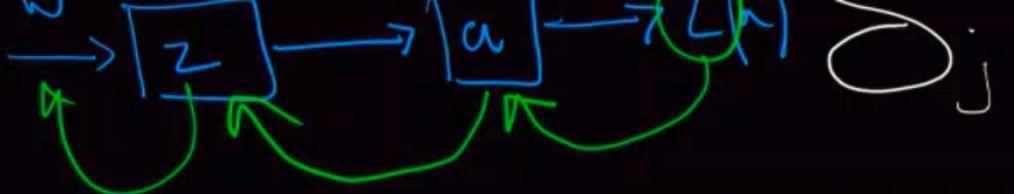
$$\left| \begin{array}{l} L = \frac{1}{2} \sum (y_i^l - a_i^l)^2 \\ \frac{\partial L}{\partial a_j} = \frac{1}{2} \times (y_j - a_j) \end{array} \right.$$

$$a_j = \sigma(z_j)$$

$$\frac{\partial a_j}{\partial z_j} = \underline{\sigma'(z_j)}$$

$$= (1 - \sigma(z_j)) \sigma(z_j)$$

rohit58405840@gmail.com



$$\textcircled{1} \quad \left\{ \begin{array}{l} \frac{\partial L}{\partial a_j} = (y_j - a_j) \\ \end{array} \right.$$

$$\textcircled{2} \quad \left\{ \begin{array}{l} \frac{\partial a_j}{\partial z_j} = \sigma'(z_j) \\ \end{array} \right.$$

$$\textcircled{3} \quad \left\{ \begin{array}{l} \frac{\partial z_j}{\partial w_{i,j}} \\ \end{array} \right.$$



$$\textcircled{1} \quad L = \frac{1}{2} \sum (y_i - a_i)^2$$

$$\frac{\partial L}{\partial a_j} = \frac{1}{2} \times (y_j - a_j)$$

rohit58405840@gmail.com

---


$$\textcircled{2}$$

$$a_j = \sigma(z_j)$$

$$\frac{\partial a_j}{\partial z_j} = \underline{\sigma'(z_j)}$$

$$= (1 - \sigma(z_j)) \sigma(z_j)$$


---

$$\textcircled{2} \quad \left\{ \begin{array}{l} \frac{\partial a_j}{\partial z_j} = \sigma'(z_j) \\ \end{array} \right.$$

rohit58405840@gmail.com

$$\textcircled{3} \quad \left\{ \begin{array}{l} \frac{\partial z_j}{\partial w_{i,j}} = a_j^{l-1} \\ \end{array} \right.$$

$$\textcircled{2} \quad \begin{aligned} a_j &= \sigma(z_j) \\ \frac{\partial a_j}{\partial z_j} &= \underline{\sigma'(z_j)} \\ &= (1 - \sigma(z_j)) \sigma(z_j) \\ \textcircled{3} \quad z_j^l &= \sum_i w_{i,j} \boxed{a_i^{l-1}} + b_j^l \end{aligned}$$

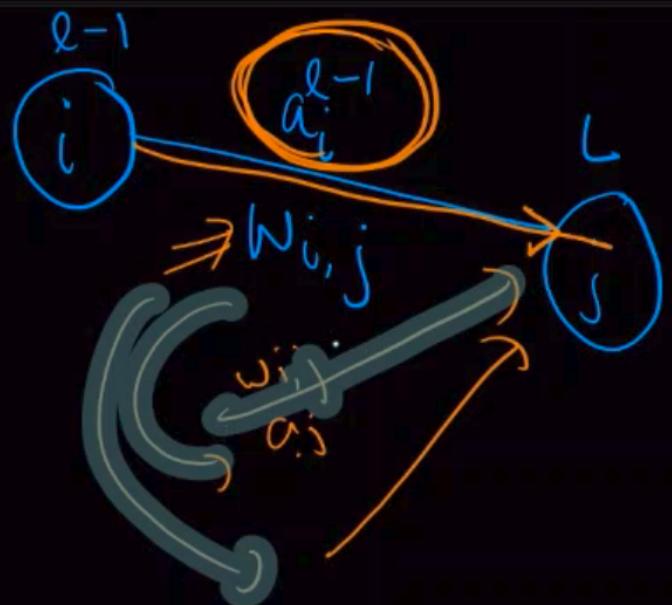
$$\frac{\partial z_j}{\partial w_{i,j}} =$$



$$③ \left\{ \begin{array}{l} \frac{\partial z_j}{\partial w_{i,j}} = a_i^{l-1} \\ \text{rohit58405840@gmail.com} \end{array} \right.$$

$$\begin{aligned} \frac{\partial a_j}{\partial z_j} &= \underline{\sigma'(z_j)} \\ &= (1 - \underline{\sigma(z_j)}) \underline{\sigma(z_j)} \end{aligned}$$

$$④ z_j^l = \sum_i w_{i,j}^l a_i^{l-1} + b_j^l$$



$$\frac{\partial z_j}{\partial w_{i,j}} = a_i^{l-1}$$



$$\frac{\partial L}{\partial w_{ij}} = \left[ (y_j - a_j) \sigma'(z_j) \right] a_i^{l-1}$$



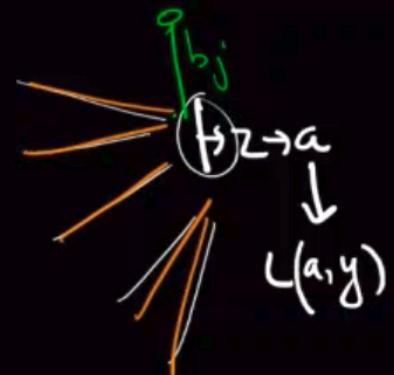
$$\Rightarrow \frac{\partial L}{\partial w_{ij}} = \sum_j^L a_i^{l-1}$$



$$\Rightarrow \frac{\partial L}{\partial w_{ij}} = \sum_j a_i^{l-1}$$



$$\Rightarrow \frac{\partial L}{\partial b_j} = \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial b}$$

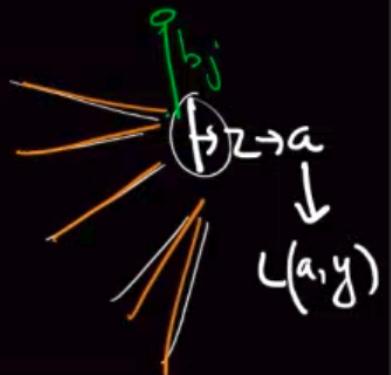




$$\Rightarrow \frac{\partial L}{\partial b_j} = \boxed{\frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial b_j}}$$

rohit58405840@gmail.com

$$= \delta_j^L$$

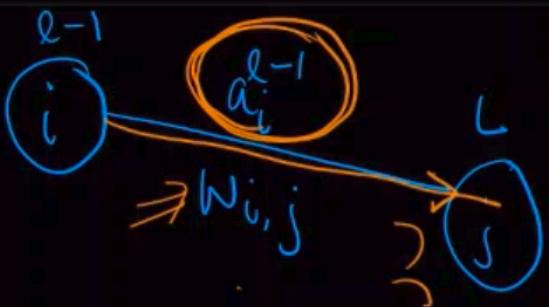


$$\textcircled{3} \quad \left\{ \frac{\partial z_j}{\partial w_{i,j}} = a_i \right.$$

$$\left. \frac{\partial a_j}{\partial z_j} = \sigma'(z_j) \right)$$

ronit58405840@gmail.com

$$\Rightarrow z_j^l = \sum_{i=1}^l w_{i,j} a_i^{l-1} + b_j^l$$



$$\frac{\partial z_j}{\partial w_{i,j}} = a_i^{l-1}$$

$$\frac{\partial z_j}{\partial b_j} = 1$$

$$\frac{\partial L}{\partial w_{i,j}} = \frac{(y_j - a_j) \sigma'(z_j)}{a_i^{l-1}}$$

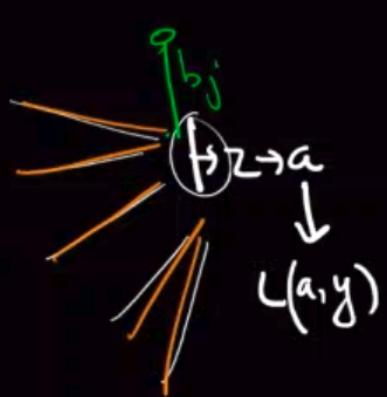
$\delta_j^l$



$$\Rightarrow \frac{\partial L}{\partial b_j} = \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial b_j}$$

$$\frac{\partial L}{\partial b_j} = \delta_j^L \cdot 1$$

Z



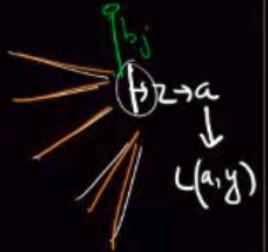


$$\Rightarrow \frac{\partial L}{\partial w_{ij}} = \sum_j \delta_i$$



$$\Rightarrow \frac{\partial L}{\partial b_j} = \frac{\partial L}{\partial a_j} \cdot \frac{\partial a_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial b_j}$$

$$\frac{\partial L}{\partial b_j} = \delta_j^L \cdot 1$$



Z

For output layer      rohit58405840@gmail.com

$$\frac{\partial L}{\partial w_{ij}} =$$



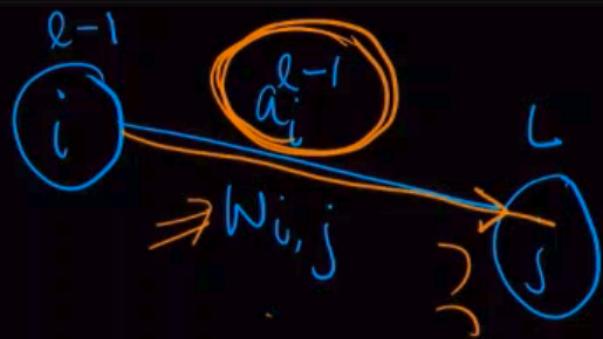


For output layer

$$\frac{\partial L}{\partial w_{ij}} = \delta_j^L \cdot \underline{a}_i^{i-1}$$

$$\frac{\partial L}{\partial b_j} = \delta_j^L \cdot 1$$





$$\frac{\partial z_j}{\partial w_{i,j}} = a_i^{l-1}$$

$$\frac{\partial z_j}{\partial b_j} = 1$$

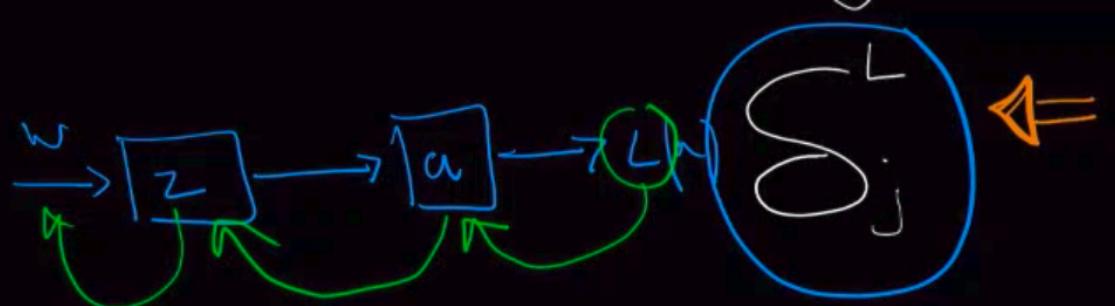
$$\frac{\partial z_j}{\partial a_i^{l-1}} = 1$$

$$\frac{\partial L}{\partial w_{i,j}} = \boxed{-(y_j - a_j) \sigma'(z_j)} \frac{\partial z_j}{\partial a_i^{l-1}}$$

rohit58405840@gmail.com

$$\Rightarrow \left( \frac{\partial L}{\partial w_{i,j}} \right) = \sum_j^L a_i^{l-1}$$

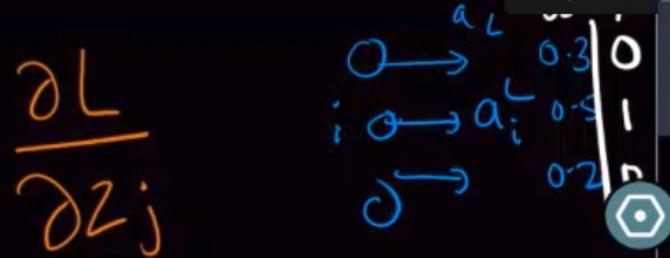




$$\textcircled{1} \quad \left\{ \begin{array}{l} \frac{\partial L}{\partial a_j} = (y_j - a_j) \end{array} \right.$$

$$\textcircled{2} \quad \left\{ \begin{array}{l} \frac{\partial a_j}{\partial z_j} = \sigma'(z_j) \end{array} \right.$$

$$\textcircled{3} \quad \left\{ \begin{array}{l} \frac{\partial z_j}{\partial w_{i,j}} = a_i^{l-1} \end{array} \right.$$



$$\textcircled{1} \quad L = \frac{1}{2} \sum (y_i - a_i)^2$$

$$\frac{\partial L}{\partial a_j} = \frac{1}{2} \times (y_j - a_j) - 1$$

$$\textcircled{2} \quad \begin{aligned} a_j &= \sigma(z_j) \\ \frac{\partial a_j}{\partial z_j} &= \underline{\sigma'(z_j)} \\ &= (1 - \sigma(z_j)) \sigma(z_j) \end{aligned}$$

$$\textcircled{3} \quad z_j = \sum_{i=1}^l w_{i,j} a_i^{l-1} + b_j$$

rohit58405840@gmail.com



$$\frac{\partial L}{\partial b_j} = \delta_j^L \cdot 1$$

z

For output layer

rishi58405840@gmail.com

$$\frac{\partial L}{\partial w_{ij}} = \delta_j^L \cdot \underline{a_i^{i-1}}$$

$$\delta_j^L = (a_j - y_j) \sigma'(z_j)$$

$$\frac{\partial L}{\partial b_j} = \delta_j^L \cdot 1$$



For output layer

$$\frac{\partial L}{\partial w_{ij}} = \delta_j^L \cdot \underline{a}_i^{i-1}$$

$$\frac{\partial L}{\partial b_j} = \delta_j^L \cdot 1$$

$$\delta_j^L = (a_j - y_j) \sigma'(z_j)$$



rohit58405840@gmail.com