



rohit58405840@gmail.com

Back propagation  
 using different loss Function

$$\delta^L = (a^L - y) \odot \sigma'(z^L)$$

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

Assumption

$$\text{MSE Loss} = \frac{1}{2} \sum_i (y_i - a_i)^2$$

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

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





using different loss Function

$$\delta^L = \underbrace{(a^L - y)}_{\text{Loss}} \odot \sigma'(z^L)$$

rohit58405840@gmail.com

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

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

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

Assumption

$$\text{MSE Loss} = \frac{1}{2} \sum_i (y_i - a_i)^2$$

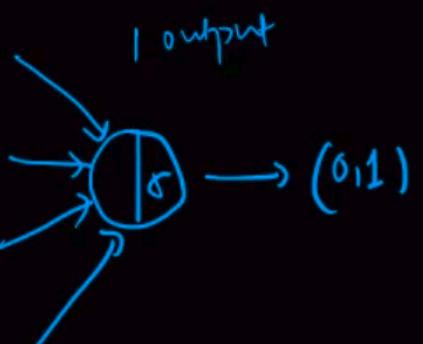
$$\frac{\partial L}{\partial z^L}$$

Vectorization for one example





## Binary classification

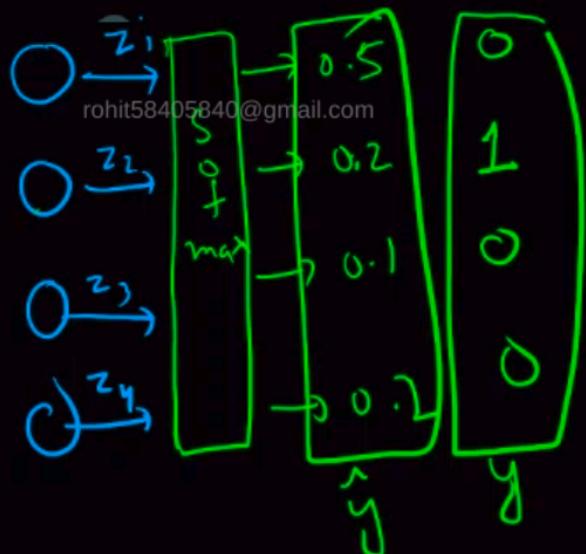


$y_i$  → True label  
 $\hat{y}_i$  → Predict

## Cross Entropy

$$-\sum_{i=1}^m \left( y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \right)$$

## K categories

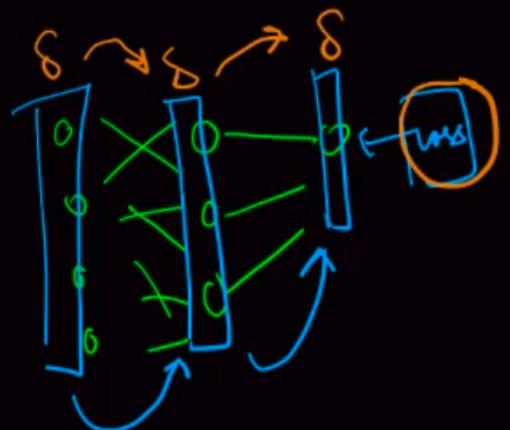
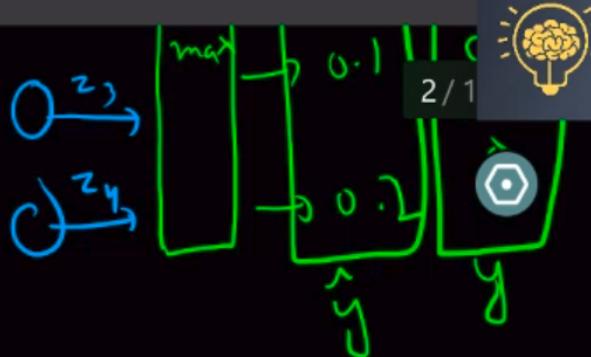




Binary Cross Entropy

$$-\sum_{i=1}^m \left( y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \right)$$

$y_i$  → True label  
 $\hat{y}_i$  → Predict





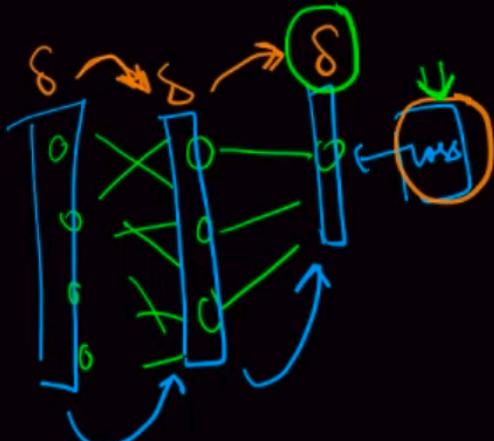
ŷ



2/1

Cross Entropy

$$-\sum_{i=1}^m \left( y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \right)$$

 $\delta_L$  change!

$$(z \rightarrow \sigma(z)) \hat{y} \rightarrow L(y, \hat{y})$$

$$\delta_L = \frac{\partial L}{\partial z^l}$$

rohit58405840@gmail.com



Coding Blocks | Online



$$\delta_L = \frac{\partial L}{\partial z^e} = \left( \frac{\partial L}{\partial a} \right) \cdot \left[ \frac{\partial a}{\partial z} \right]$$



$$(z + \sigma(a)) \rightarrow L(y, \hat{y})$$

$\hat{y} = a$



$$= \left( -\frac{y_i}{\hat{y}_i} + \frac{1-y_i}{1-\hat{y}_i} \right) \hat{y}_i \cancel{(1-\hat{y}_i)} \quad a = \sigma(z)$$

$$= \cancel{(-y_i + y_i \hat{y}_i + \hat{y}_i - \hat{y}_i \hat{y}_i)} \quad \begin{aligned} \frac{\partial a}{\partial z} &= \sigma(z)(1-\sigma(z)) \\ \hat{y}_i \cancel{(1-\hat{y}_i)} &= a(1-a) \\ &= \hat{y}_i(1-\hat{y}_i) \end{aligned}$$

$$\delta_L = \boxed{(\hat{y}_i - y_i)}$$





3 / 1

$$= \left( -\frac{y_i}{\hat{y}_i} + \frac{1-y_i}{1-\hat{y}_i} \right) \hat{y}_i \cancel{\ln(1-\hat{y}_i)}$$

$$= \cancel{\left( -y_i + y\hat{y}_i + \hat{y}_i - \cancel{\hat{y}_i y_i} \right)} \cancel{\hat{y}_i (1-\hat{y}_i)}$$

$$\begin{aligned} a &= \sigma(z) \\ \frac{\partial a}{\partial z} &= \sigma(z)(1-\sigma(z)) \\ \frac{\partial z}{\partial z} &= 1 \\ a(1-a) &= \hat{y}(1-\hat{y}) \end{aligned}$$

$$\boxed{\delta_L = (\hat{y}_i - y_i)}$$

Cross Entropy  
Loss.

rohit58405840@gmail.com



Coding Blocks | Online



0 00

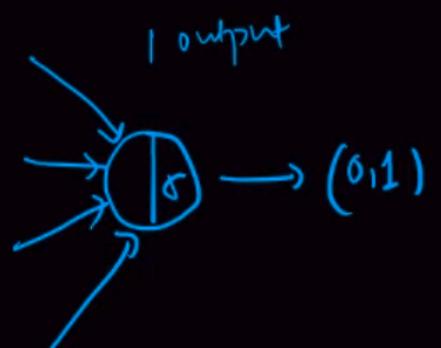


2 / 1



## Binary classification

rohit58405840@gmail.com



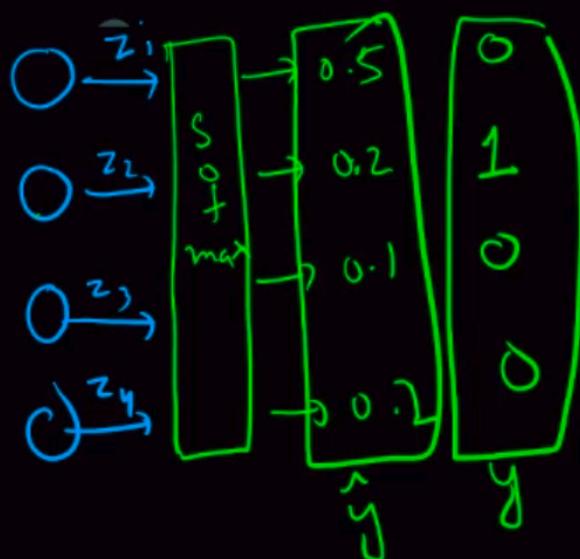
$y_i$  → True label  
 $\hat{y}_i$  → Predict

## Binary Cross Entropy

$$-\sum_{i=1}^m \left( y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \right)$$

Categorical cross entropy

## K Categories





Categorical Cross Entropy

rohit58405840@gmail.com

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i \in N} \sum_{j \in C} y_{nji} \log \hat{y}_{nji}$$



## Categorical Cross Entropy

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{n \in N} \sum_{i \in C} y_{n,i} \log \hat{y}_{n,i}$$

2 classes

$$= -\frac{1}{N} \sum_{n \in N} \left( y_{n,1} \log \hat{y}_{n,1} + \cancel{y_{n,0} \log \hat{y}_{n,0}} \right)$$

$$y_{n,0} = 1 - y_{n,1}$$

$$= -\frac{1}{N} \sum_{n \in N} \left( y_{n,1} \log y_{n,1} + \overbrace{\overbrace{y_{n,0}}^{\text{reg}}}^{\text{reg}} \right)$$



$$y_{n,0} = 1 - y_{n,1}$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \left( \frac{\partial a}{\partial z} \right)$$

rohit58405840@gmail.com

$$= (\hat{y}_i - y_i)$$

↑ Predict      ↑ Target



using different Loss Function

Tel: 9406840@gmail.com

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

$$\delta^L = (a^L - y) \odot \sigma'(z^L)$$

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

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

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

Assumption

$$\text{MSE loss} = \frac{1}{2} \sum_i (y_i - a_i)^2$$

$$\frac{\partial L}{\partial z^L} = (a^L - y) \cdot \sigma'(z^L)$$

Vectorization for one example