rohit58405840@gmail.com

Backpropagation

using different loss Function

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

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

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

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

Assumption

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

using different loss Function

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

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

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

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

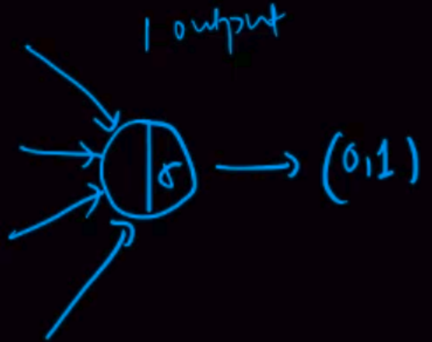
Assumption

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

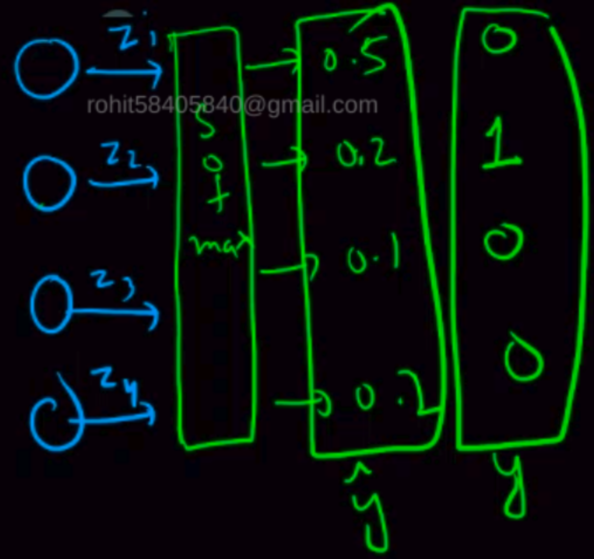
$$\delta^L = \boxed{\frac{\partial L}{\partial z^L} = (a^L - y) \cdot \sigma'(z^L)}$$

vectorization for one example

Binary classification



K categories

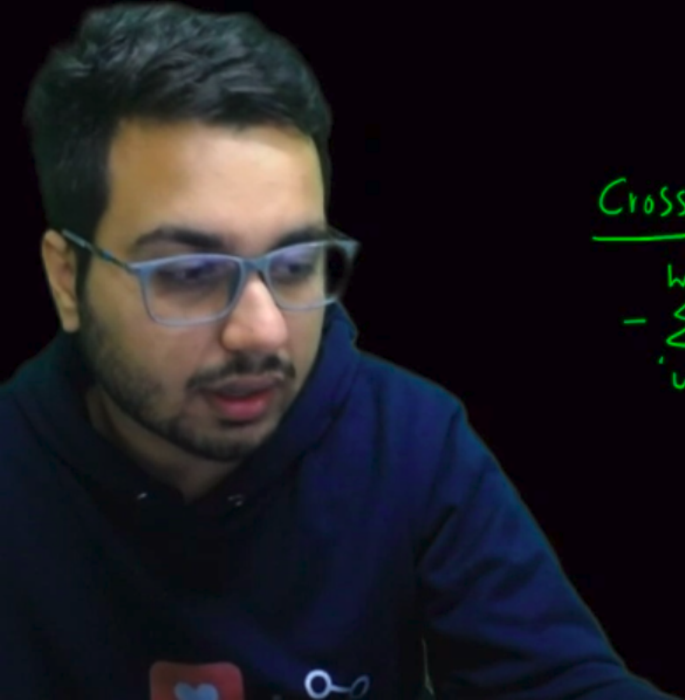


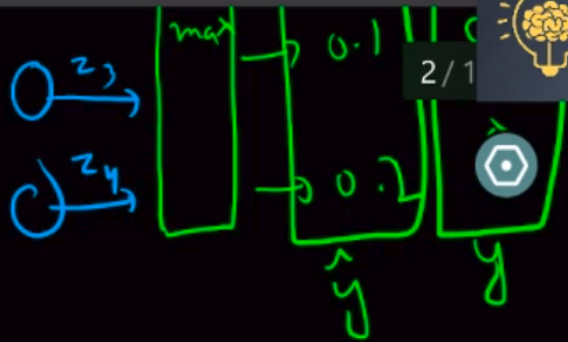
$y_i \rightarrow$ True label

$\hat{y}_i \rightarrow$ Predict

Cross Entropy

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



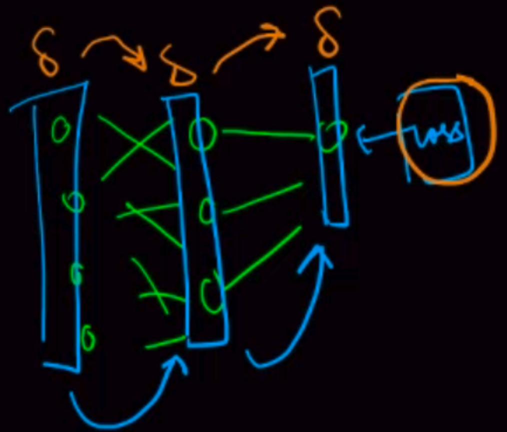


$y_i \rightarrow$ True label

$\hat{y}_i \rightarrow$ Predict

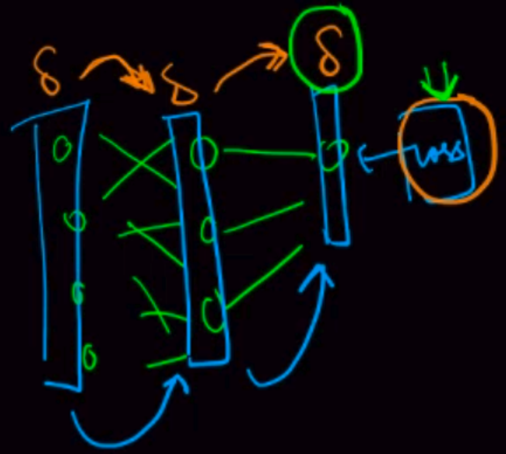
Binary Cross Entropy

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



Cross Entropy

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



δ_L change!

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

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

$$z \to a = \sigma(\,) \to \hat{y} \to L(y, \hat{y})$$

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

$$\hat{y} = a$$

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

$$a = \sigma(z)$$

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

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

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

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

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

$$a = \sigma($$

$$\frac{\partial a}{\partial z} = \sigma(z)(1-\sigma(z))$$

$$= a(1-a)$$

$$= \hat{y}(1-\hat{y})$$

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

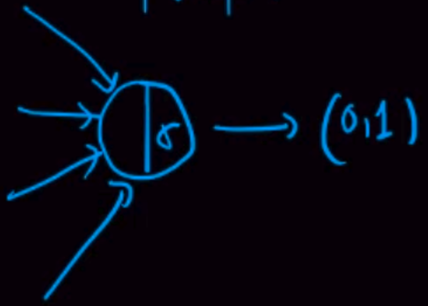
Cross Entropy Loss.

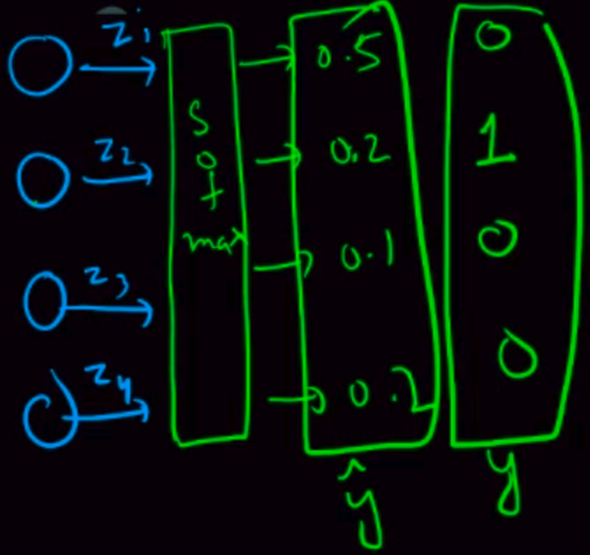
Categorical cross entropy

$\boxed{K \text{ categories}}$

Binary classification

1 output





$y_i \rightarrow$ True label

$\hat{y}_i \rightarrow$ 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

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

43 %

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

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

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

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

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

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \cdot \frac{\partial a}{\partial z}$$

softmax

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

Predict

Target

using different loss Function

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

$$\boxed{\delta^L = \underbrace{(a^L - y)}\odot \sigma'(z^l)}$$

×

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

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

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

Assumption

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

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

vectorization for one example