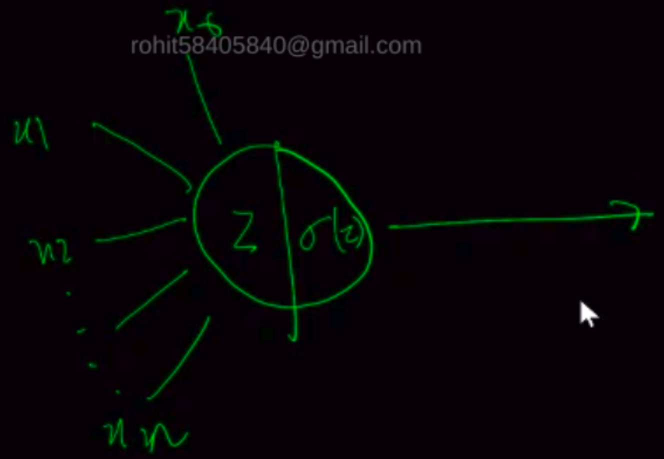
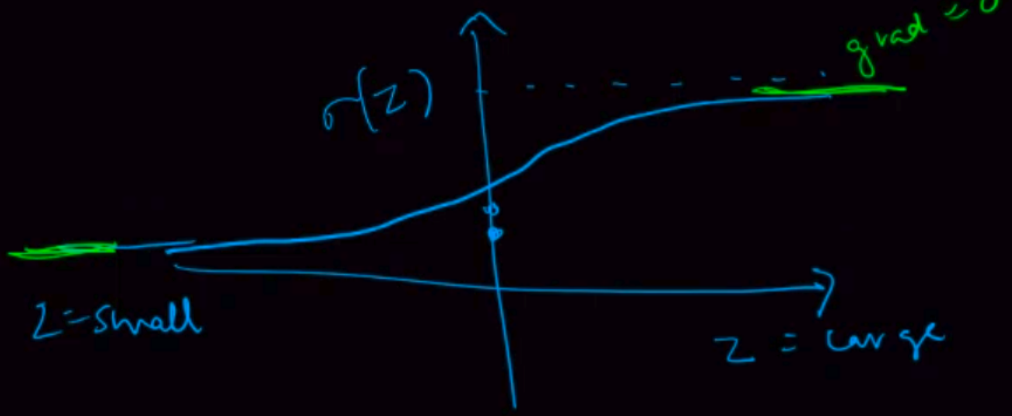
Vanishing Gradients Problem

backprop_2_chaning_loss_function.pdf - Drawboard PDF



$$Z = \sum_{i=0}^{n} \boxed{W_i x_i}$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



35 %

nn

z = small

z = large

$z \to \infty$

$\sigma(z) = 1$

$$\text{grad} = \sigma'(z) = \sigma(z)\left[1 - \sigma(z)\right]$$

$$= 0$$

$z \to -\infty$

$\sigma(z) \to$

$$= 0$$

grad = 0

CODE

For 1 example

output $\delta^L = \boxed{(a^L - y^L)}$ ✓

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

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

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

Almost

For m examples

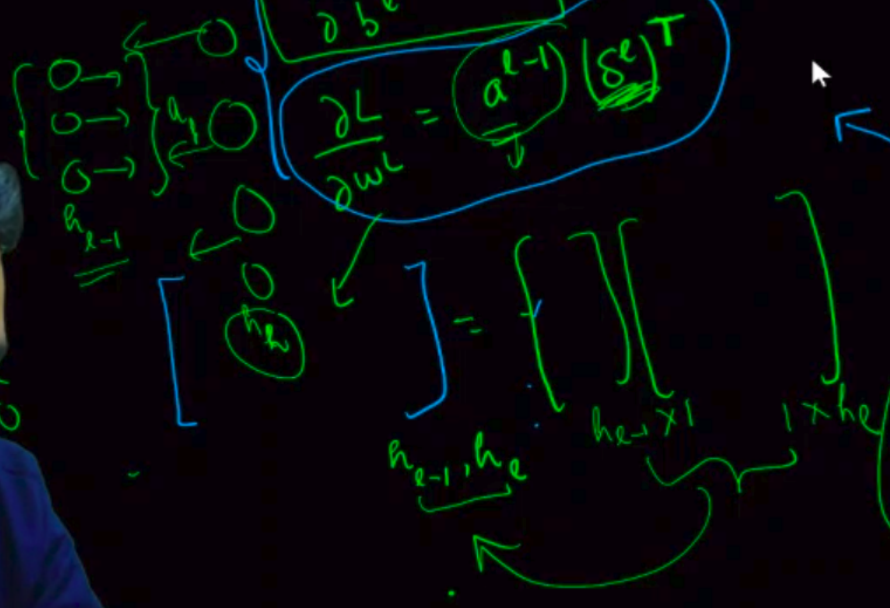
$\delta^L = \boxed{a^L - y^L}$ ✓

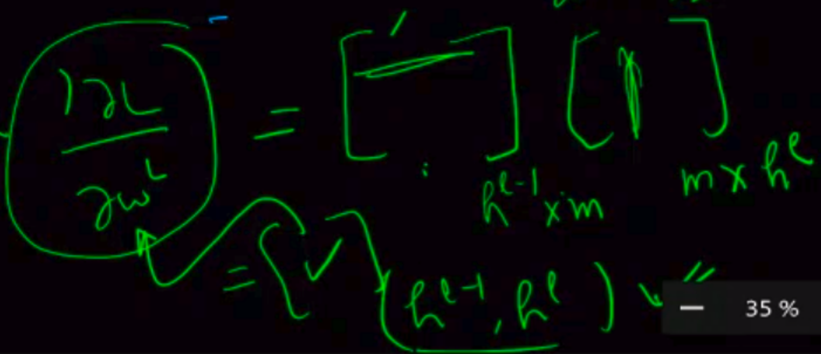
$\delta^l = (\delta^{l+1} \cdot w^{l+1^T}) \odot \sigma'(z^l)$

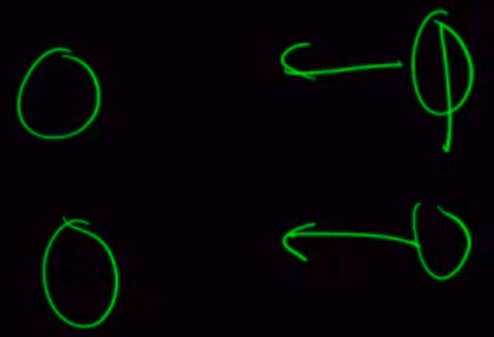
$$\frac{\partial L}{\partial b^l} = \frac{1}{m}\text{np.sum}(\delta^l, \text{axis}=0)$$

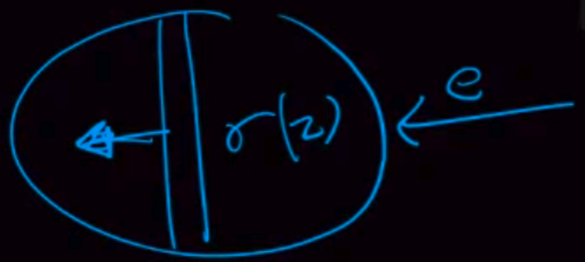
$$\frac{\partial L}{\partial w^l} = a^{l-1^T} \cdot \delta^l$$

Do make dimensions









$$e \cdot \sigma'(z)$$

$$= e \cdot \sigma(z)\,\underbrace{(1-\sigma(z))}_{0}$$

backprop_2_chaning_loss_function.pdf - Drawboard PDF

10 / 10

$e \cdot \sigma'(z)$

rohit58405840@gmail.com

$\{ w_i, x_i$

$\boxed{\text{small}}$

$[-0.5 \;, \; 0.5]$

$= e \cdot \sigma(z)\,(1-\sigma(z))$

0

$= e \cdot \sigma(z)\,(1-\sigma(z))$

0

$[2, 5]$

35 %

$\{W_i \; x_i$

↳ small

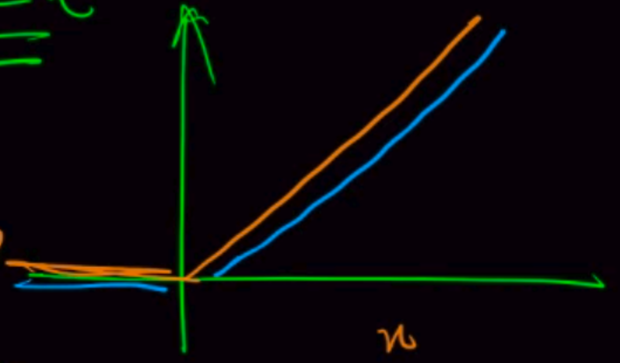
↳ $[-0.5 \;,\; +0.5]$

$= e \cdot \sigma(z)\,(1-\sigma(z))$

$= e \cdot \underbrace{\overbrace{\sigma(z)\,(1-\sigma(z))}^{= 0}}_{= 0}$

ReLu

$$\text{ReLu}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$



$$\text{ReLu}'(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$