$$C = (A_2 - Y)^2$$

$$A_2 = \text{sigmoid}(Z_2)$$

$$Z_2 = W_2 A_1 + B_2$$

$$A_1 = \text{Relu}(Z_1)$$

$$Z_1 = W_1 X + B_1$$

<u>chain rule</u>

$$\frac{dz}{dx} = \frac{dz}{dy} \times \frac{dy}{dx}$$

$$\frac{dC}{dW_2} = \frac{dZ_2}{dW_2} \times \frac{dA_2}{dZ_2} \times \frac{dC}{dA_2}$$

$$\boxed{\frac{dC}{dW_2}} = (A_1)\underbrace{\underbrace{(\sigma(x)(1 - \sigma(x)))}_{\text{sigmoid prime}}(2A_2 - 2y)}_{\text{delta}}$$

cost derivative

sigmoid

$$\boxed{\frac{dC}{dB_2}} = \frac{dZ_2}{dB_2} * \frac{dA_2}{dZ_2} * \frac{dC}{dA_2} = (1) * \underbrace{(\sigma(Z_2)(1 - \sigma(Z_2))(2A_2 - 2y))}_{\text{delta}}$$

$$\boxed{\frac{dC}{dW_1}} = \frac{dZ_1}{dW_1} * \frac{dA_1}{dZ_1} * \frac{dZ_2}{dA_1} * \frac{dA_2}{dZ_2} * \frac{dC}{dA_2} =$$

$$= \underbrace{(x)\,(Z_1 > 0) \cdot W_2}\cdot\ \sigma(Z_2)(1 - \sigma(Z_2)) \cdot (2A_2 - 2y)$$

$$\frac{dC}{dB_1} = \frac{dZ_1}{dB_1} * \frac{dA_1}{dZ_1} * \frac{dZ_2}{dA_1} * \frac{dA_2}{dZ_2} * \frac{dC}{dA_2} =$$

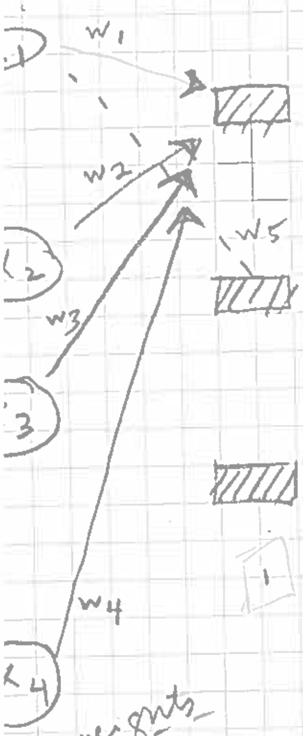
$$= \underbrace{(1)\,(Z_1 > 0) \cdot W_2}_{\text{sy}} \cdot \underbrace{\sigma(Z_2)(1 - \sigma(Z_2))}_{\text{sigmoid prime}} \cdot \underbrace{(2A_2 - 2y)}_{\text{cost derivative}}$$

$|\longleftarrow$ delta $\longrightarrow|$

previous layer.

Equation depects simple two layer neural network, A1 is the output of the first layer and input to the second layer. W and B are the weights (determining inputs importance) and bias (which shifts the activation function output). For a single neuron, this is the sum of each input multiplied by its weight plus bias and for entire layer of neurons this is done via matrix multiplication Z = WX + B

Z is the Weighted Sum, or output for a single Neuron, before Activation function
A is the result of the activation function which introduces non-linearity, allowing the network to learn complicated patterns
C is the Loss function which measures how far the network's prediction is from the target
dC/db and dC/dw are the gradient descent updates used to adjust weights and biases



$x_1 w_1 + x_2 w_2 + x_3 w_3 + x_4 w_4$

$x_1 w_5 + x_2 w_6 + x_3 w_7 + x_4 w_8$

$x_1 w_9 + x_2 w_{10} + x_3 w_{11} + x_4 w_{12}$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad [4 \times 1]$$

$[x_1, x_2, x_3, x_4]$ $[1 \times 4]$

transpose

weights

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_5 & w_6 & w_7 & w_8 \\ w_9 & w_{10} & w_{11} & w_{12} \end{bmatrix} \quad \begin{bmatrix} \_\_ \\ \_\_ \\ \_\_ \end{bmatrix}$$

$[3 \times 4]$ $\qquad$ $[3 \times 1]$

(hidden size, input size)

fe $\quad$ 325

~~2:10 pm~~ $\quad$ 325

~~3:18 pm~~

~~4:11 pm~~ $\quad$ 325

4:13 pm $\quad$ 325

~~4:49 pm~~

~~1:48~~ $\quad$ 315

~~2:48 pm~~

~~3:18 pm~~ $\quad$ 315

4:19 pm

S

$$\begin{bmatrix}0\\1\\2\\3\end{bmatrix} \quad (4,1)$$

b. unti- shape (4,1)

$[10, 20, 30, 40]$ (1,4)

$10 * 0 + 20 * 1 + 30 * 2 + 40 * 3$

$0 + 20 + 60 + 120$

$80 + 120$

200

$[0 \quad 1 \quad 2 \quad 3]$ (1x4)

$$\begin{bmatrix}10\\20\\30\\40\end{bmatrix} \begin{matrix} 0 & 10 & 20 & 30 \\ 0 & 20 & 40 & 60 \\ 0 & 30 & 60 & 90 \\ 0 & 40 & 80 & 120 \end{matrix}$$

(4x1)

$$Z_1 = W_{hidden} \; X + B_{hidden}$$

$$\begin{bmatrix} 0.05629 & 0.596 & 0.238 \\ 0.303 & 0.421 & 0.855 \\ 0.655 & 0.457 & 0.618 \end{bmatrix}$$

W_{hidden} 3x3