

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

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

$$Z_2 = W_2 A_1 + B_2$$

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

$$Z_1 = W_1 X + B_1$$

chain rule

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

$$\frac{dC}{dw_2} = \frac{dZ_2}{dw_2} \times \frac{dA_2}{dZ_2} \times \frac{dC}{dA_2}$$

$$\frac{dC}{dw_2} = (A_1) \left(\underbrace{\sigma(x)}_{\text{sigmoid}} \right) \left(\underbrace{(1 - \sigma(x))}_{\text{sigmoid prime}} \right) \underbrace{(2A_2 - 2Y)}_{\text{cost derivative}}$$

$$\frac{dC}{dB_2} = \frac{dZ_2}{dB_2} \times \frac{dA_2}{dZ_2} \times \frac{dC}{dA_2} = (1) \times (\underbrace{\sigma(Z_2)}_{\text{sigmoid}}) \underbrace{(1 - \sigma(Z_2))}_{\text{sigmoid prime}} \underbrace{(2A_2 - 2Y)}_{\text{delta}}$$

$$\frac{dC}{dw_1} = \frac{dZ_1}{dw_1} \times \frac{dA_1}{dZ_1} \times \frac{dZ_2}{dA_1} \times \frac{dA_2}{dZ_2} \times \frac{dC}{dA_2} =$$

$$= (X) \underbrace{(Z_1 > 0)}_{\text{sy}} \cdot W_2 \cdot \underbrace{\sigma(Z_2)}_{\text{sigmoid}} \underbrace{(1 - \sigma(Z_2))}_{\text{sigmoid prime}} \cdot \underbrace{(2A_2 - 2Y)}_{\text{delta}}$$

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

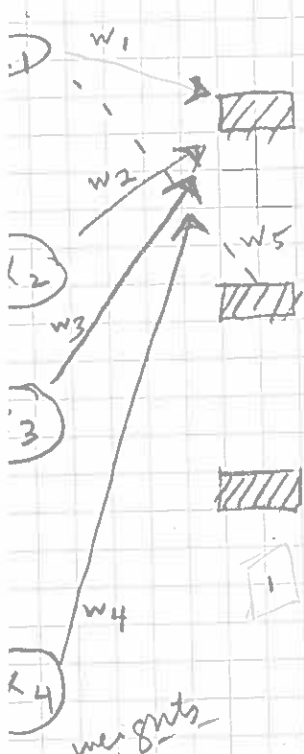
$$= (1) \underbrace{(Z_1 > 0)}_{\text{sy}} \cdot W_2 \cdot \underbrace{\sigma(Z_2)}_{\text{sigmoid prime}} \underbrace{(1 - \sigma(Z_2))}_{\text{sigmoid}} \cdot \underbrace{(2A_2 - 2Y)}_{\text{cost derivative}}$$

← delta →
previous layer.

Equation depicts 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



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

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

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

x_1
 x_2
 x_3
 x_4

$[4 \times 1]$

$[x_1, x_2, x_3, x_4]$

$[1 \times 4]$

transpose

w_1	w_2	w_3	w_4
w_5	w_6	w_7	w_8
w_9	w_{10}	w_{11}	w_{12}

$[3 \times 4]$

$[3 \times 1]$

(hidden size, input size)

$\frac{1.48}{2.48} \text{ pm}$
 $\frac{2.48}{3.18} \text{ pm}$
 $\frac{3.18}{4.19} \text{ pm}$
 $\frac{4.19}{5.19} \text{ pm}$
 $\frac{5.19}{6.19} \text{ pm}$
 $\frac{6.19}{7.19} \text{ pm}$
 $\frac{7.19}{8.19} \text{ pm}$
 $\frac{8.19}{9.19} \text{ pm}$
 $\frac{9.19}{10.19} \text{ pm}$
 $\frac{10.19}{11.19} \text{ pm}$
 $\frac{11.19}{12.19} \text{ pm}$
 $\frac{12.19}{13.19} \text{ pm}$
 $\frac{13.19}{14.19} \text{ pm}$
 $\frac{14.19}{15.19} \text{ pm}$
 $\frac{15.19}{16.19} \text{ pm}$
 $\frac{16.19}{17.19} \text{ pm}$
 $\frac{17.19}{18.19} \text{ pm}$
 $\frac{18.19}{19.19} \text{ pm}$
 $\frac{19.19}{20.19} \text{ pm}$
 $\frac{20.19}{21.19} \text{ pm}$
 $\frac{21.19}{22.19} \text{ pm}$
 $\frac{22.19}{23.19} \text{ pm}$
 $\frac{23.19}{24.19} \text{ pm}$
 $\frac{24.19}{25.19} \text{ pm}$
 $\frac{25.19}{26.19} \text{ pm}$
 $\frac{26.19}{27.19} \text{ pm}$
 $\frac{27.19}{28.19} \text{ pm}$
 $\frac{28.19}{29.19} \text{ pm}$
 $\frac{29.19}{30.19} \text{ pm}$
 $\frac{30.19}{31.19} \text{ pm}$
 $\frac{31.19}{32.19} \text{ pm}$
 $\frac{32.19}{33.19} \text{ pm}$
 $\frac{33.19}{34.19} \text{ pm}$
 $\frac{34.19}{35.19} \text{ pm}$
 $\frac{35.19}{36.19} \text{ pm}$
 $\frac{36.19}{37.19} \text{ pm}$
 $\frac{37.19}{38.19} \text{ pm}$
 $\frac{38.19}{39.19} \text{ pm}$
 $\frac{39.19}{40.19} \text{ pm}$
 $\frac{40.19}{41.19} \text{ pm}$
 $\frac{41.19}{42.19} \text{ pm}$
 $\frac{42.19}{43.19} \text{ pm}$
 $\frac{43.19}{44.19} \text{ pm}$
 $\frac{44.19}{45.19} \text{ pm}$
 $\frac{45.19}{46.19} \text{ pm}$
 $\frac{46.19}{47.19} \text{ pm}$
 $\frac{47.19}{48.19} \text{ pm}$
 $\frac{48.19}{49.19} \text{ pm}$
 $\frac{49.19}{50.19} \text{ pm}$
 $\frac{50.19}{51.19} \text{ pm}$
 $\frac{51.19}{52.19} \text{ pm}$
 $\frac{52.19}{53.19} \text{ pm}$
 $\frac{53.19}{54.19} \text{ pm}$
 $\frac{54.19}{55.19} \text{ pm}$
 $\frac{55.19}{56.19} \text{ pm}$
 $\frac{56.19}{57.19} \text{ pm}$
 $\frac{57.19}{58.19} \text{ pm}$
 $\frac{58.19}{59.19} \text{ pm}$
 $\frac{59.19}{60.19} \text{ pm}$
 $\frac{60.19}{61.19} \text{ pm}$
 $\frac{61.19}{62.19} \text{ pm}$
 $\frac{62.19}{63.19} \text{ pm}$
 $\frac{63.19}{64.19} \text{ pm}$
 $\frac{64.19}{65.19} \text{ pm}$
 $\frac{65.19}{66.19} \text{ pm}$
 $\frac{66.19}{67.19} \text{ pm}$
 $\frac{67.19}{68.19} \text{ pm}$
 $\frac{68.19}{69.19} \text{ pm}$
 $\frac{69.19}{70.19} \text{ pm}$
 $\frac{70.19}{71.19} \text{ pm}$
 $\frac{71.19}{72.19} \text{ pm}$
 $\frac{72.19}{73.19} \text{ pm}$
 $\frac{73.19}{74.19} \text{ pm}$
 $\frac{74.19}{75.19} \text{ pm}$
 $\frac{75.19}{76.19} \text{ pm}$
 $\frac{76.19}{77.19} \text{ pm}$
 $\frac{77.19}{78.19} \text{ pm}$
 $\frac{78.19}{79.19} \text{ pm}$
 $\frac{79.19}{80.19} \text{ pm}$
 $\frac{80.19}{81.19} \text{ pm}$
 $\frac{81.19}{82.19} \text{ pm}$
 $\frac{82.19}{83.19} \text{ pm}$
 $\frac{83.19}{84.19} \text{ pm}$
 $\frac{84.19}{85.19} \text{ pm}$
 $\frac{85.19}{86.19} \text{ pm}$
 $\frac{86.19}{87.19} \text{ pm}$
 $\frac{87.19}{88.19} \text{ pm}$
 $\frac{88.19}{89.19} \text{ pm}$
 $\frac{89.19}{90.19} \text{ pm}$
 $\frac{90.19}{91.19} \text{ pm}$
 $\frac{91.19}{92.19} \text{ pm}$
 $\frac{92.19}{93.19} \text{ pm}$
 $\frac{93.19}{94.19} \text{ pm}$
 $\frac{94.19}{95.19} \text{ pm}$
 $\frac{95.19}{96.19} \text{ pm}$
 $\frac{96.19}{97.19} \text{ pm}$
 $\frac{97.19}{98.19} \text{ pm}$
 $\frac{98.19}{99.19} \text{ pm}$
 $\frac{99.19}{100.19} \text{ pm}$

$$[10, 20, 30, 40]_{(1,4)} \quad \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} \leftarrow [4,1] \quad \text{b. int. shape } (4,1)$$

$$\begin{aligned} &10 \times 0 + 20 \times 1 + 30 \times 2 + 40 \times 3 \\ &\quad 0 + 20 + 60 + 120 \\ &\quad 80 + 120 \\ &\quad 200 \end{aligned}$$

$$\begin{bmatrix} 0 & 1 & 2 & 3 \end{bmatrix}_{(1 \times 4)} \quad \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \end{bmatrix}_{(4 \times 1)}$$

10	0	10	20	30
20	0	20	40	60
30	0	30	60	90
40	0	40	80	120

$$Z_1 = W_{\text{hidden}} \times B_{\text{hidden}}$$

$$W_{\text{hidden}} \quad 3 \times 3 \quad \begin{bmatrix} 0.05629 & 0.596 & 0.238 \\ 0.303 & 0.421 & 0.855 \\ 0.655 & 0.457 & 0.618 \end{bmatrix}$$