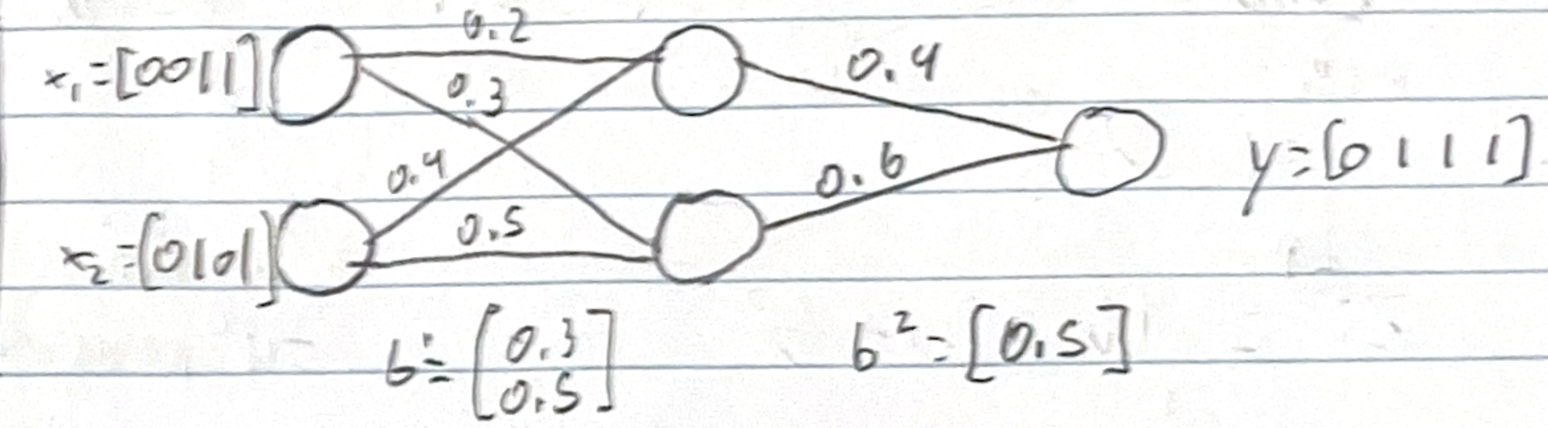
2. Math problem.

Compute the forward passes and gradient manually for a neural network with 1 hidden layer (w/ 2 neurons) and 1 output neuron.

	Input		
	x_1	x_2	y
S_1	0	0	0
S_2	0	1	1
S_3	1	0	1
S_4	1	1	1

$W^1 = \begin{bmatrix} 0.2 & 0.4 \\ 0.3 & 0.5 \end{bmatrix}$ $W^2 = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$



$b^1 = \begin{bmatrix} 0.3 \\ 0.5 \end{bmatrix}$ $b^2 = [0.5]$

* input layer size is determined by number of features

$z^{(1)}$

$$z^{(1)} = Wx + b$$

$$z^{(1)} = \underset{W}{\begin{bmatrix} 0.2 & 0.4 \\ 0.3 & 0.5 \end{bmatrix}} \cdot \underset{x}{\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}} + \underset{b}{\begin{bmatrix} 0.3 \\ 0.5 \end{bmatrix}}$$

2×2 $\quad 2 \times 4$ $\quad 2 \times 1$

$$= \begin{matrix} (0.2 \cdot 0 + 0.4 \cdot 0) & (0.2 \cdot 0 + 0.4 \cdot 1) & (0.2 \cdot 1 + 0.4 \cdot 0) & (0.2 \cdot 1 + 0.4 \cdot 1) \\ (0.3 \cdot 0 + 0.5 \cdot 0) & (0.3 \cdot 0 + 0.5 \cdot 1) & (0.3 \cdot 1 + 0.5 \cdot 0) & (0.3 \cdot 1 + 0.5 \cdot 1) \end{matrix}$$

$$z^{(1)} = \begin{bmatrix} 0 & 0.4 & 0.2 & 0.6 \\ 0 & 0.5 & 0.3 & 0.8 \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.3 & 0.7 & 0.5 & 0.9 \\ 0.5 & 1.0 & 0.8 & 1.3 \end{bmatrix}$$

$a^{(1)} = \sigma(z^{(1)})$ using Sigmoid $= \begin{bmatrix} \sigma(0.3) & \sigma(0.7) & \sigma(0.5) & \sigma(0.9) \\ \sigma(0.5) & \sigma(1.0) & \sigma(0.8) & \sigma(1.3) \end{bmatrix}$

$$\frac{1}{1+e^{-z}}$$

$$\sigma(0.3) = \frac{1}{1+e^{-(.3)}} \approx .574$$

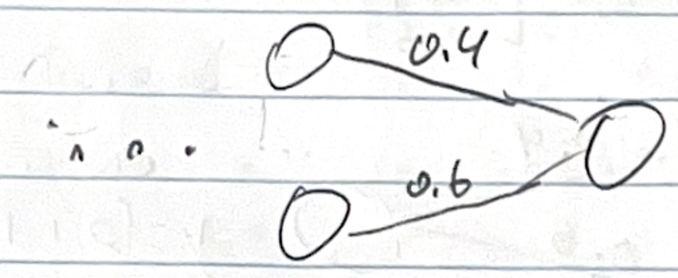
$$\sigma(0.7) = \frac{1}{1+e^{-(.7)}} \approx .668$$

$$\sigma(0.5) = \frac{1}{1+e^{-(.5)}} \approx .622$$

$$\sigma(0.6) = \frac{1}{1+e^{-(.9)}} \approx .646$$

$$a^{(1)} = \begin{bmatrix} 0.574 & 0.668 & 0.622 & 0.711 \\ .622 & .731 & .690 & .786 \end{bmatrix}$$

$$a^{(1)} = \begin{bmatrix} 0.574 & 0.668 & 0.622 & 0.711 \\ 0.622 & 0.731 & 0.690 & 0.786 \end{bmatrix}$$



$$w = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} \qquad b = [0.5]$$

$$Z^{(2)} = W^2 X^2 + b^2$$

$$= \underset{1\times2}{[0.4 \quad 0.6]} \cdot \underset{2\times4}{a^{(1)}} + 0.5$$

$$= (0.4 \cdot 0.574 + 0.6 \cdot 622)\ (0.4 \cdot 668 + 0.6 \cdot 731) - (0.4 \cdot 622 + 0.6 \cdot 690)\ (0.4 \cdot 711 + 0.6 \cdot$$

$$= [0.603 \quad 0.706 \quad 0.663 \quad 0.756] + [0.5$$

$$Z^{(2)} = [1.103 \quad 1.206 \quad 1.163 \quad 1.256]$$

- Apply Sigmoid

$$\frac{1}{1+e^{-(1.103)}} \approx 0.752 \qquad \frac{1}{1+e^{-(1.206)}} \approx 0.770 \qquad \frac{1}{1+e^{-(1.163)}} \approx 0.762 \qquad \frac{1}{1+e^{-(1.256)}} \approx 0.778$$

$$a^{(2)} = [0.752 \quad 0.770 \quad 0.762 \quad 0.778]$$

- Now we can compute the cost using mean squared error.

$$C = \frac{1}{2}\left(a^{(2)} - y_{True}\right)^2 \qquad y = [0 \quad 1 \quad 1 \quad 1]$$

$$= \underset{0.565}{(0.752 - 0)^2} + \underset{0.053}{(0.770 - 1)^2} + \underset{0.056}{(0.762 - 1)^2} + \underset{0.024}{(0.778 - 1)^2}$$

$$= 0.723$$

$$MSE = \frac{1}{n}\sum_{i=2}^{n}\left(a_i^{(2)} - y_i\right)^2 \qquad loss = \frac{1}{2n}\sum_{i=1}^{n}\left(a^{(2)} - y_i\right)^2$$

$$MSE = \frac{1}{4} * 0.723 \approx \boxed{0.181}, \quad Loss = \boxed{0.090}$$

- To simplify derivation we modify MSE to make it easier to do back propogation

Backpropegation using Loss Function $= \frac{1}{2n}\sum\left(a^{(2)} - y_i\right)^2$

Calculate the output layer error defined by $\delta^{(2)}$ and $\delta^{(1)}$

$$\delta^{(2)} = \left(a^{(2)} - y_{True}\right) \cdot \sigma'\left(z^{(2)}\right)$$ derivative of sigmoid $\sigma'(x) = \sigma(x)\cdot(1-\sigma(x))$

$$= [0.752 \quad 0.770 \quad 0.762 \quad 0.778] - \begin{bmatrix} 0 \\ \vdots \end{bmatrix}$$
$$-0 \quad -1 \quad -1 \quad -1$$

$$= [0.752 \quad -.230 \quad -.238 \quad -.222] \cdot \sigma\left(z^{(2)}\right)\cdot\left(1-\sigma\left(z^{(2)}\right)\right)$$

$$= [0.752 \quad -.230 \quad -.238 \quad -.222] \cdot \begin{bmatrix} 0.752 \\ 0.770 \\ 0.762 \\ 0.778 \end{bmatrix} \cdot \begin{bmatrix} 1-.752 = 0.248 \\ 1-.770 = 0.230 \\ 1-.762 = 0.238 \\ 1-.778 = 0.222 \end{bmatrix}$$
$1\times4 \quad * \quad$

$$= [.566 \quad -.177 \quad -.181 \quad -.173] \cdot \begin{bmatrix} 0.248 \\ 0.230 \\ 0.238 \\ 0.222 \end{bmatrix}$$

$$\delta^{(2)} = [.140 \quad -.041 \quad -.043 \quad -.038]$$

- AFTer reviewing, I realized That I should have used The Hadamard Product $\odot$ rather Than doing regular matrix multiplication
  - Luckily since matrices be were of a single dimension, the result is the same.

Now The hidden layer error for $\delta^{(1)}$ $\quad 1 = 0.579\ldots$

$$\delta^{(1)} = \left(W^{(2)}\right)^T \cdot \delta^{(2)} \quad \sigma(z^{(1)}) \cdot \left(1 - \sigma(z)'\right)$$

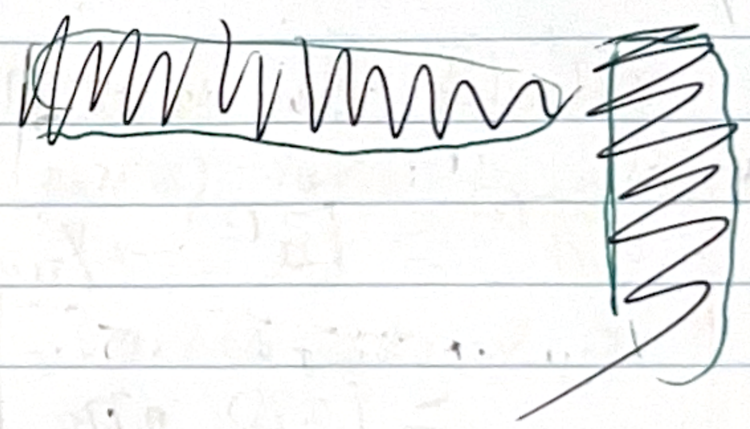
$$= [0.140 \;\; .041 \;\; -.043 \;\; -.038] \cdot \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.574 & 0.668 & 0.622 & 0.711 \\ 0.622 & 0.731 & 0.690 & 0.214 \end{bmatrix} \quad (2)$$

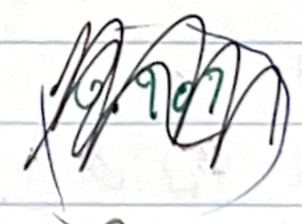
$$= \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} \cdot [0.140 \;\; -.041 \;\; .043 \;\; -.038] \quad \sigma(z^{(1)}) \cdot \begin{bmatrix} 0.426 & 0.332 & 0.378 & 0.289 \\ 0.378 & 0.269 & 0.360 & 0.24 \end{bmatrix} \quad (3)$$

$2 \times 1 = 1 \times 4$ $\qquad 1 - \sigma(z^1)$

$$= \begin{matrix} 0.4 \times 0.140 & 0.4 \cdot (-.041) & 0.4 \cdot (-.043) & 0.4 \cdot (-.038) \\ 0.6 \times 0.140 & 0.6 (-.041) & 0.6 \cdot (-.043) & 0.6 (-.238) \end{matrix}$$

$$(1) = \begin{bmatrix} .056 & -.016 & -.017 & -.015 \\ .084 & -.025 & -.026 & -.023 \end{bmatrix}$$





★ We can now multiply these 3 matrices ($W^{(2)}$), $\sigma(z^{(1)})$, and $(1-\sigma(z)')$ using the hadamard product. First I will 2 and 3 Then multiply that by 1.

$$(1) \times (2) \times (3) = \begin{bmatrix} (.056 \times 0.574 \times 0.426) & (-.056 \times 0.668 \times 0.332) \ldots \\ (.084 \times 0.622 \times 0.378) & (-.025 \times 0.731 \times 0.269) \ldots \end{bmatrix}$$

$$\delta^{(1)} = \begin{bmatrix} .0135 & -.0036 & .0040 & .0031 \\ .0197 & .0048 & .0055 & .0038 \end{bmatrix}$$

- Calculating the gradients for output layer. and hidden layer.

$\eta = 0.09$

Now that we have the loss functions for each layer we can calculate the gradient and update the biases.

$$\frac{\partial \cos t}{\partial w^{(2)}} = \frac{1}{n} a^{(1)} \left(\delta^{(2)}\right)^T$$

$$= \frac{1}{4} \begin{bmatrix} 0.574 & 0.668 & 0.622 & 0.711 \\ 0.622 & 0.731 & 0.690 & 0.780 \end{bmatrix} \times \begin{bmatrix} 0.140 \\ -0.041 \\ -0.043 \\ -0.038 \end{bmatrix}$$

2×4 ✓ 4×1

$$= \frac{1}{4} \begin{bmatrix} -.00079 \\ -.00243 \end{bmatrix}$$

$$\text{gradient}[w^{(2)}] = \begin{bmatrix} -.00020 \\ -.00060 \end{bmatrix}$$

* output nox to match shape of weights.

- To update the gradient for $bias^{(2)}$ we use: $\frac{\partial \cos t}{\partial b^{(2)}} = \frac{1}{n} \sum_{j=1} \delta^{(2)}_j$

$$\sum [0.140 - 0.041 - 0.043 - 0.038) = \frac{0.018}{4} = \boxed{0.0045 = b^{(2)}} \text{ gradient}$$

$$\frac{\partial \cos t}{\partial w^{(1)}} = \frac{1}{n} \delta^{(1)} \cdot x^T$$

$$= \delta = \begin{bmatrix} .0135 & -.0036 & .0040 & .0031 \\ .0197 & .0048 & .0055 & .0038 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

$$= \frac{1}{4} \begin{bmatrix} -0.0071 & -0.0067 \\ -0.0093 & -0.0086 \end{bmatrix}$$

$$\frac{\partial \cos t}{\partial w^{(1)}} = \begin{bmatrix} -0.0018 & -0.0017 \\ -0.0023 & -0.0022 \end{bmatrix}$$

$$\frac{\partial \cos t}{\partial b^{(1)}} = \frac{1}{n} \sum \delta^{(1)} = (.0135) + (-.0036) + (.0040) + (.0031) = \begin{bmatrix} 0.003 \\ 0.0050 \end{bmatrix} \frac{1}{4} = \begin{bmatrix} 0.0008 \\ 0.0014 \end{bmatrix}$$

$.0197 + .0048 + .0055 + .0038$

Update the parameters : $\theta_{new} = \theta_{old} - lr \times \frac{\partial L(\theta)}{\partial \theta}$

$$W^{(2)} \leftarrow W^{(2)} - \eta \frac{\partial L}{\partial W^{(2)}}$$

$lr = 0.5$

$$W^{(2)}_{new} = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} - 0.5 \times \begin{bmatrix} -.0002 \\ -.0006 \end{bmatrix} = \begin{bmatrix} 0.4001 \\ 0.6003 \end{bmatrix}$$

$$b^{(2)} \leftarrow b^{(2)} - \eta \frac{\partial L}{\partial b^{(2)}}$$

$$= 0.5 - 0.5 \times 0.0045 = 0.4978$$

$$W^{(1)} \leftarrow W^{(1)} - \eta \frac{\partial L}{\partial W^{(1)}}$$

$$W^{(1)} = \begin{bmatrix} 0.2 & 0.4 \\ 0.3 & 0.5 \end{bmatrix} - 0.5 \times \begin{bmatrix} -0.00178 & -0.0017 \\ -0.0023 & -0.0022 \end{bmatrix} = \begin{bmatrix} 0.2009 & 0.4008 \\ 0.3012 & 0.5011 \end{bmatrix}$$

$$b^{(1)} \leftarrow b^{(1)} - \eta \frac{\partial L}{\partial b^{(1)}}$$

$$= \begin{bmatrix} 0.3 \\ 0.5 \end{bmatrix} - 0.5 \times \begin{bmatrix} 0.0008 \\ 0.0014 \end{bmatrix} = \begin{bmatrix} 0.2996 \\ 0.4993 \end{bmatrix}$$

Read for another forward pass.

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1