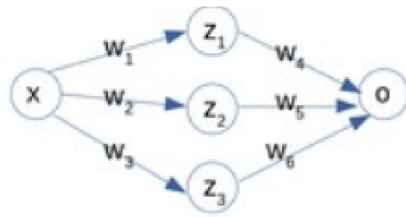
1.1 (2 points) Consider the neural network shown in the figure below.



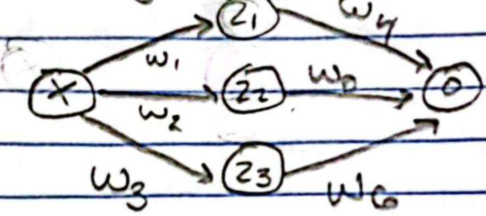
The weight matrix, W, is: $[1, 1, -1, 0.5, 1, 2]^T$. Assume that the hidden layer uses RelU and the output layer uses Sigmoid activation function. Assume squared error. The input $x = 4$, and the output $y = 0$.

Recall that RelU is defined as a function, $f(x) = \max(0, x)$. Its derivative is $f'(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$

Squared error is defined as $E(y, \hat{y}) = (y - \hat{y})^2$.

The partial derivative of E with respect to $\hat{y}$ is $-2(y - \hat{y})$.

a) Use foward propogation to compute the projected outcome



Given $x = 4$

$w_1 = 1,\ w_2 = 1,\ w_3 = -1,\ w_4 = 0.5,\ w_5 = 1,\ w_6 = 2$

$z_1 = x w_1 = 4$

$z_2 = x w_2 = 4$ $\Rightarrow$ hidden layers (z_1, z_2, z_3)

$z_3 = x w_3 = -4$ Activation function is used to get output of hidden layer $\Rightarrow$ $z_1 = 4$

$z_2 = 4$

$z_3 = 0$ (neg. val)

output $\Rightarrow (z_1 w_4) + (z_2 w_5) + (z_3 w_6) \Rightarrow (4(0.5)) + (4(1)) + (0(2)) =$

$4 + 2 + 0 = 6$

output layer, activation layer

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

$\sigma(6) = \frac{1}{1+e^{-6}} = \boxed{0.998}$

b) Now we calculate the loss

$\text{Loss} = (y - \hat{y})^2$

$L = (y - a)^2 \quad (\because \hat{y} = a)$

$L = (0 - 0.998)^2 = 0.996$

c) for back propogation we take a look at computation graph

$x \Rightarrow \text{hidden} \Rightarrow \text{Relu} \Rightarrow \text{output} \Rightarrow \sigma(x) \Rightarrow 0$

$x \Rightarrow (z_1, z_2, z_3) \to \text{Relu} \to z \to a \to (y-\hat{y})$

$$\Rightarrow \frac{\partial L}{\partial a} = -2(y-a)$$
$$= -2(0-0.998)$$
$$= 2(0.998) = 1.996$$

$$\frac{\partial a}{\partial z} = \frac{\partial(\sigma(z))}{\partial z} = a(1-a) = 0.998(1-0.998) = 0.001996$$

$$\frac{\partial L}{\partial z} = \frac{\partial L}{\partial a} \times \frac{\partial a}{\partial z} = 1.996 \times 0.001996 = 0.00398$$

$$\frac{\partial L}{\partial w_4} = \frac{\partial L}{\partial z} \times \frac{\partial z}{\partial w_4} = (0.00398)(z_1) = 0.0159$$

$$\frac{\partial L}{\partial w_5} = \frac{\partial L}{\partial z} \times \frac{\partial z}{\partial w_5} = 0.00398\,(z_2) = 0.0159$$

$$\frac{\partial L}{\partial w_6} = \frac{\partial L}{\partial z} \times \frac{\partial z}{\partial w_6} = 0.00398(z_3) = 0$$

$$z = \underset{z_2}{z_1'}\, w_4 + \underset{z_3}{z_2'}\, w_5 + \underset{z_4}{w_3'}\, w_6$$

$$\frac{\partial L}{z_1} = w_4 \quad \frac{\partial L}{z_2} = w_5 \quad \frac{\partial L}{z_3} = w_6$$

$$\frac{\partial L}{z_1} = \frac{\partial L}{\partial z} \times \frac{\partial z}{\partial w_4} = 0.00398(0.5) = 0.00199$$

$$\frac{\partial L}{z_2} = \frac{\partial L}{\partial z} \times \frac{\partial z}{\partial w_5} = 0.00398(1) = 0.00398$$

$$\frac{\partial L}{z_3} = \frac{\partial L}{\partial z} \times \frac{\partial z}{\partial w_6} = 0.00398(2) = 0.00796$$

$z_1' = relu(z_1) = \frac{dz_1'}{dz_1} = 1$

$z_2' = relu(z_2) = \frac{dz_2'}{dz_2} = 1$

$z_3' = relu(z_3) = \frac{dz_3'}{dz_3} = 1$

so $\frac{d1}{dz_1} = \frac{dL}{dx_1'} \times \frac{dz_1'}{dz_1} = 0{,}00199(1) = 0{,}00199$

$\frac{d2}{dz_2} = 0{,}00398 \times \frac{dL}{dz_3} = 0{,}00796$

$z_1 = w_1 x = 0{,}00199(4) = 0{,}00796$
$z_2 = w_2 x = 0{,}00398(4) = 0{,}01592$
$z_3 = w_3 x = 0{,}00796(4) = 0{,}03184$

$w_1 = w_1 - \alpha \frac{dL}{dw_1}$ $\quad \alpha$ = learning rate

$\Rightarrow \alpha = 1$

$w_1 = 1 - (1)(0{,}00796) = 0{,}99204$
$w_2 = 1 - (1)(0{,}01592) = 0{,}98408$
$w_3 = -1 - (1)(0{,}03184) = -1{,}03884$
$w_4 = 0{,}5\ (1)(0{,}0159) = 0{,}46816$
$w_5 = 1 - (1)(0{,}0159) = 0{,}96818$
$w_6 = 2 - (1)(0) = 2$

$\Rightarrow$ Forward computation

$z_1 = x w_1 = 4(0{,}99204) = 3{,}96816$
$z_2 = x w_2 = 4(0{,}98408) = 3{,}93632$
$z_3 = x w_3 = 4(-1{,}0384) = -4{,}1538$

$z_1' = 3{,}9616$
$z_2' = 3{,}93632$
$z_3' = 0$

$z \Rightarrow z_1' w_4 + z_2' w_5 + z_3' w_6 =$

$3{,}9616(0{,}5) + 3{,}93632(1) + 0 \times 2 =$
$1{,}9808 + 3{,}93632 + 0 = 5{,}91712$
$\boxed{\therefore z = 5{,}917}$

d) $a = \sigma(z) = \frac{1}{1+e^{-z}} = 0{,}9973142878$
$L(y, a) = (y - a)^2 =$
$(0 - 0{,}9973142878)^2 = 0{,}9946$
$\boxed{a = 0{,}997}$

e)
loss 1 = 0,9996
loss 2 = 0,9946
So (loss 2 < loss 1) and output after update is closer to target 0

1.2) Tan Chapter 4, Question 14,15

Question 14: For each of the Boolean functions given below, state whether the problem is linearly separable.
(a) A AND B AND C
Answer: Yes
(b) NOT A AND B
Answer: Yes
(c) (A OR B) AND (A OR C)
Answer: Yes

(d) (A XOR B) AND (A OR B)
Answer: No

Question 15:

a) Demonstrate how the perceptron model can be used to represent the AND and OR functions between a pair of Boolean variables.
    i) Answer: Let x1 and x2 be a pair of Boolean variables and y be the output. For AND function, a possible perceptron model is:

$$y = sgn\left[x_1 + x_2 - 1.5\right].$$

For OR function, a possible perceptron model is:

$$y = sgn\left[x_1 + x_2 - 0.5\right].$$

b) Comment on the disadvantage of using linear functions as activation functions for multilayer neural networks.
    i) Answer: Multilayer neural networks are useful for modeling nonlinear relationships between the input and output attributes. However, if linear functions are used as activation functions (instead of sigmoid or hyperbolic tangent function), the output is still a linear combination of its input attributes. Such a network is just as expressive as a perceptron.

1.3) Consider a dataset that has 8 predictors. You train a neural network with 3 hidden layers and an output layer that predicts a continuous value (a regression problem). The first hidden layer has 16 neurons, the second has 8 neurons, and the third has 4 neurons. In this network, how many total parameters will you have?

The amount of parameters is computed with the following formula, ((I x H) + (H x O) + (H + O)) = ((8 x 28) + (28 x 1) + (28 + 1)) = 281 parameters