Assignment 1
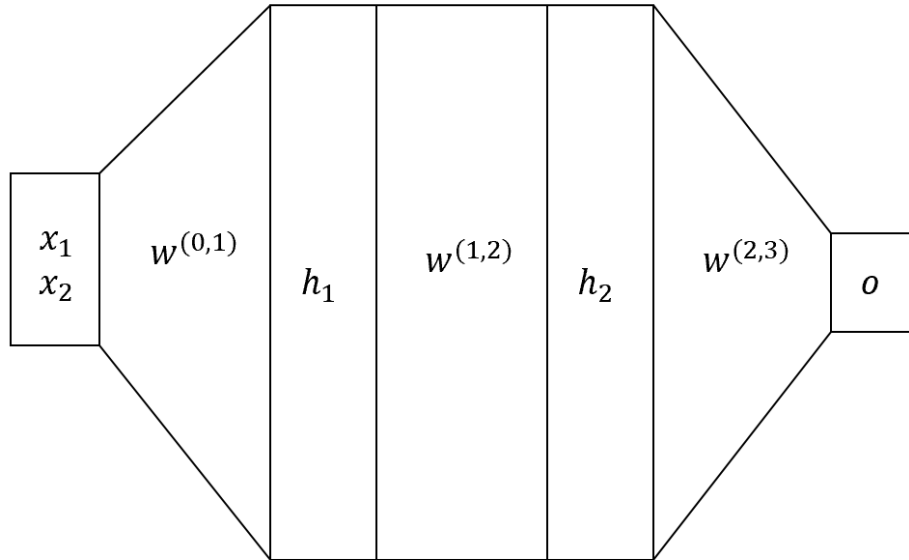
LI XINYA (G2004358J)

The neural network structure:

$$sigmoid(x) = \frac{1}{1 + \exp(-x)}$$

$h_1 = sigmoid\left(w^{(0,1)} \cdot x + b^{(0,1)}\right) \qquad (1*10) = (1*2) \cdot (2*10) + (1*10)$

$h_2 = sigmoid(w^{(1,2)} \cdot h_1 + b^{(1,2)}) \quad (1*10) = (1*10) \cdot (10*10) + (1*10)$

$o = h_2 \cdot w^{(2,3)} + b^{(2,3)} \qquad\qquad (1*1) = (1*10) \cdot (10*1) + (1*1)$

$$Loss = (y\_pred - y\_true)^2$$

Back propagation:

$$z_1 = w^{(0,1)} \cdot x + b^{(0,1)}$$
$$z_2 = w^{(1,2)} \cdot h_1 + b^{(1,2)}$$
$$z_3 = w^{(2,3)} \cdot h_2 + b^{(2,3)}$$

$$\frac{do}{dw^{(2,3)}} = \frac{do}{dz_3}\frac{dz_3}{dw^{(2,3)}}$$
$$\frac{dh_2}{dw^{(1,2)}} = \frac{dh_2}{dz_2}\frac{dz_2}{dw^{(1,2)}}$$
$$\frac{dh_1}{dw^{(0,1)}} = \frac{dh_1}{dz_1}\frac{dz_1}{dw^{(0,1)}}$$

The gradients result from pytorch autograd and scratch are the same. And the mean square error between them is 0 (as shown in the code).