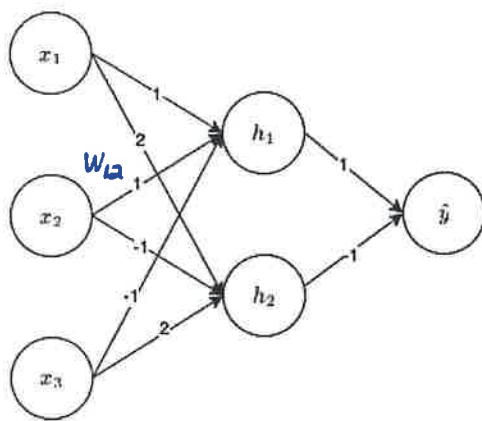1. Perform one iteration of gradient descent on w_{12}. (The weight going to h_1 from x_2.) Use input vector $x = (1, 1, 1)$, ReLU as our nonlinearity, $L = (y - \hat{y})^2$ as our loss, $y = 1$ as our current label and $\eta = 0.1$ as our learning rate. (10 points)

$$h_1 = relu(1 + 1 - 1) = 1$$

$$h_2 = relu(2 - 1 + 2) = 3$$

$$\hat{y} = 1 \cdot 1 - 1 \cdot 3 = -2$$

$$L = (y - \hat{y})^2 \qquad \frac{\partial L}{\partial \hat{y}} = -2(y - \hat{y}) = -2(1 - -2) = -6$$

$$\hat{y} = w_1^2 h_1 + w_2^2 h_2 \qquad \frac{\partial \hat{y}}{\partial h_1} = w_1^2 = 1$$

$$h_1 = relu(1x_1 + 1x_2 - 1x_3) \qquad \frac{\partial h_1}{\partial w_{12}} = 1 \cdot x_2 = 1 \cdot 1 = 1$$

$$\frac{\partial L}{\partial w_{12}} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial h_1} \frac{\partial h_1}{\partial w_{12}} = -6 \cdot 1 \cdot 1 = -6$$

$$w_{12} = w_{12} - \eta \frac{\partial L}{\partial w_{12}} = 1 - 0.1(-6) = 1.6 \quad \checkmark$$