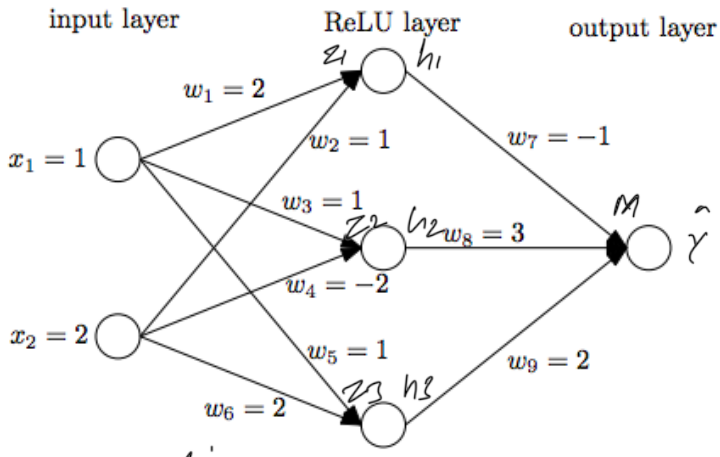
Assume the artificial neural network below, with mean square error loss and true output of 3.

All activation functions are ReLU.

Perform one round of forward propagation and calculate the loss.

Then perform a Stochastic Gradient Descent update using back propagation to get the new values of weights w_6 and w_9 with learning rate 0.01.



Forward propagation

$z_1 = x_1 w_1 + x_2 w_2 = 2 + 2 = 4$

$h_1 = ReLU(z_1) = 4$

$z_2 = x_1 w_3 + x_2 w_4 = 1 + (-4)$

$h_2 = ReLU(z_2) = 0$

$z_3 = x_1 w_5 + x_2 w_6 = 1 + 2 \cdot 2 = 5$

$h_3 = ReLU(z_3) = 5$

$M = h_1 w_7 + h_2 w_8 + h_3 w_9 = 4 \times -1 + 0 \cdot 3 + 5 \cdot 2 = 6$

$\hat{y} = ReLU(M) = 6$

$Loss = (\hat{y} - y)^2 = (6-3)^2 = 9$.

Back propagation

$$\frac{\partial Loss}{\partial w_4} = \frac{\partial Loss}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial M} \cdot \frac{\partial M}{\partial w_4}$$

$$= 2(\hat{y} - y) \cdot 1 \cdot h_3$$

$$= 2 \times 3 \times 1 \times 5$$

$$= 30$$

$$w_4^* = w_4 - \eta \cdot \frac{\partial Loss}{\partial w_4} = 2 - 0.01 \times 30$$

$$= 2 - 0.3$$

$$= 1.7$$

$$\frac{\partial Loss}{\partial w_6} = \frac{\partial Loss}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial M} \cdot \frac{\partial M}{\partial h_3} \cdot \frac{\partial h_3}{\partial z_3} \cdot \frac{\partial z_3}{\partial w_6}$$

$$= 6 \cdot 1 \cdot 2 \cdot 1 \cdot 2$$

$$= 24.$$

$$w_6^* = w_6 - \eta \frac{\partial Loss}{\partial w_6}$$

$$= 2 - 0.01 \times 24$$

$$= 2 - 0.24$$

$$= 1.76.$$