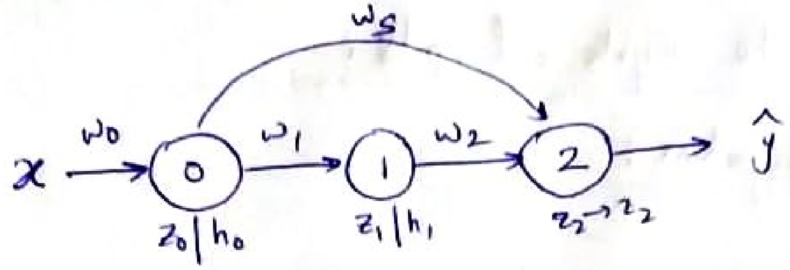

Forward Propagation :

$$L = |y - \hat{y}|$$

$$\frac{\partial L}{\partial \hat{y}} = \begin{cases} \pm 1 & \\ 0 & y = \hat{y} \end{cases}$$

$$z_0 = x w_0 + b_0 \qquad [b_0 = 0]$$

$$h_0 = \begin{cases} 0 & z_0 < 0 \\ z_0 & \text{else} \end{cases}$$

$$z_1 = h_0 w_1 + b_1 \qquad [b_1 = 0]$$

$$h_1 = \begin{cases} 0 & z_1 < 0 \\ z_1 & \text{else} \end{cases}$$

$$z_2 = h_1 w_2 + h_0 w_s + b_2 \qquad [b_2 = 0]$$

$$\hat{y} = z_2$$

Backpropagation :

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_2} = \frac{\partial L}{\partial \hat{y}} \cdot 1 \cdot h_1 = h_1 \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_1} = \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_s} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial z_2}{\partial w_s} = \frac{\partial L}{\partial \hat{y}} \cdot 1 \cdot h_0 = h_0 \cdot \frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1} = \begin{cases} h_0 \cdot \frac{\partial L}{\partial h_1} & z_1 \geq 0 \\ 0 & z_1 < 0 \end{cases}$$

$$\frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial h_1} \cdot \frac{\partial h_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial h_0} + \frac{\partial L}{\partial z_2} \cdot \frac{\partial z_2}{\partial h_0} = \begin{cases} \frac{\partial L}{\partial h_1} w_1 + \frac{\partial L}{\partial \hat{y}} w_s & \text{if } z_1 \geq 0 \\ \frac{\partial L}{\partial \hat{y}} \cdot w_s & \text{if } z_1 < 0 \end{cases}$$

$$\frac{\partial L}{\partial w_0} = \frac{\partial L}{\partial h_0} \cdot \frac{\partial h_0}{\partial z_0} \cdot \frac{\partial z_0}{\partial w_0} = \begin{cases} x \cdot \frac{\partial L}{\partial h_0} & \text{if } z_0 \geq 0 \\ 0 & \text{if } z_0 < 0 \end{cases}$$

After Using the weights and inputs:

Forward propagation:

$z_0 = x_1 \cdot w_0$
$= 1 \times 0.5 = 0.5$

$h_0 = 0.5$

$z_1 = 0.5 \times 0.5 = 0.25$

$h_1 = 0.25$

$z_2 = 0.25 \times 0.5 + 0.5 \times 0.5$
$= 0.125 + 0.250 = 0.375$

$\hat{y} = 0.375$

$\frac{\partial L}{\partial \hat{y}} = -1$, $\frac{\partial L}{\partial w_2} = -1 \cdot 0.25 = -0.25$

$\frac{\partial L}{\partial h_1} = 0.5 \times (-1) = -0.5$

$\frac{\partial L}{\partial w_3} = h_0 \cdot \frac{\partial L}{\partial \hat{y}} = 0.5 \times (-1) = -0.5$

$\frac{\partial L}{\partial w_1} = 0.5 \times 0.5 \times -1 = -0.25$

$\frac{\partial L}{\partial h_0} = -0.5 \times 0.5 + (-1) \cdot 0.5$
$= -0.75$

$\frac{\partial L}{\partial w_0} = 1 \cdot (-0.75) = -0.75$

$Loss = 3.375$

$w_0 = w_0 - \alpha \frac{\partial L}{\partial w_0} = 0.5 - (-0.75) = 1.25$

$w_1 = w_1 - \alpha \frac{\partial L}{\partial w_1} = 0.5 - (-0.25) = 0.75$

$w_2 = w_2 - \alpha \frac{\partial L}{\partial w_2} = 0.5 - (-0.25) = 0.75$

$w_3 = w_3 - \alpha \frac{\partial L}{\partial w_3} = 0.5 - (-0.5) = 1.0$

$$\frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial z_2} = -1 \quad \Rightarrow \quad b_2 = 0-(-1) = 1$$

$$\frac{\partial L}{\partial b_1} = \frac{\partial L}{\partial z_1} = -0.5 \quad \Rightarrow \quad b_1 = 0-(-0.5) = 0.5$$

$$\frac{\partial L}{\partial b_0} = \frac{\partial L}{\partial z_0} = -0.75 \quad \Rightarrow \quad b_0 = 0-(-0.75) = 0.7$$

After updating weights & biases;

$$z_0 = x w_0 + b_0 = (1)(1.25) + 0.75 = 2$$

$$h_0 = \text{Relu}(z_0) = 2$$

$$z_1 = h_0 w_1 + b_1 = 2(0.75) + 0.5 = 2$$

$$h_1 = \text{Relu}(z_1) = 2$$

$$z_2 = w_2 h_1 + w_3 h_0 + b_2$$

$$= (0.75)(2) + (1)(2) + 1 = 4.5$$

$$h_2 = 4.5 = \hat{y}$$

$$\text{loss} = |y - \hat{y}| = |-3-4.5| = 7.5$$

Question: What difference does skip connection make?

Answer: Skip connection helps in solving vanishing gradient problem. If there are many layers, the magnitude of the gradient may shrink and become very less by the time it reaches the first layer. Skip connection helps to prevent this shrinkage by directly propagating the gradient from last layer.

Feedback: Excercise 2.1 is a good pen and paper revision of what we have studied so far ~~in~~ regarding multi-layer perceptrons.