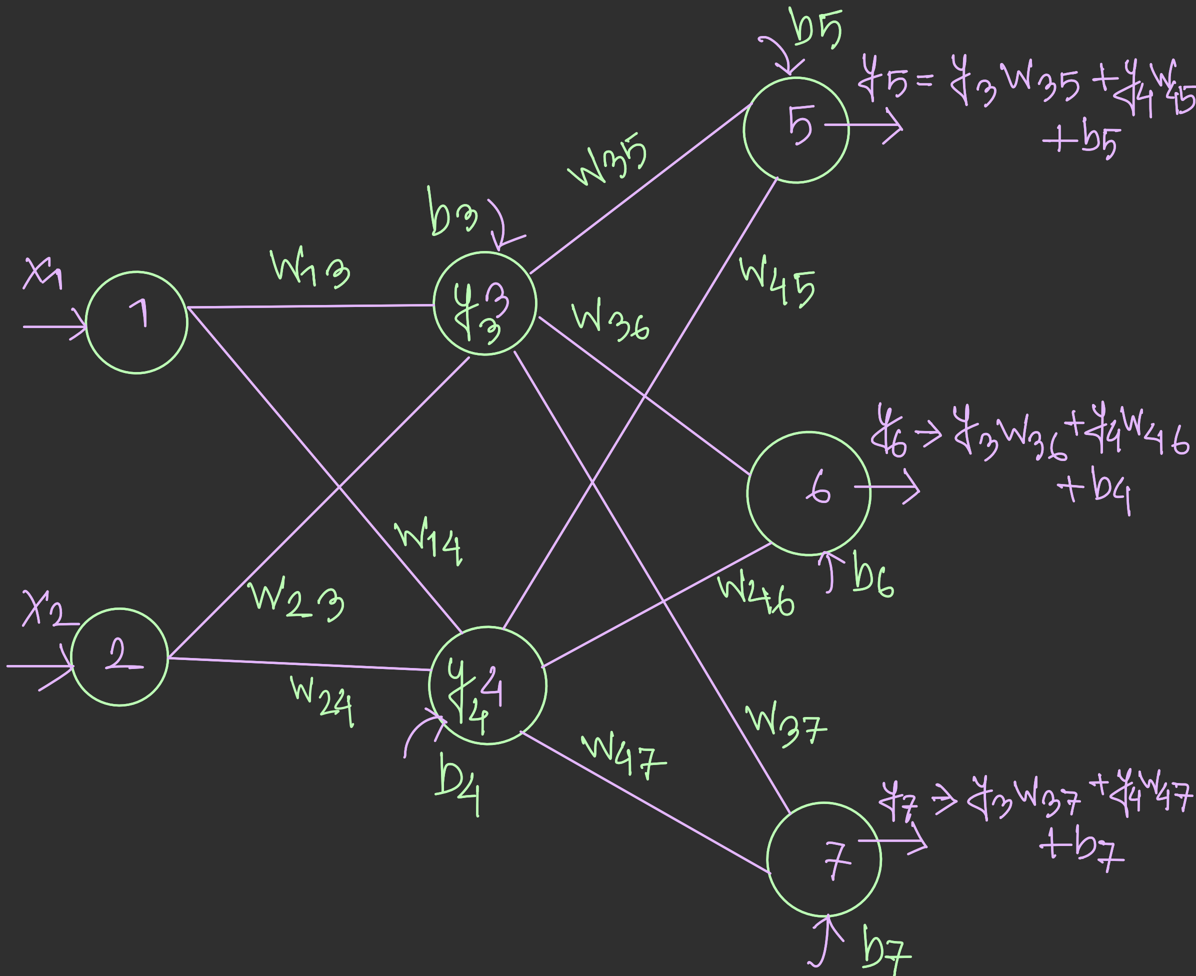
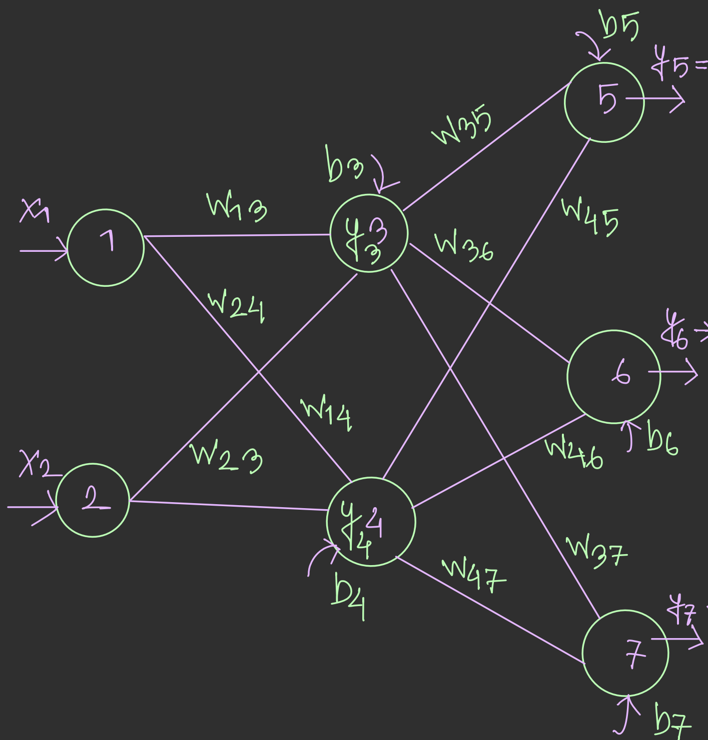
R2R

→ Parameter update without activation function



x_1
x_2
1
2
y_3 3
y_4 4
5
6
7
b_3
b_4
b_5
b_6
b_7
w_{13}
w_{14}
w_{23}
w_{24}
w_{35}
w_{36}
w_{37}
w_{45}
w_{46}
w_{47}
$y_5 = y_3 w_{35} + y_4 w_{45} + b_5$
$y_6 \rightarrow y_3 w_{36} + y_4 w_{46} + b_4$
$y_7 \rightarrow y_3 w_{37} + y_4 w_{47} + b_7$

⇒ To get final output y_5, y_6, y_7 are passed to an activation function (Softmax here).



$y_5 = y_3 W_{35} + y_4 W_{45} + b_5$ | $P_5 = \dfrac{e^{y_5}}{e^{y_5} + e^{y_6} + e^{y_7}}$

$y_6 \rightarrow y_3 W_{36} + y_4 W_{46} + b_4$ | $P_6 = \dfrac{e^{y_6}}{e^{y_5} + e^{y_6} + e^{y_7}}$

$y_7 \rightarrow y_3 W_{37} + y_4 W_{47} + b_7$ | $P_7 = \dfrac{e^{y_7}}{e^{y_5} + e^{y_6} + e^{y_7}}$

⇒ For now let's consider we are not using any activation function. Y_5, Y_6, Y_7 are the predicted output.

⇒ Considering the loss function is:

$$L = \frac{1}{2}\left\{ (ob_5 - y_5)^2 + (ob_6 - y_6)^2 + (ob_7 - y_7)^2 \right\}$$

→ Differentiate the loss function with respect to all the outputs.

$$\frac{dLoss}{dy_5} = \frac{1}{2} * 2(ob_5 - y_5)(-1) = -(ob_5 - y_5) = -Err_5$$

↳ this tells us how change in y_5 affect the Loss. (ultimately output)

$$\frac{dLoss}{dy_6} = -(ob_6 - y_6) = -Err_6.$$

$$\frac{dLoss}{dy_7} = -(ob_7 - y_7) = -Err_7.$$

⇒ How change in W_{35} changes the Loss?

$$\frac{dLoss}{dW_{35}} = \frac{dLoss}{dy_5} * \frac{dy_5}{dW_{35}} \quad \text{(chain rule)}$$

$y_5 = y_3 W_{35} + y_4 W_{45} + b_5$

After differentiation;

y_3

$$= -Err5 \cdot y_3$$

To update the weight (W_{35})

$$W_{35} = W_{35} - \eta \frac{dLoss}{dW_{35}}$$

$$= W_{35} + \eta \, Err_5 \, y_3$$

Similarly we can update the weight for W_{46}

$$\frac{dLoss}{dW_{46}} = \frac{dLoss}{dy_6} * \frac{dy_6}{dW_{46}} = -Err_6 * y_4$$

To update W_{46}:

$$W_{46} = W_{46} + Err_6 \, y_4 .$$

What about updating b_6?

$$\frac{dLoss}{db_6} = \frac{dLoss}{dy_6} * \frac{dy_6}{b_6}$$

$$= -Err_6$$

$$b_6 = b_6 + \eta \, Err_6 .$$

⇒ How do we update hidden layer's output (y_3, y_4) to reduce the loss? $y_3 = Relu(x_1 w_{13} + x_2 w_{23} + b_3)$

$$\frac{dLoss}{dy_3} = \frac{dLoss}{dy_5} * \frac{dy_5}{dy_3} + \frac{dLoss}{dy_6} * \frac{dy_6}{dy_3} + \frac{dLoss}{dy_7} * \frac{dy_7}{dy_3}$$

$$= (-Err_5 * W_{35}) + (-Err_6 * W_{36}) + (-Err_7 * W_{37})$$

Similarly $\frac{dLoss}{dy_4}$ | $y_4 = Relu(x_1 w_{14} + x_2 w_{24} + b_4)$

$$= (-Err_5 \cdot W_{45}) + (-Err_6 \cdot W_{46}) + (-Err_7 \cdot W_{47})$$

⇒ Now we can proceed to update $w_{13}, w_{23}, w_{14} \dots$

$$\frac{dLoss}{dW_{13}} = \frac{dLoss}{dy_3} * \frac{dy_3}{dW_{13}}$$

$$= \{(-Err_5 * W_{35}) + (-Err_6 * W_{36}) + (-Err_7 * W_{37})\} * Relu'(x_1 w_{13} + x_2 w_{23} + b_3) x_1$$

To update W_{13}

$$W_{13} = W_{13} - \eta \left\{ (-Err_5 * W_{35}) + (-Err_6 * W_{36}) + (-Err_7 * W_{37}) \right\} * Relu'(x_1 W_{13} + x_2 W_{23} + b_3) x_1$$

$\Rightarrow$ Similarly to update W_{24}

$$\frac{dLoss}{dW_{24}} = \frac{dLoss}{dy_4} * \frac{dy_4}{dW_{24}}$$

$$= \left\{ (-Err_5 W_{45}) + (-Err_6 W_{46}) + (Err_7 W_{47}) \right\} \cdot R'(x_1 W_{14} + x_2 W_{24} + b_4) \cdot x_2$$

$$W_{24} = W_{24} - \eta \frac{dLoss}{dW_{24}}$$

⇒ Now let's pass the final output (y_5, y_6, y_7) through activation (Relu) function.

⇒ Considering the loss function is:

$$L = \frac{1}{2}\left\{ (ob_5 - y_5)^2 + (ob_6 - y_6)^2 + (ob_7 - y_7)^2 \right\}$$

→ Differentiate the loss function with respect to all the outputs.

$$\frac{dLoss}{dy_5} = \frac{1}{2} * 2(ob_5 - y_5)(-1) = -(ob_5 - y_5) = -Err_5$$

↳ this tells us how change in y_5 affect the Loss. (ultimately output)

$$\frac{dLoss}{dy_6} = -(ob_6 - y_6) = -Err_6.$$

$$\frac{dLoss}{dy_7} = -(ob_7 - y_7) = -Err_7.$$

⇒ Now to update W_{35}

$$\frac{dLoss}{dW_{35}} = \frac{dLoss}{dy_5} * \frac{dy_5}{dW_{35}}$$

$$= -\underbrace{Err_5 * Relu'(y_3 W_{35} + y_4 W_{45} + b_5)}_{\delta_5 = Err_5 y_5'} * y_3$$

$$= -\delta_5 y_3$$

$$\therefore W_{35} = W_{35} - \eta * (-\delta_5 y_3)$$

$$= W_{35} + \eta \delta_5 y_3$$

Similarly we can compute the rest:

For instance;

$$\frac{dLoss}{dW_{47}} = -\delta_7 y_4$$

$$\therefore W_{47} = W_{47} + \eta \delta_7 y_4$$

Here δ_7

$Err_7 y_7'$

$\Rightarrow (Ob_7 - y_7) Relu'(y_3 W_{37} + y_4 W_{47} + b_7)$

⇒ Similarly

$$\frac{dLoss}{db_6} = -\delta_6$$

$\hookrightarrow Err_6 * y'_6$

$$\therefore b_6 = b_6 + \delta_6$$

P.T.O

⇒ How do we update hidden layer's output (y_3, y_4) to reduce the loss? $y_3 = Relu(x_1 w_{13} + x_2 w_{23} + b_3)$

$$\frac{dLoss}{dy_3} = \frac{dLoss}{dy_5} * \frac{dy_5}{dy_3} + \frac{dLoss}{dy_6} * \frac{dy_6}{dy_3} + \frac{dLoss}{dy_7} * \frac{dy_7}{dy_3}$$

$$= -\delta_5 w_{35} + (-\delta_6 w_{36}) + (-\delta_7 w_{37})$$

Similarly $\frac{dLoss}{dy_4}$ | $y_4 = Relu(x_1 w_{14} + x_2 w_{24} + b_4)$

$$\Rightarrow -\delta_5 w_{45} + (-\delta_6 w_{46}) + (-\delta_7 w_{47})$$

⇒ Now we can proceed to update $w_{13}, w_{23}, w_{14} \dots$

$$\frac{dLoss}{dw_{13}} = \frac{dLoss}{dy_3} * \frac{dy_3}{dw_{13}}$$

$$= \left[(-\delta_5 w_{35}) + (-\delta_6 w_{36}) + (-\delta_7 w_{37})\right] Relu'(x_1 w_{13} + x_2 w_{23} + b_3)\, x_1$$

$$= -\underbrace{\left[\delta_5 w_{35} + \delta_6 w_{36} + \delta_7 w_{37}\right] y_3'}_{\delta_3}\, x_1$$

$$= -\delta_3 x_1$$

$$\therefore w_{13} = w_{13} - \eta(-\delta_3 x_1)$$

$$= w_{13} + \eta \delta_3 x_1$$

Similarly w_{24};

$$w_{24} = w_{24} + \eta \delta_4 x_2$$

$$\delta_4 = (\delta_5 w_{45} + \delta_6 w_{46} + \delta_7 w_{47}) y_4'$$

$$b_3 = b_3 + \eta \delta_3$$