CS 797O: Neural Nets and Deep Learning
Unit 2: Assignment 2- Backpropagation

Abu Tyeb Azad
WSU ID: Q688C867

February 17, 2023

1 Training AND

1.1 Part 1

1.1.1 Training Input, Target, Parameters, Hyperparameters

Training inputs, $x_1 = 0$, $x_2 = 1$
Target, $t_1 = 1$
Hyperparameter: Learning rate, $\eta = 0.25$
Parameters:
$w_{z_1x_1} = 0.7$, $w_{z_1x_2} = -0.2$, $w_{z_10} = 0.4$
$w_{z_2x_1} = -0.4$, $w_{z_2x_2} = 0.3$, $w_{z_20} = 0.6$
$w_{yz_1} = 0.5$, $w_{yz_2} = 0.1$, $w_{y_0} = -0.3$

1.1.2 Forward Propagation

$$
\begin{aligned}
u_{z_1} &= w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10}.1 \\
&= w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10} \\
&= (0.7)0 + (-0.2)1 + 0.4 \\
&= 0 - 0.2 + 0.4 \\
&= 0.2
\end{aligned}
$$

$\boldsymbol{u_{z_1} = 0.2}$

$$
\begin{aligned}
x_{z_1} = \sigma_{z_1}(u_{z1}) &= \frac{1}{1+e^{-u_{z_1}}} \\
&= \frac{1}{1+e^{-0.2}} \\
&= \frac{1}{1+0.8187} \\
&= \frac{1}{1.8187} \\
&= 0.5498 \\
&= 0.55
\end{aligned}
$$

$\boldsymbol{x_{z_1} = 0.55}$

$$\begin{aligned} u_{z_2} &= w_{z_2 x_1} x_1 + w_{z_2 x_2} x_2 + w_{z_2 0}.1 \\ &= w_{z_2 x_1} x_1 + w_{z_2 x_2} x_2 + w_{z_2 0} \\ &= (-0.4)0 + (0.3)1 + 0.6 \\ &= 0 + 0.3 + 0.6 \\ &= 0.9 \end{aligned}$$

$\boldsymbol{u_{z_2} = 0.9}$

$$\begin{aligned} x_{z_2} = \sigma_{z_2}(u_{z}2) &= \frac{1}{1 + e^{-u_{z_2}}} \\ &= \frac{1}{1 + e^{-0.9}} \\ &= \frac{1}{1 + 0.4066} \\ &= \frac{1}{1.4066} \\ &= 0.7109 \\ &= 0.71 \end{aligned}$$

$\boldsymbol{x_{z_2} = 0.71}$

$$\begin{aligned} u_y &= w_{y z_1} x_{z_1} + w_{y z_2} x_{z_2} + w_{y_0}.1 \\ &= w_{y z_1} x_{z_1} + w_{y z_2} x_{z_2} + w_{y_0} \\ &= (0.5)(0.55) + (0.1)(0.71) + (-0.3) \\ &= 0.275 + 0.071 - 0.3 \\ &= 0.046 \end{aligned}$$

$\boldsymbol{u_y = 0.046}$

$$\begin{aligned} x_y = O_1 = \sigma_y(u_y) &= \frac{1}{1 + e^{-u_y}} \\ &= \frac{1}{1 + e^{-0.046}} \\ &= \frac{1}{1 + 0.9550} \\ &= \frac{1}{1.9550} \\ &= 0.5115 \\ &= 0.51 \end{aligned}$$

$\boldsymbol{x_y = O_1 = 0.51}$

$$
\begin{aligned}
E(w) &= \frac{1}{2}(t_1 - O_1)^2 \\
&= \frac{1}{2}(1 - 0.51)^2 \\
&= \frac{1}{2}(0.49)^2 \\
&= \frac{1}{2}(0.2401) \\
&= 0.12005 \\
&= 0.12
\end{aligned}
$$

1.1.3 Derivative of Binary Sigmoid Activation Function

$$\sigma(u) = \frac{1}{1 + e^{-u}}$$

$$
\begin{aligned}
\sigma'(u) &= \frac{\mathrm{d}}{\mathrm{d}u}\sigma(u) \\
&= \frac{\mathrm{d}}{\mathrm{d}u}(1 + e^{-u})^{-1} \\
&= -(1 + e^{-u})^{-2}(-e^{-u}) \\
&= \frac{e^{-u}}{(1 + e^{-u})^2} \\
&= \frac{1}{(1 + e^{-u})} . \frac{e^{-u}}{(1 + e^{-u})} \\
&= \frac{1}{(1 + e^{-u})} . \frac{(1 + e^{-u}) - 1}{(1 + e^{-u})} \\
&= \frac{1}{(1 + e^{-u})}\left(1 - \frac{1}{(1 + e^{-u})}\right) \\
&= \sigma(u)(1 - \sigma(u))
\end{aligned}
$$

$\boldsymbol{\sigma'(u) = \sigma(u)(1 - \sigma(u))}$

1.1.4 Backpropagation

Last Layer, w_{yz_1}:

(1) $w_{yz_1}^{new} = w_{yz_1}^{old} - \eta \frac{\partial E}{\partial w_{yz_1}}$

(2) $\frac{\partial E}{\partial w_{yz_1}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{yz_1}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{yz_1}}$

(3) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(4) $\frac{\partial O_1}{\partial w_{yz_1}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{yz_1}} = \sigma_y'(u_y).x_{z_1} = \sigma_y(u_y)(1 - \sigma_y(u_y)).x_{z_1}$

All the values in the last part of equation (4) are calculated during forward propagation. So, we can replace $\frac{\partial O_1}{\partial w_{yz_1}}$ in equation (2) and then plug in the values.

$$\begin{aligned}\frac{\partial E}{\partial w_{yz_1}} &= -(t_1 - O_1)\left[\sigma_y(u_y)(1 - \sigma_y(u_y))x_{z_1}\right]\\ &= -(1 - 0.51)\left[(0.51)(1 - 0.51)(0.55)\right]\\ &= -(0.49)(0.51)(0.49)(0.55)\\ &= -0.0673\end{aligned}$$

Now, by placing the value of $\frac{\partial E}{\partial w_{yz_1}}$ in equation (1), we can calculate the updated weight for w_{yz_1}.

$w_{yz_1}^{new} = w_{yz_1}^{old} - \eta\frac{\partial E}{\partial w_{yz_1}} = 0.5 - (0.25)(-0.0673) = 0.5 + 0.016825 = 0.516825 = 0.52$

$\boldsymbol{w_{yz_1}^{new} = 0.52}$

Last Layer, w_{yz_2}:

(5) $w_{yz_2}^{new} = w_{yz_2}^{old} - \eta \frac{\partial E}{\partial w_{yz_2}}$

(6) $\frac{\partial E}{\partial w_{yz_2}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{yz_2}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{yz_2}}$

(7) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(8) $\frac{\partial O_1}{\partial w_{yz_2}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{yz_2}} = \sigma_y'(u_y).x_{z_2} = \sigma_y(u_y)(1 - \sigma_y(u_y)).x_{z_2}$

All the values in the last part of equation (8) are calculated during forward propagation. So, we can replace $\frac{\partial O_1}{\partial w_{yz_2}}$ in equation (6) and then plug in the values.

$$\begin{aligned}\frac{\partial E}{\partial w_{yz_2}} &= -(t_1 - O_1)\left[\sigma_y(u_y)(1 - \sigma_y(u_y))x_{z_2}\right]\\ &= -(1 - 0.51)\left[(0.51)(1 - 0.51)(0.71)\right]\\ &= -(0.49)(0.51)(0.49)(0.71)\\ &= -0.0869\end{aligned}$$

Now, by placing the value of $\frac{\partial E}{\partial w_{yz_2}}$ in equation (5), we can calculate the updated weight for w_{yz_2}.

$w_{yz_2}^{new} = w_{yz_2}^{old} - \eta \frac{\partial E}{\partial w_{yz_2}} = 0.1 - (0.25)(-0.0869) = 0.1 + 0.021725 = 0.121725 = 0.12$

$\boldsymbol{w_{yz_2}^{new} = 0.12}$

Last Layer, w_{y0}:

(9) $w_{y0}^{new} = w_{y0}^{old} - \eta \frac{\partial E}{\partial w_{y0}}$

(10) $\frac{\partial E}{\partial w_{y0}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{y0}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{y0}}$

(11) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(12) $\frac{\partial O_1}{\partial w_{y0}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{y0}} = \sigma_y'(u_y).1 = \sigma_y'(u_y) = \sigma_y(u_y)(1 - \sigma_y(u_y))$

All the values in the last part of equation (12) are calculated during forward propagation. So, we can replace $\frac{\partial O_1}{\partial w_{y0}}$ in equation (10) and then plug in the values.

$$
\begin{aligned}
\frac{\partial E}{\partial w_{y0}} &= -(t_1 - O_1)\left[\sigma_y(u_y)(1 - \sigma_y(u_y))\right] \\
&= -(1 - 0.51)\left[(0.51)(1 - 0.51)\right] \\
&= -(0.49)(0.51)(0.49) \\
&= -0.1225
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{y0}}$ in equation (9), we can calculate the updated weight for w_{y0}.

$w_{y0}^{new} = w_{y0}^{old} - \eta \frac{\partial E}{\partial w_{y0}} = -0.3 - (0.25)(-0.1225) = -0.3 + 0.0306 = -0.2694 = -0.27$

$\boldsymbol{w_{y0}^{new} = -0.27}$

First Layer, $w_{z_1x_1}$:

(13) $w_{z_1x_1}^{new} = w_{z_1x_1}^{old} - \eta \frac{\partial E}{\partial w_{z_1x_1}}$

(14) $\frac{\partial E}{\partial w_{z_1x_1}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_1x_1}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{z_1x_1}}$

(15) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(16) $\frac{\partial O_1}{\partial w_{z_1x_1}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{z_1x_1}} = \sigma_y'(u_y)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1x_1}} = \sigma_y(u_y)(1 - \sigma_y(u_y))w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1x_1}}$

(17) $x_{z_1} = \sigma_{z_1}(w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10}) = \sigma_{z_1}(u_{z_1})$

(18) $\frac{\partial x_{z_1}}{\partial w_{z_1x_1}} = \sigma_{z_1}'(u_{z_1})\frac{\partial u_{z_1}}{\partial w_{z_1x_1}} = \sigma_{z_1}'(u_{z_1})x_1 = \sigma_{z_1}(u_{z_1})(1 - \sigma_{z_1}(u_{z_1}))x_1$

All the values in the last part of equation (18) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_1}}{\partial w_{z_1x_1}}$ in equation (16) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_1x_1}} &= \sigma_y(u_y)(1-\sigma_y(u_y))w_{yz_1}\sigma_{z_1}(u_{z_1})(1-\sigma_{z_1}(u_{z_1}))x_1 \\
&= (0.51)(1-0.51)(0.5)(0.55)(1-0.55)(0) \\
&= 0
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_1x_1}}$ in equation (14) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_1x_1}} &= -\left(t_1 - O_1\right).0 \\
&= 0
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_1x_1}}$ in equation (13), we can calculate the updated weight for $w_{z_1x_1}$.

$w^{new}_{z_1x_1} = w^{old}_{z_1x_1} - \eta\frac{\partial E}{\partial w_{z_1x_1}} = 0.7 - (0.25)(0) = 0.7 - 0 = 0.7$

$\boldsymbol{w^{new}_{z_1x_1} = 0.7}$

<u>First Layer, $w_{z_1x_2}$:</u>

(19) $w^{new}_{z_1x_2} = w^{old}_{z_1x_2} - \eta\frac{\partial E}{\partial w_{z_1x_2}}$

(20) $\frac{\partial E}{\partial w_{z_1x_2}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_1x_2}}\right)\right] = -\left(t_1 - O_1\right)\frac{\partial O_1}{\partial w_{z_1x_2}}$

(21) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(22) $\frac{\partial O_1}{\partial w_{z_1x_2}} = \sigma_y^{'}(u_y)\frac{\partial u_y}{\partial w_{z_1x_2}} = \sigma_y^{'}(u_y)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1x_2}} = \sigma_y(u_y)(1-\sigma_y(u_y))w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1x_2}}$

(23) $x_{z_1} = \sigma_{z_1}(w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10}) = \sigma_{z_1}(u_{z_1})$

(24) $\frac{\partial x_{z_1}}{\partial w_{z_1x_2}} = \sigma_{z_1}^{'}(u_{z_1})\frac{\partial u_{z_1}}{\partial w_{z_1x_2}} = \sigma_{z_1}^{'}(u_{z_1})x_2 = \sigma_{z_1}(u_{z_1})(1-\sigma_{z_1}(u_{z_1}))x_2$

All the values in the last part of equation (24) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_1}}{\partial w_{z_1x_2}}$ in equation (22) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_1x_2}} &= \sigma_y(u_y)(1-\sigma_y(u_y))w_{yz_1}\sigma_{z_1}(u_{z_1})(1-\sigma_{z_1}(u_{z_1}))x_2 \\
&= (0.51)(1-0.51)(0.5)(0.55)(1-0.55)(1) \\
&= (0.51)(0.49)(0.5)(0.55)(0.45)(1) \\
&= 0.0309
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_1x_2}}$ in equation (20) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_1x_2}} &= -\,(t_1 - O_1)\,(0.0309)\\
&= -\,(1 - 0.51)\,(0.0309)\\
&= -\,(0.49)\,(0.0309)\\
&= -0.0151
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_1x_2}}$ in equation (19), we can calculate the updated weight for $w_{z_1x_2}$.

$w^{new}_{z_1x_2} = w^{old}_{z_1x_2} - \eta \frac{\partial E}{\partial w_{z_1x_2}} = -0.2 - (0.25)(-0.0151) = -0.2 + 0.0038 = -0.1962 = -0.196$

$\boldsymbol{w^{new}_{z_1x_2} = -0.196}$

First Layer, w_{z_10}:

(25) $w^{new}_{z_10} = w^{old}_{z_10} - \eta \frac{\partial E}{\partial w_{z_10}}$

(26) $\frac{\partial E}{\partial w_{z_10}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_10}}\right)\right] = -\,(t_1 - O_1)\,\frac{\partial O_1}{\partial w_{z_10}}$

(27) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(28) $\frac{\partial O_1}{\partial w_{z_10}} = \sigma_y^{'}(u_y)\frac{\partial u_y}{\partial w_{z_10}} = \sigma_y^{'}(u_y)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_10}} = \sigma_y(u_y)(1 - \sigma_y(u_y))w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_10}}$

(29) $x_{z_1} = \sigma_{z_1}(w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10}) = \sigma_{z_1}(u_{z_1})$

(30) $\frac{\partial x_{z_1}}{\partial w_{z_10}} = \sigma_{z_1}^{'}(u_{z_1})\frac{\partial u_{z_1}}{\partial w_{z_10}} = \sigma_{z_1}^{'}(u_{z_1}).1 = \sigma_{z_1}^{'}(u_{z_1}) = \sigma_{z_1}(u_{z_1})(1 - \sigma_{z_1}(u_{z_1}))$

All the values in the last part of equation (30) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_1}}{\partial w_{z_10}}$ in equation (28) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_10}} &= \sigma_y(u_y)(1 - \sigma_y(u_y))w_{yz_1}\sigma_{z_1}(u_{z_1})(1 - \sigma_{z_1}(u_{z_1}))\\
&= (0.51)(1 - 0.51)(0.5)(0.55)(1 - 0.55)\\
&= (0.51)(0.49)(0.5)(0.55)(0.45)\\
&= 0.0309
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_10}}$ in equation (26) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_1 0}} &= -\,(t_1 - O_1)\,(0.0309) \\
&= -\,(1 - 0.51)\,(0.0309) \\
&= -\,(0.49)\,(0.0309) \\
&= -0.0151
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_1 0}}$ in equation (25), we can calculate the updated weight for $w_{z_1 0}$.

$w^{new}_{z_1 0} = w^{old}_{z_1 0} - \eta \frac{\partial E}{\partial w_{z_1 0}} = 0.4 - (0.25)(-0.0151) = 0.4 + 0.0038 = 0.4038 = 0.404$

$\boldsymbol{w^{new}_{z_1 0} = 0.404}$

First Layer, $w_{z_2 x_1}$:

(31) $w^{new}_{z_2 x_1} = w^{old}_{z_2 x_1} - \eta \frac{\partial E}{\partial w_{z_2 x_1}}$

(32) $\frac{\partial E}{\partial w_{z_2 x_1}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_2 x_1}}\right)\right] = -\,(t_1 - O_1)\,\frac{\partial O_1}{\partial w_{z_2 x_1}}$

(33) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(34) $\frac{\partial O_1}{\partial w_{z_2 x_1}} = \sigma'_y(u_y)\frac{\partial u_y}{\partial w_{z_2 x_1}} = \sigma'_y(u_y) w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}} = \sigma_y(u_y)(1 - \sigma_y(u_y)) w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}}$

(35) $x_{z_2} = \sigma_{z_2}(w_{z_2 x_1}x_1 + w_{z_2 x_2}x_2 + w_{z_2 0}) = \sigma_{z_2}(u_{z_2})$

(36) $\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}} = \sigma'_{z_2}(u_{z_2})\frac{\partial u_{z_2}}{\partial w_{z_2 x_1}} = \sigma'_{z_2}(u_{z_2})x_1 = \sigma_{z_2}(u_{z_2})(1 - \sigma_{z_2}(u_{z_2}))x_1$

All the values in the last part of equation (36) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}}$ in equation (34) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_2 x_1}} &= \sigma_y(u_y)(1 - \sigma_y(u_y)) w_{yz_2}\sigma_{z_2}(u_{z_2})(1 - \sigma_{z_2}(u_{z_2}))x_1 \\
&= (0.51)(1 - 0.51)(0.1)(0.71)(1 - 0.71)(0) \\
&= 0
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_2 x_1}}$ in equation (32) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_2 x_1}} &= -\,(t_1 - O_1)\,(0) \\
&= 0
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_2 x_1}}$ in equation (31), we can calculate the updated weight for $w_{z_2 x_1}$.

$w_{z_2x_1}^{new} = w_{z_2x_1}^{old} - \eta \frac{\partial E}{\partial w_{z_2x_1}} = -0.4 - (0.25)(0) = -0.4 - 0 = -0.4$

$\boldsymbol{w_{z_2x_1}^{new} = -0.4}$

First Layer, $w_{z_2x_2}$:

(37) $w_{z_2x_2}^{new} = w_{z_2x_2}^{old} - \eta \frac{\partial E}{\partial w_{z_2x_2}}$

(38) $\frac{\partial E}{\partial w_{z_2x_2}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_2x_2}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{z_2x_2}}$

(39) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(40) $\frac{\partial O_1}{\partial w_{z_2x_2}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{z_2x_2}} = \sigma_y'(u_y)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2x_2}} = \sigma_y(u_y)(1 - \sigma_y(u_y))w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2x_2}}$

(41) $x_{z_2} = \sigma_{z_2}(w_{z_2x_1}x_1 + w_{z_2x_2}x_2 + w_{z_20}) = \sigma_{z_2}(u_{z_2})$

(42) $\frac{\partial x_{z_2}}{\partial w_{z_2x_2}} = \sigma_{z_2}'(u_{z_2})\frac{\partial u_{z_2}}{\partial w_{z_2x_2}} = \sigma_{z_2}'(u_{z_2})x_2 = \sigma_{z_2}(u_{z_2})(1 - \sigma_{z_2}(u_{z_2}))x_2$

All the values in the last part of equation (42) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_2}}{\partial w_{z_2x_2}}$ in equation (40) and plug in the values.

$$\begin{aligned}
\frac{\partial O_1}{\partial w_{z_2x_2}} &= \sigma_y(u_y)(1 - \sigma_y(u_y))w_{yz_2}\sigma_{z_2}(u_{z_2})(1 - \sigma_{z_2}(u_{z_2}))x_2 \\
&= (0.51)(1 - 0.51)(0.1)(0.71)(1 - 0.71)(1) \\
&= (0.51)(0.49)(0.1)(0.71)(0.29)(1) \\
&= 0.0051
\end{aligned}$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_2x_2}}$ in equation (38) we get,

$$\begin{aligned}
\frac{\partial E}{\partial w_{z_2x_2}} &= -(t_1 - O_1)(0.0051) \\
&= -(1 - 0.51)(0.0051) \\
&= -(0.49)(0.0051) \\
&= -0.0025
\end{aligned}$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_2x_2}}$ in equation (37), we can calculate the updated weight for $w_{z_2x_2}$.

$w_{z_2x_2}^{new} = w_{z_2x_2}^{old} - \eta \frac{\partial E}{\partial w_{z_2x_2}} = 0.3 - (0.25)(-0.0025) = 0.3 + 0.0006 = 0.3006$

$\boldsymbol{w_{z_2x_2}^{new} = 0.3006}$

First Layer, w_{z_20}:

(43) $w_{z_20}^{new} = w_{z_20}^{old} - \eta \frac{\partial E}{\partial w_{z_20}}$

(44) $\frac{\partial E}{\partial w_{z_20}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_20}}\right)\right] = -\left(t_1 - O_1\right)\frac{\partial O_1}{\partial w_{z_20}}$

(45) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(46) $\frac{\partial O_1}{\partial w_{z_20}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{z_20}} = \sigma_y'(u_y)w_{yz2}\frac{\partial x_{z_2}}{\partial w_{z_20}} = \sigma_y(u_y)(1-\sigma_y(u_y))w_{yz2}\frac{\partial x_{z_2}}{\partial w_{z_20}}$

(47) $x_{z_2} = \sigma_{z2}(w_{z_2x_1}x_1 + w_{z_2x_2}x_2 + w_{z_20}) = \sigma_{z2}(u_{z2})$

(48) $\frac{\partial x_{z_2}}{\partial w_{z_20}} = \sigma_{z_2}'(u_{z_2})\frac{\partial u_{z_2}}{\partial w_{z_20}} = \sigma_{z_2}'(u_{z_2}).1 = \sigma_{z_2}'(u_{z_2}) = \sigma_{z_2}(u_{z_2})(1-\sigma_{z_2}(u_{z_2}))$

All the values in the last part of equation (48) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_2}}{\partial w_{z_20}}$ in equation (46) and plug in the values.

$$\begin{aligned}
\frac{\partial O_1}{\partial w_{z_20}} &= \sigma_y(u_y)(1-\sigma_y(u_y))w_{yz2}\sigma_{z_2}(u_{z_2})(1-\sigma_{z_2}(u_{z_2})) \\
&= (0.51)(1-0.51)(0.1)(0.71)(1-0.71) \\
&= (0.51)(0.49)(0.1)(0.71)(0.29) \\
&= 0.0051
\end{aligned}$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_20}}$ in equation (44) we get,

$$\begin{aligned}
\frac{\partial E}{\partial w_{z_20}} &= -\left(t_1 - O_1\right)(0.0051) \\
&= -(1-0.51)(0.0051) \\
&= -(0.49)(0.0051) \\
&= -0.0025
\end{aligned}$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_20}}$ in equation (43), we can calculate the updated weight for w_{z_20}.

$w_{z_20}^{new} = w_{z_20}^{old} - \eta \frac{\partial E}{\partial w_{z_20}} = 0.6 - (0.25)(-0.0025) = 0.6 + 0.0006 = 0.6006$

$\boldsymbol{w_{z_20}^{new} = 0.6006}$

<u>Updated Weights:</u>

$\boldsymbol{w_{z_1x_1} = 0.7}$, $\boldsymbol{w_{z_1x_2} = -0.196}$, $\boldsymbol{w_{z_10} = 0.404}$

$\boldsymbol{w_{z_2x_1} = -0.4}$, $\boldsymbol{w_{z_2x_2} = 0.3006}$, $\boldsymbol{w_{z_20} = 0.6006}$

$\boldsymbol{w_{yz_1} = 0.52}$, $\boldsymbol{w_{yz_2} = 0.12}$, $\boldsymbol{w_{y_0} = -0.27}$

1.2 Part 2

1.2.1 Training Input, Target, Parameters, Hyperparameters

Training inputs, $x_1 = -1$, $x_2 = 1$
Target, $t_1 = 1$
Hyperparameter: Learning rate, $\eta = 0.25$
Parameters:
$w_{z_1x_1} = 0.7$, $w_{z_1x_2} = -0.2$, $w_{z_10} = 0.4$
$w_{z_2x_1} = -0.4$, $w_{z_2x_2} = 0.3$, $w_{z_20} = 0.6$
$w_{yz_1} = 0.5$, $w_{yz_2} = 0.1$, $w_{y_0} = -0.3$

1.2.2 Forward Propagation

$$\begin{aligned}
u_{z_1} &= w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10}.1 \\
&= w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10} \\
&= (0.7)(-1) + (-0.2)(1) + 0.4 \\
&= -0.7 - 0.2 + 0.4 \\
&= -0.5
\end{aligned}$$

$\boldsymbol{u_{z_1} = -0.5}$

$$\begin{aligned}
x_{z_1} = \sigma_{z_1}(u_{z1}) &= \frac{e^{u_{z_1}} - e^{-u_{z_1}}}{e^{u_{z_1}} + e^{-u_{z_1}}} \\
&= \frac{e^{-0.5} - e^{-(-0.5)}}{e^{-0.5} + e^{-(-0.5)}} \\
&= \frac{e^{-0.5} - e^{0.5}}{e^{-0.5} + e^{0.5}} \\
&= \frac{0.6065 - 1.6487}{0.6065 + 1.6487} \\
&= \frac{-1.0422}{2.2552} \\
&= -0.4621 \\
&= -0.46
\end{aligned}$$

$\boldsymbol{x_{z_1} = -0.46}$

$$\begin{aligned}
u_{z_2} &= w_{z_2x_1}x_1 + w_{z_2x_2}x_2 + w_{z_20}.1 \\
&= w_{z_2x_1}x_1 + w_{z_2x_2}x_2 + w_{z_20} \\
&= (-0.4)(-1) + (0.3)1 + 0.6 \\
&= 0.4 + 0.3 + 0.6 \\
&= 1.3
\end{aligned}$$

$\boldsymbol{u_{z_2} = 1.3}$

$$
\begin{aligned}
x_{z_2} = \sigma_{z_2}(u_{z2}) &= \frac{e^{u_{z_2}} - e^{-u_{z_2}}}{e^{u_{z_2}} + e^{-u_{z_2}}} \\
&= \frac{e^{1.3} - e^{-1.3}}{e^{1.3} + e^{-1.3}} \\
&= \frac{3.6693 - 0.2725}{3.6693 + 0.2725} \\
&= \frac{3.3968}{3.9418} \\
&= 0.8617 \\
&= 0.86
\end{aligned}
$$

$\boldsymbol{x_{z_2} = 0.86}$

$$
\begin{aligned}
u_y &= w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}.1 \\
&= w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0} \\
&= (0.5)(-0.46) + (0.1)(0.86) + (-0.3) \\
&= -0.23 + 0.086 - 0.3 \\
&= -0.444
\end{aligned}
$$

$\boldsymbol{u_y = -0.044}$

$$
\begin{aligned}
x_y = O_1 = \sigma_y(u_y) &= \frac{e^{u_y} - e^{-u_y}}{e^{u_y} + e^{-u_y}} \\
&= \frac{e^{-0.444} - e^{-(-0.444)}}{e^{-0.444} + e^{-(-0.444)}} \\
&= \frac{e^{-0.444} - e^{0.444}}{e^{-0.444} + e^{0.444}} \\
&= \frac{0.6415 - 1.5589}{0.6415 + 1.5589} \\
&= \frac{-0.9174}{2.2004} \\
&= -0.4169 \\
&= -0.42
\end{aligned}
$$

$\boldsymbol{x_y = O_1 = -0.42}$

$$
\begin{aligned}
E(w) &= \frac{1}{2}(t_1 - O_1)^2 \\
&= \frac{1}{2}(1 - (-0.42))^2 \\
&= \frac{1}{2}(1 + 0.42))^2 \\
&= \frac{1}{2}(1.42)^2 \\
&= \frac{1}{2}(2.0164) \\
&= 1.0082 \\
&= 1.01
\end{aligned}
$$

1.2.3 Derivative of Bipolar Sigmoid Activation Function

$$
\sigma(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}
$$

$$
\begin{aligned}
\sigma'(u) &= \frac{\mathrm{d}}{\mathrm{d}u}\sigma(u) \\
&= \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{e^u - e^{-u}}{e^u + e^{-u}}\right) \\
&= \frac{1}{(e^u + e^{-u})^2}\left[(e^u + e^{-u})\frac{\mathrm{d}}{\mathrm{d}u}(e^u - e^{-u}) - (e^u - e^{-u})\frac{\mathrm{d}}{\mathrm{d}u}(e^u + e^{-u})\right] \\
&= \frac{1}{(e^u + e^{-u})^2}\left[(e^u + e^{-u})(e^u + e^{-u}) - (e^u - e^{-u})(e^u - e^{-u})\right] \\
&= \frac{1}{(e^u + e^{-u})^2}\left[(e^u + e^{-u})^2 - (e^u - e^{-u})^2\right] \\
&= 1 - \frac{(e^u - e^{-u})^2}{(e^u + e^{-u})^2} \\
&= 1 - \left(\frac{e^u - e^{-u}}{e^u + e^{-u}}\right)^2 \\
&= 1 - \sigma(u)^2
\end{aligned}
$$

$\boldsymbol{\sigma'(u) = 1 - \sigma(u)^2}$

1.2.4 Backpropagation

Last Layer, w_{yz_1}:

(1) $w_{yz_1}^{new} = w_{yz_1}^{old} - \eta \frac{\partial E}{\partial w_{yz_1}}$

(2) $\frac{\partial E}{\partial w_{yz_1}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{yz_1}}\right)\right] = -\left(t_1 - O_1\right)\frac{\partial O_1}{\partial w_{yz_1}}$

(3) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(4) $\frac{\partial O_1}{\partial w_{yz_1}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{yz_1}} = \sigma_y'(u_y)x_{z_1} = (1 - \sigma_y(u_y)^2)x_{z_1}$

All the values in the last part of equation (4) are calculated during forward propagation. So, we can replace $\frac{\partial O_1}{\partial w_{yz_1}}$ in equation (2) and then plug in the values.

$$
\begin{aligned}
\frac{\partial E}{\partial w_{yz_1}} &= -\left(t_1 - O_1\right)\left[(1 - \sigma_y(u_y)^2)x_{z_1}\right] \\
&= -\left(1 - (-0.42)\right)\left[(1 - (-0.42)^2)(-0.46)\right] \\
&= -(1 + 0.42)(1 - 0.1764)(-0.46) \\
&= -(1.42)(0.8236)(-0.46) \\
&= 0.5380
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{yz_1}}$ in equation (1), we can calculate the updated weight for w_{yz_1}.

$w_{yz_1}^{new} = w_{yz_1}^{old} - \eta\frac{\partial E}{\partial w_{yz_1}} = 0.5 - (0.25)(0.5380) = 0.5 - 0.1345 = 0.3655 = 0.37$

$\boldsymbol{w_{yz_1}^{new} = 0.37}$

<u>Last Layer, w_{yz_2}:</u>

(5) $w_{yz_2}^{new} = w_{yz_2}^{old} - \eta \frac{\partial E}{\partial w_{yz_2}}$

(6) $\frac{\partial E}{\partial w_{yz_2}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{yz_2}}\right)\right] = -\left(t_1 - O_1\right)\frac{\partial O_1}{\partial w_{yz_2}}$

(7) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(8) $\frac{\partial O_1}{\partial w_{yz_2}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{yz_2}} = \sigma_y'(u_y)x_{z_2} = (1 - \sigma_y(u_y)^2)x_{z_2}$

All the values in the last part of equation (8) are calculated during forward propagation. So, we can replace $\frac{\partial O_1}{\partial w_{yz_2}}$ in equation (6) and then plug in the values.

$$
\begin{aligned}
\frac{\partial E}{\partial w_{yz_2}} &= -\left(t_1 - O_1\right)\left[(1 - \sigma_y(u_y)^2)x_{z_2}\right] \\
&= -\left(1 - (-0.42)\right)\left[(1 - (-0.42)^2)(0.86)\right] \\
&= -(1 + 0.42)(1 - 0.1764)(0.86) \\
&= -(1.42)(0.8236)(0.86) \\
&= -1.0058
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{yz_2}}$ in equation (5), we can calculate the updated weight for w_{yz_2}.

$w_{yz_2}^{new} = w_{yz_2}^{old} - \eta \frac{\partial E}{\partial w_{yz_2}} = 0.1 - (0.25)(-1.0058) = 0.1 + 0.25145 = 0.35145 = 0.35$

$\boldsymbol{w_{yz_2}^{new} = 0.35}$

<u>Last Layer, w_{y0}:</u>

(9) $w_{y0}^{new} = w_{y0}^{old} - \eta \frac{\partial E}{\partial w_{y0}}$

(10) $\frac{\partial E}{\partial w_{y0}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{y0}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{y0}}$

(11) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y0}) = \sigma_y(u_y)$

(12) $\frac{\partial O_1}{\partial w_{y0}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{y0}} = \sigma_y'(u_y).1 = \sigma_y'(u_y) = 1 - \sigma_y(u_y)^2$

All the values in the last part of equation (12) are calculated during forward propagation. So, we can replace $\frac{\partial O_1}{\partial w_{y0}}$ in equation (10) and then plug in the values.

$$\begin{aligned}
\frac{\partial E}{\partial w_{y0}} &= -(t_1 - O_1)(1 - \sigma_y(u_y)^2) \\
&= -(1 - (-0.42))(1 - (-0.42)^2) \\
&= -(1 + 0.42)(1 - 0.1764) \\
&= -(1.42)(0.8236) \\
&= -1.1695
\end{aligned}$$

Now, by placing the value of $\frac{\partial E}{\partial w_{y0}}$ in equation (9), we can calculate the updated weight for w_{y0}.

$w_{y0}^{new} = w_{y0}^{old} - \eta \frac{\partial E}{\partial w_{y0}} = -0.3 - (0.25)(-1.1695) = -0.3 + 0.2924 = -0.0076 = -0.008$

$\boldsymbol{w_{y0}^{new} = -0.008}$

<u>First Layer, $w_{z_1x_1}$:</u>

(13) $w_{z_1x_1}^{new} = w_{z_1x_1}^{old} - \eta \frac{\partial E}{\partial w_{z_1x_1}}$

(14) $\frac{\partial E}{\partial w_{z_1x_1}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_1x_1}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{z_1x_1}}$

(15) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y0}) = \sigma_y(u_y)$

(16) $\frac{\partial O_1}{\partial w_{z_1x_1}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{z_1x_1}} = \sigma_y'(u_y)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1x_1}} = (1 - \sigma_y(u_y)^2)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1x_1}}$

(17) $x_{z_1} = \sigma_{z_1}(w_{z_1x_1}x_1 + w_{z_1x_2}x_2 + w_{z_10}) = \sigma_{z_1}(u_{z_1})$

(18) $\frac{\partial x_{z_1}}{\partial w_{z_1 x_1}} = \sigma'_{z_1}(u_{z_1})\frac{\partial u_{z_1}}{\partial w_{z_1 x_1}} = \sigma'_{z_1}(u_{z_1})x_1 = (1-\sigma_{z_1}(u_{z_1})^2)x_1$

All the values in the last part of equation (18) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_1}}{\partial w_{z_1 x_1}}$ in equation (16) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_1 x_1}} &= (1-\sigma_y(u_y)^2)w_{yz_1}(1-\sigma_{z_1}(u_{z_1})^2)x_1 \\
&= (1-(-0.42)^2)(0.5)(1-(-0.46)^2)(-1) \\
&= (1-0.1764)(0.5)(1-0.2116)(-1) \\
&= (0.8236)(0.5)(0.7884)(-1) \\
&= -0.3247
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_1 x_1}}$ in equation (14) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_1 x_1}} &= -\,(t_1 - O_1)\,(-0.3247) \\
&= -(1-(-0.42))(-0.3247) \\
&= -(1+0.42)(-0.3247) \\
&= (1.42)(0.3247) \\
&= 0.4611
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_1 x_1}}$ in equation (13), we can calculate the updated weight for $w_{z_1 x_1}$.

$w^{new}_{z_1 x_1} = w^{old}_{z_1 x_1} - \eta\frac{\partial E}{\partial w_{z_1 x_1}} = 0.7 - (0.25)(0.4611) = 0.7 - 0.1153 = 0.5847 = 0.58$

$\boldsymbol{w^{new}_{z_1 x_1} = 0.58}$

First Layer, $w_{z_1 x_2}$:

(19) $w^{new}_{z_1 x_2} = w^{old}_{z_1 x_2} - \eta\frac{\partial E}{\partial w_{z_1 x_2}}$

(20) $\frac{\partial E}{\partial w_{z_1 x_2}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_1 x_2}}\right)\right] = -\,(t_1 - O_1)\,\frac{\partial O_1}{\partial w_{z_1 x_2}}$

(21) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(22) $\frac{\partial O_1}{\partial w_{z_1 x_2}} = \sigma'_y(u_y)\frac{\partial u_y}{\partial w_{z_1 x_2}} = \sigma'_y(u_y)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1 x_2}} = (1-\sigma_y(u_y)^2)w_{yz_1}\frac{\partial x_{z_1}}{\partial w_{z_1 x_2}}$

(23) $x_{z_1} = \sigma_{z_1}(w_{z_1 x_1}x_1 + w_{z_1 x_2}x_2 + w_{z_1 0}) = \sigma_{z_1}(u_{z_1})$

(24) $\frac{\partial x_{z_1}}{\partial w_{z_1 x_2}} = \sigma'_{z_1}(u_{z_1})\frac{\partial u_{z_1}}{\partial w_{z_1 x_2}} = \sigma'_{z_1}(u_{z_1})x_2 = (1-\sigma_{z_1}(u_{z_1})^2)x_2$

All the values in the last part of equation (24) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_1}}{\partial w_{z_1 x_2}}$ in equation (22) and plug in the values.

$$\begin{aligned}\frac{\partial O_1}{\partial w_{z_1 x_2}} &= (1-\sigma_y(u_y)^2) w_{yz_1} (1-\sigma_{z_1}(u_{z_1})^2) x_2 \\ &= (1-(-0.42)^2)(0.5)(1-(-0.46)^2)(1) \\ &= (1-0.1764)(0.5)(1-0.2116)(1) \\ &= (0.8236)(0.5)(0.7884)(1) \\ &= 0.3247\end{aligned}$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_1 x_2}}$ in equation (20) we get,

$$\begin{aligned}\frac{\partial E}{\partial w_{z_1 x_2}} &= -(t_1 - O_1)(0.3247) \\ &= -(1-(-0.42))(0.3247) \\ &= -(1.42)(0.3247) \\ &= -0.4611\end{aligned}$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_1 x_2}}$ in equation (19), we can calculate the updated weight for $w_{z_1 x_2}$.

$w^{new}_{z_1 x_2} = w^{old}_{z_1 x_2} - \eta \frac{\partial E}{\partial w_{z_1 x_2}} = -0.2 - (0.25)(-0.4611) = -0.2 + 0.1153 = -0.0847 = -0.08$

$\boldsymbol{w^{new}_{z_1 x_2} = -0.08}$

First Layer, $w_{z_1 0}$:

(25) $w^{new}_{z_1 0} = w^{old}_{z_1 0} - \eta \frac{\partial E}{\partial w_{z_1 0}}$

(26) $\frac{\partial E}{\partial w_{z_1 0}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_1 0}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{z_1 0}}$

(27) $O_1 = x_y = \sigma_y(w_{yz_1} x_{z_1} + w_{yz_2} x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(28) $\frac{\partial O_1}{\partial w_{z_1 0}} = \sigma'_y(u_y)\frac{\partial u_y}{\partial w_{z_1 0}} = \sigma'_y(u_y) w_{yz_1} \frac{\partial x_{z_1}}{\partial w_{z_1 0}} = (1-\sigma_y(u_y)^2) w_{yz_1} \frac{\partial x_{z_1}}{\partial w_{z_1 0}}$

(29) $x_{z_1} = \sigma_{z_1}(w_{z_1 x_1} x_1 + w_{z_1 x_2} x_2 + w_{z_1 0}) = \sigma_{z_1}(u_{z_1})$

(30) $\frac{\partial x_{z_1}}{\partial w_{z_1 0}} = \sigma'_{z_1}(u_{z_1})\frac{\partial u_{z_1}}{\partial w_{z_1 0}} = \sigma'_{z_1}(u_{z_1}).1 = \sigma'_{z_1}(u_{z_1}) = (1-\sigma_{z_1}(u_{z_1})^2)$

All the values in the last part of equation (30) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_1}}{\partial w_{z_1 0}}$ in equation (28) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_1 0}} &= (1-\sigma_y(u_y)^2)w_{yz_1}(1-\sigma_{z_1}(u_{z_1})^2) \\
&= (1-(-0.42)^2)(0.5)(1-(-0.46)^2)) \\
&= (1-0.1764)(0.5)(1-0.2116) \\
&= (0.8236)(0.5)(0.7884) \\
&= 0.3247
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_1 0}}$ in equation (26) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_1 0}} &= -(t_1 - O_1)(0.3247) \\
&= -(1-(-0.42))(0.3247) \\
&= -(1.42)(0.3247) \\
&= -0.4611
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_1 0}}$ in equation (25), we can calculate the updated weight for $w_{z_1 0}$.

$w_{z_1 0}^{new} = w_{z_1 0}^{old} - \eta \frac{\partial E}{\partial w_{z_1 0}} = 0.4 - (0.25)(-0.4611) = 0.4 + 0.1153 = 0.5153 = 0.52$

$\boldsymbol{w_{z_1 0}^{new} = 0.52}$

First Layer, $w_{z_2 x_1}$:

(31) $w_{z_2 x_1}^{new} = w_{z_2 x_1}^{old} - \eta \frac{\partial E}{\partial w_{z_2 x_1}}$

(32) $\frac{\partial E}{\partial w_{z_2 x_1}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_2 x_1}}\right)\right] = -(t_1 - O_1)\frac{\partial O_1}{\partial w_{z_2 x_1}}$

(33) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(34) $\frac{\partial O_1}{\partial w_{z_2 x_1}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{z_2 x_1}} = \sigma_y'(u_y)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}} = (1-\sigma_y(u_y)^2)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}}$

(35) $x_{z_2} = \sigma_{z_2}(w_{z_2 x_1}x_1 + w_{z_2 x_2}x_2 + w_{z_2 0}) = \sigma_{z_2}(u_{z_2})$

(36) $\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}} = \sigma_{z_2}'(u_{z_2})\frac{\partial u_{z_2}}{\partial w_{z_2 x_1}} = \sigma_{z_2}'(u_{z_2})x_1 = (1-\sigma_{z_2}(u_{z_2})^2)x_1$

All the values in the last part of equation (36) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_2}}{\partial w_{z_2 x_1}}$ in equation (34) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_2x_1}} &= (1-\sigma_y(u_y)^2)w_{yz_2}(1-\sigma_{z_2}(u_{z_2})^2)x_1\\
&= (1-(-0.42)^2)(0.1)(1-(0.86)^2))(-1)\\
&= (1-0.1764)(0.1)(1-0.7396)(-1)\\
&= (0.8236)(0.1)(0.2604)(-1)\\
&= -0.0214
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_2x_1}}$ in equation (32) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_2x_1}} &= -(t_1-O_1)(-0.0214)\\
&= -(1-(-0.42))(-0.0214)\\
&= -(1+0.42)(-0.0214)\\
&= -(1.42)(-0.0214)\\
&= (1.42)(0.0214)\\
&= 0.0304
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_2x_1}}$ in equation (31), we can calculate the updated weight for $w_{z_2x_1}$.

$w_{z_2x_1}^{new} = w_{z_2x_1}^{old} - \eta\frac{\partial E}{\partial w_{z_2x_1}} = -0.4-(0.25)(0.0304) = -0.4-0.0076 = -0.4076 = -0.41$

$\boldsymbol{w_{z_2x_1}^{new} = -0.41}$

First Layer, $w_{z_2x_2}$:

(37) $w_{z_2x_2}^{new} = w_{z_2x_2}^{old} - \eta\frac{\partial E}{\partial w_{z_2x_2}}$

(38) $\frac{\partial E}{\partial w_{z_2x_2}} = \frac{1}{2}\left[2(t_1-O_1)\left(-\frac{\partial O_1}{\partial w_{z_2x_2}}\right)\right] = -(t_1-O_1)\frac{\partial O_1}{\partial w_{z_2x_2}}$

(39) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1}+w_{yz_2}x_{z_2}+w_{y_0}) = \sigma_y(u_y)$

(40) $\frac{\partial O_1}{\partial w_{z_2x_2}} = \sigma_y'(u_y)\frac{\partial u_y}{\partial w_{z_2x_2}} = \sigma_y'(u_y)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2x_2}} = (1-\sigma_y(u_y)^2)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_2x_2}}$

(41) $x_{z_2} = \sigma_{z_2}(w_{z_2x_1}x_1+w_{z_2x_2}x_2+w_{z_20}) = \sigma_{z_2}(u_{z_2})$

(42) $\frac{\partial x_{z_2}}{\partial w_{z_2x_2}} = \sigma_{z_2}'(u_{z_2})\frac{\partial u_{z_2}}{\partial w_{z_2x_2}} = \sigma_{z_2}'(u_{z_2})x_2 = (1-\sigma_{z_2}(u_{z_2})^2)x_2$

All the values in the last part of equation (42) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_2}}{\partial w_{z_2x_2}}$ in equation (40) and plug in the values.

$$
\begin{aligned}
\frac{\partial O_1}{\partial w_{z_2x_2}} &= (1-\sigma_y(u_y)^2)w_{yz_2}(1-\sigma_{z_2}(u_{z_2})^2)x_2 \\
&= (1-(-0.42)^2)(0.1)(1-(0.86)^2))(1) \\
&= (1-0.1764)(0.1)(1-0.7396)(1) \\
&= (0.8236)(0.1)(0.2604)(1) \\
&= 0.0214
\end{aligned}
$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_2x_2}}$ in equation (38) we get,

$$
\begin{aligned}
\frac{\partial E}{\partial w_{z_2x_2}} &= -\left(t_1 - O_1\right)(0.0214) \\
&= -(1-(-0.42))(0.0214) \\
&= -(1+0.42)(0.0214) \\
&= -(1.42)(0.0214) \\
&= -0.0304
\end{aligned}
$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_2x_2}}$ in equation (37), we can calculate the updated weight for $w_{z_2x_2}$.

$w^{new}_{z_2x_2} = w^{old}_{z_2x_2} - \eta\frac{\partial E}{\partial w_{z_2x_2}} = 0.3 - (0.25)(-0.0304) = 0.3 + 0.0076 = 0.3076 = 0.31$

$\boldsymbol{w^{new}_{z_2x_2} = 0.31}$

<u>First Layer, w_{z_20}:</u>

(43) $w^{new}_{z_20} = w^{old}_{z_20} - \eta\frac{\partial E}{\partial w_{z_20}}$

(44) $\frac{\partial E}{\partial w_{z_20}} = \frac{1}{2}\left[2(t_1 - O_1)\left(-\frac{\partial O_1}{\partial w_{z_20}}\right)\right] = -\left(t_1 - O_1\right)\frac{\partial O_1}{\partial w_{z_20}}$

(45) $O_1 = x_y = \sigma_y(w_{yz_1}x_{z_1} + w_{yz_2}x_{z_2} + w_{y_0}) = \sigma_y(u_y)$

(46) $\frac{\partial O_1}{\partial w_{z_20}} = \sigma_y^{'}(u_y)\frac{\partial u_y}{\partial w_{z_20}} = \sigma_y^{'}(u_y)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_20}} = (1-\sigma_y(u_y)^2)w_{yz_2}\frac{\partial x_{z_2}}{\partial w_{z_20}}$

(47) $x_{z_2} = \sigma_{z_2}(w_{z_2x_1}x_1 + w_{z_2x_2}x_2 + w_{z_20}) = \sigma_{z_2}(u_{z_2})$

(48) $\frac{\partial x_{z_2}}{\partial w_{z_20}} = \sigma_{z_2}^{'}(u_{z_2})\frac{\partial u_{z_2}}{\partial w_{z_20}} = \sigma_{z_2}^{'}(u_{z_2}).1 = \sigma_{z_2}^{'}(u_{z_2}) = (1-\sigma_{z_2}(u_{z_2})^2)$

All the values in the last part of equation (48) are calculated during forward propagation. So, we can replace $\frac{\partial x_{z_2}}{\partial w_{z_20}}$ in equation (46) and plug in the values.

$$\frac{\partial O_1}{\partial w_{z_2 0}} = (1 - \sigma_y(u_y)^2) w_{yz_2} (1 - \sigma_{z_2}(u_{z_2})^2)$$
$$= (1 - (-0.42)^2)(0.1)(1 - (0.86)^2))$$
$$= (1 - 0.1764)(0.1)(1 - 0.7396)$$
$$= (0.8236)(0.1)(0.2604)$$
$$= 0.0214$$

Plugging in the value of $\frac{\partial O_1}{\partial w_{z_2 0}}$ in equation (44) we get,

$$\frac{\partial E}{\partial w_{z_2 0}} = -\left(t_1 - O_1\right)\left(0.0214\right)$$
$$= -(1 - (-0.42))(0.0214)$$
$$= -(1 + 0.42)(0.0214)$$
$$= -(1.42)(0.0214)$$
$$= -0.0304$$

Now, by placing the value of $\frac{\partial E}{\partial w_{z_2 0}}$ in equation (43), we can calculate the updated weight for $w_{z_2 0}$.

$w_{z_2 0}^{new} = w_{z_2 0}^{old} - \eta \frac{\partial E}{\partial w_{z_2 0}} = 0.6 - (0.25)(-0.0304) = 0.6 + 0.0076 = 0.6076 = 0.61$

$\boldsymbol{w_{z_2 0}^{new} = 0.61}$

Updated Weights:

$\boldsymbol{w_{z_1 x_1} = 0.58}$, $\boldsymbol{w_{z_1 x_2} = -0.08}$, $\boldsymbol{w_{z_1 0} = 0.52}$

$\boldsymbol{w_{z_2 x_1} = -0.41}$, $\boldsymbol{w_{z_2 x_2} = 0.31}$, $\boldsymbol{w_{z_2 0} = 0.61}$

$\boldsymbol{w_{y z_1} = 0.37}$, $\boldsymbol{w_{y z_2} = 0.35}$, $\boldsymbol{w_{y_0} = -0.008}$

2 Backpropagation Calculation

$$E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{m}(t_k - O_k)^2 \tag{1}$$

Equation (1) is the error function for the network. In the equation, m is the number of output neurons. Since the given feedforward network has one output neuron 5, here m=1. So, we can rewrite equation (1) like following:

$$E(\mathbf{w}) = \frac{1}{2}(t_1 - O_1)^2 \tag{2}$$

The derivative of error function, E with respect to weight w_{42} can be written like this:

$$\frac{\partial E}{\partial w_{42}} = -\left(t_1 - O_1\right) \frac{\partial O_1}{\partial w_{42}} \tag{3}$$

$O_1 = x_5 = \sigma_5(w_{53}x_3 + w_{54}x_4)$, where $w_{53}x_3 + w_{54}x_4 = u_5$.

So, the partial derivative of O_1 with respect to weight w_{42} can be written like this:

$$\frac{\partial O_1}{\partial w_{42}} = \sigma_5^{'}(u_5) w_{54} \frac{\partial x_4}{\partial w_{42}} \tag{4}$$

Now, $x_4 = \sigma_4(w_{41}x_1 + w_{42}x_2)$, where $w_{41}x_1 + w_{42}x_2 = u_4$.

So, the partial derivative of x_4 with respect to weight w_{42} can be written like this:

$$\frac{\partial x_4}{\partial w_{42}} = \sigma_4^{'}(u_4) x_2 \tag{5}$$

If we replace the value of $\frac{\partial x_4}{\partial w_{42}}$ in equation (4) using equation (5), equation (4) takes the following form:

$$\frac{\partial O_1}{\partial w_{42}} = \sigma_5^{'}(u_5) w_{54} \sigma_4^{'}(u_4) x_2 \tag{6}$$

Now, if we replace the value of $\frac{\partial O_1}{\partial w_{42}}$ in equation (3) using equation (6), equation (3) takes the following form:

$$\frac{\partial E}{\partial w_{42}} = -\left(t_1 - O_1\right) \sigma_5^{'}(u_5) w_{54} \sigma_4^{'}(u_4) x_2 \tag{7}$$

The error in neuron 5 can be written in the following form:

$$\delta_5 = \left(t_1 - O_1\right) \sigma_5^{'}(u_5) \tag{8}$$

If we place the expression of error δ_5 from equation (8) in equation (7), equation (7) takes the following form:

$$\frac{\partial E}{\partial w_{42}} = -\delta_5 w_{54} \sigma_4^{'}(u_4) x_2 \tag{9}$$

The error in neuron 4 can be written in the following form:

$$\delta_4 = \delta_5 w_{54} \sigma_4^{'}(u_4) \tag{10}$$

If we place the expression of error δ_4 from equation (10) in equation (9), equation (9) takes the following form:

$$\frac{\partial E}{\partial w_{42}} = -\delta_4 x_2 \tag{11}$$

We know the update rule of weights for backpropagation algorithm is the following one:

$$w_{ij}^{new} = w_{ij}^{old} - \eta \frac{\partial E}{\partial w_{ij}} \tag{12}$$

For weight w_{42}, equation (12) takes the following form:

$$w_{42}^{new} = w_{42}^{old} - \eta \frac{\partial E}{\partial w_{42}} \tag{13}$$

If we place the value of $\frac{\partial E}{\partial w_{42}}$ from equation (11) in equation (13), we will find the following final formula for weight update for weight w_{42},

$$w_{42}^{new} = w_{42}^{old} - \eta(-\delta_4 x_2) = w_{42}^{old} + \eta(\delta_4 x_2) \tag{14}$$