# Assignment 4

Lokesh Mohanty (SR no: 21014)

November 2022

## Question 1

- Predicted outputs as a function of inputs,

$$
\begin{aligned}
h_1 &= sigmoid(w_{21}^1 x_2 + B) \\
h_2 &= sigmoid(w_{12}^1 x_1 + w_{22}^1 x_2 + B) \\
h_3 &= ReLU(w_{23}^1 x_2 + w_{33}^1 x_3 + B) \\
h_4 &= ReLU(w_{24}^1 x_2 + B) \\
\hat{y_1} &= w_{11}^2 h_1 + w_{31}^2 h_3 \\
&= w_{11}^2 sigmoid(w_{21}^1 x_2 + B) + w_{31} ReLU(w_{23}^1 x_2 + w_{33}^1 x_3 + B) \\
\hat{y_2} &= w_{22}^2 h_2 + w_{42}^2 h_4 \\
&= w_{22}^2 sigmoid(w_{12}^1 x_1 + w_{22}^1 x_2 + B) + w_{42}^2 ReLU(w_{24}^1 x_2 + B)
\end{aligned}
$$

- After one forward pass,

$$
\begin{aligned}
h_1 &= sigmoid(w_{21}^1 x_2 + B) = sigmoid(0.25 * 0.1 + 0.6) \\
&= sigmoid(0.625) = 0.6513 \\
h_2 &= sigmoid(w_{12}^1 x_1 + w_{22}^1 x_2 + B) = sigmoid(0.55 * 0.15 + 0.6) \\
&= sigmoid(0.7025) = 0.6687 \\
h_3 &= ReLU(w_{23}^1 x_2 + w_{33}^1 x_3 + B) = ReLU(0.1 * 0.1 + 0.6 * 0.05 + 0.6) \\
&= ReLU(0.64) = 0.64 \\
h_4 &= ReLU(w_{24}^1 x_2 + B) = ReLU(0.05 * 0.1 + 0.6) \\
&= ReLU(0.605) = 0.605 \\
\hat{y_1} &= w_{11}^2 h_1 + w_{31}^2 h_3 = 0.7 * 0.6513 + 0.33 * 0.64 = 0.6671 \\
\hat{y_2} &= w_{22}^2 h_2 + w_{42}^2 h_4 = 0.45 * 0.6687 + 0.8 * 0.605 = 0.7849
\end{aligned}
$$

- Mean-squared error,

$$
\begin{aligned}
mse &= \frac{1}{2}(y_1 - \hat{y_1})^2 + \frac{1}{2}(y_2 - \hat{y_2})^2 \\
&= \frac{1}{2}(0.31 - 0.6671)^2 + \frac{1}{2}(0.27 - 0.7849)^2 = 0.1963
\end{aligned}
$$

- One pass of backpropagation,

$$\begin{aligned}
L(loss) &= mse = \frac{1}{2}(0.31-\hat{y}_1)^2 + \frac{1}{2}(0.27-\hat{y}_2)^2 \\
w_{31}^2 &= w_{31}^2 - \alpha\frac{\partial L}{\partial w_{31}^2} \\
w_{21}^1 &= w_{21}^1 - \alpha\frac{\partial L}{\partial w_{21}^1} \\
\frac{\partial L}{\partial w_{31}^2} &= \frac{\partial L}{\partial \hat{y}_1}\frac{\partial \hat{y}_1}{\partial w_{31}^2} = (0.6671-0.31)(0.64) = 0.2285 \\
\frac{\partial L}{\partial w_{21}^1} &= \frac{\partial \hat{y}_2}{\partial \hat{y}_1}\frac{\partial \hat{y}_1}{\partial h_1}\frac{\partial h_1}{\partial w_{21}^1} = (0.6671-0.31)(0.7)(0.02271) = 0.0056 \\
\therefore w_{31}^2 &= 0.33 - 0.5 * 0.2285 = 0.2157 \\
w_{21}^1 &= 0.25 - 0.5 * 0.0056 = 0.2472
\end{aligned}$$

# Question 2

$$f(x_1, x_2) = tanh\left(\frac{x_1}{x_2}\right) + sigmoid(x_1)$$

The above function can be considered as neural network with the type,
Input Layer: $x_1$, $x_2$
Hidden Layer: $h_1$, $h_2$, where $h_1 = x_1/x_2$, $h_2 = x_1$
Output Layer: $y$, where $y = \phi_1(h_1) + \phi_2(h_2)$, $\phi_1(x) = tanh(x)$, $\phi_2(x) = sigmoid(x)$

**Reverse mode automatic differentiation**

1. Derivative of output node with respect to the nearest layer:

$$\frac{\partial y}{\partial \phi_1(h_1)} = 1, \frac{\partial y}{\partial \phi_2(h_2)} = 1 \text{ as the output node activation function is the identity function}$$

2. Derivative of activation function with respect to corresponding node:

$$\frac{\partial \phi_1(h_1)}{\partial h_1} = \frac{\partial tanh(h_1)}{\partial h_1}$$
$$\frac{\partial \phi_2(h_2)}{\partial h_2} = \frac{\partial sigmoid(h_2)}{\partial h_2}$$

3. Derivative of hidden layers with respect to input:

$$\frac{\partial h_1}{\partial x_1} = \frac{1}{x_2}, \frac{\partial h_1}{\partial x_2} = -\frac{x_1}{x_2^2}$$
$$\frac{\partial h_2}{\partial x_1} = 1, \frac{\partial h_2}{\partial x_2} = 0$$

The final derivative can be computed by

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial \phi_1(h_1)}.\frac{\partial \phi_1(h_1)}{\partial h_1}.\frac{\partial h_1}{\partial x_1} + \frac{\partial y}{\partial \phi_2(h_2)}.\frac{\partial \phi_2(h_2)}{\partial h_2}.\frac{\partial h_2}{\partial x_2}$$
$$\frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{\partial y}{\partial x_2} = \frac{\partial y}{\partial \phi_1(h_1)}.\frac{\partial \phi_1(h_1)}{\partial h_1}.\frac{\partial h_1}{\partial x_2}\left(\text{as } \frac{\partial h_2}{\partial x_2} = 0\right)$$