Deep Learning for Visual Recognition
Assignment 2: Neural Networks

Bilge Ulusay
Ewald Bindereif
Ulvi Shukurzade
Cüneyt Erem

November 29, 2020

1 Theoretical Exercises

1.1

We first derive an defining equation for each line of the four lines describing the given trapezoid.

1.

$$\begin{aligned}
\text{slope}_1 &= \frac{2-(-2)}{8-(-4)} = \frac{1}{3} \\
x_2 &= \frac{1}{3}x_1 + \text{intercept} \\
\text{intercept} &= -\frac{2}{3} \text{ (by applying one of the two points)} \\
0 &= 2 - x_1 + 3x_2
\end{aligned} \tag{1}$$

2.

$$\begin{aligned}
\text{slope}_2 &= \frac{2-(-2)}{-2-(-4)} = 2 \\
x_2 &= 2x_1 + \text{intercept} \\
\text{intercept} &= 6 \text{ (by applying one of the two points)} \\
0 &= 6 + 2x_1 - x_2
\end{aligned} \tag{2}$$

3.

$$\begin{aligned}
\text{slope}_3 &= \frac{4-2}{2-(-2)} = \frac{1}{2} \\
x_2 &= \frac{1}{2}x_1 + \text{intercept} \\
\text{intercept} &= 3 \text{ (by applying one of the two points)} \\
0 &= 6 + x_1 - 2x_2
\end{aligned} \tag{3}$$

4.

$$\begin{aligned}
\text{slope}_4 &= \frac{2-4}{8-2} = -\frac{1}{3} \\
x_2 &= -\frac{1}{3}x_1 + \text{intercept} \\
\text{intercept} &= \frac{14}{3} \text{ (by applying one of the two points)} \\
0 &= 14 - x_1 - 3x_2
\end{aligned} \tag{4}$$

The right hand sides of equations (1), (2), (3) and (4) of points on different sides of a single line have opposite signs. If a point belongs to class A (the region inside) all these right hand sides are positive. If a point belongs to class B (the region outside) at least one of these right hand sides are negative.
We choose the following architecture:

- 2 input units: coordinate x_1 and coordinate x_2;
- 4 units in the hidden layer: h_1, h_2, h_3 and h_4;
- 1 output unit y ($y = 1$ if the input belongs to class A and 0 if the input belongs to class B).

We choose a function σ as transfer function which equals 1 for input arguments greater or equal 0 and equals 0 otherwise. The four hidden units are given by:

$$\begin{aligned} h_1 &= \sigma(2 - x_1 + 3x_2), \\ h_2 &= \sigma(6 + 2x_1 - x_2), \\ h_3 &= \sigma(6 + x_1 - 2x_2), \\ h_4 &= \sigma(14 - x_1 - 3x_2). \end{aligned}$$

The output unit can be chosen as follows:

$$y = \sigma(-3.5 + h_1 + h_2 + h_3 + h_4).$$

All relevant parameters/weights can be taken directly from the equations above.

1.2

First Forward Pass

We first calculate the total net input for h_1 and h_2:

$$\begin{aligned} \text{net}_{h_1} &= b_{h_1} + w_1 x_1 + w_2 x_2 = 0.3 + 0.1 \cdot 0.1 + 0.2 \cdot 0.4 = 0.39, \\ \text{net}_{h_2} &= b_{h_2} + w_3 x_1 + w_4 x_2 = 0.3 + 0.2 \cdot 0.1 + 0.3 \cdot 0.4 = 0.44. \end{aligned}$$

We then insert it into the Sigmoid function σ to get the output of h_1 and h_2:

$$\begin{aligned} \text{out}_{h_1} &= \sigma(\text{net}_{h_1}) = \frac{1}{1 + \exp(-0.39)} = 0.5963, \\ \text{out}_{h_2} &= \sigma(\text{net}_{h_2}) = \frac{1}{1 + \exp(-0.44)} = 0.6083. \end{aligned}$$

We repeat this process for the output layer neurons using out_{h_1} and out_{h_2} as inputs and the Softmax function:

$$\begin{aligned} \text{net}_{o_1} &= b_{o_1} + w_5 \text{out}_{h_1} + w_6 \text{out}_{h_2} = 0.6 + 0.4 \cdot 0.5963 + 0.5 \cdot 0.6083 = 1.1427, \\ \text{net}_{o_2} &= b_{o_2} + w_7 \text{out}_{h_1} + w_8 \text{out}_{h_2} = 0.6 + 0.5 \cdot 0.5963 + 0.6 \cdot 0.6083 = 1.2631; \end{aligned}$$

$$\begin{aligned} \text{out}_{o_1} &= \text{softmax}(\text{net}_{o_1}) = \frac{\exp(1.1427)}{\exp(1.1427) + \exp(1.2631)} = 0.4699, \\ \text{out}_{o_2} &= \text{softmax}(\text{net}_{o_2}) = \frac{\exp(1.2631)}{\exp(1.1427) + \exp(1.2631)} = 0.5301. \end{aligned}$$

The total error equals

$$E_{\text{total}} = \text{MSE} = E_1 + E_2 = \frac{1}{2}\Big((0.1 - 0.4699)^2 + (0.9 - 0.5301)^2\Big) = 0.1368.$$

Backpropagation – Output Layer

First, we consider w_5. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_5} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial w_5}.$$

$$\frac{\partial E_{\text{total}}}{\partial \text{out}_{o_1}} = -(0.1 - 0.4699) = 0.3966$$

$$\frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} = \text{softmax}(\text{net}_{o_1}) \cdot (1 - \text{softmax}(\text{net}_{o_1})) = 0.2491$$

$$\frac{\partial \text{net}_{o_1}}{\partial w_5} = \text{out}_{h_1} = 0.5963$$

$$\frac{\partial E_{\text{total}}}{\partial w_5} = 0.3966 \cdot 0.2491 \cdot 0.5963 = 0.0589$$

$$w_5 = w_5 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_5} = 0.4 - 0.5 \cdot 0.0589 = 0.3706$$

Now we consider w_6. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_6} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial w_6}.$$

$$\frac{\partial E_{\text{total}}}{\partial \text{out}_{o_1}} = -(0.1 - 0.4699) = 0.3966$$

$$\frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} = \text{softmax}(\text{net}_{o_1}) \cdot (1 - \text{softmax}(\text{net}_{o_1})) = 0.2491$$

$$\frac{\partial \text{net}_{o_1}}{\partial w_6} = \text{out}_{h_2} = 0.6083$$

$$\frac{\partial E_{\text{total}}}{\partial w_5} = 0.3966 \cdot 0.2491 \cdot 0.6083 = 0.0601$$

$$w_6 = w_6 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_6} = 0.5 - 0.5 \cdot 0.0601 = 0.4700$$

Now we consider w_7. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_7} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial w_7}.$$

$$\frac{\partial E_{\text{total}}}{\partial \text{out}_{o_2}} = -(0.9 - 0.5301) = -0.3699$$

$$\frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} = \text{softmax}(\text{net}_{o_2}) \cdot (1 - \text{softmax}(\text{net}_{o_2})) = 0.2491$$

$$\frac{\partial \text{net}_{o_2}}{\partial w_7} = \text{out}_{h_1} = 0.5963$$

$$\frac{\partial E_{\text{total}}}{\partial w_5} = -0.3699 \cdot 0.2491 \cdot 0.5963 = -0.0549$$

$$w_7 = w_7 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_7} = 0.5 - 0.5 \cdot (-0.0549) = 0.5275$$

Now we consider w_8. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_8} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial w_8}.$$

$$\frac{\partial E_{\text{total}}}{\partial \text{out}_{o_2}} = -(0.9 - 0.5301) = -0.3699$$

$$\frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} = \text{softmax}(\text{net}_{o_2}) \cdot (1 - \text{softmax}(\text{net}_{o_2})) = 0.2491$$

$$\frac{\partial \text{net}_{o_2}}{\partial w_8} = \text{out}_{h_2} = 0.6083$$

$$\frac{\partial E_{\text{total}}}{\partial w_8} = -0.3966 \cdot 0.2491 \cdot 0.6083 = -0.0601$$

$$w_8 = w_8 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_8} = 0.6 - 0.5 \cdot (-0.0601) = 0.6301$$

Now we consider b_{o_1} (connected with o_1). By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial b_{o_1}} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial b_{o_1}}.$$

$$\frac{\partial E_{\text{total}}}{\partial \text{out}_{o_1}} = -(0.1 - 0.4699) = 0.3966$$

$$\frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} = \text{softmax}(\text{net}_{o_1}) \cdot (1 - \text{softmax}(\text{net}_{o_1})) = 0.2491$$

$$\frac{\partial \text{net}_{o_1}}{\partial b_{o_1}} = 1$$

$$\frac{\partial E_{\text{total}}}{\partial w_8} = 0.3966 \cdot 0.2491 \cdot 1 = 0.0988$$

$$b_{o_1} = b_{o_1} - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial b_{o_1}} = 0.6 - 0.5 \cdot 0.0988 = 0.5506$$

Now we consider b_{o_2} (connected with o_2). By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial b_{o_2}} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial b_{o_2}}.$$

$$\frac{\partial E_{\text{total}}}{\partial \text{out}_{o_2}} = -(0.9 - 0.5301) = -0.3699$$

$$\frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} = \text{softmax}(\text{net}_{o_2}) \cdot (1 - \text{softmax}(\text{net}_{o_2})) = 0.2491$$

$$\frac{\partial \text{net}_{o_2}}{\partial b_{o_2}} = 1$$

$$\frac{\partial E_{\text{total}}}{\partial w_8} = -0.3699 \cdot 0.2491 \cdot 1 = 0.0921$$

$$b_{o_2} = b_{o_2} - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial b_{o_2}} = 0.6 - 0.5 \cdot 0.0921 = 0.5540$$

Backpropagation – Hidden Layer

First, we consider w_1. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_1} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{h_1}} \cdot \frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} \cdot \frac{\partial \text{net}_{h_1}}{\partial w_1} = \left(\frac{\partial E_1}{\partial \text{out}_{h_1}} + \frac{\partial E_2}{\partial \text{out}_{h_1}}\right) \cdot \frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} \cdot \frac{\partial \text{net}_{h_1}}{\partial w_1}.$$

$$\begin{aligned}
\frac{\partial E_1}{\partial \text{out}_{h_1}} &= \frac{\partial E_1}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial \text{out}_{h_1}} \\
&= \frac{\partial E_1}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot w_5 \\
&= -(0.1 - 0.4699) \cdot 0.2491 \cdot 0.4 = 0.0369
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E_2}{\partial \text{out}_{h_1}} &= \frac{\partial E_2}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial \text{out}_{h_1}} \\
&= \frac{\partial E_2}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot w_7 \\
&= -(0.9 - 0.5301) \cdot 0.2491 \cdot 0.5 = 0.0461
\end{aligned}$$

$$\frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} = \text{out}_{h_1} \cdot (1 - \text{out}_{h_1}) = 0.2407$$

$$\frac{\partial \text{net}_{h_1}}{\partial w_1} = x_1 = 0.1$$

$$\frac{\partial E_{\text{total}}}{\partial w_1} = (0.0369 + 0.0461) \cdot 0.2407 \cdot 0.1 = 0.0020$$

$$w_1 = w_1 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_1} = 0.1 - 0.5 \cdot 0.0020 = 0.099$$

Now we consider w_2. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_2} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{h_1}} \cdot \frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} \cdot \frac{\partial \text{net}_{h_1}}{\partial w_2} = \left(\frac{\partial E_1}{\partial \text{out}_{h_1}} + \frac{\partial E_2}{\partial \text{out}_{h_1}} \right) \cdot \frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} \cdot \frac{\partial \text{net}_{h_1}}{\partial w_2}.$$

$$\begin{aligned}
\frac{\partial E_1}{\partial \text{out}_{h_1}} &= \frac{\partial E_1}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial \text{out}_{h_1}} \\
&= \frac{\partial E_1}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot w_5 \\
&= -(0.1 - 0.4699) \cdot 0.2491 \cdot 0.4 = 0.0369
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E_2}{\partial \text{out}_{h_1}} &= \frac{\partial E_2}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial \text{out}_{h_1}} \\
&= \frac{\partial E_2}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot w_7 \\
&= -(0.9 - 0.5301) \cdot 0.2491 \cdot 0.5 = 0.0461
\end{aligned}$$

$$\frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} = \text{out}_{h_1} \cdot (1 - \text{out}_{h_1}) = 0.2407$$

$$\frac{\partial \text{net}_{h_1}}{\partial w_2} = x_2 = 0.4$$

$$\frac{\partial E_{\text{total}}}{\partial w_2} = (0.0369 + 0.0461) \cdot 0.2407 \cdot 0.4 = 0.0080$$

$$w_2 = w_2 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_2} = 0.2 - 0.5 \cdot 0.0080 = 0.196$$

Now we consider w_3. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_3} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{h_2}} \cdot \frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} \cdot \frac{\partial \text{net}_{h_2}}{\partial w_3} = \left(\frac{\partial E_1}{\partial \text{out}_{h_2}} + \frac{\partial E_2}{\partial \text{out}_{h_2}} \right) \cdot \frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} \cdot \frac{\partial \text{net}_{h_2}}{\partial w_3}.$$

$$\begin{aligned}
\frac{\partial E_1}{\partial \text{out}_{h_2}} &= \frac{\partial E_1}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial \text{out}_{h_2}} \\
&= \frac{\partial E_1}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot w_6 \\
&= -(0.1 - 0.5963) \cdot 0.2491 \cdot 0.5 = 0.0618
\end{aligned}$$

$$\begin{aligned}\frac{\partial E_2}{\partial \text{out}_{h_2}} &= \frac{\partial E_2}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial \text{out}_{h_2}} \\ &= \frac{\partial E_2}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot w_8 \\ &= -(0.9 - 0.5301) \cdot 0.2491 \cdot 0.6 = -0.0553\end{aligned}$$

$$\frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} = \text{out}_{h_2} \cdot (1 - \text{out}_{h_2}) = 0.2383$$

$$\frac{\partial \text{net}_{h_2}}{\partial w_3} = x_1 = 0.1$$

$$\frac{\partial E_{\text{total}}}{\partial w_3} = (0.0618 - 0.0553) \cdot 0.2383 \cdot 0.1 = 0.0002$$

$$w_3 = w_3 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_3} = 0.2 - 0.5 \cdot 0.0002 = 0.1999$$

Now we consider w_4. By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial w_4} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{h_2}} \cdot \frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} \cdot \frac{\partial \text{net}_{h_2}}{\partial w_4} = \left(\frac{\partial E_1}{\partial \text{out}_{h_2}} + \frac{\partial E_2}{\partial \text{out}_{h_2}}\right) \cdot \frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} \cdot \frac{\partial \text{net}_{h_2}}{\partial w_4}.$$

$$\begin{aligned}\frac{\partial E_1}{\partial \text{out}_{h_2}} &= \frac{\partial E_1}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial \text{out}_{h_2}} \\ &= \frac{\partial E_1}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot w_6 \\ &= -(0.1 - 0.5963) \cdot 0.2491 \cdot 0.5 = 0.0618\end{aligned}$$

$$\begin{aligned}\frac{\partial E_2}{\partial \text{out}_{h_2}} &= \frac{\partial E_2}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial \text{out}_{h_2}} \\ &= \frac{\partial E_2}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot w_8 \\ &= -(0.9 - 0.5301) \cdot 0.2491 \cdot 0.6 = -0.0553\end{aligned}$$

$$\frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} = \text{out}_{h_2} \cdot (1 - \text{out}_{h_2}) = 0.2383$$

$$\frac{\partial \text{net}_{h_2}}{\partial w_4} = x_2 = 0.4$$

$$\frac{\partial E_{\text{total}}}{\partial w_4} = (0.0618 - 0.0553) \cdot 0.2383 \cdot 0.4 = 0.0006$$

$$w_4 = w_4 - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial w_4} = 0.3 - 0.5 \cdot 0.0006 = 0.2997$$

Now we consider b_{h_1} (connected with h_1). By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial b_{h_1}} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{h_1}} \cdot \frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} \cdot \frac{\partial \text{net}_{h_1}}{\partial b_{h_1}} = \left(\frac{\partial E_1}{\partial \text{out}_{h_1}} + \frac{\partial E_2}{\partial \text{out}_{h_1}}\right) \cdot \frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} \cdot \frac{\partial \text{net}_{h_1}}{\partial b_{h_1}}.$$

$$\begin{aligned}\frac{\partial E_1}{\partial \text{out}_{h_1}} &= \frac{\partial E_1}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial \text{out}_{h_1}} \\ &= \frac{\partial E_1}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot w_5 \\ &= -(0.1 - 0.4699) \cdot 0.2491 \cdot 0.4 = 0.0369\end{aligned}$$

$$\begin{aligned}\frac{\partial E_2}{\partial \text{out}_{h_1}} &= \frac{\partial E_2}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial \text{out}_{h_1}} \\ &= \frac{\partial E_2}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot w_7 \\ &= -(0.9 - 0.5301) \cdot 0.2491 \cdot 0.5 = -0.0461\end{aligned}$$

$$\frac{\partial \text{out}_{h_1}}{\partial \text{net}_{h_1}} = \text{out}_{h_1} \cdot (1 - \text{out}_{h_1}) = 0.2407$$

$$\frac{\partial \text{net}_{h_1}}{\partial b_{h_1}} = 1$$

$$\frac{\partial E_{\text{total}}}{\partial b_{h_1}} = (0.0369 - 0.0461) \cdot 0.2407 = -0.0022$$

$$b_{h_1} = b_{h_1} - 0.5 \cdot \frac{\partial E_{\text{total}}}{\partial b_{h_1}} = 0.3 - 0.5 \cdot (-0.0022) = 0.3011$$

Now we consider b_{h_2} (connected with h_2). By applying the chain rule we get:

$$\frac{\partial E_{\text{total}}}{\partial b_{h_2}} = \frac{\partial E_{\text{total}}}{\partial \text{out}_{h_2}} \cdot \frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} \cdot \frac{\partial \text{net}_{h_2}}{\partial b_{h_2}} = \left(\frac{\partial E_1}{\partial \text{out}_{h_2}} + \frac{\partial E_2}{\partial \text{out}_{h_2}}\right) \cdot \frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} \cdot \frac{\partial \text{net}_{h_2}}{\partial b_{h_2}}.$$

$$\begin{aligned}\frac{\partial E_1}{\partial \text{out}_{h_2}} &= \frac{\partial E_1}{\partial \text{net}_{o_1}} \cdot \frac{\partial \text{net}_{o_1}}{\partial \text{out}_{h_2}} \\ &= \frac{\partial E_1}{\partial \text{out}_{o_1}} \cdot \frac{\partial \text{out}_{o_1}}{\partial \text{net}_{o_1}} \cdot w_6 \\ &= -(0.1 - 0.5963) \cdot 0.2491 \cdot 0.5 = 0.0618\end{aligned}$$

$$\begin{aligned}\frac{\partial E_2}{\partial \text{out}_{h_2}} &= \frac{\partial E_2}{\partial \text{net}_{o_2}} \cdot \frac{\partial \text{net}_{o_2}}{\partial \text{out}_{h_2}} \\ &= \frac{\partial E_2}{\partial \text{out}_{o_2}} \cdot \frac{\partial \text{out}_{o_2}}{\partial \text{net}_{o_2}} \cdot w_8 \\ &= -(0.9 - 0.5301) \cdot 0.2491 \cdot 0.6 = -0.0553\end{aligned}$$

$$\frac{\partial \text{out}_{h_2}}{\partial \text{net}_{h_2}} = \text{out}_{h_2} \cdot (1 - \text{out}_{h_2}) = 0.2383$$

$$\frac{\partial \mathrm{net}_{h_2}}{\partial b_{h_2}} = 1$$

$$\frac{\partial E_{\mathrm{total}}}{\partial b_{h_2}} = (0.0618 - 0.0553) \cdot 0.2383 = 0.0015$$

$$b_{h_2} = b_{h_2} - 0.5 \cdot \frac{\partial E_{\mathrm{total}}}{\partial b_{h_2}} = 0.3 - 0.5 \cdot 0.0015 = 0.2993$$

Second Forward Pass

We first calculate the total net input for h_1 and h_2:

$$\begin{aligned} \mathrm{net}_{h_1} &= b_{h_1} + w_1 x_1 + w_2 x_2 = 0.3011 + 0.099 \cdot 0.1 + 0.196 \cdot 0.4 = 0.3894, \\ \mathrm{net}_{h_2} &= b_{h_2} + w_3 x_1 + w_4 x_2 = 0.2993 + 0.1999 \cdot 0.1 + 0.2997 \cdot 0.4 = 0.4392. \end{aligned}$$

We then insert it into the Sigmoid function to get the output of h_1 and h_2:

$$\begin{aligned} \mathrm{out}_{h_1} &= \sigma(\mathrm{net}_{h_1}) = \frac{1}{1 + \exp(-0.3894)} = 0.5961, \\ \mathrm{out}_{h_2} &= \sigma(\mathrm{net}_{h_2}) = \frac{1}{1 + \exp(-0.4392)} = 0.6081. \end{aligned}$$

We repeat this process for the output layer neurons using out_{h_1} and out_{h_2} as inputs and the Softmax function:

$$\begin{aligned} \mathrm{net}_{o_1} &= b_{o_1} + w_5 \mathrm{out}_{h_1} + w_6 \mathrm{out}_{h_2} = 0.5506 + 0.3706 \cdot 0.5961 + 0.4700 \cdot 0.6081 = 1.0573, \\ \mathrm{net}_{o_2} &= b_{o_2} + w_7 \mathrm{out}_{h_1} + w_8 \mathrm{out}_{h_2} = 0.5540 + 0.5275 \cdot 0.5961 + 0.6301 \cdot 0.6081 = 1.2516; \end{aligned}$$

$$\begin{aligned} \mathrm{out}_{o_1} &= \mathrm{softmax}(\mathrm{net}_{o_1}) = \frac{\exp(1.0573)}{\exp(1.0573) + \exp(1.2516)} = 0.4516, \\ \mathrm{out}_{o_2} &= \mathrm{softmax}(\mathrm{net}_{o_2}) = \frac{\exp(1.2516)}{\exp(1.0573) + \exp(1.2516)} = 0.5484. \end{aligned}$$

The total error equals

$$E_{\mathrm{total}} = \mathrm{MSE} = E_1 + E_2 = \frac{1}{2}\Big((0.1 - 0.4516)^2 + (0.9 - 0.5484)^2\Big) = 0.1236,$$

which is lower than in the first forward pass.