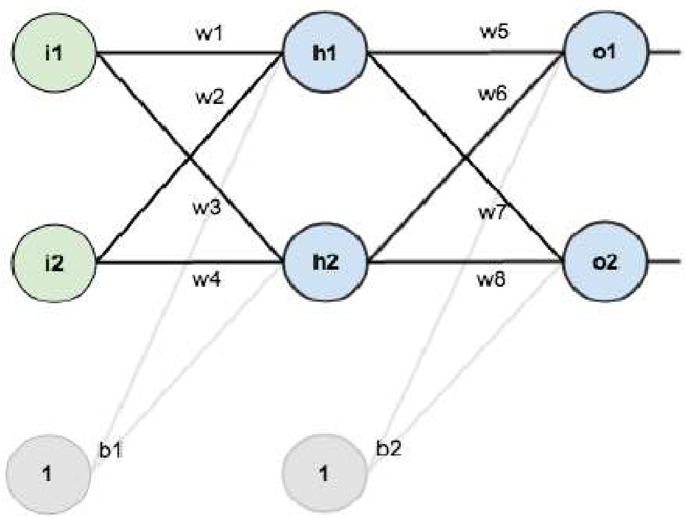
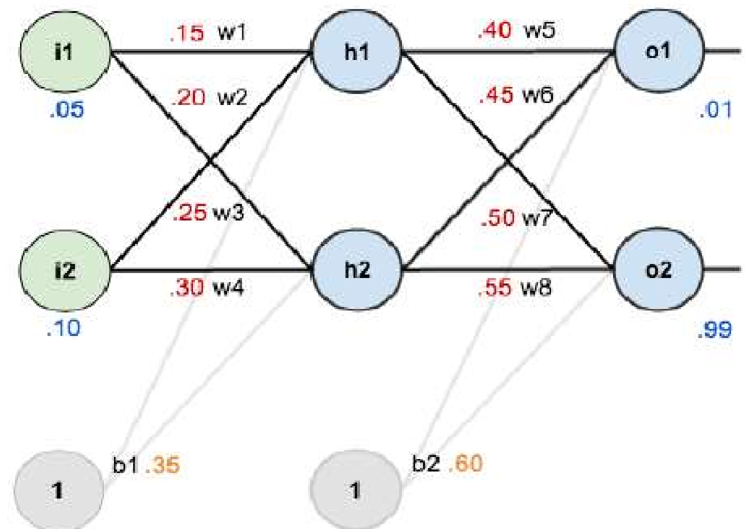
Train the model and compute gradient





Train the model and compute gradient

Forward pass (first layer):

$$net_{h_1} = w_1 * i_1 + w_2 * i_2 + b_1 * 1$$

$$net_{h_1} = 0.15 * 0.05 + 0.2 * 0.1 + 0.35 * 1 = 0.3775$$

$$out_{h_1} = \frac{1}{1+e^{-net_{h_1}}} = \frac{1}{1+e^{-0.3775}} = 0.593269992$$

$$out_{h_2} = 0.596884378$$

Forward pass (second layer):

$$net_{o_1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$net_{o_1} = 0.4 * 0.593269992 + 0.45 * 0.596884378 + 0.6 * 1 = 1.105905967$$

$$out_{o_1} = \frac{1}{1+e^{-net_{o_1}}} = \frac{1}{1+e^{-1.105905967}} = 0.75136507$$

Train the model and compute gradient

Calculating the total error:

$$E_{total} = \sum \frac{1}{2}(target - output)^2$$

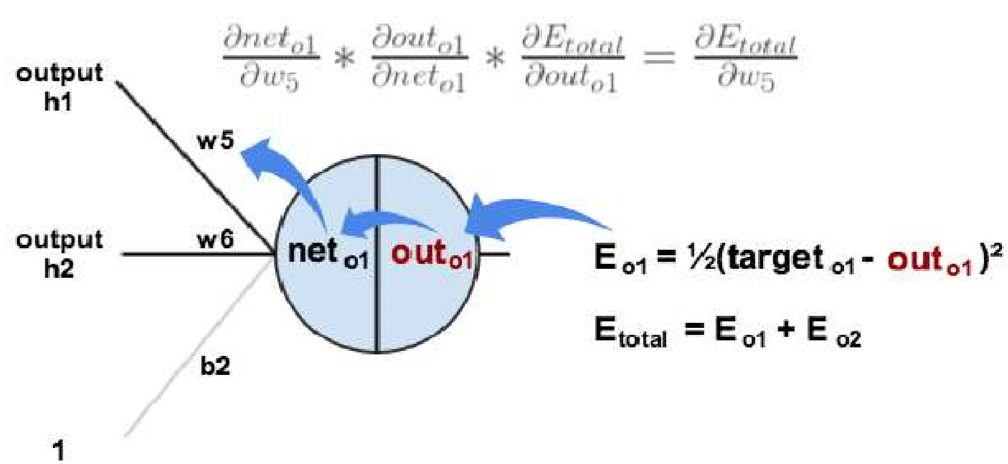
$$E_{total} = \sum \frac{1}{2}(target_{o_1} - output_{o_1})^2 = \frac{1}{2}(0.01 - 0.75136507)^2 = 0.274811083$$

$$E_{o_2} = 0.023560026$$

$$E_{total} = E_{o_2} + E_{o_2} = 0.274811083 + 0.023560026 = 0.298371109$$

Train the model and compute gradient

Backpropagation - gradient computation



Train the model and compute gradient

Backpropagation - gradient computation

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial out_{o_1}} * \frac{\partial out_{o_1}}{\partial net_{o_1}} * \frac{\partial net_{o_1}}{\partial w_5}$$

$$\frac{\partial E_{total}}{\partial out_{o_1}} = 2 * \frac{1}{2}(target_{o_1} - out_{o_1})^{2-1} * -1 + 0 \quad \rightarrow \quad \frac{\partial E_{total}}{\partial out_{o_1}} = -\left(target_{o_1} - out_{o_1}\right) = -(0.01 - 0.75136507) = 0.741365$$

$$out_{o_1} = \frac{1}{1 + e^{-net_{o_1}}} \quad \rightarrow \quad \frac{\partial out_{o_1}}{\partial net_{o_1}} = out_{o_1}\left(1 - out_{o_1}\right) = 0.75136507 * (1 - 0.75136507) = 0.186815602$$

$$net_{o_1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1 \quad \rightarrow \quad \frac{\partial net_{o_1}}{\partial w_5} = 1 * out_{h_1} * w_5^{(1-1)} + 0 + 0 = out_{h_1} = 0.593269992$$

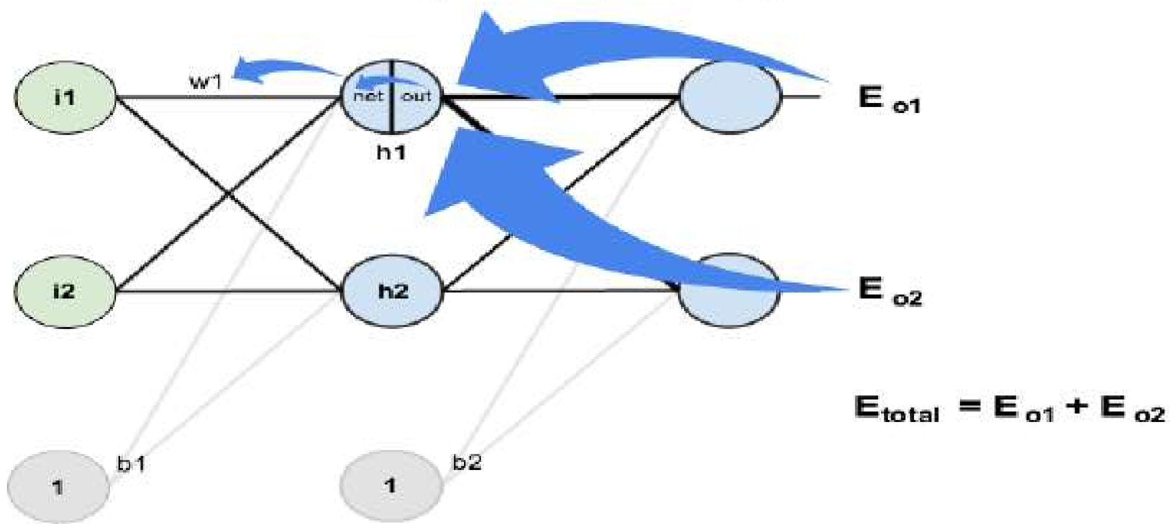
$$\frac{\partial E_{total}}{\partial w_5} = 0.74136507 * 0.186815602 * 0.593269992 = 0.082167041$$

$$w_5^{+} = w_5 - \alpha * \frac{\partial E_{total}}{\partial w_5} = 0.4 - 0.5 * 0.082167041 = 0.35891648$$

Train the model and compute gradient

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h1}} * \frac{\partial out_{h1}}{\partial net_{h1}} * \frac{\partial net_{h1}}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial out_{h1}} = \frac{\partial E_{o1}}{\partial out_{h1}} + \frac{\partial E_{o2}}{\partial out_{h1}}$$



Train the model and compute gradient

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h_1}} * \frac{\partial out_{h_1}}{\partial net_{h_1}} * \frac{\partial net_{h_1}}{\partial w_1}$$

$$\frac{\partial E_{total}}{\partial out_{h_1}} = \frac{\partial E_{o_1}}{\partial out_{h_1}} + \frac{\partial E_{o_2}}{\partial out_{h_1}}$$

$$\frac{\partial E_{o_1}}{\partial out_{h_1}} = \frac{\partial E_{o_1}}{\partial net_{o_1}} * \frac{\partial net_{o_1}}{\partial out_{h_1}}$$

$$\frac{\partial E_{o_1}}{\partial net_{o_1}} = \frac{\partial E_{o_1}}{\partial out_{o_1}} * \frac{\partial out_{o_1}}{\partial net_{o_1}} = 0.741365 * 0.186815602 = 0.138498562$$

Train the model and compute gradient

$$net_{o_1} = w_5 * out_{h1} + w_6 * out_{h2} + b_2 * 1$$

$$\frac{\partial net_{o_1}}{\partial out_{h_1}} = w_5 = 0.40$$

$$\frac{\partial E_{o_1}}{\partial out_{h_1}} = \frac{\partial E_{o_1}}{\partial net_{o_1}} * \frac{\partial net_{o_1}}{\partial out_{h_1}} = 0.138498562 * 0.40 = 0.055399425$$

$$\frac{\partial E_{o_2}}{\partial out_{h_1}} = -0.019049119$$

$$\frac{\partial E_{total}}{\partial out_{h_1}} = \frac{\partial E_{o_1}}{\partial out_{h_1}} + \frac{\partial E_{o_2}}{\partial out_{h_1}} = 0.055399425 - 0.019049119$$

Train the model and compute gradient

$$out_{h_1} = \frac{1}{1 + e^{-net_{h_1}}}$$

$$\frac{\partial out_{h_1}}{\partial net_{h_1}} = out_{h_1}\left(1 - out_{h_1}\right) = 0.59326999 * (1 - 0.59326999) = 0.241300709$$

$$net_{h_1} = w_1 * i_1 + w_3 * i_2 + b_1 * 1 \quad \rightarrow \quad \frac{\partial net_{h_1}}{\partial w_1} = i_1 = 0.05$$

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial out_{h_1}} * \frac{\partial out_{h_1}}{\partial net_{h_1}} * \frac{\partial net_{h_1}}{\partial w_1} \quad \rightarrow \quad \frac{\partial E_{total}}{\partial w_1} = 0.036350306 * 0.241300709 * 0.05 = 0.000438568$$

$$w_1^+ = w_1 - \alpha * \frac{\partial E_{total}}{\partial w_1} = 0.15 - 0.5 * 0.000438568 = 0.149780716$$

Problems in training the DL models

1. **Vanishing Gradient:** Text generation, machine translation, and stock market prediction are just a few examples of the time-dependent and sequential data. The gradient problem makes training difficult.
2. **Exploding Gradient:** An Exploding Gradient occurs when a neural network is being trained and the slope tends to grow exponentially rather than decay. Large error gradients lead to very large updates to the model weights