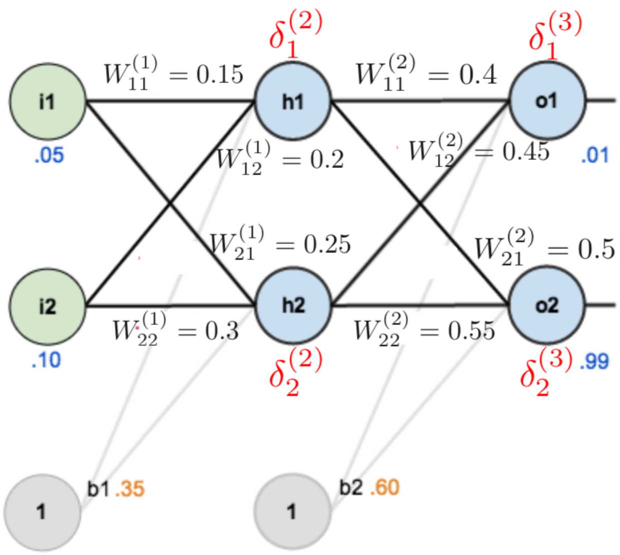
Practice Problem

Perform one iteration of forward propagation and backpropagation for the following neural network, assuming learning rate $\alpha = 0.5$ and a single training sample $\mathcal{X} = \{(i_1, i_2), (y_1, y_2)\}$, where $i_1 = 0.05$, $i_2 = 0.1$, $y_1 = 0.01$, and $y_2 = 0.99$. We also assume a sigmoid activation function $g(x) = \frac{1}{1+e^{-x}}$ for the hidden nodes, where $g'(x) = g(x)[1 - g(x)]$.



Forward propagation

$$h_1 = g(z_1^{(2)}) = g(W_{11}^{(1)} i_1 + W_{12}^{(1)} i_2 + b_1) = g(0.15 \cdot 0.05 + 0.2 \cdot 0.10 + 0.35) = 0.5932$$

$$h_2 = g(z_2^{(2)}) = g(W_{21}^{(1)} i_1 + W_{22}^{(1)} i_2 + b_1) = g(0.25 \cdot 0.05 + 0.3 \cdot 0.10 + 0.35) = 0.5968$$

$$o_1 = g(z_1^{(3)}) = g(W_{11}^{(2)} h_1 + W_{12}^{(2)} h_2 + b_2) = g(0.4 \cdot 0.5932 + 0.45 \cdot 0.5968 + 0.60) = 0.7513$$

$$o_2 = g(z_2^{(3)}) = g(W_{21}^{(2)} h_1 + W_{22}^{(2)} h_2 + b_2) = g(0.5 \cdot 0.5932 + 0.55 \cdot 0.5968 + 0.60) = 0.7729$$

Backpropagation Outline

- For each node i in output layer L: $\delta_i^{(L)} = (\alpha_i^{(L)} - y_n)f'(z_i^{(L)})$
- For each (hidden) node i in layer $l = L-1, L-2, \ldots, 2$: $\delta_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)}\right) f'(z_i^{(l)})$
- Compute the desired partial derivatives as: $\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{ij}^{(l)}} = \alpha_j^{(l)} \delta_i^{(l+1)}$, $\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta b_i^{(l)}} = \delta_i^{(l+1)}$
- Update the weights as: $W_{ij}^{(l)} := W_{ij}^{(l)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{ij}^{(l)}}$, $b_i^{(l)} := b_i^{(l)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta b_i^{(l)}}$

Backpropagation

$$\delta_1^{(3)} = (o_1 - y_1)g'(z_1^{(3)}) = (0.7513 - 0.01)g'(0.4 \cdot 0.5932 + 0.45 \cdot 0.5968 + 0.60) = 0.1384$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{11}^{(2)}} = h_1 \delta_1^{(3)} = 0.082167$$

$$W_{11}^{(2)} := W_{11}^{(2)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{11}^{(2)}} = 0.4 - 0.5 \cdot 0.082167 = 0.3589$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{12}^{(2)}} = h_2 \delta_1^{(3)} = 0.082667$$

$$W_{12}^{(2)} := W_{12}^{(2)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{12}^{(2)}} = 0.45 - 0.5 \cdot 0.082667 = 0.4086$$

$$\delta_2^{(3)} = (o_2 - y_2)g'(z_2^{(3)}) = (0.7729 - 0.99)g'(0.5 \cdot 0.5932 + 0.55 \cdot 0.5968 + 0.60) = -0.038098$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{21}^{(2)}} = h_1 \delta_2^{(3)} = -0.02260$$

$$W_{21}^{(2)} := W_{21}^{(2)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{21}^{(2)}} = 0.5 - 0.5 \cdot (-0.02260) = 0.51130$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{22}^{(2)}} = h_2 \delta_2^{(3)} = -0.02244$$

$$W_{22}^{(2)} := W_{22}^{(2)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{22}^{(2)}} = 0.55 - 0.55 \cdot (-0.02244) = 0.561370$$

$$\begin{aligned}\delta_1^{(2)} &= (W_{11}^{(2)} \delta_1^{(3)} + W_{21}^{(2)} \delta_2^{(3)})g'(z_1^{(2)}) \\ &= (0.4 \cdot 0.138498 + 0.5 \cdot (-0.038098)) \cdot g'(0.15 \cdot 0.05 + 0.2 \cdot 0.1 + 0.35) \\ &= 0.008771\end{aligned}$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{11}^{(1)}} = i_1 \delta_1^{(2)} = 0.05 \cdot 0.008771 = 0.0004385$$

$$W_{11}^{(1)} := W_{11}^{(1)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{11}^{(1)}} = 0.15 - 0.5 \cdot 0.0004385 = 0.14978071$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{12}^{(1)}} = i_2 \delta_1^{(2)} = 0.1 \cdot 0.008771 = 0.0008771$$

$$W_{12}^{(1)} := W_{12}^{(1)} - \alpha \frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{12}^{(1)}} = 0.2 - 0.5 \cdot 0.0008771 = 0.19956143$$

$$
\begin{aligned}
\delta_2^{(1)} &= (W_{12}^{(2)}\delta_1^{(3)} + W_{22}^{(2)}\delta_2^{(3)})g'(z_2^{(2)}) \\
&= (0.45 \cdot 0.138498 + 0.55 \cdot (-0.038098)) \cdot g'(0.25 \cdot 0.05 + 0.3 \cdot 0.1 + 0.35) \\
&= 0.009954
\end{aligned}
$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{21}^{(1)}} = i_1\delta_2^{(2)} = 0.05 \cdot 0.009954 = 0.000497712$$

$$W_{21}^{(1)} := W_{21}^{(1)} - \alpha\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{21}^{(1)}} = 0.25 - 0.5 \cdot 0.000497712 = 0.249751$$

$$\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{22}^{(1)}} = i_2\delta_2^{(2)} = 0.1 \cdot 0.009954 = 0.0009954$$

$$W_{22}^{(1)} := W_{22}^{(1)} - \alpha\frac{\vartheta J(\mathbf{W},\mathbf{b})}{\vartheta W_{22}^{(1)}} = 0.3 - 0.5 \cdot 0.0009954 = 0.29950229$$