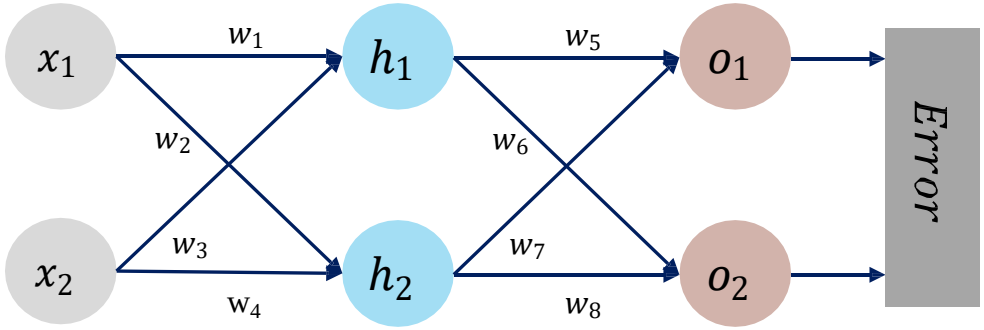
误差反向传播：前馈神经网络

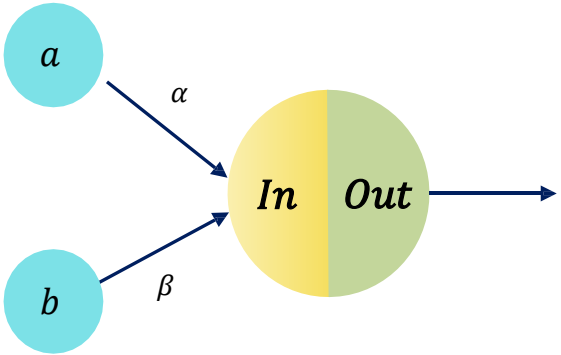


Input Layer

Hidden Layer

Output Layer

w_i Weight

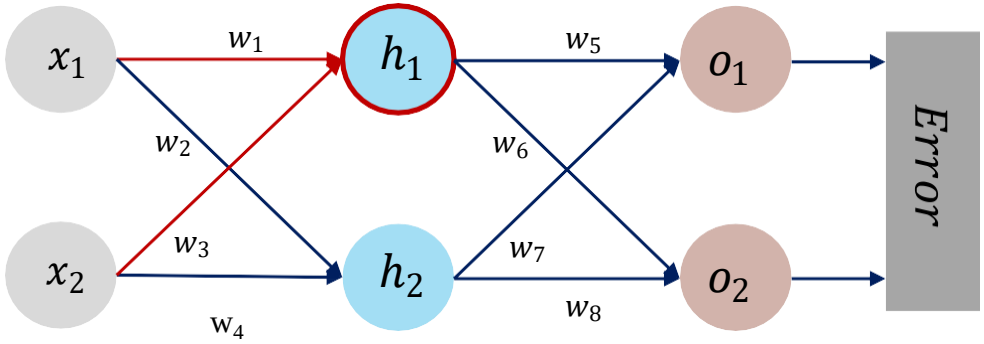


$$In = \alpha \times a + \beta \times b$$

$$Out = Sigmoid(In)$$

误差反向传播：前馈神经网络

正向传播

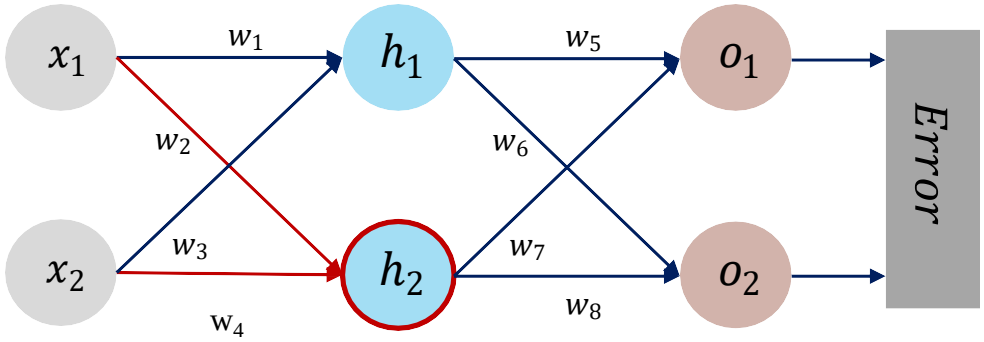


$$In_{h_1} = w_1 \times x_1 + w_3 \times x_2$$

$$h_1 = Out_{h_1} = Sigmoid(In_{h_1})$$

误差反向传播：前馈神经网络

正向传播

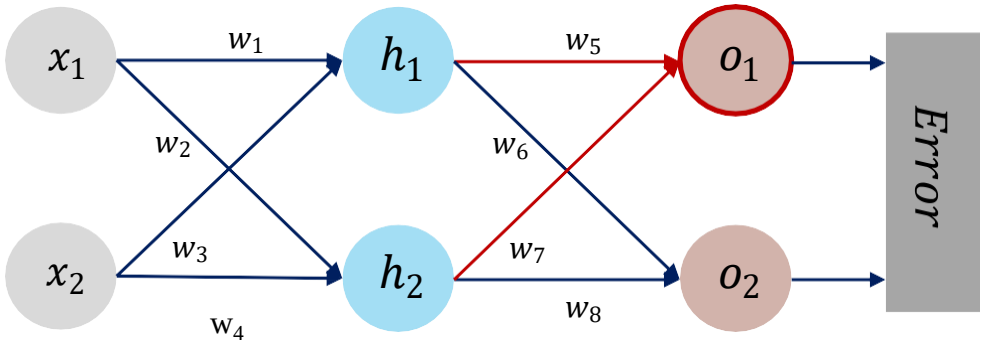


$$In_{h_2} = w_2 \times x_1 + w_4 \times x_2$$

$$h_2 = Out_{h_2} = Sigmoid(In_{h_2})$$

误差反向传播：前馈神经网络

正向传播

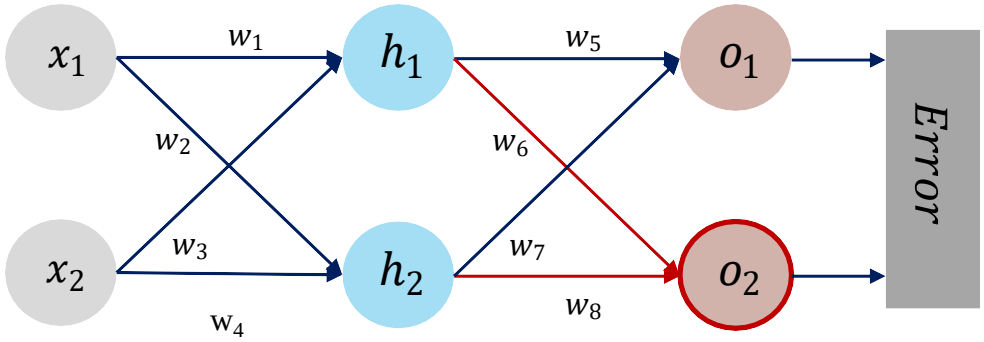


$$In_{o_1} = w_5 \times h_1 + w_7 \times h_2$$

$$o_1 = Out_{o_1} = Sigmoid(In_{o_1})$$

误差反向传播：前馈神经网络

正向传播

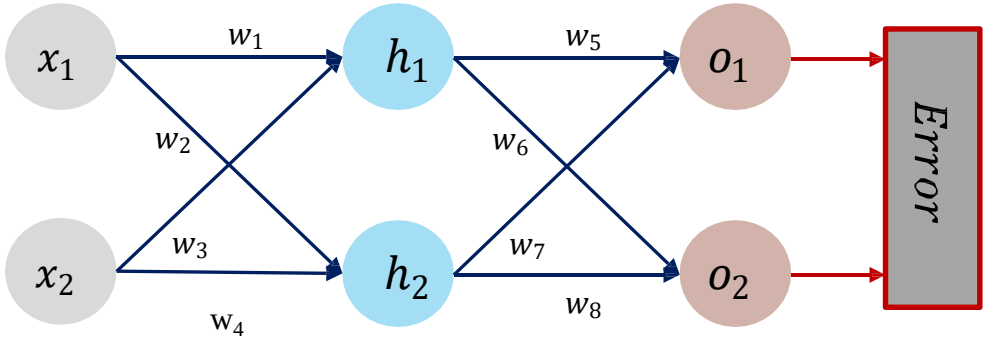


$$In_{o_2} = w_6 \times h_1 + w_8 \times h_2$$

$$o_2 = Out_{o_2} = Sigmoid(In_{o_2})$$

误差反向传播：前馈神经网络

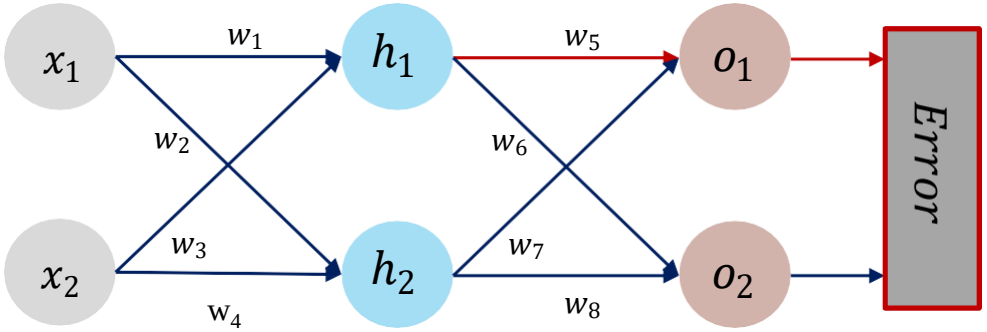
正向传播



$$Error = \frac{1}{2}\sum_{i=1}^{2}(o_i - y_i)^2$$

误差反向传播：前馈神经网络

反向传播：梯度计算



1. 计算w_5~ w_8的梯度(以w_5为例)

$$\delta_5 = \frac{\partial Error}{\partial w_5} = \frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial In_{o_1}} \times \frac{\partial In_{o_1}}{\partial w_5}$$

where,

$$\frac{\partial Error}{\partial o_1} = o_1 - y_1 \quad \Leftarrow \quad Error = \frac{1}{2}\sum_{i=1}^{2}(o_i - y_i)^2$$

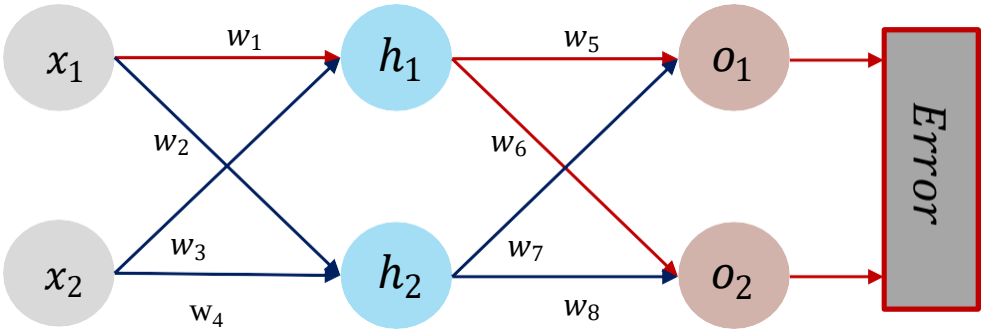
$$\frac{\partial o_1}{\partial In_{o_1}} = o_1 \times (1 - o_1) \quad \Leftarrow \quad o_1 = Out_{o_1} = Sigmoid(In_{o_1})$$

$$\frac{\partial In_{o_1}}{\partial w_5} = h_1 \quad \Leftarrow \quad In_{o_1} = w_5 \times h_1 + w_7 \times h_2$$

误差反向传播：前馈神经网络

反向传播：梯度计算

$$\delta_5 = \frac{\partial Error}{\partial w_5} = \frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial In_{o_1}} \times \frac{\partial In_{o_1}}{\partial w_5}$$



1. 计算w_1~ w_4的梯度(以w_1为例)

$$\delta_1 = \frac{\partial Error}{\partial w_1} = \frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial w_1} + \frac{\partial Error}{\partial o_2} \times \frac{\partial o_2}{\partial w_1}$$

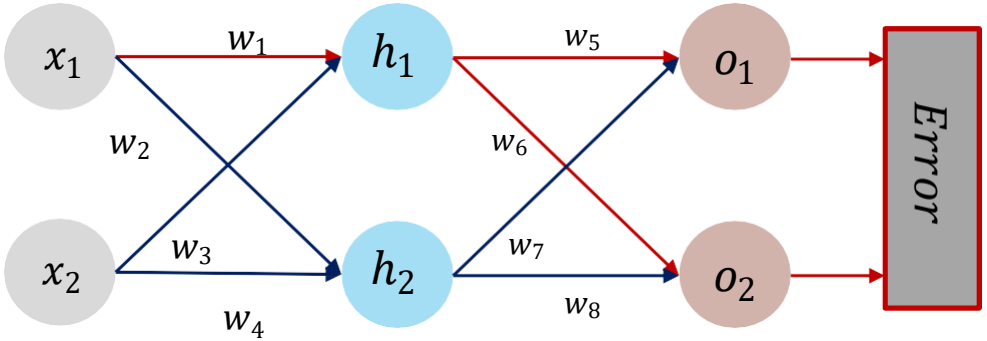
$$= \frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial In_{o_1}} \times \frac{\partial In_{o_1}}{\partial h_1} \times \frac{\partial h_1}{\partial In_{h_1}} \times \frac{\partial In_{h_1}}{\partial w_1} + \frac{\partial Error}{\partial o_2} \times \frac{\partial o_2}{\partial In_{o_2}} \times \frac{\partial In_{o_2}}{\partial h_1} \quad \frac{\partial h_1}{\partial In_{h_1}} \times \frac{\partial In_{h_1}}{\partial w_1}$$

$$= \left(\frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial In_{o_1}} \times \frac{\partial In_{o_1}}{\partial h_1} + \frac{\partial Error}{\partial o_2} \times \frac{\partial o_2}{\partial In_{o_2}} \times \frac{\partial In_{o_2}}{\partial h_1}\right) \times \frac{\partial h_1}{\partial In_{h_1}} \times \frac{\partial In_{h_1}}{\partial w_1}$$

$$= (\delta_5 + \delta_6) \times \frac{\partial h_1}{\partial In_{h_1}} \times \frac{\partial In_{h_1}}{\partial w_1}$$

误差反向传播：前馈神经网络

反向传播：参数更新



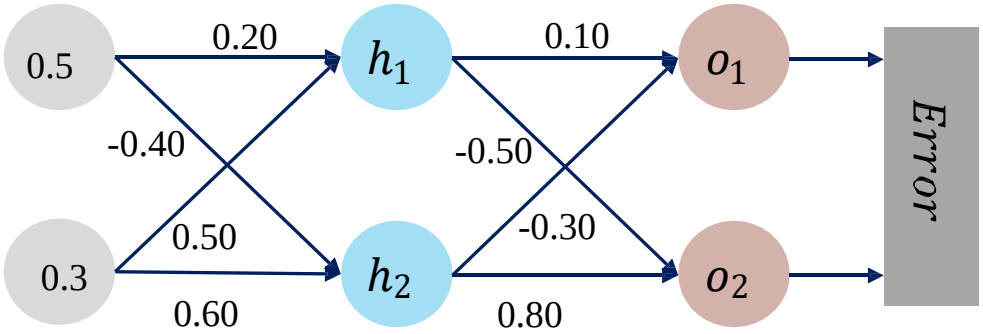
2. 更新参数，其中η 被称为学习率

$$w_i' = w_i - \eta \times \delta_i$$

原有参数　步长　传递误差

误差反向传播：前馈神经网络

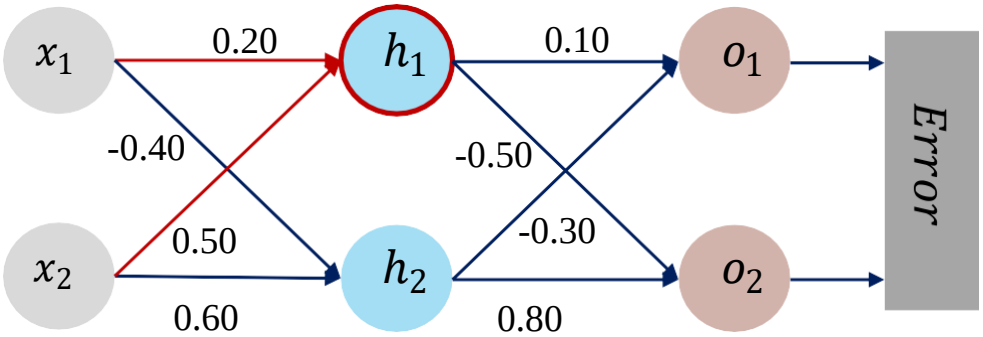
参数初始化



其中，$Error = 0.5 \times (o_1 - 0.23)^2 + 0.5 \times (o_2 - (-0.07))^2$，这里0.23和-0.07是对输入样本数据$(0.5, 0.3)$的标注信息。

误差反向传播：前馈神经网络

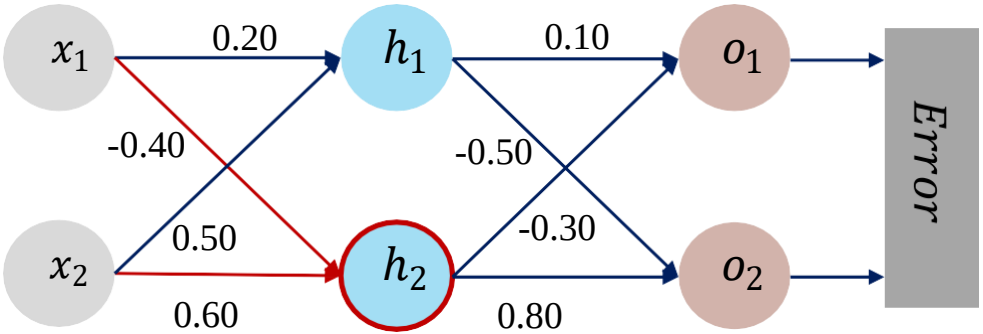
正向传播



$$h_1 = sigmoid\,(0.20 \times 0.50 + 0.50 \times 0.30) = 0.56$$

误差反向传播：前馈神经网络

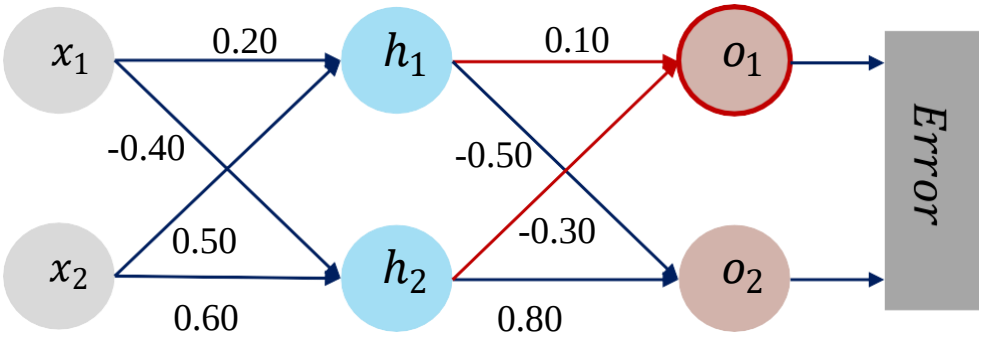
正向传播



$$h_2 = sigmoid\,(-0.40 \times 0.50 + 0.60 \times 0.30) = 0.50$$

误差反向传播：前馈神经网络

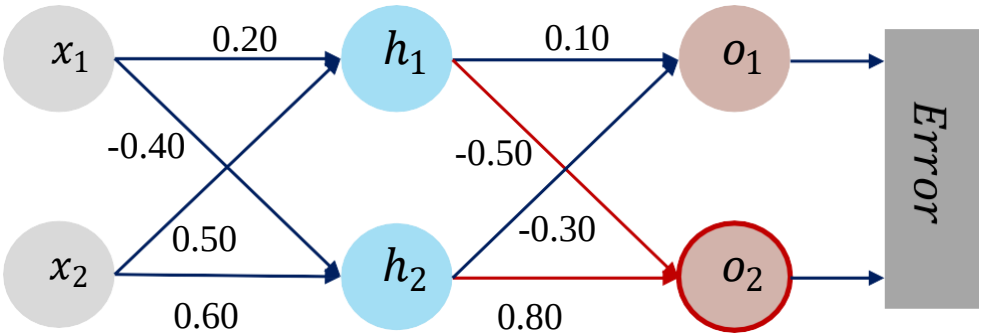
正向传播



$$o_1 = sigmoid\,(0.10 \times 0.56 + -0.30 \times 0.50) = 0.48$$

误差反向传播：前馈神经网络

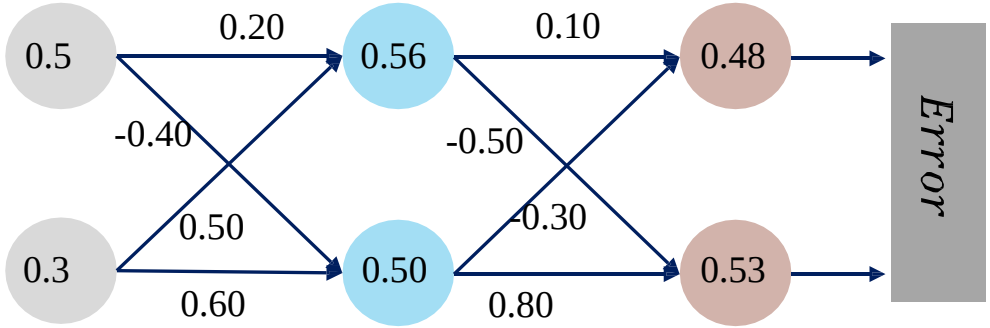
正向传播



$$o_2 = sigmoid\,(-0.50 \times 0.56 + 0.80 \times 0.50) = 0.53$$

误差反向传播：前馈神经网络

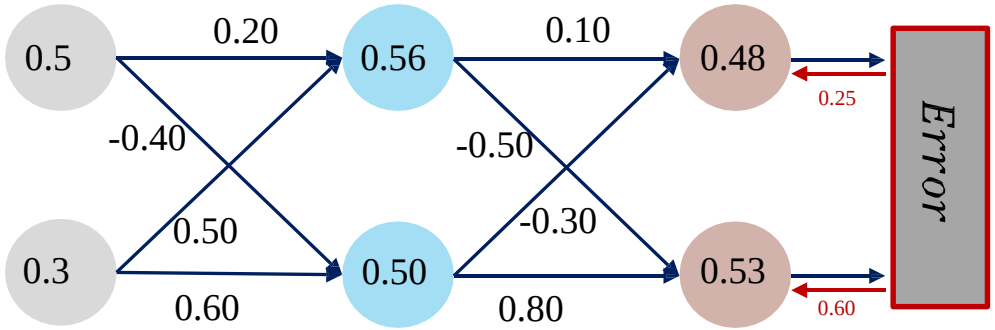
反向传播（假设学习速率Learning Rate $\eta = 1$）



$$Error = 0.5\times(0.48-0.23)^2+0.5\times(0.53-(-0.07))^2=0.21$$

误差反向传播：前馈神经网络

梯度计算

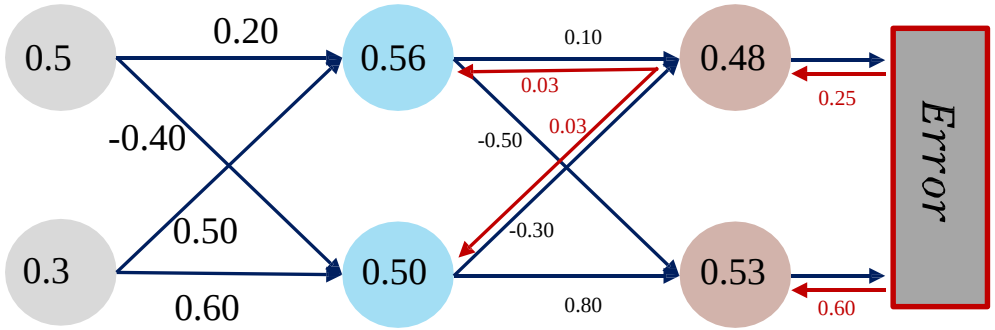


$$\delta_{o^1} = \frac{\partial Error}{\partial o_1} = o_1 - 0.23 = 0.25$$

$$\delta_{o^2} = \frac{\partial Error}{\partial o_2} = o_2 - (-0.07) = 0.60$$

误差反向传播：前馈神经网络

梯度计算

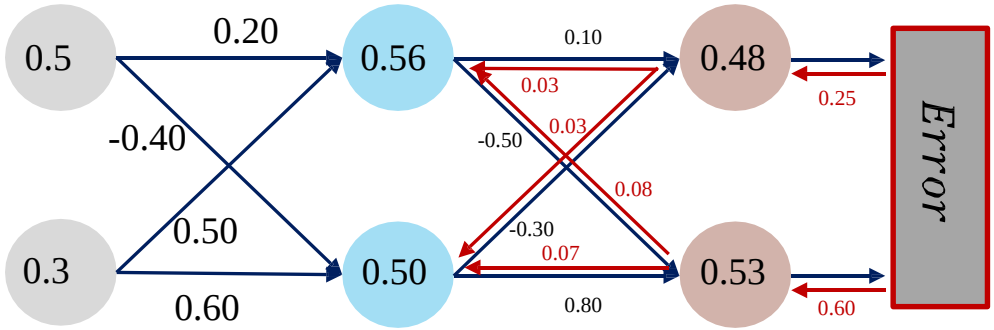


$$\delta_5 = \frac{\partial Error}{\partial w_5} = \frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial In_{o_1}} \times \frac{\partial In_{o_1}}{\partial w_5} = 0.25 \times 0.48 \times (1 - 0.48) \times 0.56 = 0.03$$

$$\delta_7 = \frac{\partial Error}{\partial w_7} = \frac{\partial Error}{\partial o_1} \times \frac{\partial o_1}{\partial In_{o_1}} \times \frac{\partial In_{o_1}}{\partial w_7} = 0.25 \times 0.48 \times (1 - 0.48) \times 0.50 = 0.03$$

误差反向传播：前馈神经网络

梯度计算

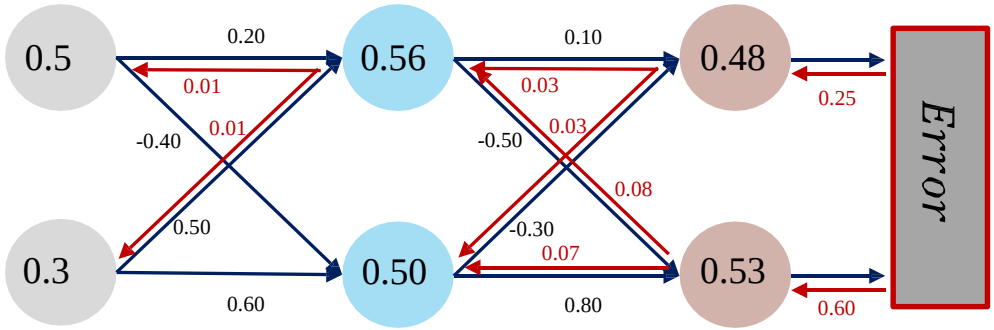


$$\delta_6 = \frac{\partial Error}{\partial w_6} = \frac{\partial Error}{\partial o_2} \times \frac{\partial o_2}{\partial In_{o_2}} \times \frac{\partial In_{o_2}}{\partial w_6} = 0.60 \times 0.53 \times (1 - 0.53) \times 0.56 = 0.08$$

$$\delta_8 = \frac{\partial Error}{\partial w_8} = \frac{\partial Error}{\partial o_2} \times \frac{\partial o_2}{\partial In_{o_2}} \times \frac{\partial In_{o_2}}{\partial w_8} = 0.60 \times 0.53 \times (1 - 0.53) \times 0.50 = 0.07$$

误差反向传播：前馈神经网络

梯度计算

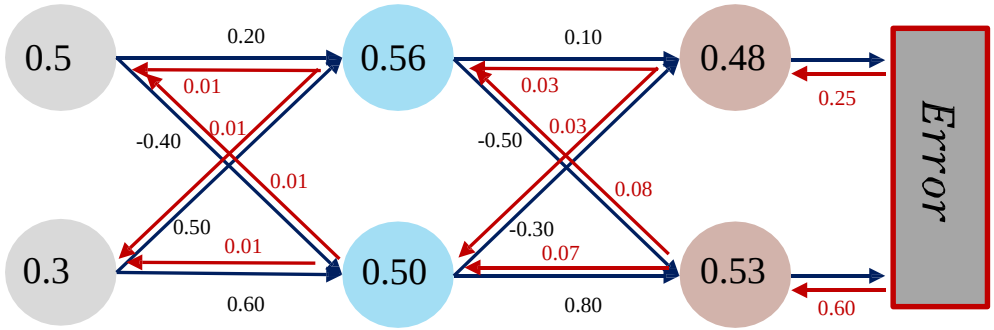


$$\delta_1 = \frac{\partial Error}{\partial w_1} = (\delta_5 + \delta_6) \times \frac{\partial h_1}{\partial In_{h_1}} \times \frac{\partial In_{h_1}}{\partial w_1} = (0.03 + 0.08) \times 0.56 \times (1 - 0.56) \times 0.50 = 0.01$$

$$\delta_3 = \frac{\partial Error}{\partial w_3} = (\delta_5 + \delta_6) \times \frac{\partial h_1}{\partial In_{h_1}} \times \frac{\partial In_{h_1}}{\partial w_3} = (0.03 + 0.08) \times 0.56 \times (1 - 0.56) \times 0.30 = 0.01$$

误差反向传播：前馈神经网络

梯度计算



$$\delta_2 = \frac{\partial Error}{\partial w_2} = (\delta_7 + \delta_8) \times \frac{\partial h_2}{\partial In_{h_2}} \times \frac{\partial In_{h_2}}{\partial w_2} = (0.03 + 0.07) \times 0.50 \times (1 - 0.50) \times 0.5 = 0.01$$

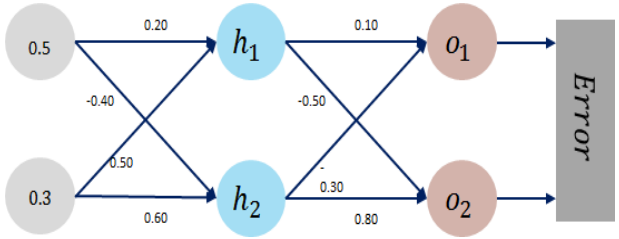
$$\delta_4 = \frac{\partial Error}{\partial w_4} = (\delta_7 + \delta_8) \times \frac{\partial h_2}{\partial In_{h_2}} \times \frac{\partial In_{h_2}}{\partial w_4} = (0.03 + 0.07) \times 0.50 \times (1 - 0.50) \times 0.3 = 0.01$$

误差反向传播：前馈神经网络

参数更新

$$w_i' = w_i - \eta \times \delta_i$$

原有参数　步长　传递误差



更新