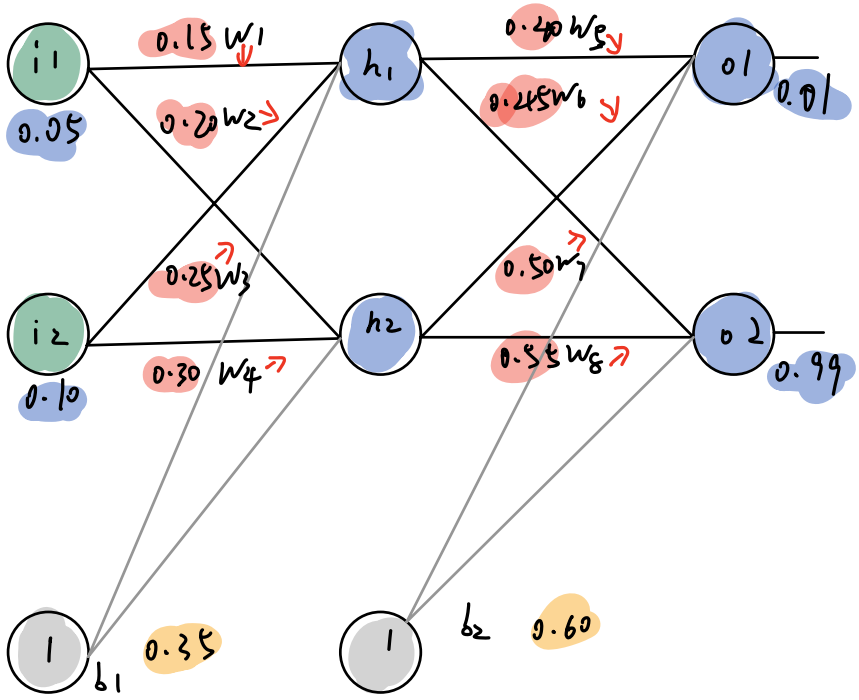
BP神经网络



用 sigmoid $\frac{x}{1+e^{-x}}$ 激活 a表示输出，z表示加权输入

step 1. 前向传播. (输入→隐藏→输出)

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

$$= 0.15 \times 0.05 + 0.20 \times 0.10 + 0.35 \times 1$$

$$= 0.3775$$

$$a_{h_1} = \frac{1}{1+e^{-0.3775}} = 0.59326992$$

$$z_{h_2} = w_3 * i_1 + w_4 * i_2 + b_1 * 1$$

$$= 0.25 \times 0.05 + 0.10 \times 0.30 + 0.35 \times 1$$

$$= 0.3925$$

$$a_{h_2} = \frac{1}{1+e^{-0.3925}} = 0.596884378$$

$$z_{o_1} = w_5 * a_{h_1} + w_6 * a_{h_2} + b_2 * 1$$
$$= 1.105905967$$

$$a_{o_1} = 0.751365069$$

$$z_{o_2} = w_7 * a_{h_1} + w_8 * a_{h_2} + b_2 * 1$$
$$= 1.224921679$$

$$a_{o_2} = 0.772928465$$

step 2 反向传播

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

$$E_{total} = \frac{1}{2}((0.01 - 0.751365069)^2 + (0.99 - 0.772928465)^2) = 0.298371109$$

② 隐藏层 → 输出层权重更新

以 w_5 为例，整体 loss 对 w_5 求偏导

$$\frac{\partial E_{total}}{\partial w_5} = \frac{\partial E_{total}}{\partial a_{o1}} \cdot \frac{\partial a_{o1}}{\partial z_{o1}} \cdot \frac{\partial z_{o1}}{\partial w_5}$$

($w_5 \Rightarrow z_{o1} \rightarrow E_{total}$)

$$\frac{\partial E_{total}}{\partial a_{o1}} = (target_{o1} - a_{o1})(-1) = 0.741365069$$

$$\frac{\partial a_{o1}}{\partial z_{o1}} = a_{o1}(1 - a_{o1}) = 0.186815602$$

$$\frac{\partial z_{o1}}{\partial w_5} = a_{h_1} = 0.593269992$$

$$\frac{E_{total}}{\partial w_5} = 0.08216704$$

$$\frac{\partial E_{total}}{\partial w_5} = -(target_{o1} - a_{o1}) * a_{o1} * (1-a_{o1}) * a_{h_1}$$

用 δ_{o1} 来表示输出层的误差

$$\delta_{o1} = \frac{\partial E_{total}}{\partial a_{o1}} * \frac{\partial a_{o1}}{\partial z_{o1}} = \frac{\partial E_{total}}{\partial z_{o1}}$$

$$= -(target_{o1} - a_{o1}) * a_{o1} * (1-a_{o1})$$

$$\frac{\partial E_{total}}{\partial w_5} = \delta_{o1} * a_{h_1}$$

假设学习率为 θ, $\theta = 0.5$, 我们更新 w_5 得到

$$w_5^+ = w_5 - \theta * \frac{\partial E_{total}}{\partial w_5} = 0.4 - 0.5 \times 0.08216704$$

$$= 0.35891648$$

$$w_6^+ = 0.408666186$$

$$w_7^+ = 0.511301270$$

$$w_8^+ = 0.561370121$$

③ 隐藏层 → 隐藏层的权值更新

$$\frac{\partial E_{total}}{\partial a_{h_1}} = \frac{\partial E_{o1}}{\partial a_{h_1}} + \frac{\partial E_{o2}}{\partial a_{h_2}}$$

$$= \frac{\partial E_{o1}}{\partial a_{o1}} * \frac{\partial a_{o1}}{\partial z_{o1}} * \frac{\partial z_{o1}}{\partial a_{h_1}} + \frac{\partial E_{o2}}{\partial a_{o2}} * \frac{\partial a_{o2}}{\partial z_{o2}} * \frac{\partial z_{o2}}{\partial a_{h_2}}$$

$$= 0.036350306$$

$$\frac{\partial a_{h1}}{\partial z_{h1}} = a_{h1} * (1 - a_{h1}) = 0.241300786$$

$$\frac{\partial z_{h1}}{\partial w_1} = 0.05$$

$$\frac{\partial E_{total}}{\partial w_1} = \frac{\partial E_{total}}{\partial a_{h1}} * \frac{\partial a_{h1}}{\partial z_{h1}} * \frac{\partial z_{h1}}{\partial w_1} = 0.000438568$$

$$w_1^+ = w_1 - \theta * \frac{\partial E_{total}}{\partial w_1} = 0.14978071$$

$$w_2^+ = 0.19956143$$

$$w_3^+ = 0.24975114$$

$$w_4^+ = 0.29950229.$$