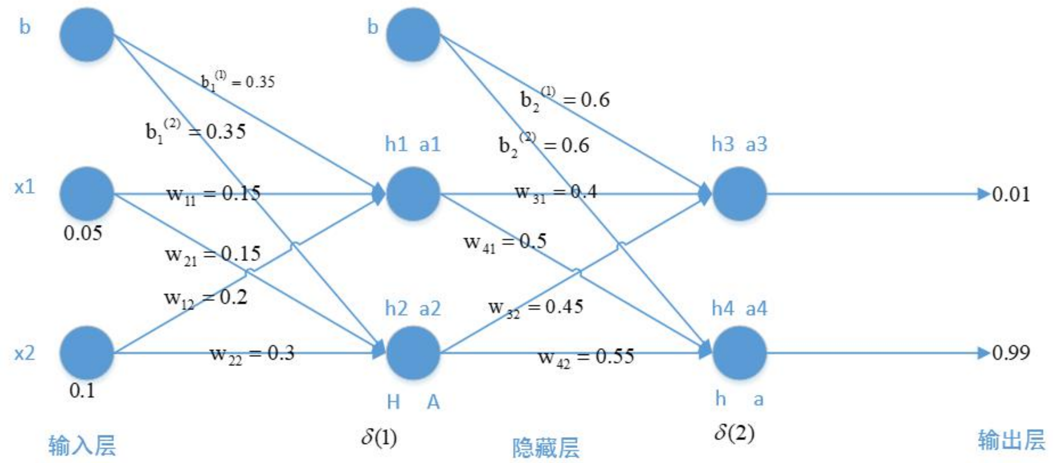
1、手撕推导神经网络的前向传播与反向传播过程---数学之美



1.1 用矩阵计算前向传播过程：

$$X=\begin{pmatrix} b1 \\ x1 \\ x2 \end{pmatrix}=\begin{pmatrix} 1 \\ 0.05 \\ 0.1 \end{pmatrix}$$

$$W_1=\begin{pmatrix} b_1^{(1)} & w_{11} & w_{12} \\ b_2^{(1)} & w_{21} & w_{22} \end{pmatrix}=\begin{pmatrix} 0.35 & 0.15 & 0.2 \\ 0.35 & 0.25 & 0.3 \end{pmatrix}$$

$$H=\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}=W_1\cdot X=\begin{pmatrix} 0.35 & 0.15 & 0.2 \\ 0.35 & 0.25 & 0.3 \end{pmatrix}\begin{pmatrix} 1 \\ 0.05 \\ 0.1 \end{pmatrix}=\begin{pmatrix} 0.3775 \\ 0.3925 \end{pmatrix}$$

$$A=\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}=\sigma(H)=\begin{pmatrix} 0.5933 \\ 0.5969 \end{pmatrix}$$

$$W_2=\begin{pmatrix} b_1^{(2)} & w_{31} & w_{32} \\ b_2^{(2)} & w_{41} & w_{42} \end{pmatrix}=\begin{pmatrix} 0.6 & 0.4 & 0.45 \\ 0.6 & 0.5 & 0.55 \end{pmatrix}$$

$$A'=\begin{pmatrix} 1 \\ a_1 \\ a_2 \end{pmatrix}=\begin{pmatrix} 1 \\ 0.5933 \\ 0.5969 \end{pmatrix}$$

$$\mathrm{h}=\begin{pmatrix}h_3\\h_4\end{pmatrix}=W_2\cdot A'=\begin{pmatrix}0.6&0.4&0.45\\0.6&0.5&0.55\end{pmatrix}\begin{pmatrix}1\\0.5933\\0.5969\end{pmatrix}=\begin{pmatrix}1.106\\1.225\end{pmatrix}$$

$$\mathrm{a}=\begin{pmatrix}a_3\\a_4\end{pmatrix}=\sigma(h)=\begin{pmatrix}0.751\\0.773\end{pmatrix}$$

1.2 反向传播计算矩阵计算方法：

反向传播计算的过程，就是计算两个误差项：

$$\text{误差项}\begin{cases}\dfrac{\partial \mathrm{J}_{总}}{\partial \mathrm{h}}=\delta(2)\\\dfrac{\partial \mathrm{J}_{总}}{\partial \mathrm{H}}=\delta(1)\end{cases}$$

$$\mathrm{J}_{总}=\frac{1}{2}(\mathrm{target}-a)^2$$

$$\mathrm{target}=\begin{pmatrix}0.1\\0.99\end{pmatrix}$$

$$\mathrm{a}=\begin{pmatrix}a_3\\a_4\end{pmatrix}$$

梯度下降法：

$$\mathrm{w}=w-\alpha\frac{\partial J}{\partial w}$$

$$b=b-\alpha\frac{\partial J}{\partial b}$$

$\partial\mathrm{J}$是损失函数，α是学习率

本次使用的激活函数是 sigmoid： $\mathrm{f(x)}=\dfrac{1}{1+\mathrm{e}^{\text{-x}}}$

$$f(x)'=f(x)[1-f(x)]$$

1.2.1 计算$\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{h}}$误差项

$$\begin{aligned}
\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{h}} &= \frac{\partial J_{总}}{\partial \mathrm{a}} \cdot \frac{\partial \mathrm{a}}{\partial \mathrm{h}} \\
&= (a - \mathrm{target}) \cdot a \cdot (1-a) \\
&= [\begin{pmatrix} 0.751 \\ 0.773 \end{pmatrix} - \begin{pmatrix} 0.01 \\ 0.99 \end{pmatrix}] \otimes \{\begin{pmatrix} 0.751 \\ 0.773 \end{pmatrix} \otimes [E - \begin{pmatrix} 0.751 \\ 0.773 \end{pmatrix}]\} \\
&= \begin{pmatrix} 0.74 \\ -0.217 \end{pmatrix} \otimes \begin{pmatrix} 0.249 \\ 0.127 \end{pmatrix} \\
&= \begin{pmatrix} 0.1386 \\ -0.0386 \end{pmatrix} \\
&= \delta(2)
\end{aligned}$$

1.2.2 计算$\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{H}}$误差项

$$\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{H}} = \frac{\partial J_{总}}{\partial \mathrm{h}} \cdot \frac{\partial h}{\partial A} \frac{\partial A}{\partial \mathrm{H}} \qquad \mathrm{h} = W_2 A + b^{(2)}$$

$$\frac{\partial \mathrm{h}}{\partial \mathrm{A}} = \mathrm{W}_2^{\ T} = \begin{pmatrix} w_{31} & w_{32} \\ w_{41} & w_{42} \end{pmatrix}^T = \begin{pmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{pmatrix}^T = \begin{pmatrix} 0.4 & 0.5 \\ 0.45 & 0.55 \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial \mathrm{A}}{\partial \mathrm{H}} &= A(1-A) \\
&= \begin{pmatrix} 0.5933 \\ 0.5969 \end{pmatrix} \otimes [E - \begin{pmatrix} 0.5933 \\ 0.5969 \end{pmatrix}] \\
&= \begin{pmatrix} 0.5933 \\ 0.5969 \end{pmatrix} \otimes \begin{pmatrix} 0.4067 \\ 0.4031 \end{pmatrix} = \begin{pmatrix} 0.2413 \\ 0.24061 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\delta(1) = \frac{\partial \mathrm{J}_{总}}{\partial \mathrm{H}} &= \delta(2) \cdot \frac{\partial h}{\partial A} \frac{\partial A}{\partial \mathrm{H}} = \frac{\partial h}{\partial A} \cdot \delta(2) \cdot \frac{\partial A}{\partial \mathrm{H}} \\
&= (W_2^{\ T} \cdot \delta(2)) \otimes A(1-A) \qquad \mathrm{h} = W_2 A + \mathrm{b}^{(2)} \\
&= \begin{pmatrix} 0.0087 \\ 0.0099 \end{pmatrix}
\end{aligned}$$

1.2.3 更新输出层到隐藏层 w，b

$$\mathrm{h} = W_2 A + \mathrm{b}^{(2)}$$

$$\begin{aligned}\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{W}_2} &= \frac{\partial J_{总}}{\partial \mathrm{h}} \cdot \frac{\partial h}{\partial W_2} \\ &= \delta(2) \cdot \frac{\partial h}{\partial W_2} \\ &= \delta(2) \cdot A^T \\ &= \begin{pmatrix} 0.1386 \\ -0.0386 \end{pmatrix} \begin{pmatrix} 0.5933 & 0.5969 \end{pmatrix} \\ &= \begin{pmatrix} 0.0822 & 0.0827 \\ -0.0229 & -0.023 \end{pmatrix}\end{aligned}$$

所以：

$$\begin{aligned}\mathrm{W}_2{}^* &= W_2 - \alpha \frac{\partial J_{总}}{\partial W_2} \\ &= \begin{pmatrix} 0.4 & 0.45 \\ 0.5 & 0.55 \end{pmatrix} - 0.5 \cdot \begin{pmatrix} 0.0822 & 0.0827 \\ -0.0229 & -0.023 \end{pmatrix} \\ &= \begin{pmatrix} 0.3589 & 0.4086 \\ 0.5115 & 0.5615 \end{pmatrix}\end{aligned}$$

$$\mathrm{h} = W_2 A + \mathrm{b}^{(2)}$$

$$\begin{aligned}\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{b}^{(2)}} &= \frac{\partial J_{总}}{\partial \mathrm{h}} \cdot \frac{\partial h}{\partial b^{(2)}} \\ &= \delta(2) \cdot E \\ &= \delta(2) \\ &= \begin{pmatrix} 0.1386 \\ -0.0386 \end{pmatrix}\end{aligned}$$

$$\begin{aligned}\mathrm{b}^{(2)*} &= \mathrm{b}^{(2)} - \alpha \frac{\partial J_{总}}{\partial \mathrm{b}^{(2)}} \\ &= \begin{pmatrix} 0.6 \\ 0.6 \end{pmatrix} - 0.5 \cdot \begin{pmatrix} 0.1386 \\ -0.0386 \end{pmatrix} \\ &= \begin{pmatrix} 0.5307 \\ 0.6193 \end{pmatrix}\end{aligned}$$

1.2.4 更新隐藏层到输入层 w，b

$$\mathrm{H} = W_1 X + \mathrm{b}^{(1)}$$

$$\begin{aligned}\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{W}_1} &= \frac{\partial J_{总}}{\partial \mathrm{H}} \cdot \frac{\partial H}{\partial W_1} \\ &= \delta(1) \cdot \frac{\partial H}{\partial W_1} \\ &= \delta(1) \cdot X^T \\ &= \begin{pmatrix} 0.0087 \\ 0.0099 \end{pmatrix} \begin{pmatrix} 0.05 & 0.1 \end{pmatrix} \\ &= \begin{pmatrix} 0.00435 & 0.0087 \\ 0.00495 & 0.0099 \end{pmatrix}\end{aligned}$$

所以：

$$\begin{aligned}\mathrm{W}_1^* &= W_1 - \alpha \frac{\partial J_{总}}{\partial W_1} \\ &= \begin{pmatrix} 0.15 & 0.2 \\ 0.25 & 0.3 \end{pmatrix} - 0.5 \cdot \frac{\partial J_{总}}{\partial W_1} \\ &= \begin{pmatrix} 0.1498 & 0.1996 \\ 0.2498 & 0.2995 \end{pmatrix}\end{aligned}$$

$$\mathrm{H} = W_1 X + \mathrm{b}^{(1)}$$

$$\begin{aligned}\frac{\partial \mathrm{J}_{总}}{\partial \mathrm{b}^{(1)}} &= \frac{\partial J_{总}}{\partial \mathrm{H}} \cdot \frac{\partial H}{\partial \mathrm{b}^{(1)}} \\ &= \delta(1) \cdot E \\ &= \begin{pmatrix} 0.0087 \\ 0.0099 \end{pmatrix}\end{aligned}$$

$$\begin{aligned}\mathrm{b}^{(1)^*} &= \mathrm{b}^{(1)} - \alpha \frac{\partial J_{总}}{\partial \mathrm{b}^{(1)}} \\ &= \begin{pmatrix} 0.35 \\ 0.35 \end{pmatrix} - 0.5 \cdot \begin{pmatrix} 0.0087 \\ 0.0099 \end{pmatrix} \\ &= \begin{pmatrix} 0.3456 \\ 0.345 \end{pmatrix}\end{aligned}$$