1、正向传播

(1) 输入层 → 隐藏层

$$z_{11} = x_1 \cdot w_{11} + x_2 \cdot w_{12}$$

$$a_{11} = \frac{1}{1+e^{-z_{11}}}$$

$$z_{12} = x_1 \cdot w_{13} + x_2 \cdot w_{14}$$

$$a_{12} = \frac{1}{1+e^{-z_{12}}}$$

(2) 隐藏层 → 输出层

$$z_{21} = a_{11} \cdot w_{21} + a_{12} \cdot w_{22}$$

$$a_{21} = \frac{1}{1+e^{-z_{21}}}$$

$$z_{22} = a_{11} \cdot w_{23} + a_{12} \cdot w_{24}$$

$$a_{22} = \frac{1}{1+e^{-z_{22}}}$$

输出

$$\hat{y}_1 = a_{21}, \quad \hat{y}_2 = a_{22}$$

2、反向传播

(1) 计算损失函数

$$E = \sum \frac{1}{n}(Y - \hat{y})^2$$

(Y:目标值, $\hat{y}$:预测值)

$$E_1 = \frac{1}{2}(Y_1 - \hat{y}_1)^2$$

$$E_2 = \frac{1}{2}(Y_2 - \hat{y}_2)^2$$

$$E = E_1 + E_2$$

(2) 隐藏层(h1) → 输出层权值更新

$$\frac{\partial E}{\partial w_{21}} = \frac{\partial E}{\partial a_{21}} \cdot \frac{\partial a_{21}}{\partial z_{21}} \cdot \frac{\partial z_{21}}{\partial w_{21}}$$

① 计算 $\frac{\partial E}{\partial a_{21}}$

$$E = \frac{1}{2}(Y_1 - \hat{y}_1)^2 + \frac{1}{2}(Y_2 - \hat{y}_2)^2$$

$$= \frac{1}{2}(Y_1 - a_{21})^2 + \frac{1}{2}(Y_2 - a_{22})^2$$

$$\frac{\partial E}{\partial a_{21}} = 2 \cdot \frac{1}{2}(Y_1 - a_{21}) \cdot (-1) + 0 = a_{21} - Y_1$$

② $$\frac{\partial a_{21}}{\partial z_{21}} = \frac{-e^{-z_{21}}}{(1+e^{-z_{21}})^2} = \frac{1}{1+e^{-z_{21}}} \cdot \frac{-e^{-z_{21}}}{1+e^{-z_{21}}}$$

$$= a_{21} \cdot (1 - a_{21})$$

$$\frac{\partial E}{\partial w} = \begin{bmatrix} \frac{\partial E}{\partial w_{21}} & \frac{\partial E}{\partial w_{22}} \\ \frac{\partial E}{\partial w_{23}} & \frac{\partial E}{\partial w_{24}} \end{bmatrix}$$ 矩阵运算

$$= \begin{bmatrix} (a_{21} - Y_1) a_{21}(1 - a_{21}) \\ (a_{22} - Y_2) a_{22}(1 - a_{22}) \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \end{bmatrix}$$

③ $$\frac{\partial z_{21}}{\partial w_{21}} = a_{11}$$

$$\therefore \frac{\partial E}{\partial w_{21}} = (a_{21} - Y_1) \cdot [a_{21} \cdot (1 - a_{21})] \cdot a_{11}$$

⇓

$$\frac{\partial E}{\partial w_{22}} = (a_{21} - Y_1) \cdot [a_{21} \cdot (1 - a_{21})] \cdot a_{12}$$

$$\frac{\partial E}{\partial w_{23}} = (a_{22} - Y_2) \cdot [a_{22} \cdot (1 - a_{22})] \cdot a_{11}$$

$$\frac{\partial E}{\partial w_{24}} = (a_{22} - Y_2) [a_{22} \cdot (1 - a_{22})] a_{12}$$

$$w_{21}^{+} = w_{21} - \eta \cdot \frac{\partial E}{\partial w_{21}}, \quad w_{22}^{+} = w_{22} - \eta \cdot \frac{\partial E}{\partial w_{22}}$$

$$w_{23}^{+} = w_{23} - \eta \cdot \frac{\partial E}{\partial w}, \quad w_{24}^{+} = w_{24} - \eta \cdot \frac{\partial E}{\partial w_{24}}$$

(3) 输入层→隐藏层的权值更新

$$\frac{\partial E}{\partial w_{11}}=\frac{\partial E}{\partial a_{11}}\cdot\frac{\partial a_{11}}{\partial z_{11}}\cdot\frac{\partial z_{11}}{\partial w_{11}}$$

① $\frac{\partial E}{\partial a_{11}}=\frac{\partial E_1}{\partial a_{11}}+\frac{\partial E_2}{\partial a_{11}}$

$$\frac{\partial E_1}{\partial a_{11}}=\frac{\partial E_1}{\partial a_{21}}\cdot\frac{\partial a_{21}}{\partial z_{21}}\cdot\frac{\partial z_{21}}{\partial a_{11}}$$

$$=(a_{21}-Y_1)\cdot[a_{21}\cdot(1-a_{21})]\cdot w_{21}$$

$$\frac{\partial E_2}{\partial a_{11}}=\frac{\partial E_2}{\partial a_{22}}\cdot\frac{\partial a_{22}}{\partial z_{22}}\cdot\frac{\partial z_{22}}{\partial a_{11}}$$

$$=(a_{22}-Y_2)\cdot[a_{22}\cdot(1-a_{22})]\cdot w_{23}$$

② $\frac{\partial a_{11}}{\partial z_{11}}=a_{11}(1-a_{11})$

③ $\frac{\partial z_{11}}{\partial w_{11}}=x_1+0=x_1$

由①②③可得 $\frac{\partial E}{\partial w_{11}}=[(a_{21}-Y_1)a_{21}(1-a_{21})w_{21}+(a_{22}-Y_2)a_{22}(1-a_{22})w_{23}]a_{11}(1-a_{11})x_1$

⇓

$$\frac{\partial E}{\partial w_{12}}=[(a_{21}-Y_1)a_{21}(1-a_{21})w_{21}+(a_{22}-Y_2)a_{22}(1-a_{22})w_{23}]a_{11}(1-a_{11})x_2$$

$$\frac{\partial E}{\partial w_{13}}=[(a_{21}-Y_1)a_{21}(1-a_{21})w_{22}+(a_{22}-Y_2)a_{22}(1-a_{22})w_{24}]a_{12}(1-a_{12})x_1$$

$$\frac{\partial E}{\partial w_{14}}=[(a_{21}-Y_1)a_{21}(1-a_{21})w_{22}+(a_{22}-Y_2)a_{22}(1-a_{22})w_{24}]a_{12}(1-a_{12})x_2$$

⇓

$w_{11}^{+}=w_{11}-\eta\frac{\partial E}{\partial w_{11}}$, $\quad w_{12}^{+}=w_{12}-\eta\frac{\partial E}{\partial w_{12}}$

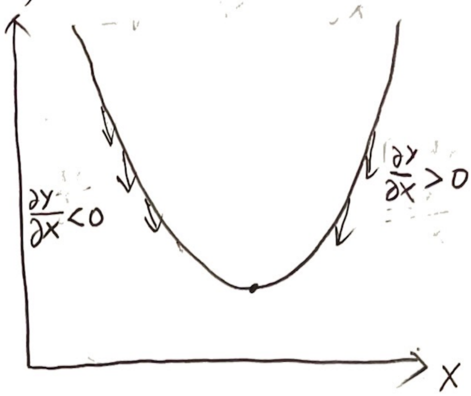
$w_{13}^{+}=w_{13}-\eta\frac{\partial E}{\partial w_{13}}$, $\quad w_{14}^{+}=w_{14}-\eta\frac{\partial E}{\partial w_{14}}$

$$\frac{\partial E}{\partial w}=\begin{bmatrix}\frac{\partial E}{\partial w_{11}} & \frac{\partial E}{\partial w_{12}}\\ \frac{\partial E}{\partial w_{13}} & \frac{\partial E}{\partial w_{14}}\end{bmatrix}$$

矩降运算

$$\begin{bmatrix}w_{21} & w_{23}\\ w_{22} & w_{24}\end{bmatrix}\begin{bmatrix}(a_{21}-Y_1)a_{21}(1-a_{21})\\ (a_{22}-Y_2)a_{22}(1-a_{22})\end{bmatrix}\begin{bmatrix}a_{11}(1-a_{11})\\ a_{12}(1-a_{12})\end{bmatrix}[x_1\ \ x_2]$$

梯度下降法



$$\frac{\partial y}{\partial x}=\frac{\partial y}{\partial x}-\eta\frac{\partial y}{\partial x}$$

$\frac{\partial y}{\partial x}>0$, → $\frac{\partial y}{\partial x}$向左移动

$\frac{\partial y}{\partial x}<0$, → $\frac{\partial y}{\partial x}$向右移动