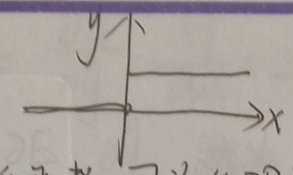
1. $f(x)=\begin{cases}0, & x<0\\ 1, & x\geq 0\end{cases}$



该函数不适合做激活函数，因为在BP过程中其导数为0.

2. $E=\sum_{i=1}^{n}(\hat{y}_i-y_i)^2$，$\hat{y}_i$为预测值，$y_i$为实际标签。

原神经网络结构：

$$o_{11}=w_1\cdot i_1+w_3\cdot i_2+b_1$$
$$o_{12}=w_2\cdot i_1+w_4\cdot i_2+b_2$$
$$h_1=\max(0,o_{11})\longrightarrow \text{ReLU}$$
$$h_2=\max(0,o_{12})\longrightarrow \text{ReLU}$$
~~$o_{21}=w_5\cdot h_1$~~ ~~$o_{22}=w_6\cdot h_2$~~ $o_2=w_5\cdot h_1+w_6\cdot h_2$

$\hat{y}=$ ~~$\hat{y}=$~~ $\dfrac{1}{1+e^{-o_2}}\longrightarrow \text{Sigmoid}$

$$\frac{\partial E}{\partial \hat{y}}=2\sum_{i=1}^{n}(\hat{y}_i-y_i)\qquad \frac{\partial \hat{y}}{\partial w_5}=o_2(1-o_2)\cdot h_1$$

$$\frac{\partial \hat{y}}{\partial h_1}=o_2(1-o_2)\cdot w_5\qquad \frac{\partial h_1}{\partial w_1}=\begin{cases}i_1, & w_1\cdot i_1+w_3\cdot i_2+b_1\geq 0\\ 0, & \text{otherwise}\end{cases}$$

$$\frac{\partial \hat{y}}{\partial h_2}=o_2(1-o_2)\cdot w_6\qquad \frac{\partial h_2}{\partial b_2}=\begin{cases}1, & w_2\cdot i_1+w_4\cdot i_2+b_2\geq 0\\ 0, & \text{otherwise}\end{cases}$$

地址:南京市仙林大道163号 邮编:210023 9101117

联立可得：

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_5} = 2\sum_{i=1}^{n}(\hat{y}_i - y_i)\sigma_2(1-\sigma_2)\cdot h_1$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1} = \begin{cases} 2\sum_{i=1}^{n}(\hat{y}_i - y_i)\sigma_2(1-\sigma_2)\cdot w_5 \cdot \vec{v}_1, & w_1\cdot\vec{v}_1 + w_3\cdot\vec{v}_2 + b_1 \geq 0 \\ 0, \text{ otherwise.} \end{cases}$$

$$\frac{\partial E}{\partial b_2} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial h_2} \cdot \frac{\partial h_2}{\partial b_2} = \begin{cases} 2\sum_{i=1}^{n}(\hat{y}_i - y_i)\sigma_2(1-\sigma_2)\cdot w_6, & w_2\cdot\vec{v}_1 + w_4\cdot\vec{v}_2 + b_2 \geq 0 \\ 0, \text{ otherwise.} \end{cases}$$

其中

$$\sigma_2 = w_5\cdot h_1 + w_6\cdot h_2$$

$$\hat{y} = \frac{1}{1+e^{-\sigma_2}}$$

(将0代入可得具体形式具体计算)

当 $X_1 = [0.3, 2.8]$，$Y_1 = 1$，且 $w_1, w_2, w_3, w_4, w_5, w_6 =$ ——

$h_1 = w_1\vec{v}_1 + w_3\cdot\vec{v}_2 + b_1 = 0.98$

$h_2 = w_2\cdot\vec{v}_1 + w_4\cdot\vec{v}_2 + b_2 = 2.0$

$\sigma = w_5\cdot h_1 + w_6\cdot h_2 = 4.672$

$\hat{y} = \frac{1}{1+e^{-\sigma}} = 0.990$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_5} = 2(\hat{y}-y)\cdot\sigma(1-\sigma)\cdot w_5\cdot\vec{v}_2$$

$$= 2w_5^2 h_1(\hat{y}-y)(1-w_5\cdot h_1)\cdot\vec{v}_2$$

$$= 752.99386 \approx 753.$$

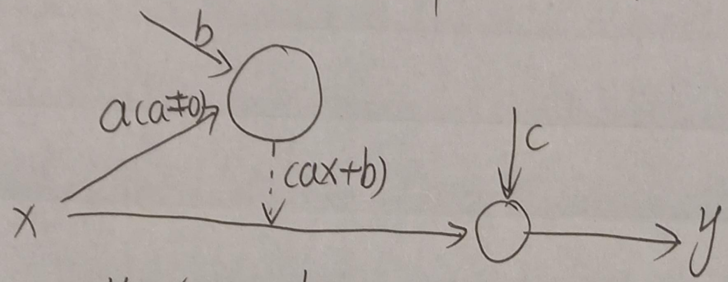
3. 不能用一个神经元拟合二次曲线.

至少用2个神经元才行.

对于任意一个二次函数.有如下形式 $y=ax^2+bx+c,\ a\neq 0$.

~~可构造神~~ $y = ax^2+bx+c =$ ~~xcax~~ $(ax+b)\cdot x+c,\ a\neq 0$.

因此可以构造含有2个神经元的结构.



$y=(ax+b)\cdot x+c,\ a\neq 0$. ~~两个神经元均不需激活函数~~

两个神经元均不需激活函数. 其中第一个神经元的输出要作为第二个神经元~~的输~~ ~~求输~~对应输入x的权值.

此时. BP公式为 ~~$\frac{\partial y}{\partial}$~~ $E=(\hat{y}-y)^2\cdot\frac{1}{2}$

$$\frac{\partial E}{\partial c}=\frac{\partial E}{\partial y}\cdot\frac{\partial y}{\partial c}=(\hat{y}-y)$$

$$\frac{\partial E}{\partial a}=\frac{\partial E}{\partial y}\cdot\frac{\partial y}{\partial w}\cdot\frac{\partial w}{\partial a}=(\hat{y}-y)x^2$$

~~$(\hat{y}-y)(ax+b)$~~ ~~$(\hat{y}-y)(ax+b)\cdot a$~~

$$\frac{\partial E}{\partial b}=\frac{\partial E}{\partial y}\cdot\frac{\partial y}{\partial w}\cdot\frac{\partial w}{\partial b}=(\hat{y}-y)\cdot x$$

上述结构可导出BP公式. 因此该结构理论上能够work.

(如果考虑传播时延. 还需要在传播x的途径上加一个神经元 ~~来做~~ x→o

地址:南京市仙林大道163号 邮编: 210023 9101117