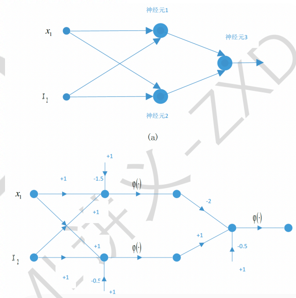
1. 在节 9.1.2 的异或神经网络中，神经网络如下图所示，对应的第一层神经网络权系数矩阵和偏置分别为



$$\boldsymbol{W}^{(1)}=\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2\end{bmatrix}=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\quad \boldsymbol{w}_0=\begin{bmatrix}0 & -1\end{bmatrix}^{\mathrm{T}}$$

第二层的权向量和偏置为

$$\boldsymbol{w}^{(2)}=\begin{bmatrix}1\\ -2\end{bmatrix}\qquad w_0^{(2)}=0$$

每个神经元的激活函数是 ReLU 函数 $\varphi(a)=\max\{0,a\}$。通过计算验证该网络可正确计算异或输出。

$x_1=\begin{pmatrix}0\\0\end{pmatrix}$

$z_1^{(1)}=W^{(1)}x_1+w_0=\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}0\\0\end{pmatrix}+\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}0\\-1\end{pmatrix}$

$a_1^{(1)}=\varphi(z_1^{(1)})=\begin{pmatrix}0\\0\end{pmatrix}$

$z_1^{(2)}=w^{(2)}a_1^{(1)}+w_0^{(2)}=(1\ \ -2)\begin{pmatrix}0\\0\end{pmatrix}+0=0$

$o_1=\varphi(z_1^{(2)})=0$

$x_2=\begin{pmatrix}1\\1\end{pmatrix}$

$z_2^{(1)}=W^{(1)}x_2+w_0=\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}2\\1\end{pmatrix}$

$a_2^{(1)}=\varphi(z_2^{(1)})=\begin{pmatrix}2\\1\end{pmatrix}$

$z_2^{(2)}=w^{(2)}a_2^{(1)}+w_0^{(2)}=(1\ \ -2)\begin{pmatrix}2\\1\end{pmatrix}+0=0$

$o_2=\varphi(z_2^{(2)})=0$

$x_3=\begin{pmatrix}1\\0\end{pmatrix}$

$z_3^{(1)}=W^{(1)}x_3+w_0=\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}+\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$

$a_3^{(1)}=\varphi(z_3^{(1)})=\begin{pmatrix}1\\0\end{pmatrix}$

$z_3^{(2)}=w^{(2)}a_3^{(1)}+w_0^{(2)}=(1\ \ -2)\begin{pmatrix}1\\0\end{pmatrix}+0=1$

$o_3=\varphi(z_3^{(2)})=1$

$x_4=\begin{pmatrix}0\\1\end{pmatrix}$

$z_4^{(1)}=W^{(1)}x_4+w_0=\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}+\begin{pmatrix}0\\-1\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}$

$a_4^{(1)}=\varphi(z_4^{(1)})=\begin{pmatrix}1\\0\end{pmatrix}$

$$z_4^{(2)} = w^{(2)} a_4^{(1)} + w_0^{(2)} = (1\ \ -2)\begin{pmatrix}1\\0\end{pmatrix} + 0 = 1$$

$$O_4 = \varphi(z_4^{(2)}) = 1$$

∴ 可以正确分类异或

3. 对于神经网络做 K 分类的输出，其样本损失函数为

$$L_n = L(\hat{\boldsymbol{y}}_n, \boldsymbol{y}_n) = -\sum_{k=1}^{K}[y_{nk} \ln \hat{y}_{nk}]$$

证明其对输出激活 $a_{nk}^{(L)}$ 的导数为

$$\frac{\partial L_n}{\partial a_{nk}^{(L)}} = \frac{\partial L(\hat{\boldsymbol{y}}_n, \boldsymbol{y}_n)}{\partial a_{nk}^{(L)}} = \hat{y}_{nk} - y_{nk}$$

$$L_n = L(\hat{y}_n, y_n) = -\sum_{k=1}^{K}(y_{nk} \ln \hat{y}_{nk})$$

$$\hat{y}_{nk} = \sigma(a_{nk}^{(L)}) = \frac{\exp(a_{nk}^{(L)})}{\sum_{j=1}^{K}\exp(a_{nj}^{(L)})}$$

$$\frac{\partial L_n}{\partial \hat{y}_{nk}} = \frac{\partial}{\partial \hat{y}_{nk}} - \sum_{k=1}^{K}(y_{nk} \ln \hat{y}_{nk}) = -y_{nk}\frac{1}{\hat{y}_{nk}} = -\frac{y_{nk}}{\hat{y}_{nk}}$$

$$\frac{\partial \hat{y}_{ni}}{\partial a_{nk}^{(L)}} = \frac{\partial}{\partial a_{nk}^{(L)}} \frac{\exp(a_{ni}^{(L)})}{\sum_{j=1}^{K}\exp(a_{nj}^{(L)})}$$

$i=k$ 时

$$\frac{\partial \hat{y}_{ni}}{\partial a_{nk}^{(L)}} = \frac{\exp(a_{nk}^{(L)})\sum_{j=1}^{K}\exp(a_{nj}^{(L)}) - \exp(a_{nk}^{(L)})\exp(a_{nk}^{(L)})}{\left(\sum_{j=1}^{K}\exp(a_{nj}^{(L)})\right)^2} = \hat{y}_{nk}(1-\hat{y}_{nk})$$

$i \neq k$ 时

$$\frac{\partial \hat{y}_{ni}}{\partial a_{nk}^{(L)}} = \frac{-\exp(a_{ni}^{(L)})\exp(a_{nk}^{(L)})}{\left(\sum_{j=1}^{K}\exp(a_{nj}^{(L)})\right)^2} = -\hat{y}_{ni}\hat{y}_{nk}$$

$$\frac{\partial L_n}{\partial a_{nk}^{(L)}} = \sum_{i=1}^{k}\frac{\partial L_n}{\partial \hat{y}_{ni}}\frac{\partial \hat{y}_{ni}}{\partial a_{nk}^{(L)}} = \sum_{i=1}^{k} -\frac{y_{ni}}{\hat{y}_{ni}}\frac{\partial \hat{y}_{ni}}{\partial a_{nk}^{(L)}} = \sum_{i\neq k}\frac{y_{ni}}{\hat{y}_{ni}}\hat{y}_{ni}\hat{y}_{nk} - \frac{y_{nk}}{\hat{y}_{nk}}\hat{y}_{nk}(1-\hat{y}_{nk})$$

$$= \sum_{i\neq k} y_{ni}\hat{y}_{nk} + y_{nk}\hat{y}_{nk} - y_{nk}$$

$$= \hat{y}_{nk} - y_{nk}$$

1. 对于如下的二维数据和卷积核

24	76	78	45	32
44	45	89	56	30
15	32	98	98	35
17	44	110	89	40
22	54	128	98	34

1	1
−1	−1

（1） 求有效卷积的卷积输出；

（2） 对卷积输出通过 ReLu 激活函数后，按照步幅 2 进行 2X2 窗口的最大池化，写出池化输出。

(1) 11 20 -22 -9

42 4 -51 -47

-14 -24 -3 4

-7 -28 -27 -3

(2) 42 0

0 4

3. 设一个简单的 RNN 网络的计算表示为

$$\boldsymbol{a}^{(t)}=\boldsymbol{b}+\boldsymbol{W}\boldsymbol{h}^{(t-1)}+\boldsymbol{U}\boldsymbol{x}^{(t)}$$

$$\boldsymbol{h}^{(t)}=\tanh(\boldsymbol{a}^{(t)})$$

$$o^{(t)}=c+\boldsymbol{V}\boldsymbol{h}^{(t)}$$

$$\hat{y}^{(t)}=o^{(t)}$$

（1） 若 $\boldsymbol{h}$ 和 $\boldsymbol{x}$ 均为 2 维向量，网络参数分别为

$\boldsymbol{b}=[0.2,-0.1]^{\mathrm{T}}$， $c=0.25$， $\boldsymbol{V}=[0.5,\quad 1]$

$$\boldsymbol{W}=\begin{bmatrix}0.8 & -0.1\\ -0.12 & 0.8\end{bmatrix},\ \boldsymbol{U}=\begin{bmatrix}2 & -1\\ 1 & 1\end{bmatrix}$$

若输入 $\boldsymbol{x}^{(t)}=[\sin 0.2\pi t\quad \cos 0.5\pi t]^{\mathrm{T}}$，请借助计算机，计算在 $1\le t\le 10$ 范围内的输出 $\hat{y}^{(t)}$ 序列。

（2） 对于该网络，若采用误差平方作为目标函数，以（1）中给出的系数为初始值，设状态初始值为 $\boldsymbol{h}^{(0)}=\boldsymbol{0}$，若给出一个三个时刻的序列样本集：

$$\left\{\left(\boldsymbol{x}^{(t)},\hat{y}^{(t)}\right)\right\}_{t=1}^{3}=\left\{\left((1,2)^{\mathrm{T}},-1\right),\left((-1,0)^{\mathrm{T}},1\right),\left((1,-1)^{\mathrm{T}},2\right)\right\}$$

请利用通过时间的反向传播算法，对网络参数进行更新。

(1) $y^{(1)} = 1.14$

$y^{(2)} = 0.86$

$y^{(3)} = 1.42$

$y^{(4)} = 1.60$

$y^{(5)} = 1.08$

$y^{(6)} = -0.40$

$y^{(7)} = -1.12$

$y^{(8)} = -0.86$

$y^{(9)} = -1.00$

$y^{(10)} = -0.43$

(2) 前向传播计算得到.

$h^{(1)} = \begin{pmatrix} 0.197 \\ 0.994 \end{pmatrix}$ $h^{(2)} = \begin{pmatrix} -0.940 \\ -0.317 \end{pmatrix}$ $h^{(3)} = \begin{pmatrix} 0.986 \\ -0.236 \end{pmatrix}$

$\hat{y}^{(1)} = 1.34$ $y^{(2)} = -0.54$ $y^{(3)} = 0.51$

由 $J^{(t)} = (y^{(t)} - \hat{y}^{(t)})^2$

$$\frac{\partial J^{(t)}}{\partial y^{(t)}} = -2(y^{(t)} - \hat{y}^{(t)}) = 2(\hat{y}(t) - y(t))$$

$$\frac{\partial J}{\partial O^{(t)}} = \frac{\partial J}{\partial y}\frac{\partial y}{\partial O^{(t)}} = -2(y^{(t)} - \hat{y}^{(t)}) = \{4.685, -3.075, -2.987\}$$

$$\frac{\partial J}{\partial h^{(t)}} = \frac{\partial J}{\partial O^{(t)}}\frac{\partial O^{(t)}}{\partial h^{(t)}} + \frac{\partial J}{\partial O^{(t+1)}}\frac{\partial h^{(t+1)}}{\partial h^{(t)}}$$

$$= 2V^T(y^{(t)} - \hat{y}^{(t)}) + W^T(y^{(t+1)} - \hat{y}^{(t+1)})\,\mathrm{diag}\{1 - h^{(t+1)2}\}$$

$$= \left\{\begin{pmatrix} 2.343 \\ 4.685 \end{pmatrix}, \begin{pmatrix} -3.075 \\ 1.743 \end{pmatrix}, \begin{pmatrix} -1.497 \\ -2.990 \end{pmatrix}\right\}$$

$$\therefore \frac{\partial J}{\partial c} = \sum_t \frac{\partial J}{\partial O^{(t)}} = 0.688$$

$$\frac{\partial J}{\partial b} = \sum_t \mathrm{diag}\{1 - h^{(t)2}\}\frac{\partial J}{\partial h^{(t)}} = \begin{pmatrix} -0.456 \\ -7.438 \end{pmatrix}$$

$$\frac{\partial J}{\partial V} = \sum_t \frac{\partial J}{\partial O^{(t)}} h^{(t)} = \begin{pmatrix} -0.436 \\ -3.169 \end{pmatrix}$$

$$\frac{\partial J}{\partial W} = \sum_t \mathrm{diag}\{1 - h^{(t)2}\} \frac{\partial J}{\partial h^{(t)}} h^{(t-1)T} = \begin{pmatrix} 0.008 & -0.142 \\ 1.728 & -3.760 \end{pmatrix}$$

$$\frac{\partial J}{\partial U} = \sum_t \mathrm{diag}\{1 - h^{(t)2}\} \frac{\partial J}{\partial h^{(t)}} x^{(t)T} = \begin{pmatrix} 1.374 & 2.559 \\ 1.920 & 2.935 \end{pmatrix}$$