Lecture10-11 作业

1，假设$g_0(\vec{x})=1$，以下哪一组（$\alpha_0,\alpha_1,\alpha_2$）允许$G(\vec{x})=sign(\sum_{t=0}^{2}\alpha_t g_t(\vec{x}))$实现$OR(g_1,g_2)$的功能。（a）（-3，+1，+1）；（b）（-1，+1，+1）；（c）（+1，+1，+1）；（d）（+3，+1，+1）。

解：$OR(g_1,g_2)$的关系意味着只要有一个$g_i=1$，输出即为1，当两个都为“-1”时，输出才为-1。

根据题目条件:

(a)$G(\vec{x})=sign(-3+g_1(\vec{x})+g_2(\vec{x}))$，当$g_1(\vec{x})=g_2(\vec{x})=1$时，$G(\vec{x})=-1$，不满足定义；

(b) $G(\vec{x})=sign(-1+g_1(\vec{x})+g_2(\vec{x}))$，当$g_1(\vec{x})=-1$，$g_2(\vec{x})=1$，*或者*$g_1(\vec{x})=1$，$g_2(\vec{x})=-1$时，$G(\vec{x})=-1$，不满足定义；

(c) $G(\vec{x})=sign(1+g_1(\vec{x})+g_2(\vec{x}))$，$g_1(\vec{x})$*与*$g_2(\vec{x})$取任意的+1和-1时，均能满足$OR(g_1,g_2)$的定义；

(d) $G(\vec{x})=sign(3+g_1(\vec{x})+g_2(\vec{x}))$，当$g_1(\vec{x})=g_2(\vec{x})=-1$时，$G(\vec{x})=1$，不满足定义。

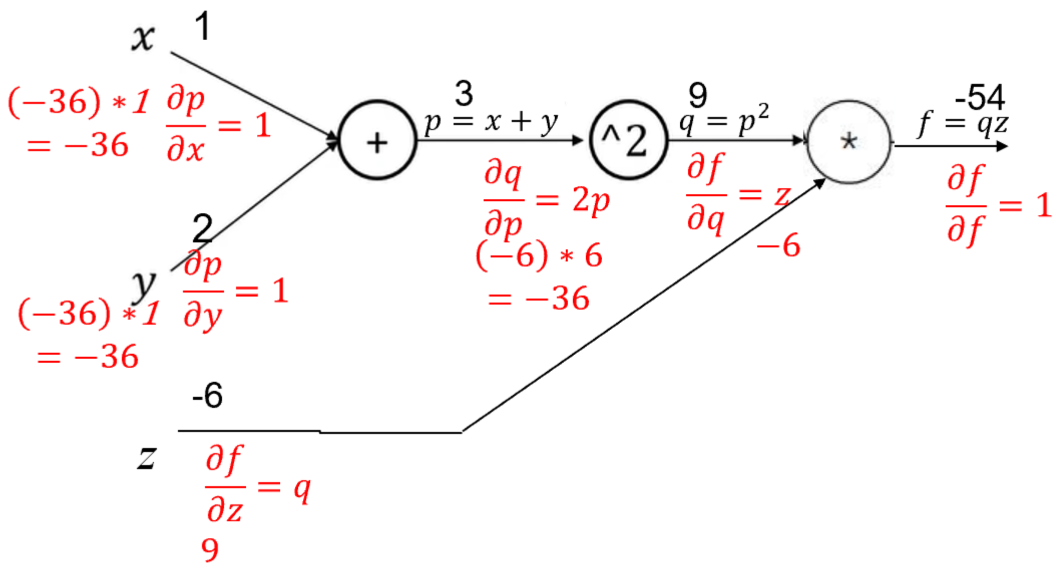
所以，只有(c)满足定义。

2，在3-5-1的神经网络中，网络参数有多少？

解：在第一层3-5中的参数为：（3+1（常数项））*5=20；在第二层5-1中的参数为（5+1（常数项））*1=6，所以，网络参数一共为26。

3，画出（$x+y$）^2z的计算图，当 x=1,y=2,z=-6 时，写出前向传播的数值和反向传播的梯度值。

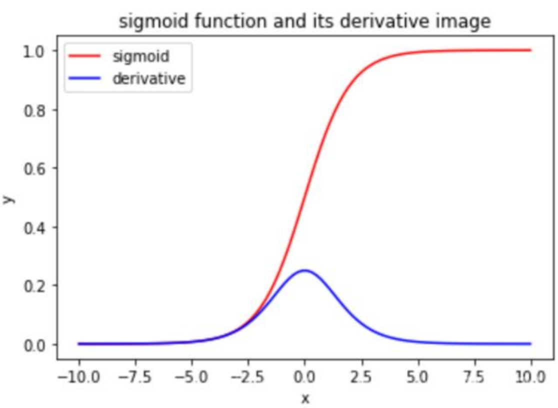
解：



4，计算 Sigmoid 函数、双曲正切函数和 ReLU 函数的导数函数，分析这三个函数作为激活函数时的优缺点。

解：（1）对于 Sigmoid 函数，$\theta(x)=\frac{1}{1+e^{-x}}$

$$\frac{\partial\theta(x)}{\partial x}=\frac{e^{-x}}{(1+e^{-x})^2}=\frac{1+e^{-x}-1}{1+e^{-x}}\frac{1}{1+e^{-x}}=(1-\theta(x))\theta(x)$$

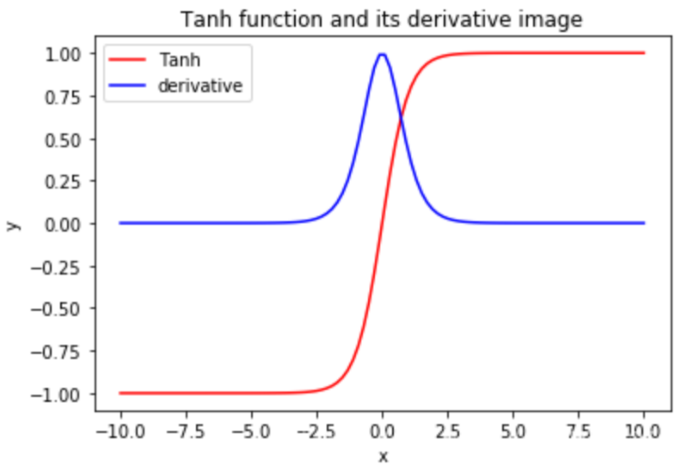


Sigmoid 函数作为激活函数的优点是连续、单调、可导且具有非线性特点。但其缺点是非中心对称，输出不是 0 均值的，同时它的导数函数曲线见右图，x 变化很小的范围内，导数才有值且不大，在神经

网络应用中，训练过程使用的是反向传播算法，通过链式法则回传的梯度不断相乘，因此，Sigmoid 函数作为激活函数时会导致梯度消失问题，尤其在网络比较深时会达不到训练效果。

（2）对于双曲正切函数，$tanh(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}}$

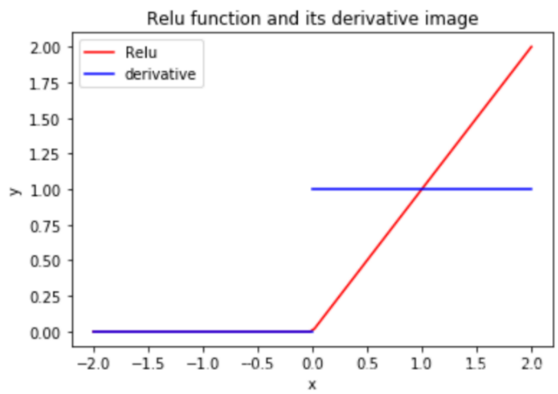
$$\frac{\partial \tanh(x)}{\partial x}=\frac{(e^x+e^{-x})(e^x+e^{-x})-(e^x-e^{-x})(e^x-e^{-x})}{(e^x+e^{-x})^2}=1-tanh^2(x)$$



双曲正切函数作为激活函数的优点是中心对称、单调、连续、可导且具有非线性特点。但其缺点是它的导数函数曲线见右图，x 变化更小的范围内，导数才有值且不大于 1，在神经网络应用中，训练过程使用的是反向传播算法，通过链式法则回传的梯度不断相乘，因此，双曲正切函数作为激活函数时同样会导致梯度消失问题，尤其在网络比较深时会达不到训练效果。

（3）对于 ReLU 函数，$ReLU(x)=\max(0,x)$

$$\frac{\partial \mathrm{ReLU}(x)}{\partial x}=\begin{cases}1 & x>0\\ 0 & x\leq 0\end{cases}$$



ReLU 函数作为激活函数的优点是梯度不会饱和，解决了梯度消失问题，没有指数计算，计算复杂度低。它的缺点是非中心对称，负数部分的梯度为

0，导致相应参数不会更新。

5，对于一幅 300*300 大小的彩色（RGB）图像，（1）如果输入端与有 100 个神经元的第一层隐含层用全链接方式（Fully Connected neural Network）连接时，请问这一层会包含多少参数？（2）如果用 100 个 5*5*3 大小的滤波器作卷积操作，那么这一层的参数为多少？如果滤波器移动步长（stride=1）为 1，经过卷积计算后的输出端神经元个数有多少？

解：（1）因为是 RGB 图像，所以共有 3 个通道，全连接情况下包含的参数是：300*300（图像大小）*3（颜色通道数）*100（第一层神经元个数）+1*100（每个神经元都要与输入层的常数项连接）=27000100；

（2）卷积操作时，第一层的参数是由卷积滤波器大小和常数项共同确定的，因此其包含的参数为：（5*5（滤波器大小）*3（颜色通道数）+1（常数项））*100（滤波器个数）=7600；（即：F=5，K=100，D1=3，参数量=（F*F*D1）*K+K=5*5*3*100+100=7600）

因为卷积核为 5*5*3，且填充值为 0，卷积后第一层神经元的个数为（（300（图像长或宽）-5（滤波器大小）/1（移动步长））+1）^2*100（滤波器个数）=8761600（相当于 W1=300，H1=300，D1=3，F=5，K=100，S=1，P=0，W2=（W1-F+2P）/S+1=（300-5+0）/1+1=296；H2=（H1-F+2P）/S+1=（300-5+0）/1+1=296；D2=K=100，神经元个数为：W2*H2*D2=296*296*100=8761600）

6，某一个卷积神经网络结构如下：

（i）输入层 Input 的 RGB 图像大小是 227*227*3。

（ii）第 1 层卷积层 Conv-1 是通过对输入图像用 96 个 11*11*3 大小的滤波器通过步长(stride)为 4，不做边缘填充（padding）得到的。

（iii）接下来是池化层 MaxPool-1，它用 3*3 尺寸、步长为 2 对 Conv-1 做 Max Pooling 操作。

（iv）然后我们对图像进行边缘填充，填充值为 2（如原来图像大小为 7*7 时，做填充值为 2 的填充后，图像大小变为 11*11），用 256 个 5*5 大小的滤波器按步长为 1，做第二次卷积操作，得到 Conv-2 层。

（v）再接一个池化层 MaxPool-2，它用 3*3 尺寸、步长为 2 做一次 Max Pooling 操作。

（vi）MaxPool-2 层输出去接一个有 4096 个神经元的全连接层 FC-1。

（vii）再接一个全连接层 FC-2 实现对 1000 个类别的分类。

请计算：（1）输入层到 Conv-1 层的参数量有多少？（2）经过池化层 MaxPool-1 后的神经元是多少？（3）经过第二次卷积操作后的图像大小为多少？（4）MaxPool-2 层到 FC-1 层的参数量是多少？（5）FC-1 层到 FC-2 层的参数量是多少？

解：输入图像大小为 227*227*3（即：W1=227，H1=227，D1=3），第一层卷积核为 11*11（即：F=11），共 96 个滤波器（即：K=96），步长为 4（即：S=4），边缘填充为 0（即：P=0），则卷积以后的图像边

长为：（（227-11+2*0）/4）+1=55，大小为 55*55（即 W2=H2=55），与 96 个滤波器构成特征图，所以卷积层 Conv-1 的神经元个数为 55*55*96=290400，（1）输入层到 Conv-1 层的参数量为 F*F*D1*K+1*K，即：11*11*3*96+1*96=34944；对 55*55 大小的图像做第一次 Maxpooling，这时候通道数 96 保持不变，因为它用 3*3 大小的尺寸以步长为 2 做 Maxpooling，则得到的图像边长为（（55-3）/2）+1=27，图像大小为 27*27，（2）经过池化层 MaxPool-1 后的神经元为 27*27*96；再做第二次卷积，此时是对 27*27 大小的图像，用 5*5 大小的滤波器按步长为 1，边缘填充值为 2 做卷积，滤波器个数为 256 个，所以，卷积后图像的边长为（（27-5+2*2）/1）+1=27，大小为 27*27，与 256 个滤波器构成特征图，所以卷积层 Conv-2 的神经元个数为 27*27*256=186624，（3）经过第二次卷积操作后的图像大小为 27*27，MaxPool-1 层到 Conv-2 层的参数量为：5*5*96（池化层的通道数）*256（滤波器个数）+256（常数项）=614656；再经过池化层 MaxPool-2，3*3 尺寸、步长为 2，则得到的图像边长为（（27-3）/2）+1=13，图像大小为 13*13，上一层的通道数是 256，所以，MaxPool-2 层的神经元个数为 13*13*256=43264；MaxPool-2 层输出去接一个有 4096 个神经元的全连接层 FC-1，所以，（4）MaxPool-2 层到 FC-1 层的参数量是 13*13*256*4096+4096（常数项）=177213440；最后的输出层要对 1000 个类别进行分类，即 FC-2 层的神经元个数是 1000 个，而输入是 4096 个神经元，所以，（5）FC-1 层到 FC-2 层的参数量是 4096*1000+1000=4097000

7，有训练样本集为：$\mathrm{D}=\{(\vec{x}_1,y_1)=((1,1)^T,1),(\vec{x}_2,y_2)=((-1,-1)^T,1),(\vec{x}_3,y_3)=((-1,1)^T,-1),(\vec{x}_4,y_4)=((1,-1)^T,-1)\}$，假设某神经网络结构为第一层有两个神经元，第二层有三个神经元，第三层有一个神经元，前两层每个神经元的激活函数为ReLU（即$x_d^{(l)}=\max(0,s_d^{(l)})$，这里$s_d^{(l)}$代表第l层第d个神经元的输入，$x_d^{(l)}$代表该神经元的输出），第三层为线性输出，即$\hat{y}=\mathrm{s}_1^{(3)}$。误差函数为：$E_{in}=\frac{1}{N}\sum_n(y_n-\hat{y}_n)^2$，学习率为0.01。假设初始权系数矩阵定义如下：

$$\mathbf{w}_0^{(1)}=\begin{pmatrix}1&1\\1&1\\1&1\end{pmatrix},\ \mathbf{w}_0^{(2)}=\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix},\ \mathbf{w}_0^{(3)}=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

其中w的下标0代表迭代次数为0（即初始状态），上标数字分别代表第1、2、3层。要求将上述训练样本集的样本用反向传播法按顺序进行一轮训练，写出每一次迭代时各层的权系数矩阵，即：t=1时，进入样本$\vec{x}_1$，得到$\mathbf{w}_1^{(1)}$、$\mathbf{w}_1^{(2)}$和$\mathbf{w}_1^{(3)}$；t=2时，进入样本$\vec{x}_2$，得到$\mathbf{w}_2^{(1)}$、$\mathbf{w}_2^{(2)}$和$\mathbf{w}_2^{(3)}$；t=3时，进入样本$\vec{x}_3$，得到$\mathbf{w}_3^{(1)}$、$\mathbf{w}_3^{(2)}$和$\mathbf{w}_3^{(3)}$；t=4时，进入样本$\vec{x}_4$，得到$\mathbf{w}_4^{(1)}$、$\mathbf{w}_4^{(2)}$和$\mathbf{w}_4^{(3)}$

解：

（1）算法步骤描述：

假设训练样本集有 N 个样本$\{\vec{x}_1,\dots\vec{x}_n,\dots\vec{x}_N\}$，每个样本有 d 维特征，写成增广向量后是 d+1 维，$\vec{x}_n=(1,x_{n1},\dots x_{nd})^T$，将神经网络的输入层当第 0 层，所以写为：$\vec{x}_n^{(0)}=(1,x_{n1}^{(0)},\dots x_{nd}^{(0)})^T$，当 d=2 时，$\vec{x}_n^{(0)}=(1,x_{n1}^{(0)},x_{n2}^{(0)})^T$

假设第一层有两个神经元，第二层有三个神经元，第三层有一个神经元。

第一层、第二层和第三层的权系数矩阵分别为：

$$\mathbf{w}^{(1)} = \begin{pmatrix} w_{01}^{(1)} & w_{02}^{(1)} \\ w_{11}^{(1)} & w_{12}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} \end{pmatrix}, \quad \mathbf{w}^{(2)} = \begin{pmatrix} w_{01}^{(2)} & w_{02}^{(2)} & w_{03}^{(2)} \\ w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \end{pmatrix}, \quad \mathbf{w}^{(3)} = \begin{pmatrix} w_{01}^{(3)} \\ w_{11}^{(3)} \\ w_{21}^{(3)} \\ w_{31}^{(3)} \end{pmatrix}$$

则第一层神经元的输入为:

$$\begin{pmatrix} s_1^{(1)} \\ s_2^{(1)} \end{pmatrix} = (\mathbf{w}^{(1)})^T \vec{x}_n^{(0)}$$

假设第一层神经元的激活函数为ReLU，即：$x^{(1)} = \max(0, s^{(1)})$，则:

$$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} \max(0, s_1^{(1)}) \\ \max(0, s_2^{(1)}) \end{pmatrix}$$

第二层神经元的输入为:

$$\begin{pmatrix} s_1^{(2)} \\ s_2^{(2)} \\ s_3^{(2)} \end{pmatrix} = (\mathbf{w}^{(2)})^T \begin{pmatrix} 1 \\ x_1^{(1)} \\ x_2^{(1)} \end{pmatrix}$$

假设第二层神经元的激活函数为ReLU，即：$x^{(2)} = \max(0, s^{(2)})$，则:

$$\begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} = \begin{pmatrix} \max(0, s_1^{(2)}) \\ \max(0, s_2^{(2)}) \\ \max(0, s_3^{(2)}) \end{pmatrix}$$

则第三层的输入为:

$$\mathrm{s}_1^{(3)} = (\mathbf{w}^{(3)})^T \begin{pmatrix} 1 \\ x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix}$$

因为第三层是线性操作，即输出$\hat{y} = \mathrm{s}_1^{(3)}$

对于输入样本$\vec{x}_n$，假设其标签为y_n，采用平方误差函数。即：$e_n = (y_n - \hat{y}_n)^2$

$$\delta_1^{(3)} = -2(y_n - \mathrm{s}_1^{(3)})$$

运用反向传播法，于是：$\delta_j^{(2)} = \sum_k (\delta_k^{(3)})(w_{jk}^{(3)})(x_j^{(2)})'$

对于ReLU来说，其导数为：$(x_j^{(L)})' = \left[\!\left[s_j^{(L-1)} \geq 0 \right]\!\right]$

所以：$\delta_j^{(2)}=\sum_k(\delta_k^{(3)})(w_{jk}^{(3)})[\![s_j^{(2)}\geq 0]\!]=\delta_1^{(3)}w_{j1}^{(3)}[\![s_j^{(2)}\geq 0]\!]$

即：$\vec{\delta}^{(2)}=\begin{pmatrix}\delta_1^{(2)}\\ \delta_2^{(2)}\\ \delta_3^{(2)}\end{pmatrix}=\begin{pmatrix}\delta_1^{(3)}w_{11}^{(3)}[\![s_1^{(2)}\geq 0]\!]\\ \delta_1^{(3)}w_{21}^{(3)}[\![s_2^{(2)}\geq 0]\!]\\ \delta_1^{(3)}w_{31}^{(3)}[\![s_3^{(2)}\geq 0]\!]\end{pmatrix}$

继续运用反向传播法，于是：$\delta_j^{(1)}=\sum_k(\delta_k^{(2)})(w_{jk}^{(2)})(x_j^{(1)})'$，所以：

$$\delta_j^{(1)}=\sum\nolimits_k(\delta_k^{(2)})(w_{jk}^{(2)})[\![s_j^{(1)}\geq 0]\!]=(\delta_1^{(2)}w_{j1}^{(2)}+\delta_2^{(2)}w_{j2}^{(2)}+\delta_3^{(2)}w_{j3}^{(2)})[\![s_j^{(1)}\geq 0]\!]$$

由此可以得到：

$$\vec{\delta}^{(1)}=\begin{pmatrix}\delta_1^{(1)}\\ \delta_2^{(1)}\end{pmatrix}=\begin{pmatrix}[\![s_1^{(1)}\geq 0]\!] & 0\\ 0 & [\![s_2^{(1)}\geq 0]\!]\end{pmatrix}\begin{pmatrix}w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}\begin{pmatrix}\delta_1^{(2)}\\ \delta_2^{(2)}\\ \delta_3^{(2)}\end{pmatrix}$$

假定t表示迭代次数，η为学习步长，利用梯度下降法进行权系数更新：

$$\mathbf{w}_{t+1}^{(1)}=\mathbf{w}_t^{(1)}-\eta\vec{x}_n^{(0)}\overrightarrow{(\delta^{(1)})}^T=\mathbf{w}_t^{(1)}-\eta\begin{pmatrix}1\\ x_{n1}^{(0)}\\ x_{n2}^{(0)}\end{pmatrix}\left(\delta_1^{(1)},\delta_2^{(1)}\right)$$

$$=\begin{pmatrix}w_{01}^{(1)} & w_{02}^{(1)}\\ w_{11}^{(1)} & w_{12}^{(1)}\\ w_{21}^{(1)} & w_{22}^{(1)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(1)} & \delta_2^{(1)}\\ x_{n1}^{(0)}\delta_1^{(1)} & x_{n1}^{(0)}\delta_2^{(1)}\\ x_{n2}^{(0)}\delta_1^{(1)} & x_{n2}^{(0)}\delta_2^{(1)}\end{pmatrix}$$

$$\mathbf{w}_{t+1}^{(2)}=\mathbf{w}_t^{(2)}-\eta\vec{x}_n^{(1)}\overrightarrow{(\delta^{(2)})}^T=\mathbf{w}_t^{(2)}-\eta\begin{pmatrix}1\\ x_1^{(1)}\\ x_2^{(1)}\end{pmatrix}\left(\delta_1^{(2)},\delta_2^{(2)},\delta_3^{(2)}\right)$$

$$=\begin{pmatrix}w_{01}^{(2)} & w_{02}^{(2)} & w_{03}^{(2)}\\ w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(2)} & \delta_2^{(2)} & \delta_3^{(2)}\\ x_1^{(1)}\delta_1^{(2)} & x_1^{(1)}\delta_2^{(2)} & x_1^{(1)}\delta_3^{(2)}\\ x_2^{(1)}\delta_1^{(2)} & x_2^{(1)}\delta_2^{(2)} & x_2^{(1)}\delta_3^{(2)}\end{pmatrix}$$

$$\mathbf{w}_{t+1}^{(3)}=\mathbf{w}_t^{(3)}-\eta\vec{x}_n^{(2)}\overrightarrow{(\delta^{(3)})}^T=\mathbf{w}_t^{(3)}-\eta\begin{pmatrix}1\\ x_1^{(2)}\\ x_2^{(2)}\\ x_3^{(2)}\end{pmatrix}\delta_1^{(3)}=\begin{pmatrix}w_{01}^{(3)}\\ w_{11}^{(3)}\\ w_{21}^{(3)}\\ w_{31}^{(3)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(3)}\\ x_1^{(2)}\delta_1^{(3)}\\ x_2^{(2)}\delta_1^{(3)}\\ x_3^{(2)}\delta_1^{(3)}\end{pmatrix}$$

反复迭代至T次。

（2）代入习题数据的解答流程：

t=0

$$\mathbf{w}_0^{(1)}=\begin{pmatrix}1&1\\1&1\\1&1\end{pmatrix},\ \mathbf{w}_0^{(2)}=\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix},\ \mathbf{w}_0^{(3)}=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$$

t=1时，对于第一个样本$\vec{x}_1=(1,1)^T$，则第一层神经元的输入为:

$$\begin{pmatrix}s_1^{(1)}\\s_2^{(1)}\end{pmatrix}=(\mathbf{w}^{(1)})^T\vec{x}_n^{(0)}=\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}3\\3\end{pmatrix}$$

则: $\begin{pmatrix}x_1^{(1)}\\x_2^{(1)}\end{pmatrix}=\begin{pmatrix}\max(0,s_1^{(1)})\\\max(0,s_2^{(1)})\end{pmatrix}=\begin{pmatrix}3\\3\end{pmatrix}$

第二层神经元的输入为:

$$\begin{pmatrix}s_1^{(2)}\\s_2^{(2)}\\s_3^{(2)}\end{pmatrix}=(\mathbf{w}^{(2)})^T\begin{pmatrix}1\\x_1^{(1)}\\x_2^{(1)}\end{pmatrix}=\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}\begin{pmatrix}1\\3\\3\end{pmatrix}=\begin{pmatrix}7\\7\\7\end{pmatrix}$$

则: $\begin{pmatrix}x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}=\begin{pmatrix}7\\7\\7\end{pmatrix}$

则第三层的输入为:

$$s_1^{(3)}=(\mathbf{w}^{(3)})^T\begin{pmatrix}1\\x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}=\begin{pmatrix}1&1&1&1\end{pmatrix}\begin{pmatrix}1\\7\\7\\7\end{pmatrix}=22$$

即输出$\hat{y}=s_1^{(3)}=22$

对于样本$\vec{x}_1$，其标签为1，采用平方误差函数：$e_n=(y_n-\hat{y}_n)^2$，则:

$$\delta_1^{(3)}=-2\left(y_n-s_1^{(3)}\right)=-2(1-22)=42$$

运用反向传播法，于是:

$$\delta_j^{(2)}=\sum\nolimits_k(\delta_k^{(3)})(w_{jk}^{(3)})\left[\!\left[s_j^{(2)}\geq 0\right]\!\right]=\delta_1^{(3)}w_{j1}^{(3)}\left[\!\left[s_j^{(2)}\geq 0\right]\!\right]$$

即：$\vec{\delta}^{(2)}=\begin{pmatrix}\delta_1^{(2)}\\ \delta_2^{(2)}\\ \delta_3^{(2)}\end{pmatrix}=\begin{pmatrix}\delta_1^{(3)}w_{11}^{(3)}[\![s_1^{(2)}\geq 0]\!]\\ \delta_1^{(3)}w_{21}^{(3)}[\![s_2^{(2)}\geq 0]\!]\\ \delta_1^{(3)}w_{31}^{(3)}[\![s_3^{(2)}\geq 0]\!]\end{pmatrix}=\begin{pmatrix}42*1*1\\ 42*1*1\\ 42*1*1\end{pmatrix}=\begin{pmatrix}42\\ 42\\ 42\end{pmatrix}$

继续运用反向传播法，于是：$\delta_j^{(1)}=\sum_k(\delta_k^{(2)})(w_{jk}^{(2)})(x_j^{(1)})'$，所以：

$$\delta_j^{(1)}=\sum\nolimits_k(\delta_k^{(2)})(w_{jk}^{(2)})[\![s_j^{(1)}\geq 0]\!]=(\delta_1^{(2)}w_{j1}^{(2)}+\delta_2^{(2)}w_{j2}^{(2)}+\delta_3^{(2)}w_{j3}^{(2)})[\![s_j^{(1)}\geq 0]\!]$$

由此可以得到：

$$\vec{\delta}^{(1)}=\begin{pmatrix}\delta_1^{(1)}\\ \delta_2^{(1)}\end{pmatrix}=\begin{pmatrix}[\![s_1^{(1)}\geq 0]\!] & 0\\ 0 & [\![s_2^{(1)}\geq 0]\!]\end{pmatrix}\begin{pmatrix}w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}\begin{pmatrix}\delta_1^{(2)}\\ \delta_2^{(2)}\\ \delta_3^{(2)}\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}42\\ 42\\ 42\end{pmatrix}=\begin{pmatrix}126\\ 126\end{pmatrix}$$

令$\eta=0.01$，利用梯度下降法进行权系数更新：

$$\mathbf{w}_1^{(1)}=\mathbf{w}_0^{(1)}-\eta\vec{x}_n^{(0)}(\overrightarrow{\delta}^{(1)})^T=\mathbf{w}_t^{(1)}-\eta\begin{pmatrix}1\\ x_{n1}^{(0)}\\ x_{n2}^{(0)}\end{pmatrix}\left(\delta_1^{(1)},\delta_2^{(1)}\right)$$

$$=\begin{pmatrix}w_{01}^{(1)} & w_{02}^{(1)}\\ w_{11}^{(1)} & w_{12}^{(1)}\\ w_{21}^{(1)} & w_{22}^{(1)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(1)} & \delta_2^{(1)}\\ x_{n1}^{(0)}\delta_1^{(1)} & x_{n1}^{(0)}\delta_2^{(1)}\\ x_{n2}^{(0)}\delta_1^{(1)} & x_{n2}^{(0)}\delta_2^{(1)}\end{pmatrix}$$

$$=\begin{pmatrix}1 & 1\\ 1 & 1\\ 1 & 1\end{pmatrix}-0.01\begin{pmatrix}126 & 126\\ 1*126 & 1*126\\ 1*126 & 1*126\end{pmatrix}=\begin{pmatrix}-0.26 & -0.26\\ -0.26 & -0.26\\ -0.26 & -0.26\end{pmatrix}$$

$$\mathbf{w}_1^{(2)}=\mathbf{w}_0^{(2)}-\eta\vec{x}_n^{(1)}(\overrightarrow{\delta}^{(2)})^T=\mathbf{w}_0^{(2)}-\eta\begin{pmatrix}1\\ x_1^{(1)}\\ x_2^{(1)}\end{pmatrix}\left(\delta_1^{(2)},\delta_2^{(2)},\delta_3^{(2)}\right)$$

$$=\begin{pmatrix}w_{01}^{(2)} & w_{02}^{(2)} & w_{03}^{(2)}\\ w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(2)} & \delta_2^{(2)} & \delta_3^{(2)}\\ x_1^{(1)}\delta_1^{(2)} & x_1^{(1)}\delta_2^{(2)} & x_1^{(1)}\delta_3^{(2)}\\ x_2^{(1)}\delta_1^{(2)} & x_2^{(1)}\delta_2^{(2)} & x_2^{(1)}\delta_3^{(2)}\end{pmatrix}$$

$$=\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}-0.01\begin{pmatrix}42 & 42 & 42\\ 3*42 & 3*42 & 3*42\\ 3*42 & 3*42 & 3*42\end{pmatrix}$$

$$=\begin{pmatrix}0.58 & 0.58 & 0.58\\ -0.26 & -0.26 & -0.26\\ -0.26 & -0.26 & -0.26\end{pmatrix}$$

$$\mathbf{w}_1^{(3)}=\mathbf{w}_0^{(3)}-\eta\vec{x}_n^{(2)}\overrightarrow{(\delta^{(3)})}^T=\mathbf{w}_0^{(3)}-\eta\begin{pmatrix}1\\x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}\delta_1^{(3)}=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}-0.01\begin{pmatrix}42\\7*42\\7*42\\7*42\end{pmatrix}$$

$$=\begin{pmatrix}0.58\\-1.94\\-1.94\\-1.94\end{pmatrix}$$

t=2，对于第二个样本$\vec{x}_2=(-1,-1)^T$，则第一层神经元的输入为：

$$\begin{pmatrix}s_1^{(1)}\\s_2^{(1)}\end{pmatrix}=(\mathbf{w}^{(1)})^T\vec{x}_n^{(0)}=\begin{pmatrix}-0.26&-0.26&-0.26\\-0.26&-0.26&-0.26\end{pmatrix}\begin{pmatrix}1\\-1\\-1\end{pmatrix}=\begin{pmatrix}0.26\\0.26\end{pmatrix}$$

则：$\begin{pmatrix}x_1^{(1)}\\x_2^{(1)}\end{pmatrix}=\begin{pmatrix}\max(0,s_1^{(1)})\\\max(0,s_2^{(1)})\end{pmatrix}=\begin{pmatrix}0.26\\0.26\end{pmatrix}$

第二层神经元的输入为：

$$\begin{pmatrix}s_1^{(2)}\\s_2^{(2)}\\s_3^{(2)}\end{pmatrix}=(\mathbf{w}^{(2)})^T\begin{pmatrix}1\\x_1^{(1)}\\x_2^{(1)}\end{pmatrix}=\begin{pmatrix}0.58&-0.26&-0.26\\0.58&-0.26&-0.26\\0.58&-0.26&-0.26\end{pmatrix}\begin{pmatrix}1\\0.26\\0.26\end{pmatrix}=\begin{pmatrix}0.44\\0.44\\0.44\end{pmatrix}$$

则：$\begin{pmatrix}x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}=\begin{pmatrix}0.44\\0.44\\0.44\end{pmatrix}$

则第三层的输入为：

$$\mathrm{s}_1^{(3)}=(\mathbf{w}^{(3)})^T\begin{pmatrix}1\\x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}=\begin{pmatrix}0.58&-1.94&-1.94&-1.94\end{pmatrix}\begin{pmatrix}1\\0.44\\0.44\\0.44\end{pmatrix}=-1.98$$

即输出$\hat{y}=\mathrm{s}_1^{(3)}=-1.98$

对于样本$\vec{x}_2$，其标签为1，采用平方误差函数：$e_n=(y_n-\hat{y}_n)^2$，则：

$$\delta_1^{(3)}=-2\left(y_n-\mathrm{s}_1^{(3)}\right)=-2\big(1-(-1.98)\big)=-5.96$$

运用反向传播法，于是：

$$\delta_j^{(2)}=\sum_k(\delta_k^{(3)})(w_{jk}^{(3)})[\![s_j^{(2)}\geq 0]\!]=\delta_1^{(3)}w_{j1}^{(3)}[\![s_j^{(2)}\geq 0]\!]$$

即：$\vec{\delta}^{(2)}=\begin{pmatrix}\delta_1^{(2)}\\ \delta_2^{(2)}\\ \delta_3^{(2)}\end{pmatrix}=\begin{pmatrix}\delta_1^{(3)}w_{11}^{(3)}[\![s_1^{(2)}\geq 0]\!]\\ \delta_1^{(3)}w_{21}^{(3)}[\![s_2^{(2)}\geq 0]\!]\\ \delta_1^{(3)}w_{31}^{(3)}[\![s_3^{(2)}\geq 0]\!]\end{pmatrix}=\begin{pmatrix}-5.96*(-1.94)*1\\ -5.96*(-1.94)*1\\ -5.96*(-1.94)*1\end{pmatrix}=\begin{pmatrix}11.56\\ 11.56\\ 11.56\end{pmatrix}$

继续运用反向传播法，于是：$\delta_j^{(1)}=\sum_k(\delta_k^{(2)})(w_{jk}^{(2)})(x_j^{(1)})'$，所以：

$$\delta_j^{(1)}=\sum_k(\delta_k^{(2)})(w_{jk}^{(2)})[\![s_j^{(1)}\geq 0]\!]=(\delta_1^{(2)}w_{j1}^{(2)}+\delta_2^{(2)}w_{j2}^{(2)}+\delta_3^{(2)}w_{j3}^{(2)})[\![s_j^{(1)}\geq 0]\!]$$

由此可以得到：

$$\vec{\delta}^{(1)}=\begin{pmatrix}\delta_1^{(1)}\\ \delta_2^{(1)}\end{pmatrix}=\begin{pmatrix}[\![s_1^{(1)}\geq 0]\!] & 0\\ 0 & [\![s_2^{(1)}\geq 0]\!]\end{pmatrix}\begin{pmatrix}w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}\begin{pmatrix}\delta_1^{(2)}\\ \delta_2^{(2)}\\ \delta_3^{(2)}\end{pmatrix}$$

$$=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}-0.26 & -0.26 & -0.26\\ -0.26 & -0.26 & -0.26\end{pmatrix}\begin{pmatrix}11.56\\ 11.56\\ 11.56\end{pmatrix}=\begin{pmatrix}-9.02\\ -9.02\end{pmatrix}$$

令$\eta=0.01$，利用梯度下降法进行权系数更新：

$$\mathbf{w}_2^{(1)}=\mathbf{w}_1^{(1)}-\eta\vec{x}_n^{(0)}(\overrightarrow{\delta}^{(1)})^T=\mathbf{w}_1^{(1)}-\eta\begin{pmatrix}1\\ x_{n1}^{(0)}\\ x_{n2}^{(0)}\end{pmatrix}\left(\delta_1^{(1)},\delta_2^{(1)}\right)$$

$$=\begin{pmatrix}w_{01}^{(1)} & w_{02}^{(1)}\\ w_{11}^{(1)} & w_{12}^{(1)}\\ w_{21}^{(1)} & w_{22}^{(1)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(1)} & \delta_2^{(1)}\\ x_{n1}^{(0)}\delta_1^{(1)} & x_{n1}^{(0)}\delta_2^{(1)}\\ x_{n2}^{(0)}\delta_1^{(1)} & x_{n2}^{(0)}\delta_2^{(1)}\end{pmatrix}$$

$$=\begin{pmatrix}-0.26 & -0.26\\ -0.26 & -0.26\\ -0.26 & -0.26\end{pmatrix}-0.01\begin{pmatrix}-9.02 & -9.02\\ (-1)*(-9.02) & (-1)*(-9.02)\\ (-1)*(-9.02) & (-1)*(-9.02)\end{pmatrix}$$

$$=\begin{pmatrix}-0.17 & -0.17\\ -0.35 & -0.35\\ -0.35 & -0.35\end{pmatrix}$$

$$\mathbf{w}_2^{(2)} = \mathbf{w}_1^{(2)} - \eta \vec{x}_n^{(1)} (\overrightarrow{\delta}^{(2)})^T = \mathbf{w}_1^{(2)} - \eta \begin{pmatrix} 1 \\ x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} \left(\delta_1^{(2)}, \delta_2^{(2)}, \delta_3^{(2)}\right)$$

$$= \begin{pmatrix} w_{01}^{(2)} & w_{02}^{(2)} & w_{03}^{(2)} \\ w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \end{pmatrix} - \eta \begin{pmatrix} \delta_1^{(2)} & \delta_2^{(2)} & \delta_3^{(2)} \\ x_1^{(1)}\delta_1^{(2)} & x_1^{(1)}\delta_2^{(2)} & x_1^{(1)}\delta_3^{(2)} \\ x_2^{(1)}\delta_1^{(2)} & x_2^{(1)}\delta_2^{(2)} & x_2^{(1)}\delta_3^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} 0.58 & 0.58 & 0.58 \\ -0.26 & -0.26 & -0.26 \\ -0.26 & -0.26 & -0.26 \end{pmatrix}$$

$$- 0.01 \begin{pmatrix} 11.56 & 11.56 & 11.56 \\ 0.26 * 11.56 & 0.26 * 11.56 & 0.26 * 11.56 \\ 0.26 * 11.56 & 0.26 * 11.56 & 0.26 * 11.56 \end{pmatrix}$$

$$= \begin{pmatrix} 0.46 & 0.46 & 0.46 \\ -0.29 & -0.29 & -0.29 \\ -0.29 & -0.29 & -0.29 \end{pmatrix}$$

$$\mathbf{w}_2^{(3)} = \mathbf{w}_1^{(3)} - \eta \vec{x}_n^{(2)} (\overrightarrow{\delta}^{(3)})^T = \mathbf{w}_1^{(3)} - \eta \begin{pmatrix} 1 \\ x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} \delta_1^{(3)}$$

$$= \begin{pmatrix} 0.58 \\ -1.94 \\ -1.94 \\ -1.94 \end{pmatrix} - 0.01 \begin{pmatrix} -5.96 \\ 0.44 * (-5.96) \\ 0.44 * (-5.96) \\ 0.44 * (-5.96) \end{pmatrix} = \begin{pmatrix} 0.64 \\ -1.91 \\ -1.91 \\ -1.91 \end{pmatrix}$$

t=3，对于第三个样本$\vec{x}_3 = (-1,1)^T$，则第一层神经元的输入为：

$$\begin{pmatrix} s_1^{(1)} \\ s_2^{(1)} \end{pmatrix} = (\mathbf{w}^{(1)})^T \vec{x}_n^{(0)} = \begin{pmatrix} -0.17 & -0.35 & -0.35 \\ -0.17 & -0.35 & -0.35 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.17 \\ -0.17 \end{pmatrix}$$

则：$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} \max(0, s_1^{(1)}) \\ \max(0, s_2^{(1)}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

第二层神经元的输入为：

$$\begin{pmatrix} s_1^{(2)} \\ s_2^{(2)} \\ s_3^{(2)} \end{pmatrix} = (\mathbf{w}^{(2)})^T \begin{pmatrix} 1 \\ x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} 0.46 & -0.29 & -0.29 \\ 0.46 & -0.29 & -0.29 \\ 0.46 & -0.29 & -0.29 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.46 \\ 0.46 \\ 0.46 \end{pmatrix}$$

则：$\begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} = \begin{pmatrix} 0.46 \\ 0.46 \\ 0.46 \end{pmatrix}$

则第三层的输入为:

$$\mathrm{s}_1^{(3)}=(\mathbf{w}^{(3)})^T\begin{pmatrix}1\\x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}=(0.64\quad -1.91\quad -1.91\quad -1.91)\begin{pmatrix}1\\0.46\\0.46\\0.46\end{pmatrix}=-2.00$$

即输出$\hat{y}=\mathrm{s}_1^{(3)}=-2.00$

对于样本$\vec{x}_3$，其标签为-1，采用平方误差函数：$e_n=(y_n-\hat{y}_n)^2$，则:

$$\delta_1^{(3)}=-2\left(y_n-\mathrm{s}_1^{(3)}\right)=-2\left(-1-(-2.00)\right)=-2.00$$

运用反向传播法，于是:

$$\delta_j^{(2)}=\sum\nolimits_k(\delta_k^{(3)})(w_{jk}^{(3)})\left[\!\left[s_j^{(2)}\geq 0\right]\!\right]=\delta_1^{(3)}w_{j1}^{(3)}\left[\!\left[s_j^{(2)}\geq 0\right]\!\right]$$

即: $\vec{\delta}^{(2)}=\begin{pmatrix}\delta_1^{(2)}\\\delta_2^{(2)}\\\delta_3^{(2)}\end{pmatrix}=\begin{pmatrix}\delta_1^{(3)}w_{11}^{(3)}\left[\!\left[s_1^{(2)}\geq 0\right]\!\right]\\\delta_1^{(3)}w_{21}^{(3)}\left[\!\left[s_2^{(2)}\geq 0\right]\!\right]\\\delta_1^{(3)}w_{31}^{(3)}\left[\!\left[s_3^{(2)}\geq 0\right]\!\right]\end{pmatrix}=\begin{pmatrix}(-2.00)*(-1.91)*1\\(-2.00)*(-1.91)*1\\(-2.00)*(-1.91)*1\end{pmatrix}=\begin{pmatrix}3.82\\3.82\\3.82\end{pmatrix}$

继续运用反向传播法，于是：$\delta_j^{(1)}=\sum_k(\delta_k^{(2)})(w_{jk}^{(2)})(x_j^{(1)})'$，所以:

$$\delta_j^{(1)}=\sum\nolimits_k(\delta_k^{(2)})(w_{jk}^{(2)})\left[\!\left[s_j^{(1)}\geq 0\right]\!\right]=(\delta_1^{(2)}w_{j1}^{(2)}+\delta_2^{(2)}w_{j2}^{(2)}+\delta_3^{(2)}w_{j3}^{(2)})\left[\!\left[s_j^{(1)}\geq 0\right]\!\right]$$

由此可以得到:

$$\vec{\delta}^{(1)}=\begin{pmatrix}\delta_1^{(1)}\\\delta_2^{(1)}\end{pmatrix}=\begin{pmatrix}\left[\!\left[s_1^{(1)}\geq 0\right]\!\right] & 0\\0 & \left[\!\left[s_2^{(1)}\geq 0\right]\!\right]\end{pmatrix}\begin{pmatrix}w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}\begin{pmatrix}\delta_1^{(2)}\\\delta_2^{(2)}\\\delta_3^{(2)}\end{pmatrix}$$

$$=\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}\begin{pmatrix}-0.29 & -0.29 & -0.29\\-0.29 & -0.29 & -0.29\end{pmatrix}\begin{pmatrix}3.82\\3.82\\3.82\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

令$\eta=0.01$，利用梯度下降法进行权系数更新:

$$\mathbf{w}_3^{(1)}=\mathbf{w}_2^{(1)}-\eta\vec{x}_n^{(0)}(\overrightarrow{\delta}^{(1)})^T=\mathbf{w}_2^{(1)}-\eta\begin{pmatrix}1\\x_{n1}^{(0)}\\x_{n2}^{(0)}\end{pmatrix}\left(\delta_1^{(1)},\delta_2^{(1)}\right)$$

$$=\begin{pmatrix}w_{01}^{(1)} & w_{02}^{(1)}\\w_{11}^{(1)} & w_{12}^{(1)}\\w_{21}^{(1)} & w_{22}^{(1)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(1)} & \delta_2^{(1)}\\x_{n1}^{(0)}\delta_1^{(1)} & x_{n1}^{(0)}\delta_2^{(1)}\\x_{n2}^{(0)}\delta_1^{(1)} & x_{n2}^{(0)}\delta_2^{(1)}\end{pmatrix}$$

$$=\begin{pmatrix}-0.17 & -0.17\\-0.35 & -0.35\\-0.35 & -0.35\end{pmatrix}-0.01\begin{pmatrix}0 & 0\\(-1)*0 & (-1)*0\\(+1)*0 & (+1)*0\end{pmatrix}$$

$$=\begin{pmatrix}-0.17 & -0.17\\-0.35 & -0.35\\-0.35 & -0.35\end{pmatrix}$$

$$\mathbf{w}_3^{(2)}=\mathbf{w}_2^{(2)}-\eta\vec{x}_n^{(1)}(\overrightarrow{\delta}^{(2)})^T=\mathbf{w}_2^{(2)}-\eta\begin{pmatrix}1\\x_1^{(1)}\\x_2^{(1)}\end{pmatrix}\left(\delta_1^{(2)},\delta_2^{(2)},\delta_3^{(2)}\right)$$

$$=\begin{pmatrix}w_{01}^{(2)} & w_{02}^{(2)} & w_{03}^{(2)}\\w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)}\\w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)}\end{pmatrix}-\eta\begin{pmatrix}\delta_1^{(2)} & \delta_2^{(2)} & \delta_3^{(2)}\\x_1^{(1)}\delta_1^{(2)} & x_1^{(1)}\delta_2^{(2)} & x_1^{(1)}\delta_3^{(2)}\\x_2^{(1)}\delta_1^{(2)} & x_2^{(1)}\delta_2^{(2)} & x_2^{(1)}\delta_3^{(2)}\end{pmatrix}$$

$$=\begin{pmatrix}0.46 & 0.46 & 0.46\\-0.29 & -0.29 & -0.29\\-0.29 & -0.29 & -0.29\end{pmatrix}-0.01\begin{pmatrix}3.82 & 3.82 & 3.82\\0*3.82 & 0*3.82 & 0*3.82\\0*3.82 & 0*3.82 & 0*3.82\end{pmatrix}$$

$$=\begin{pmatrix}0.42 & 0.42 & 0.42\\-0.29 & -0.29 & -0.29\\-0.29 & -0.29 & -0.29\end{pmatrix}$$

$$\mathbf{w}_3^{(3)}=\mathbf{w}_2^{(3)}-\eta\vec{x}_n^{(2)}(\overrightarrow{\delta}^{(3)})^T=\mathbf{w}_2^{(3)}-\eta\begin{pmatrix}1\\x_1^{(2)}\\x_2^{(2)}\\x_3^{(2)}\end{pmatrix}\delta_1^{(3)}$$

$$=\begin{pmatrix}0.64\\-1.91\\-1.91\\-1.91\end{pmatrix}-0.01\begin{pmatrix}-2.00\\0.46*(-2.00)\\0.46*(-2.00)\\0.46*(-2.00)\end{pmatrix}=\begin{pmatrix}0.66\\-1.90\\-1.90\\-1.90\end{pmatrix}$$

t=4，对于第四个样本$\vec{x}_2=(1,-1)^T$，则第一层神经元的输入为：

$$\begin{pmatrix}s_1^{(1)}\\s_2^{(1)}\end{pmatrix}=(\mathbf{w}^{(1)})^T\vec{x}_n^{(0)}=\begin{pmatrix}-0.17 & -0.35 & -0.35\\-0.17 & -0.35 & -0.35\end{pmatrix}\begin{pmatrix}1\\1\\-1\end{pmatrix}=\begin{pmatrix}-0.17\\-0.17\end{pmatrix}$$

则：$\begin{pmatrix} x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} \max(0, s_1^{(1)}) \\ \max(0, s_2^{(1)}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

第二层神经元的输入为：

$$\begin{pmatrix} s_1^{(2)} \\ s_2^{(2)} \\ s_3^{(2)} \end{pmatrix} = (\mathbf{w}^{(2)})^T \begin{pmatrix} 1 \\ x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} = \begin{pmatrix} 0.42 & -0.29 & -0.29 \\ 0.42 & -0.29 & -0.29 \\ 0.42 & -0.29 & -0.29 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.42 \\ 0.42 \\ 0.42 \end{pmatrix}$$

则：$\begin{pmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} = \begin{pmatrix} 0.42 \\ 0.42 \\ 0.42 \end{pmatrix}$

则第三层的输入为：

$$\mathrm{s}_1^{(3)} = (\mathbf{w}^{(3)})^T \begin{pmatrix} 1 \\ x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} = (0.66 \quad -1.90 \quad -1.90 \quad -1.90) \begin{pmatrix} 1 \\ 0.42 \\ 0.42 \\ 0.42 \end{pmatrix} = -1.73$$

即输出$\hat{y} = \mathrm{s}_1^{(3)} = -1.73$

对于样本$\vec{x}_4$，其标签为-1，采用平方误差函数：$e_n = (y_n - \hat{y}_n)^2$，则：

$$\delta_1^{(3)} = -2\left(y_n - \mathrm{s}_1^{(3)}\right) = -2\left(-1 - (-1.73)\right) = -1.46$$

运用反向传播法，于是：

$$\delta_j^{(2)} = \sum\nolimits_k (\delta_k^{(3)})(w_{jk}^{(3)})\left[\!\left[s_j^{(2)} \geq 0\right]\!\right] = \delta_1^{(3)} w_{j1}^{(3)} \left[\!\left[s_j^{(2)} \geq 0\right]\!\right]$$

即：$\vec{\delta}^{(2)} = \begin{pmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \\ \delta_3^{(2)} \end{pmatrix} = \begin{pmatrix} \delta_1^{(3)} w_{11}^{(3)} \left[\!\left[s_1^{(2)} \geq 0\right]\!\right] \\ \delta_1^{(3)} w_{21}^{(3)} \left[\!\left[s_2^{(2)} \geq 0\right]\!\right] \\ \delta_1^{(3)} w_{31}^{(3)} \left[\!\left[s_3^{(2)} \geq 0\right]\!\right] \end{pmatrix} = \begin{pmatrix} (-1.46) * (-1.90) * 1 \\ (-1.46) * (-1.90) * 1 \\ (-1.46) * (-1.90) * 1 \end{pmatrix} = \begin{pmatrix} 2.77 \\ 2.77 \\ 2.77 \end{pmatrix}$

继续运用反向传播法，于是：$\delta_j^{(1)} = \sum_k (\delta_k^{(2)})(w_{jk}^{(2)})(x_j^{(1)})'$，所以：

$$\delta_j^{(1)} = \sum\nolimits_k (\delta_k^{(2)})(w_{jk}^{(2)})\left[\!\left[s_j^{(1)} \geq 0\right]\!\right] = (\delta_1^{(2)} w_{j1}^{(2)} + \delta_2^{(2)} w_{j2}^{(2)} + \delta_3^{(2)} w_{j3}^{(2)})\left[\!\left[s_j^{(1)} \geq 0\right]\!\right]$$

由此可以得到：

$$\vec{\delta}^{(1)} = \begin{pmatrix} \delta_1^{(1)} \\ \delta_2^{(1)} \end{pmatrix} = \begin{pmatrix} [\![s_1^{(1)} \geq 0]\!] & 0 \\ 0 & [\![s_2^{(1)} \geq 0]\!] \end{pmatrix} \begin{pmatrix} w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \end{pmatrix} \begin{pmatrix} \delta_1^{(2)} \\ \delta_2^{(2)} \\ \delta_3^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -0.29 & -0.29 & -0.29 \\ -0.29 & -0.29 & -0.29 \end{pmatrix} \begin{pmatrix} 2.77 \\ 2.77 \\ 2.77 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

令$\eta = 0.01$，利用梯度下降法进行权系数更新：

$$\mathbf{w}_4^{(1)} = \mathbf{w}_3^{(1)} - \eta \vec{x}_n^{(0)} (\overrightarrow{\delta}^{(1)})^T = \mathbf{w}_3^{(1)} - \eta \begin{pmatrix} 1 \\ x_{n1}^{(0)} \\ x_{n2}^{(0)} \end{pmatrix} \left(\delta_1^{(1)}, \delta_2^{(1)} \right)$$

$$= \begin{pmatrix} w_{01}^{(1)} & w_{02}^{(1)} \\ w_{11}^{(1)} & w_{12}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} \end{pmatrix} - \eta \begin{pmatrix} \delta_1^{(1)} & \delta_2^{(1)} \\ x_{n1}^{(0)} \delta_1^{(1)} & x_{n1}^{(0)} \delta_2^{(1)} \\ x_{n2}^{(0)} \delta_1^{(1)} & x_{n2}^{(0)} \delta_2^{(1)} \end{pmatrix}$$

$$= \begin{pmatrix} -0.17 & -0.17 \\ -0.35 & -0.35 \\ -0.35 & -0.35 \end{pmatrix} - 0.01 \begin{pmatrix} 0 & 0 \\ (+1) * 0 & (+1) * 0 \\ (-1) * 0 & (-1) * 0 \end{pmatrix}$$

$$= \begin{pmatrix} -0.17 & -0.17 \\ -0.35 & -0.35 \\ -0.35 & -0.35 \end{pmatrix}$$

$$\mathbf{w}_4^{(2)} = \mathbf{w}_3^{(2)} - \eta \vec{x}_n^{(1)} (\overrightarrow{\delta}^{(2)})^T = \mathbf{w}_3^{(2)} - \eta \begin{pmatrix} 1 \\ x_1^{(1)} \\ x_2^{(1)} \end{pmatrix} \left(\delta_1^{(2)}, \delta_2^{(2)}, \delta_3^{(2)} \right)$$

$$= \begin{pmatrix} w_{01}^{(2)} & w_{02}^{(2)} & w_{03}^{(2)} \\ w_{11}^{(2)} & w_{12}^{(2)} & w_{13}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & w_{23}^{(2)} \end{pmatrix} - \eta \begin{pmatrix} \delta_1^{(2)} & \delta_2^{(2)} & \delta_3^{(2)} \\ x_1^{(1)} \delta_1^{(2)} & x_1^{(1)} \delta_2^{(2)} & x_1^{(1)} \delta_3^{(2)} \\ x_2^{(1)} \delta_1^{(2)} & x_2^{(1)} \delta_2^{(2)} & x_2^{(1)} \delta_3^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} 0.42 & 0.42 & 0.42 \\ -0.29 & -0.29 & -0.29 \\ -0.29 & -0.29 & -0.29 \end{pmatrix} - 0.01 \begin{pmatrix} 2.77 & 2.77 & 2.77 \\ 0 * 2.77 & 0 * 2.77 & 0 * 2.77 \\ 0 * 2.77 & 0 * 2.77 & 0 * 2.77 \end{pmatrix}$$

$$= \begin{pmatrix} 0.39 & 0.39 & 0.39 \\ -0.29 & -0.29 & -0.29 \\ -0.29 & -0.29 & -0.29 \end{pmatrix}$$

$$\mathbf{w}_4^{(3)} = \mathbf{w}_3^{(3)} - \eta \vec{x}_n^{(2)} (\overrightarrow{\delta}^{(3)})^T = \mathbf{w}_3^{(3)} - \eta \begin{pmatrix} 1 \\ x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{pmatrix} \delta_1^{(3)}$$

$$= \begin{pmatrix} 0.66 \\ -1.90 \\ -1.90 \\ -1.90 \end{pmatrix} - 0.01 \begin{pmatrix} -1.46 \\ 0.42 * (-1.46) \\ 0.42 * (-1.46) \\ 0.42 * (-1.46) \end{pmatrix} = \begin{pmatrix} 0.67 \\ -1.89 \\ -1.89 \\ -1.89 \end{pmatrix}$$