7，有训练样本集为：$D=\{(\vec{x}_1,y_1)=((1,1)^T,1),(\vec{x}_2,y_2)=((-1,-1)^T,1),(\vec{x}_3,y_3)=((-1,1)^T,-1),(\vec{x}_4,y_4)=((1,-1)^T,-1)\}$，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$\hat{y}=s_1^{(3)}$。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：

版权归SMART-AIA-HUST所有，严禁复制