Lecture 9 习题作业

1，假设有如下训练样本：$\vec{x}_1 = (0,0)^T$属于第一类，$\vec{x}_2 = (1,1)^T$属于第二类，$\vec{x}_3 = (-1,1)^T$属于第三类，请用多类分类中的 OVO（One-versus-one）策略，设计上述三类别的两两分类器，并分析测试样本$\vec{\mathrm{x}} = (1,-2)^T$属于那个类别。

解：利用 OVO 策略，对三个类别两两求分类面：

（1）用感知器算法求第一类和第二类之间的分类面

样本增广后为：$\vec{x}_1 = (1,0,0)^T$，$y_1 = 1$，$\vec{x}_2 = (1,1,1)^T$，$y_2 = -1$，初始化权重：$\vec{w}_{[1,2]}^{(0)} = (0,0,0)^T$

$sign\left(\vec{w}_{[1,2]}^{(0)T}\vec{x}_1\right) = 0 \neq y_1$，$\therefore$ $\vec{w}_{[1,2]}^{(1)} = \vec{w}_{[1,2]}^{(0)} + y_1\vec{x}_1 = (1,0,0)^T$，

$sign\left(\vec{w}_{[1,2]}^{(1)T}\vec{x}_2\right) = 1 \neq y_2$，$\therefore$ $\vec{w}_{[1,2]}^{(2)} = \vec{w}_{[1,2]}^{(1)} + y_2\vec{x}_2 = (0,-1,-1)^T$

$sign\left(\vec{w}_{[1,2]}^{(2)T}\vec{x}_1\right) = 0 \neq y_1$，$\therefore$ $\vec{w}_{[1,2]}^{(3)} = \vec{w}_{[1,2]}^{(2)} + y_1\vec{x}_1 = (1,-1,-1)^T$

$sign\left(\vec{w}_{[1,2]}^{(3)T}\vec{x}_2\right) = -1 = y_2$，且 $sign\left(\vec{w}_{[1,2]}^{(3)T}\vec{x}_1\right) = 1 = y_1$

$\therefore$ $\vec{w}_{[1,2]} = (1,-1,-1)^T$，分类面为：$1 - x_1 - x_2 = 0$

（2）用感知器算法求第一类和第三类之间的分类面

样本增广后为：$\vec{x}_1 = (1,0,0)^T$，$y_1 = 1$，$\vec{x}_3 = (1,-1,1)^T$，$y_3 = -1$，初始化权重：$\vec{w}_{[1,3]}^{(0)} = (0,0,0)^T$

$sign\left(\vec{w}_{[1,3]}^{(0)T}\vec{x}_1\right) = 0 \neq y_1$，$\therefore$ $\vec{w}_{[1,3]}^{(1)} = \vec{w}_{[1,3]}^{(0)} + y_1\vec{x}_1 = (1,0,0)^T$，

$sign\left(\vec{w}_{[1,3]}^{(1)T}\vec{x}_3\right) = 1 \neq y_3$，$\therefore$ $\vec{w}_{[1,3]}^{(2)} = \vec{w}_{[1,3]}^{(1)} + y_3\vec{x}_3 = (0,1,-1)^T$

$sign\left(\vec{w}_{[1,3]}^{(2)T}\vec{x}_1\right) = 0 \neq y_1$，$\therefore$ $\vec{w}_{[1,3]}^{(3)} = \vec{w}_{[1,3]}^{(2)} + y_1\vec{x}_1 = (1,1,-1)^T$

$sign\left(\vec{w}_{[1,3]}^{(3)T}\vec{x}_3\right) = -1 = y_3$，且 $sign\left(\vec{w}_{[1,3]}^{(3)T}\vec{x}_1\right) = 1 = y_1$

$\therefore\ \vec{w}_{[1,3]} = (1,1,-1)^T$，分类面为：$1 + x_1 - x_2 = 0$

（3）用感知器算法求第二类和第三类之间的分类面

样本增广后为：$\vec{x}_2 = (1,1,1)^T$，$y_2 = 1$，$\vec{x}_3 = (1,-1,1)^T$，$y_3 = -1$，初始化权重：$\vec{w}_{[2,3]}^{(0)} = (0,0,0)^T$

$sign\left(\vec{w}_{[2,3]}^{(0)T}\vec{x}_2\right) = 0 \neq y_2$，$\therefore\ \vec{w}_{[2,3]}^{(1)} = \vec{w}_{[2,3]}^{(0)} + y_2\vec{x}_2 = (1,1,1)^T$，

$sign\left(\vec{w}_{[2,3]}^{(1)T}\vec{x}_3\right) = 1 \neq y_3$，$\therefore\ \vec{w}_{[2,3]}^{(2)} = \vec{w}_{[2,3]}^{(1)} + y_3\vec{x}_3 = (0,2,0)^T$

$sign\left(\vec{w}_{[2,3]}^{(2)T}\vec{x}_2\right) = 1 = y_2$, 且 $sign\left(\vec{w}_{[2,3]}^{(2)T}\vec{x}_3\right) = -1 = y_3$

$\therefore\ \vec{w}_{[2,3]} = (0,2,0)^T$，分类面为：$x_1 = 0$

对测试样本进行增广，$\vec{\mathrm{x}} = (1,1,-2)^T$，分别代入上述三个分类面：

第一类和第二类：

$sign\left(\vec{w}_{[1,2]}^T\vec{\mathrm{x}}\right) = sign((1,-1,-1)(1,1,-2)^T = 1,\quad \therefore\ \vec{x} \in$ 第一类

第一类和第三类：

$sign\left(\vec{w}_{[1,3]}^T\vec{\mathrm{x}}\right) = sign((1,1,-1)(1,1,-2)^T = 1,\quad \therefore\ \vec{x} \in$ 第一类

第二类和第三类：

$sign\left(\vec{w}_{[2,3]}^T\vec{\mathrm{x}}\right) = sign((0,2,0)(1,1,-2)^T = 1,\quad \therefore\ \vec{x} \in$ 第二类

最终的投票结果是测试样本属于第一类。

2，现有四个样本，假设样本（3，0）和（3，6）属于第一类，样本（0，3）属于第二类，样本（-3，0）属于第三类，请用 Softmax 算法设计出这三个类别的分类器（假设这三个类别的初始权向量均为零向量，迭代步长取 1，需要写出计算过程）。

解：

（1）梯度的计算

假设输入样本$\vec{x}$ 属于K个类别$Y=\{1,2,\dots k,\dots K\}$中的某个类别k时，在Softmax中，我们按照式（1）计算其内积，按照式（2）计算其属于类别j的概率：

$$s_j=\vec{w}_j^T\vec{x} \qquad (1)$$

$$\hat{y}_j=\frac{e^{s_j}}{\sum_k e^{s_k}} \qquad (2)$$

经过Softmax函数后，得到的输出为K个类别的概率列向量：$\hat{Y}=(\hat{y}_1,\dots\hat{y}_j,\dots\hat{y}_K)^T$，假设理想的各个类别标签对应的概率为列向量：$Y=\{y_1,\dots y_j,\dots y_K\}$，且该列向量的一个元素为1，其他均为0，代表样本属于这个类别。我们选择用交叉熵作为误差函数其为表达式：

$$E_{in}(\vec{w}_k)=-\sum_{k=1}^{K}y_k ln\hat{y}_k=-ln\hat{y}_k \qquad (3)$$

我们可以计算E_{in}对于$\vec{w}_j(j=1,2,\dots,K)$的梯度：

$$\frac{\partial E_{in}}{\partial\vec{w}_j}=\frac{\partial E_{in}}{\partial\hat{y}_k}\frac{\partial\hat{y}_k}{\partial s_j}\frac{\partial s_j}{\partial\vec{w}_j}=-\frac{1}{\hat{y}_k}\frac{\partial\hat{y}_k}{\partial s_j}\vec{x} \qquad (4)$$

我们再来计算$\frac{\partial\hat{y}_k}{\partial s_j}$:

$$\frac{\partial\hat{y}_k}{\partial s_j}=\frac{\partial}{\partial s_j}\left(\frac{e^{s_k}}{\sum_k e^{s_k}}\right)=\frac{(e^{s_k})'\sum_k e^{s_k}-(\sum_k e^{s_k})'e^{s_k}}{(\sum_k e^{s_k})^2}=$$
$$\begin{cases}\frac{e^{s_j}\sum_k e^{s_k}-e^{s_j}e^{s_k}}{(\sum_k e^{s_k})^2}=\frac{e^{s_j}}{\sum_k e^{s_k}}-\frac{e^{s_j}}{\sum_k e^{s_k}}\frac{e^{s_k}}{\sum_k e^{s_k}}=\hat{y}_j(1-\hat{y}_k) & j=k\\ \frac{0\sum_k e^{s_k}-e^{s_j}e^{s_k}}{(\sum_k e^{s_k})^2}=0-\frac{e^{s_j}}{\sum_k e^{s_k}}\frac{e^{s_k}}{\sum_k e^{s_k}}=-\hat{y}_k\hat{y}_j & j\neq k\end{cases} \qquad (5)$$

将式（5）代入到式（4），我们得到E_{in}对于$\vec{w}_j$的梯度：

$$\frac{\partial E_{in}}{\partial\vec{w}_j}=\frac{\partial E_{in}}{\partial\hat{y}_k}\frac{\partial\hat{y}_k}{\partial s_j}\frac{\partial s_j}{\partial\vec{w}_j}=-\frac{1}{\hat{y}_k}\frac{\partial\hat{y}_k}{\partial s_j}\vec{x}=\begin{cases}(\hat{y}_j-1)\vec{x} & j=k\\ \hat{y}_j\vec{x} & j\neq k\end{cases} \qquad (6)$$

针对N个训练样本，将上述推导及求解过程写成矩阵或向量形式如下：

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，$\vec{x}_n = (x_{n0}, x_{n1}, \dots x_{nd})^T$，所有的训练样本我们用X来表示成一个N*(d+1)维的矩阵：

$$\boldsymbol{X} = \begin{pmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \\ \vdots \\ \vec{x}_N^T \end{pmatrix} = \begin{pmatrix} x_{10} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{N0} & \cdots & x_{Nd} \end{pmatrix} \qquad (7)$$

所有训练样本标签对应的概率输出用N*K维矩阵表示，其中K是类别数，样本只能属于其中一个类别且概率取1，其他类别概率为0，假设如下表示的第一个样本属于类别1，第N个样本属于类别K：

$$\boldsymbol{Y} = \begin{pmatrix} \vec{y}_1 \\ \vdots \\ \vec{y}_n \\ \vdots \\ \vec{y}_N \end{pmatrix} = \begin{pmatrix} y_{11} & \cdots & y_{1K} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{NK} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \qquad (8)$$

经过式(1)、式(2)后，我们得到的样本类别的概率估计值为N*K维矩阵$\widehat{\boldsymbol{Y}}$：

$$\widehat{\boldsymbol{Y}} = \begin{pmatrix} \vec{\hat{y}}_1 \\ \vdots \\ \vec{\hat{y}}_n \\ \vdots \\ \vec{\hat{y}}_N \end{pmatrix} = \begin{pmatrix} \hat{y}_{11} & \cdots & \hat{y}_{1K} \\ \vdots & \ddots & \vdots \\ \hat{y}_{N1} & \cdots & \hat{y}_{NK} \end{pmatrix} \qquad (9)$$

根据式（6）得到E_{in}的梯度可以写为：

$$\nabla E_{in} = (\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^T \boldsymbol{X} = \left(\vec{\hat{y}}_1 - \vec{y}_1, \dots \vec{\hat{y}}_n - \vec{y}_n, \dots \vec{\hat{y}}_N - \vec{y}_N\right) \begin{pmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \\ \vdots \\ \vec{x}_N^T \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^N (\hat{y}_{n1} - y_{n1}) \vec{x}_n^T \\ \vdots \\ \sum_{n=1}^N (\hat{y}_{nj} - y_{nj}) \vec{x}_n^T \\ \vdots \\ \sum_{n=1}^N (\hat{y}_{nK} - y_{nK}) \vec{x}_n^T \end{pmatrix} \qquad (10)$$

这相当于K*N维的矩阵与N*(d+1)维的矩阵做内积，得到K*(d+1)维的梯度，这里y_{nj}只会取0或者1。

假设类别对应的权系数向量用$\vec{w}$表示，加上常数项，它也是(d+1)维，一共K个类别，可以写成K*(d+1)维矩阵形式：

$$\boldsymbol{W}=\begin{pmatrix}\vec{w}_1\\ \vdots\\ \vec{w}_j\\ \vdots\\ \vec{w}_K\end{pmatrix}=\begin{pmatrix}w_{10} & \cdots & w_{1d}\\ \vdots & \ddots & \vdots\\ w_{K0} & \cdots & w_{Kd}\end{pmatrix} \qquad (11)$$

假设学习率为η，迭代次数用上标t表示，利用梯度下降法得到权重的更新式：

$$\boldsymbol{W}^{(\mathrm{t+1})}=\boldsymbol{W}^{(\mathrm{t})}-\eta\nabla E_{in}=\begin{pmatrix}\vec{w}_1{}^{(t)}-\eta\sum_{n=1}^{N}\left(\hat{y}_{n1}-y_{n1}\right)\vec{x}_n^T\\ \vdots\\ \vec{w}_j{}^{(t)}-\eta\sum_{n=1}^{N}\left(\hat{y}_{nj}-y_{nj}\right)\vec{x}_n^T\\ \vdots\\ \vec{w}_K{}^{(t)}-\eta\sum_{n=1}^{N}\left(\hat{y}_{nK}-y_{nK}\right)\vec{x}_n^T\end{pmatrix} \qquad (12)$$

根据更新后的权重，我们可以重新计算每个样本在每个类别权系数向量下的内积S，同样，我们也可以把S写成矩阵形式，它是N*K维矩阵：

$$\boldsymbol{S}=\boldsymbol{X}(\boldsymbol{W}^{(\mathrm{t+1})})^{\boldsymbol{T}}=\begin{pmatrix}\vec{x}_1^T\\ \vdots\\ \vec{x}_n^T\\ \vdots\\ \vec{x}_N^T\end{pmatrix}\left(\vec{w}_1{}^{(t+1)},\ldots,\vec{w}_j{}^{(t+1)},\ldots\vec{w}_K{}^{(t+1)}\right)=$$

$$\begin{pmatrix}(\vec{w}_1{}^{(t+1)})^T\vec{x}_1 & \cdots & (\vec{w}_K{}^{(t+1)})^T\vec{x}_1\\ \vdots & \ddots & \vdots\\ (\vec{w}_1{}^{(t+1)})^T\vec{x}_N & \cdots & (\vec{w}_K{}^{(t+1)})^T\vec{x}_N\end{pmatrix}=\begin{pmatrix}s_{11} & \cdots & s_{1j} & \cdots & s_{1K}\\ \vdots & \cdots & \cdots & \cdots & \vdots\\ s_{n1} & \cdots & s_{nj} & \cdots & s_{nK}\\ \vdots & \cdots & \cdots & \cdots & \vdots\\ s_{N1} & \cdots & s_{Nj} & \cdots & s_{NK}\end{pmatrix} \qquad (13)$$

利用Softmax可以得到：

$$\widehat{\boldsymbol{Y}}=\begin{pmatrix}\vec{\hat{y}}_1\\ \vdots\\ \vec{\hat{y}}_n\\ \vdots\\ \vec{\hat{y}}_N\end{pmatrix}=\begin{pmatrix}\hat{y}_{11} & \cdots & \hat{y}_{1K}\\ \vdots & \ddots & \vdots\\ \hat{y}_{N1} & \cdots & \hat{y}_{NK}\end{pmatrix} \qquad (14)$$

因为对于一个样本的误差函数为式（3），所以，对于所有样本其误差函数（损失函数）为：

$$E_{in}=\frac{1}{N}\sum_{n=1}^{N}(-ln\hat{y}_{nk}) \qquad (15)$$

（2）习题的求解

首先，将样本变为增广向量：$\vec{x}_1=(1,3,0)^T,\vec{x}_2=(1,3,6)^T,\vec{x}_3=(1,0,3)^T,\vec{x}_4=(1,-3,0)^T$，得到：

$$\boldsymbol{X}=\begin{pmatrix}1&3&0\\1&3&6\\1&0&3\\1&-3&0\end{pmatrix}$$

四个样本对应的理想概率值为$\vec{Y}_1=(1,0,0)^T$，$\vec{Y}_2=(1,0,0)^T$，$\vec{Y}_3=(0,1,0)^T$，$\vec{Y}_4=(0,0,1)^T$，即：

$$\boldsymbol{Y}=\begin{pmatrix}1&0&0\\1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

假设三个类别的初始权向量为：$\vec{w}_1^{(0)}=(0,0,0)^T$，$\vec{w}_2^{(0)}=(0,0,0)^T$，$\vec{w}_3^{(0)}=(0,0,0)^T$，即：

$$\boldsymbol{W}^{(0)}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}$$

令$\eta=1$。

第一次迭代：t=0，将$\vec{x}_n$，$(n=1,2,3,4)$，$\vec{w}_k^{(0)},(k=1,2,3)$代入到式（13），得到：

$$\boldsymbol{S}=\boldsymbol{X}(\boldsymbol{W}^{(0)})^{\boldsymbol{T}}=\begin{pmatrix}1&3&0\\1&3&6\\1&0&3\\1&-3&0\end{pmatrix}\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\\0&0&0\end{pmatrix}$$

利用式(2)和式(14)得到：

$$\widehat{\boldsymbol{Y}}=\begin{pmatrix}\vec{\hat{y}}_1\\\vec{\hat{y}}_2\\\vec{\hat{y}}_3\\\vec{\hat{y}}_4\end{pmatrix}=\begin{pmatrix}\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\\\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\\\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\\\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\end{pmatrix}$$

显然所有样本都没有正确分类，按照式(15)，每一个样本任意选择一个类别获得

其概率，计算$E_{in}=\frac{1}{4}\sum_{n=1}^{4}(-ln\frac{1}{3})=1.099$

所以，我们按照式（10）求得梯度：

$$\nabla E_{in}=(\widehat{\boldsymbol{Y}}-\boldsymbol{Y})^T\boldsymbol{X}=\begin{pmatrix}\frac{1}{3}-1 & \frac{1}{3}-1 & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}-1 & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}-1\end{pmatrix}\begin{pmatrix}1 & 3 & 0\\ 1 & 3 & 6\\ 1 & 0 & 3\\ 1 & -3 & 0\end{pmatrix}=\begin{pmatrix}-\frac{2}{3} & -5 & -3\\ \frac{1}{3} & 1 & 0\\ \frac{1}{3} & 4 & 3\end{pmatrix}$$

用梯度下降法式(12)进行权系数向量更新：

$$\boldsymbol{W}^{(1)}=\boldsymbol{W}^{(0)}-\eta\nabla E_{in}=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}-\begin{pmatrix}-\frac{2}{3} & -5 & -3\\ \frac{1}{3} & 1 & 0\\ \frac{1}{3} & 4 & 3\end{pmatrix}=\begin{pmatrix}\frac{2}{3} & 5 & 3\\ -\frac{1}{3} & -1 & 0\\ -\frac{1}{3} & -4 & -3\end{pmatrix}$$

根据式(13)得到S矩阵：

$$\boldsymbol{S}=\boldsymbol{X}(\boldsymbol{W}^{(1)})^{\boldsymbol{T}}=\begin{pmatrix}1 & 3 & 0\\ 1 & 3 & 6\\ 1 & 0 & 3\\ 1 & -3 & 0\end{pmatrix}\begin{pmatrix}\frac{2}{3} & -\frac{1}{3} & -\frac{1}{3}\\ 5 & -1 & -4\\ 3 & 0 & -3\end{pmatrix}=\begin{pmatrix}15.67 & -3.33 & -12.33\\ 33.67 & -3.33 & -30.33\\ 9.67 & -0.33 & -9.33\\ -14.33 & 2.67 & 11.67\end{pmatrix}$$

利用Softmax得到：

$$\widehat{\boldsymbol{Y}}=\begin{pmatrix}\vec{\hat{y}}_1\\ \vec{\hat{y}}_2\\ \vec{\hat{y}}_3\\ \vec{\hat{y}}_4\end{pmatrix}=\begin{pmatrix}1.00 & 0.00 & 0.00\\ 1.00 & 0.00 & 0.00\\ 1.00 & 0.00 & 0.00\\ 0.00 & 0.00 & 1.00\end{pmatrix}$$

第三个样本错分，计算$E_{in}=(-ln1-ln1-ln0-ln1)/4=\infty$

第二次迭代：

我们按照式（10）求得梯度：

$$\nabla E_{in}=(\widehat{\boldsymbol{Y}}-\boldsymbol{Y})^T\boldsymbol{X}=\begin{pmatrix}1-1 & 1-1 & 1 & 0\\ 0 & 0 & 0-1 & 0\\ 0 & 0 & 0 & 1-1\end{pmatrix}\begin{pmatrix}1 & 3 & 0\\ 1 & 3 & 6\\ 1 & 0 & 3\\ 1 & -3 & 0\end{pmatrix}=\begin{pmatrix}1 & 0 & 3\\ -1 & 0 & -3\\ 0 & 0 & 0\end{pmatrix}$$

用梯度下降法式(12)进行权系数向量更新：

$$\boldsymbol{W}^{(2)} = \boldsymbol{W}^{(1)} - \eta\nabla E_{in} = \begin{pmatrix} \frac{2}{3} & 5 & 3 \\ -\frac{1}{3} & -1 & 0 \\ -\frac{1}{3} & -4 & -3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 3 \\ -1 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} -0.33 & 5 & 0 \\ 0.67 & -1 & 3 \\ -0.33 & -4 & -3 \end{pmatrix}$$

根据式(13)得到S矩阵:

$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{W}^{(2)})^{\boldsymbol{T}} =$

$$\begin{pmatrix} 1 & 3 & 0 \\ 1 & 3 & 6 \\ 1 & 0 & 3 \\ 1 & -3 & 0 \end{pmatrix} \begin{pmatrix} -0.33 & 0.67 & -0.33 \\ 5 & -1 & -4 \\ 0 & 3 & -3 \end{pmatrix} = \begin{pmatrix} 14.67 & -2.33 & -12.33 \\ 14.67 & 15.67 & -30.33 \\ -0.33 & 9.67 & -9.33 \\ -15.33 & 3.67 & 11.27 \end{pmatrix}$$

利用Softmax得到:

$$\widehat{\boldsymbol{Y}} = \begin{pmatrix} \vec{\hat{y}}_1 \\ \vec{\hat{y}}_2 \\ \vec{\hat{y}}_3 \\ \vec{\hat{y}}_4 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 0.27 & 0.73 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{pmatrix}$$

第二个样本错分，计算$E_{in} = (-ln1 - ln0.27 - ln1 - ln1)/4 = 0.33$

第三次迭代:

我们按照式（10）求得梯度:

$$\nabla E_{in} = (\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^T \boldsymbol{X} = \begin{pmatrix} 1-1 & 0.27-1 & 0 & 0 \\ 0 & 0.73 & 1-1 & 0 \\ 0 & 0 & 0 & 1-1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 0 \\ 1 & 3 & 6 \\ 1 & 0 & 3 \\ 1 & -3 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} -0.73 & -2.19 & -4.38 \\ 0.73 & 2.19 & 4.38 \\ 0 & 0 & 0 \end{pmatrix}$$

用梯度下降法式(12)进行权系数向量更新:

$$\boldsymbol{W}^{(3)} = \boldsymbol{W}^{(2)} - \eta\nabla E_{in} = \begin{pmatrix} -0.33 & 5 & 0 \\ 0.67 & -1 & 3 \\ -0.33 & -4 & -3 \end{pmatrix} - \begin{pmatrix} -0.73 & -2.19 & -4.38 \\ 0.73 & 2.19 & 4.38 \\ 0 & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 0.40 & 7.19 & 4.38 \\ -0.06 & -3.19 & -1.38 \\ -0.33 & -4 & -3 \end{pmatrix}$$

根据式(13)得到S矩阵:

$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{W}^{(3)})^{\boldsymbol{T}} =$

$$\begin{pmatrix} 1 & 3 & 0 \\ 1 & 3 & 6 \\ 1 & 0 & 3 \\ 1 & -3 & 0 \end{pmatrix}\begin{pmatrix} 0.40 & -0.06 & -0.33 \\ 7.19 & -3.19 & -4 \\ 4.38 & -1.38 & -3 \end{pmatrix} = \begin{pmatrix} 21.97 & -9.63 & -12.33 \\ 48.25 & -17.91 & -30.33 \\ 13.54 & -4.20 & -9.33 \\ -21.17 & 9.51 & 11.67 \end{pmatrix}$$

利用Softmax得到:

$$\widehat{\boldsymbol{Y}} = \begin{pmatrix} \vec{\hat{y}}_1 \\ \vec{\hat{y}}_2 \\ \vec{\hat{y}}_3 \\ \vec{\hat{y}}_4 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 1.00 & 0.00 & 0.00 \\ 1.00 & 0.00 & 0.00 \\ 0.00 & 0.11 & 0.89 \end{pmatrix}$$

第三个样本错分，计算$E_{in} = (-ln1 - ln1 - ln0 - ln0.89)/4 = \infty$

第四次迭代:

我们按照式（10）求得梯度:

$$\nabla E_{in} = (\widehat{\boldsymbol{Y}} - \boldsymbol{Y})^T\boldsymbol{X} = \begin{pmatrix} 1-1 & 1-1 & 1 & 0 \\ 0 & 0 & 0-1 & 0.11 \\ 0 & 0 & 0 & 0.89-1 \end{pmatrix}\begin{pmatrix} 1 & 3 & 0 \\ 1 & 3 & 6 \\ 1 & 0 & 3 \\ 1 & -3 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 3 \\ -0.89 & -0.33 & -3 \\ -0.11 & 0.33 & 0 \end{pmatrix}$$

用梯度下降法式(12)进行权系数向量更新:

$$\boldsymbol{W}^{(4)} = \boldsymbol{W}^{(3)} - \eta\nabla E_{in} = \begin{pmatrix} 0.40 & 7.19 & 4.38 \\ -0.06 & -3.19 & -1.38 \\ -0.33 & -4 & -3 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 3 \\ -0.89 & -0.33 & -3 \\ -0.11 & 0.33 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} -0.60 & 7.19 & 1.38 \\ 0.83 & -2.86 & 1.62 \\ -0.22 & -4.33 & -3 \end{pmatrix}$$

根据式(13)得到S矩阵:

$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{W}^{(4)})^{\boldsymbol{T}} =$

$$\begin{pmatrix} 1 & 3 & 0 \\ 1 & 3 & 6 \\ 1 & 0 & 3 \\ 1 & -3 & 0 \end{pmatrix}\begin{pmatrix} -0.60 & 0.83 & -0.22 \\ 7.19 & -2.86 & -4.33 \\ 1.38 & 1.62 & -3 \end{pmatrix}=\begin{pmatrix} 20.97 & -7.75 & -13.21 \\ 29.25 & 1.97 & -31.21 \\ 3.54 & 5.69 & -9.22 \\ -22.17 & 9.41 & 12.77 \end{pmatrix}$$

利用Softmax得到：

$$\widehat{\boldsymbol{Y}} = \begin{pmatrix} \vec{\hat{y}}_1 \\ \vec{\hat{y}}_2 \\ \vec{\hat{y}}_3 \\ \vec{\hat{y}}_4 \end{pmatrix} = \begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 1.00 & 0.00 & 0.00 \\ 0.10 & 0.90 & 0.00 \\ 0.00 & 0.02 & 0.98 \end{pmatrix}$$

所有样本均正确分类，计算$E_{in} = (-ln1 - ln1 - ln0.90 - ln0.98)/4 = 0.03$

此时求得的权系数向量矩阵为：

$$\boldsymbol{W}^{(4)} = \begin{pmatrix} -0.60 & 7.19 & 1.38 \\ 0.83 & -2.86 & 1.62 \\ -0.22 & -4.33 & -3 \end{pmatrix}$$

不习惯看矩阵的，可以看如下求解过程：

第一次迭代：将$\vec{x}_n$，$(n = 1,2,3,4)$，$\vec{w}_k^{(0)}, (k = 1,2,3)$代入到式（1），对每一个样本均得到$s_1 = s_2 = s_3 = 0$，代入式（2）得到：$\hat{Y} = (\hat{y}_1, \hat{y}_2, \hat{y}_3)^T = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^T$，显然这个样本没有正确分类，所以，我们按照式（6）求得梯度去计算新的$\vec{w}_k$，我们以计算$\vec{w}_1$为例，先用式（6）计算梯度：

$$\frac{\partial E_{in}}{\partial \vec{w}_1} = \sum_{n=1}^{4} \frac{\partial E_{in}(\vec{x}_n)}{\partial \vec{w}_1} = (\hat{y}_1 - 1)\vec{x}_1 + (\hat{y}_1 - 1)\vec{x}_2 + \hat{y}_2\vec{x}_3 + \hat{y}_3\vec{x}_4$$

$$= \left(\frac{1}{3} - 1\right)\vec{x}_1 + \left(\frac{1}{3} - 1\right)\vec{x}_2 + \frac{1}{3}\vec{x}_3 + \frac{1}{3}\vec{x}_4 = (-\frac{2}{3}, -5, -3)^T$$

同理，我们可以得到：$\frac{\partial E_{in}}{\partial \vec{w}_2} = (\frac{1}{3}, 1, 0)^T$，$\frac{\partial E_{in}}{\partial \vec{w}_3} = (\frac{1}{3}, 4, 3)^T$

用梯度下降法对$\vec{w}_k$进行更新：

$\vec{w}_1^{(1)} = \vec{w}_1^{(0)} - \dfrac{\partial E_{in}}{\partial \vec{w}_1} = (0,0,0)^T - \left(-\dfrac{2}{3}, -5, -3\right)^T = (\dfrac{2}{3}, 5,3)^T$

$\vec{w}_2^{(1)} = \vec{w}_2^{(0)} - \dfrac{\partial E_{in}}{\partial \vec{w}_2} = (0,0,0)^T - \left(\dfrac{1}{3}, 1,0\right)^T = (-\dfrac{1}{3}, -1,0)^T$

$\vec{w}_3^{(1)} = \vec{w}_3^{(0)} - \dfrac{\partial E_{in}}{\partial \vec{w}_3} = (0,0,0)^T - \left(\dfrac{1}{3}, 4,3\right)^T = (-\dfrac{1}{3}, -4, -3)^T$

根据$\vec{w}_1^{(1)}$，$\vec{w}_2^{(1)}$和$\vec{w}_3^{(1)}$，我们用式（1）得到：

对于$\vec{x}_1$，我们有：$s_1 = \vec{w}_1^T \vec{x}_1 = \left(\frac{2}{3}, 5,3\right)\begin{pmatrix}1\\3\\0\end{pmatrix} = 15.67$，$s_2 = \vec{w}_2^T \vec{x}_1 = \left(-\frac{1}{3}, -1,0\right)\begin{pmatrix}1\\3\\0\end{pmatrix} = -3.33$，$s_3 = \vec{w}_3^T \vec{x}_1 = \left(-\frac{1}{3}, -4, -3\right)\begin{pmatrix}1\\3\\0\end{pmatrix} = -12.33$

利用式（2），我们可以得到：$\hat{y}_1 = \frac{e^{s_1}}{e^{s_1}+e^{s_2}+e^{s_3}} = 1.00$，$\hat{y}_2 = \frac{e^{s_2}}{e^{s_1}+e^{s_2}+e^{s_3}} = 0.00$，$\hat{y}_3 = \frac{e^{s_3}}{e^{s_1}+e^{s_2}+e^{s_3}} = 0.00$，即，$\vec{\hat{Y}}_1 = (1.00,0.00,0.00)^T$，对照$\vec{Y}_1 = (1,0,0)^T$，此时对于样本$\vec{x}_1$分类是正确的。

同理：对于$\vec{x}_2$，我们有$s_1 = 33.67$，$s_2 = -3.33$，$s_3 = -30.33$，对应的我们可以计算出$\vec{\hat{Y}}_2 = (1.00,0.00,0.00)^T$，对照$\vec{Y}_2 = (1,0,0)^T$，此时对于样本$\vec{x}_2$分类是正确的。

对于$\vec{x}_3$，我们有$s_1 = 9.67$，$s_2 = -0.33$，$s_3 = -9.33$，对应的我们可以计算出$\vec{\hat{Y}}_3 = (1.00,0.00,0.00)^T$，对照$\vec{Y}_3 = (0,1,0)^T$，此时对于样本$\vec{x}_3$分类是错误的。

对于$\vec{x}_4$，我们有$s_1 = -14.33$，$s_2 = 2.67$，$s_3 = 11.67$，对应的我们可以计算出$\vec{\hat{Y}}_4 = (0.00,0.00,1.00)^T$，对照$\vec{Y}_4 = (0,0,1)^T$，此时对于样本$\vec{x}_4$分类是正确的。

第三个样本错分，计算$E_{in} = (-ln1 - ln1 - ln0 - ln1)/4 = \infty$

第二次迭代：我们需要按照式（6）重新计算梯度去得到新的$\vec{w}_k$，仍以计算$\vec{w}_1$为例，先用式（6）计算梯度：

$$\frac{\partial E_{in}}{\partial \vec{w}_1}=\sum_{n=1}^{4}\frac{\partial E_{in}(\vec{x}_n)}{\partial \vec{w}_1}=(\hat{y}_1-1)\vec{x}_1+(\hat{y}_1-1)\vec{x}_2+\hat{y}_2\vec{x}_3+\hat{y}_3\vec{x}_4$$

$$=(1-1)\vec{x}_1+(1-1)\vec{x}_2+1\vec{x}_3+0\vec{x}_4=(1,0,3)^T$$

同理，我们可以得到：$\frac{\partial E_{in}}{\partial \vec{w}_2}=0\vec{x}_1+0\vec{x}_2+(0-1)\vec{x}_3+0\vec{x}_4=(-1,0,-3)^T$，$\frac{\partial E_{in}}{\partial \vec{w}_3}=0\vec{x}_1+0\vec{x}_2+0\vec{x}_3+(1-1)\vec{x}_4=(0,0,0)^T$

用梯度下降法对$\vec{w}_k$进行更新：

$$\vec{w}_1^{(2)}=\vec{w}_1^{(1)}-\frac{\partial E_{in}}{\partial \vec{w}_1}=(0.67,5,3)^T-(1,0,3)^T=(-0.33,5,0)^T$$

$$\vec{w}_2^{(2)}=\vec{w}_2^{(1)}-\frac{\partial E_{in}}{\partial \vec{w}_2}=(-0.33,-1,0)^T-(-1,0,-3)^T=(0.67,-1,3)^T$$

$$\vec{w}_3^{(2)}=\vec{w}_3^{(1)}-\frac{\partial E_{in}}{\partial \vec{w}_3}=(-0.33,-4,-3)^T-(0,0,0)^T=(-0.33,-4,-3)^T$$

根据$\vec{w}_1^{(2)}$，$\vec{w}_2^{(2)}$和$\vec{w}_3^{(2)}$，我们用式（1）得到：

对于$\vec{x}_1$，我们有：$s_1=\vec{w}_1^T\vec{x}_1=(-0.33,5,0)\begin{pmatrix}1\\3\\0\end{pmatrix}=14.67$，$s_2=\vec{w}_2^T\vec{x}_1=(0.67,-1,3)\begin{pmatrix}1\\3\\0\end{pmatrix}=-2.33$，$s_3=\vec{w}_3^T\vec{x}_1=(-0.33,-4,-3)\begin{pmatrix}1\\3\\0\end{pmatrix}=-12.33$

利用式（2），我们可以得到：$\hat{y}_1=\frac{e^{s_1}}{e^{s_1}+e^{s_2}+e^{s_3}}=1.00$，$\hat{y}_2=\frac{e^{s_2}}{e^{s_1}+e^{s_2}+e^{s_3}}=0.00$，$\hat{y}_3=\frac{e^{s_3}}{e^{s_1}+e^{s_2}+e^{s_3}}=0.00$，即，$\vec{\hat{Y}}_1=(1.00,0.00,0.00)^T$，对照$\vec{Y}_1=(1,0,0)^T$，此时对于样本$\vec{x}_1$分类是正确的。

同理：对于$\vec{x}_2$，我们有$s_1=14.67$，$s_2=15.67$，$s_3=-30.33$，对应的我们可以计算出$\vec{\hat{Y}}_2=(0.27,0.73,0.00)^T$，对照$\vec{Y}_2=(1,0,0)^T$，此时对于样本$\vec{x}_2$分类是错误的。

对于$\vec{x}_3$，我们有$s_1=-0.33$，$s_2=9.67$，$s_3=-9.33$，对应的我们可以计算出$\vec{\hat{Y}}_3=(0.00,1.00,0.00)^T$，对照$\vec{Y}_3=(0,1,0)^T$，此时对于样本$\vec{x}_3$分类是正确的。

对于$\vec{x}_4$，我们有$s_1 = -15.33$，$s_2 = 3.67$，$s_3 = 11.27$，对应的我们可以计算出$\vec{Y}_4 = (0.00,0.00,1.00)^T$，对照$\vec{Y}_4 = (0,0,1)^T$，此时对于样本$\vec{x}_4$分类是正确的。

第二个样本错分，计算$E_{in} = (-ln1 - ln0.27 - ln1 - ln1)/4 = 0.33$

第三次迭代：我们需要按照式（6）重新计算梯度去得到新的$\vec{w}_k$，仍以计算$\vec{w}_1$为例，先用式（6）计算梯度：

$$\frac{\partial E_{in}}{\partial \vec{w}_1} = \sum_{n=1}^{4} \frac{\partial E_{in}(\vec{x}_n)}{\partial \vec{w}_1} = (\hat{y}_1 - 1)\vec{x}_1 + (\hat{y}_1 - 1)\vec{x}_2 + \hat{y}_2\vec{x}_3 + \hat{y}_3\vec{x}_4$$

$$= (1-1)\vec{x}_1 + (0.27-1)\vec{x}_2 + 0\vec{x}_3 + 0\vec{x}_4$$

$$= (-0.73, -2.19, -4.38)^T$$

同 理 ， 我 们 可 以 得 到 ： $\frac{\partial E_{in}}{\partial \vec{w}_2} = 0\vec{x}_1 + 0.73\vec{x}_2 + (1-1)\vec{x}_3 + 0\vec{x}_4 = (0.73,2.19,4.38)^T$，$\frac{\partial E_{in}}{\partial \vec{w}_3} = 0\vec{x}_1 + 0\vec{x}_2 + 0\vec{x}_3 + (1-1)\vec{x}_4 = (0,0,0)^T$

用梯度下降法对$\vec{w}_k$进行更新：

$$\vec{w}_1^{(3)} = \vec{w}_1^{(2)} - \frac{\partial E_{in}}{\partial \vec{w}_1} = (-0.33,5,0)^T - (-0.73,-2.19,-4.38)^T$$

$$= (0.40, 7.19, 4.38)^T$$

$$\vec{w}_2^{(3)} = \vec{w}_2^{(2)} - \frac{\partial E_{in}}{\partial \vec{w}_2} = (0.67,-1,3)^T - (0.73,2.19,4.38)^T$$

$$= (-0.06, -3.19, -1.38)^T$$

$$\vec{w}_3^{(3)} = \vec{w}_3^{(2)} - \frac{\partial E_{in}}{\partial \vec{w}_3} = (-0.33,-4,-3)^T - (0,0,0)^T = (-0.33,-4,-3)^T$$

根据$\vec{w}_1^{(3)}$，$\vec{w}_2^{(3)}$和$\vec{w}_3^{(3)}$，我们用式（1）得到：

对于$\vec{x}_1$，我们有：$s_1 = \vec{w}_1^T\vec{x}_1 = (0.40, 7.19, 4.38)\begin{pmatrix}1\\3\\0\end{pmatrix} = 21.97$，$s_2 = \vec{w}_2^T\vec{x}_1 =$

$(-0.06,-3.19,-1.38)\begin{pmatrix}1\\3\\0\end{pmatrix}=-9.63$ ， $s_3=\vec{w}_3^T\vec{x}_1=(-0.33,-4,-3)\begin{pmatrix}1\\3\\0\end{pmatrix}=-12.33$

利用式（2），我们可以得到：$\hat{y}_1=\frac{e^{s_1}}{e^{s_1}+e^{s_2}+e^{s_3}}=1.0000$，$\hat{y}_2=\frac{e^{s_2}}{e^{s_1}+e^{s_2}+e^{s_3}}=0.0000$，$\hat{y}_3=\frac{e^{s_3}}{e^{s_1}+e^{s_2}+e^{s_3}}=0.0000$，即，$\vec{Y}_1=(1.00,0.00,0.00)^T$，对照$\vec{Y}_1=(1,0,0)^T$，此时对于样本$\vec{x}_1$分类是正确的。

同理：对于$\vec{x}_2$，我们有$s_1=48.25$，$s_2=-17.91$，$s_3=-30.33$，对应的我们可以计算出$\vec{Y}_2=(1.00,0.00,0.00)^T$，对照$\vec{Y}_2=(1,0,0)^T$，此时对于样本$\vec{x}_2$分类是正确的。

对于$\vec{x}_3$，我们有$s_1=13.54$，$s_2=-4.20$，$s_3=-9.33$，对应的我们可以计算出$\vec{Y}_3=(1.00,0.00,0.00)^T$，对照$\vec{Y}_3=(0,1,0)^T$，此时对于样本$\vec{x}_3$分类是错误的。

对于$\vec{x}_4$，我们有$s_1=-21.17$，$s_2=9.51$，$s_3=11.67$，对应的我们可以计算出$\vec{Y}_4=(0.0000,0.11,0.89)^T$，对照$\vec{Y}_4=(0,0,1)^T$，此时对于样本$\vec{x}_4$分类是正确的。

第三个样本错分，计算$E_{in}=(-ln1-ln1-ln0-ln0.89)/4=\infty$

第四次迭代：我们需要按照式（6）重新计算梯度去得到新的$\vec{w}_k$，仍以计算$\vec{w}_1$为例，先用式（6）计算梯度：

$$\frac{\partial E_{in}}{\partial\vec{w}_1}=\sum_{n=1}^{4}\frac{\partial E_{in}(\vec{x}_n)}{\partial\vec{w}_1}=(\hat{y}_1-1)\vec{x}_1+(\hat{y}_1-1)\vec{x}_2+\hat{y}_2\vec{x}_3+\hat{y}_3\vec{x}_4$$

$$=(1-1)\vec{x}_1+(1-1)\vec{x}_2+1\vec{x}_3+0\vec{x}_4=(1,0,3)^T$$

同理，我们可以得到：

$\frac{\partial E_{in}}{\partial \vec{w}_2} = 0\vec{x}_1 + 0\vec{x}_2 + (0-1)\vec{x}_3 + 0.11\vec{x}_4 = (-0.89, -0.33, -3)^T$,

$$\frac{\partial E_{in}}{\partial \vec{w}_3} = 0\vec{x}_1 + 0\vec{x}_2 + 0\vec{x}_3 + (0.89-1)\vec{x}_4 = (-0.11, 0.33, 0)^T$$

用梯度下降法对$\vec{w}_k$进行更新:

$$\vec{w}_1^{(4)} = \vec{w}_1^{(3)} - \frac{\partial E_{in}}{\partial \vec{w}_1} = (0.40, 7.19, 4.38)^T - (1,0,3)^T = (-0.60, 7.19, 1.38)^T$$

$$\vec{w}_2^{(4)} = \vec{w}_2^{(3)} - \frac{\partial E_{in}}{\partial \vec{w}_2} = (-0.06, -3.19, -1.38)^T - (-0.89, -0.33, -3)^T$$
$$= (0.83, -2.86, 1.62)^T$$

$$\vec{w}_3^{(4)} = \vec{w}_3^{(3)} - \frac{\partial E_{in}}{\partial \vec{w}_3} = (-0.33, -4, -3)^T - (-0.11, 0.33, 0)^T$$
$$= (-0.22, -4.33, -3)^T$$

根据$\vec{w}_1^{(4)}$，$\vec{w}_2^{(4)}$和$\vec{w}_3^{(4)}$，我们用式（1）得到:

对于$\vec{x}_1$，我们有: $s_1 = \vec{w}_1^T\vec{x}_1 = (-0.60, 7.19, 1.38)\begin{pmatrix}1\\3\\0\end{pmatrix} = 20.97$，$s_2 = \vec{w}_2^T\vec{x}_1 = (0.83, -2.86, 1.62)\begin{pmatrix}1\\3\\0\end{pmatrix} = -7.75$，$s_3 = \vec{w}_3^T\vec{x}_1 = (-0.18, -4.45, -3)\begin{pmatrix}1\\3\\0\end{pmatrix} = -13.21$

利用式（2），我们可以得到: $\hat{y}_1 = \frac{e^{s_1}}{e^{s_1}+e^{s_2}+e^{s_3}} = 1.00$，$\hat{y}_2 = \frac{e^{s_2}}{e^{s_1}+e^{s_2}+e^{s_3}} = 0.00$，$\hat{y}_3 = \frac{e^{s_3}}{e^{s_1}+e^{s_2}+e^{s_3}} = 0.00$，即，$\vec{\hat{Y}}_1 = (1.00,0.00,0.00)^T$，对照$\vec{Y}_1 = (1,0,0)^T$，此时对于样本$\vec{x}_1$分类是正确的。

同理: 对于$\vec{x}_2$，我们有$s_1 = 29.25$，$s_2 = 1.97$，$s_3 = -31.21$，对应的我们可以计算出$\vec{\hat{Y}}_2 = (1.00,0.00,0.00)^T$，对照$\vec{Y}_2 = (1,0,0)^T$，此时对于样本$\vec{x}_2$分类是正确的。

对于$\vec{x}_3$，我们有$s_1 = 3.54$，$s_2 = 5.69$，$s_3 = -9.22$，对应的我们可以计算出$\vec{\hat{Y}}_3 = (0.10,0.90,0.00)^T$，对照$\vec{Y}_3 = (0,1,0)^T$，此时对于样本$\vec{x}_3$分类是正确的。

对于$\vec{x}_4$，我们有$s_1 = -22.17$，$s_2 = 9.41$，$s_3 = 12.77$，对应的我们可以计算出$\vec{Y}_4 = (0.00,0.02,0.98)^T$，对照$\vec{Y}_4 = (0,0,1)^T$，此时对于样本$\vec{x}_4$分类是正确的。

计算$E_{in} = (-ln1 - ln1 - ln0.90 - ln0.98)/4 = 0.03$

于是我们最终得到的是：

$\vec{w}_1 = (-0.60, 7.19, 1.38)^T$

$\vec{w}_2 = (0.83, -2.86, 1.62)^T$

$\vec{w}_3 = (-0.22, -4.33, -3)^T$