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

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

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：