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

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：

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

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：

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

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

利用Softmax可以得到：

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

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：

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

（2）习题的求解