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

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：$(x_j^{(L)})'=\left[\!\left[s_j^{(L-1)}\geq 0\right]\!\right]$