由于对 L 层的每个输出 $x_j^{(L)}$(也是第 $L+1$ 层的输入), loss 对其的导数可以用下式计算:

$$\frac{\partial loss}{\partial x_j^{(L)}} = \sum_{k=1}^{d_{L+1}} \delta_k^{(L+1)} w_{j,k}^{(L+1)}$$

因此, 在已知第 $L+1$ 层的 delta 值 $\delta_k^{(L)}(k=1,\ldots,d_{L+1}$, 其中 d_{L+1} 表示第 $L+1$ 层中节点的个数) 时, 根据链式法则, 我们有:

$$\begin{aligned}
\delta_j^{(L)} = \frac{\partial loss}{\partial z_j^{(L)}} &= \frac{\partial loss}{\partial x_j^{(L)}} \frac{\partial x_j^{(L)}}{\partial z_j^{(L)}} \\
&= \left(\sum_{k=1}^{d_{L+1}} \delta_k^{(L+1)} w_{j,k}^{(L+1)} \right) \cdot f'(z_j^{(L)})
\end{aligned}$$

这样也就把前一层 (即第 L 层) 的 delta 值 $\delta_j^{(L)}(j=1,\ldots,d_L$, 其中 d_L 表示第 L 层的节点个数) 给求出来了.

知晓每一层的 $\delta_j^{(L)}$ 之后, 为了求出 $loss$ 对权重的导数, 同样采用链式法则:

$$\begin{aligned}
\frac{\partial loss}{\partial w_{j,k}^{(L+1)}} &= \sum_{k=1}^{d_{L+1}} \frac{\partial loss}{\partial z_k^{(L+1)}} \frac{\partial z_k^{(L+1)}}{\partial w_{j,k}^{(L+1)}} \\
&= \sum_{k=1}^{d_{L+1}} \delta_k^{(L+1)} x_j^{(L)}
\end{aligned}$$