

l layer (j)

l+1 layer (i)

l+2 layer (k)

：第l+1层第i个单元的计算误差（delta）

：第l层的单元数（不包括偏置单元）。

假设当前为第l+1层，单元下标为i。其中l层单元下标为j，l+1层单元下标为k。

①

若l+1层为输出层：

，为第i个样本标签

若l+1层为隐藏层：

②

③

总结：

= ① \* ② \* ③

①：传播到隐藏层的误差（errors propagated to the hidden layer）

②：隐藏层梯度（hidden layer gradients）

① \* ②：隐藏层计算误差（hidden layer delta）

若l+1层为输出层：

= ① \* ② =

若l+1层为隐藏层：

= ① \* ② =

，是学习率（learning\_rate）

参考索引

<https://my.oschina.net/findbill/blog/529001>

<http://blog.csdn.net/qrlhl/article/details/50885527>