人工神经网络结构如图。

参数说明：

L代表神经网络层数。l代表第l层。代表第l层的节点个数。C代表损失函数。代表第l层和第l+1层的连接权。代表第l层的偏置。代表第l层的激活值，同时是第l+1层的输入值。y代表真实的标签，X代表输入的样本。。。代表非线性激活函数。每一层的线性变换。代表学习率。代表第l层的残差。

用梯度下降法更新模型：

由于可先求，再通过链式法则求得，。

分为以下两个步骤进行：

1. 计算C对输出层的偏导数

一般分类问题使用交叉熵函数作为损失函数,其中n代表类别个数。

一般输出层的激活函数使用softmax函数， 。

求损失函数对最后一层线性变换的偏导数

当i=j时：

当ij时：

代入得：

由于y使用的是one-hot编码，只有一项为1，其他项为0，所以可化简为

二、计算隐藏层偏导数递推式

由链式法则得到，

即为残差的递推公式。

根据每一层的残差即可求得每一层的梯度：