

**前向传播：**

**梯度下降：**记损失函数为L

先引入中间变量：

对于输出层：

对于中间层：

权重矩阵梯度：

阈值向量梯度：

**根据前向传播和上述4个公式计算梯度并更新权重和阈值，直到达到停机条件**