# BP神经网络

人工神经网络是在生物神经系统的启发下，由许多简单的神经元互相连接而形成的网络【1】，人工神经网络可以学习复杂的非线性关系，从而可以表征真实社会中更加复杂的问题。神经网络中的一个神经元所起到的作用是接收来自其他神经元的加权输入，然后结合自身的阈值(偏置)，最后经过非线性函数的处理，得到输出结果【2】。典型的神经元模型如图1.1所示。



误差逆向传播(error BackPropagation, 简称BP)是常用的训练神经网络的方法，BP网络指使用BP算法进行训练的多层前馈神经网络【3】，是一种有监督学习算法【4】。但是，BP算法不仅仅可以用来训练前馈神经网络，它还可以用来训练递归神经网络等【5】。

# 梯度下降法

梯度下降法【6】是一种经典的最优化方法，其主要思想是不断沿着负梯度方向进行搜索【6】。给定目标函数如公式(1-1)【6】所示，是要学习的参数，是第个输入特征向量的第个分量，，表示偏置，共有维特征。



采用均方误差损失函数，如公式(1-2)【6】所示，共有个训练样本，表示第个训练样本的真实类标向量，表示第个训练样本的预测类标向量，当损失函数的值最小时，说明所训练出的模型参数最能拟合训练样本，因此求解参数的过程就是最小化损失函数。



首先求出损失函数对参数的导数，如公式(1-3)所示。然后，根据损失函数对参数的负梯度方向更新参数，如公式(1-4)【6】所示。是梯度下降法的学习速率，一般情况下，随着学习次数的增加，参数逐渐减小，即参数在学习过程中的变化越来越小。







由公式(1-4)可以看出，对每一个参数，都需要使用全部样本来学习该参数的变化量，将这种梯度下降法的实现方式称作”批梯度下降法”。在实际操作中，由于样本个数较大，所以这种参数更新方法会使导致训练过程缓慢，难以应用于实际问题。

为了克服批梯度下降法的缺点而出现了随机梯度下降法【6】和小批量梯度下降法【6】，这两种方法使用样本全集中的一个或部分样本来更新参数，这样操作使得每一次并不是按照严格意义上的最优方向来更新参数，但是从整体来看，依旧是朝着负梯度的方向更新参数，这两种梯度更新方式使参数学习的速度大大提高，适用于大规模训练样本的情况。

# BP网络前向传播和反向传播

一个典型的三层BP网络如图1.2所示，第一层是输入层，共有个输入神经元；第二层是隐层，共有个隐层神经元，第三层是输出层，共有个输出神经元。表示第个输入神经元与第个隐层神经元的连接权值，表示第个隐层神经元的偏置，表示第个隐层神经元与第个输出神经元的连接权值，表示第个输出神经元的偏置，表示第个输入神经元的输入特征值，表示第个隐层神经元的输出，表示第个输出神经元的输出，表示第个隐层神经元的输入，表示第个输出神经元的输入。

BP网络的前向传播是指将上一层的输出与层间的对应权值相乘并求和，最终经非线性函数的处理，得到下一层对应神经元的输出值。常用的非线性函数有sigmoid函数(如公式(1-5)所示)、tant函数(如公式(1-6)所示)、ReLU函数(如公式(1-7)所示)等。







如图1.2所示，前向传播过程如公式(1-8)和公式(1-9)所示，设使用sigmoid函数。







对于第个训练样本，，，设神经网络对于该样本的输出为，使用均方误差损失函数，则对样本的损失函数如公式(1-10)所示，对整个样本的损失函数如公式(1-11)所示，其中第一项表示在所有样本上的平均损失，第二项表示正则项，表示所有参数组成的向量，正则项可以避免过拟合。Sigmoid函数对参数求导结果如公式(1-12)所示。BP网络反向传播的目的是求得使损失函数最小时的参数，使用梯度下降法进行求解。







根据链式求导法则，损失函数对参数和的导数分别如公式(1-13)和公式(1-14)所示。





在样本上，参数的更新如公式(1-15)所示。



BP误差反向传播算法的基本过程为：

**输入：**训练集，学习率.

**输出：**连接权值与偏置值确定的多层前馈神经网络.

**过程：**

产生(0,1)范围内的随机数，初始化所有连接权值和偏置值；

Repeat

For all ：

根据当前参数执行前馈传播，计算当前样本的输出值；

根据公式(1.15)进行反向传播，更新当前参数；

End for

Until 达到终止条件

[1]T Kohonen. An introduction to neural computing[J]. Neural Networks. 1998 1(1):3-16.

[2] McCulloch，S Warren，Pitts，Walter. A logical calculus of the ideas immanent in nervous activity[J]. Bulletin of Mathematical Biology. 1990 5 (4):115-133.

[3]

[4]

[5]

[6] I Sutskever，J Martens，G Dahl，G Hinton. On the importance of initialization and momentum in deep learning[J] International Conference on International Conference on Machine Learning. 2013, 38(5): 3-11.