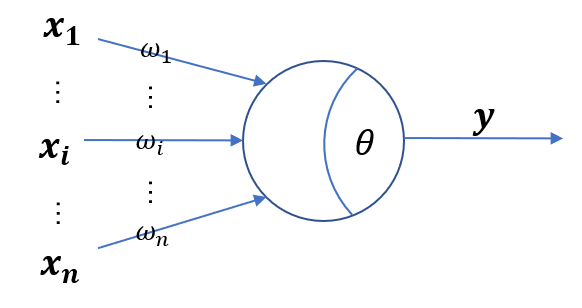
# 一、单个神经元

神经网络算法，是使用计算机模拟生物神经系统，来模拟人类思维方式的算法。它的基本单位就是人工神经元。通过相互连接形成一张神经网络。



生物神经网络中，每个神经元与其他神经元连接，当它“激活”时，会传递化学物质到相连的神经元，改变其他神经元的电位，当电位达到一定“阈值”，那么这个神经元也会被激活。

单个人工神经元的计算公式：

其中：

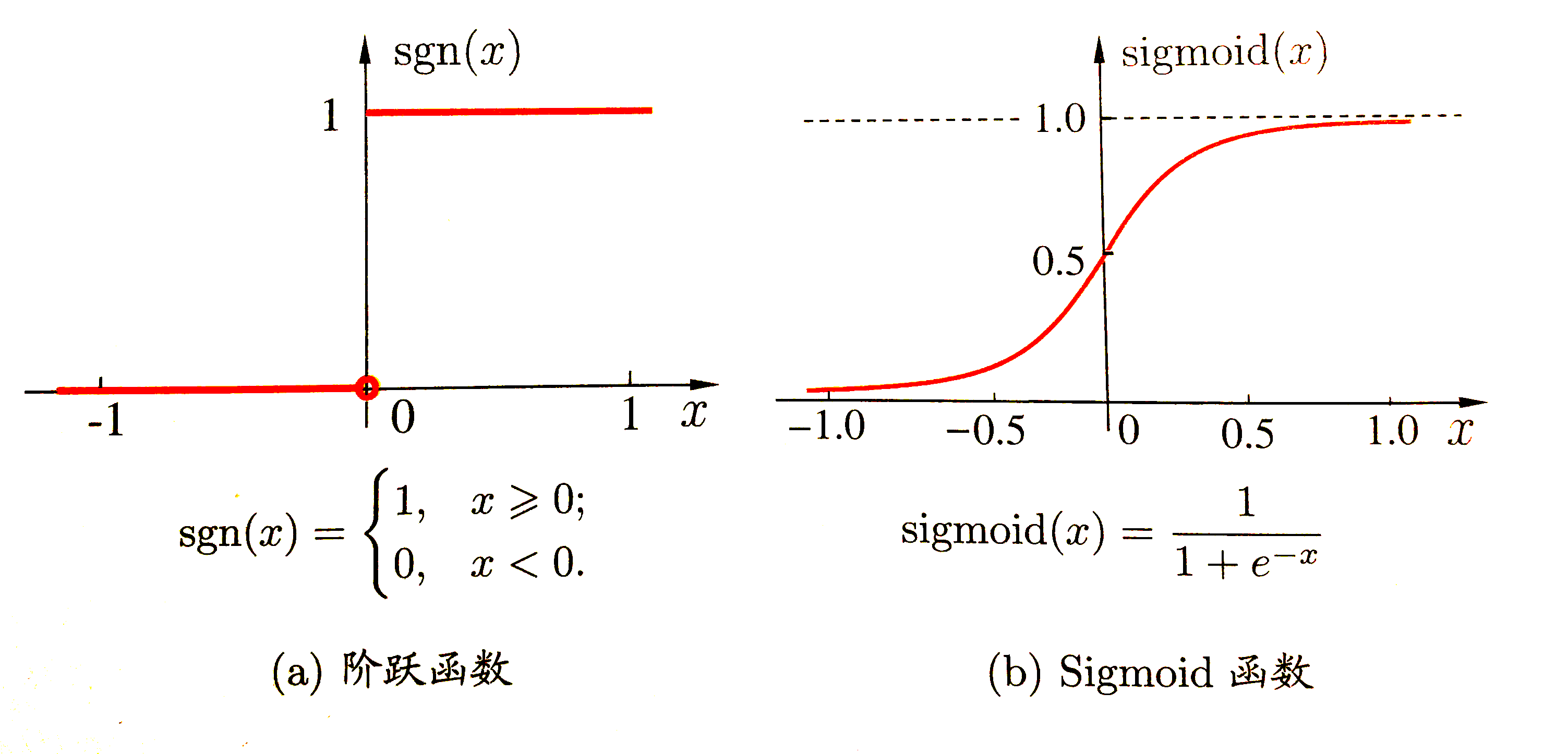
为输入参数向量，表示其他神经元输入的信号。

为每个输入参数的权重值，表示对应神经元信号的权重。

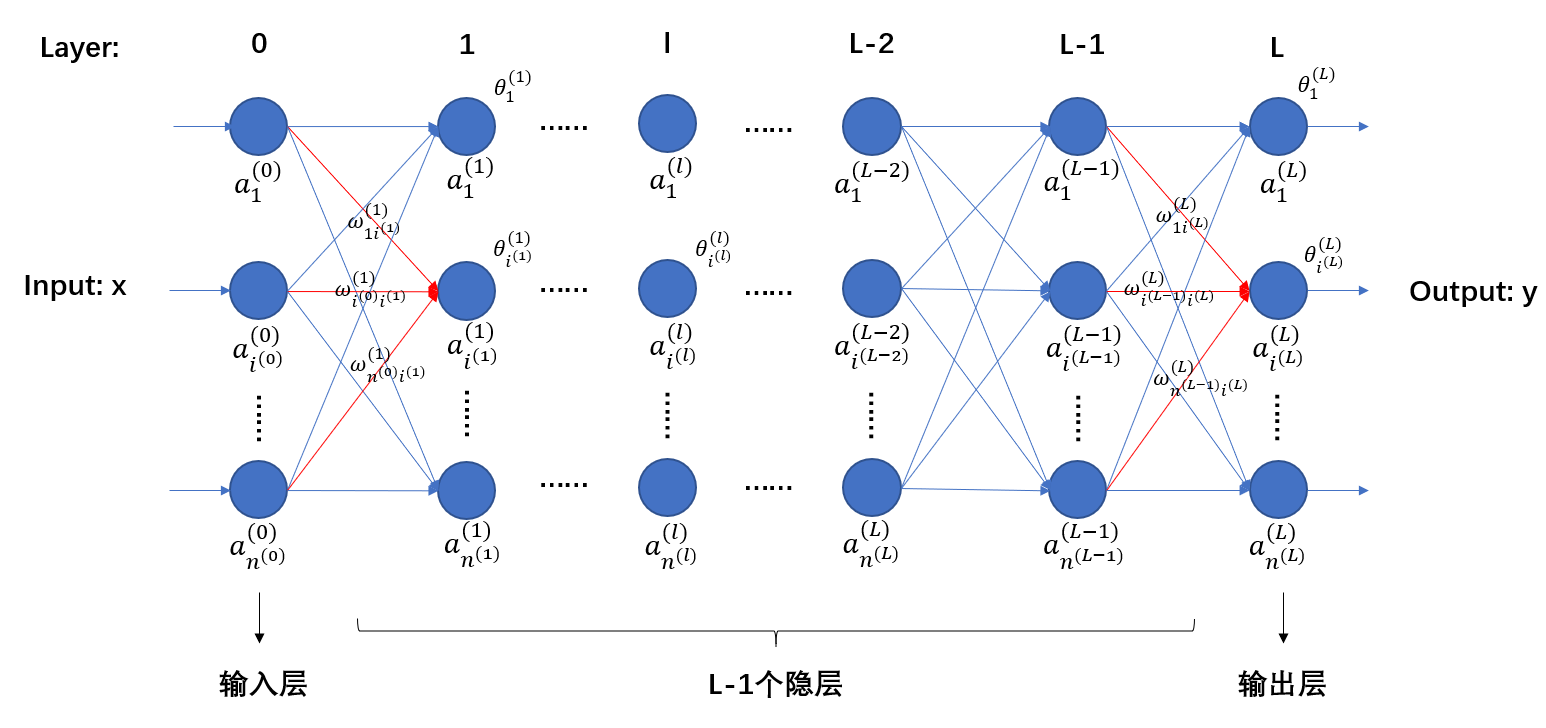
*θ*为阈值或者偏差值，是指该激活神经元的难易程度。

y为神经元的输出值，表示该神经元是否被激活。

**Act()**为激活函数，理想的激活函数是下图（a）中的跃阶函数，“1”为神经元兴奋，“0”为神经元抑制，但由于跃阶函数具有不是连续可导等不好的性质，因此一般采用下图（b）的Sigmoid函数作为激活函数。



# 二、全连接神经网络结构



我们来定义一个全连接神经网络：

全连接神经网络，就是指每一层的每个神经元都和下一层的每个神经元相连接。

Layer：0为输入层，

Layer：L为输出层

其他L-1个Layer为隐层

输入

我们称一个输入值**x**为一个**样本**

输出

变量的上标（0）（L），表示该变量处于神经网络的哪一层。

表示第L层编号为i的神经元。表示第L层的神经元数量。

更好的理解神经网络，可观看此视频：<https://www.bilibili.com/video/av15532370>

# 三、反向传播算法（BP算法）

下面来说明如何调整一个神经网络的参数，也就是误差反向传播算法（BP算法）。以得到一个能够根据输入，预测正确输出的模型。

## 1、首先我们要了解优化的目标

根据人工神经元的定义，有以下三个公式：

其中，Act()是激活函数，之前已经说过。

根据公式(2)和公式(3)，可以得出各个神经元之间的通用计算公式，如下所示：

公式(4)是人工神经网络正向传播的核心公式。

那么，我们根据什么来调整神经网络的参数，以得到一个能够正确预测结果的模型呢？请看下面的公式：

公式(5)用来计算我们期望的输出和实际输出的“差别”，其中cost()叫做损失函数。我们的期望是损失值达到最小。

但是，只根据一次输出的损失值，对参数进行调整，无法使模型适应所有输入样本。我们需要的是，调整参数，使得所有输入样本，得到输出的总损失值最小，而不是只让其中一个样本的损失值最小，导致其他样本损失值增大。因此有如下公式：

公式(6)表示一个batch的所有样本输出的总损失值的平均值。其中，bn表示一个batch中样本的数量。

为什么不用所有的样本计算损失值，而将所有样本分成一个个的batch呢？因为所有的训练样本数量太大了，可能有数以百万计，将所有的样本损失值都一起进行运算，计算量过于庞大，大大降低了模型计算的速度。

公式(6)中计算总的损失值C，其实是一个以所有的连接权值ω和所有的阈值θ未为变量的多元函数。我们想要的模型就是求得C最小时，所有ω和θ的值。直接计算显然是不可能的，因为对于一个大的深度神经网络，所有的参数变量，可能数以万计。

在这里我们使用**梯度下降算法**来逐步逼近C的最小值，也就是先随机得到一组参数变量的值，然后计算参数变量当前的**梯度**，向梯度的反方向，也就是C变小最快的方向，逐步调整参数值，最终得到C的最小值，或者近似最小值。

而将所有样本，随机分成一个个固定长度的batch，以得到近似的梯度方向，叫做**随机梯度下降算法**。

更好理解梯度下降算法，逐步求得最优的参数值，可观看此视频：<https://www.bilibili.com/video/av16144388>

## 2、开始求梯度

那么，根据梯度的定义，接下来的任务，就是求取各个参数变量相对于C的偏导数。我们将使用**误差反向传播算法**来求取各个参数变量的偏导数。

这里先剧透一下，求取参数偏导数的方法，和神经网络正向传播（根据样本计算输出值）的方式类似，也是逐层求解，只是方向正好相反，从最后一层开始，逐层向前。

更好的理解**误差反向传播算法**，可观看此视频：<https://www.bilibili.com/video/av16577449>

首先，我们先求神经网络最后一层，也就是输出层的相关参数的偏导数。为了降低推导的复杂性，我们只计算相对于一个样本的损失值函数Cbi的偏导数，因为相对于总损失值函数C的偏导数值，也不过是把某个参数的所有相对于Cbi偏导数值加起来而已。

根据公式(2)、公式(3)、公式(5)，以及“复合函数求导法则”，可以得到输出层（L层）某个神经元的权值参数ω的偏导数，计算公式如下：

根据公式(5)可以得到：

根据公式(2)可以得到：

根据公式(3)可以得到：

将公式(8)(9)(10)，带入公式(7)，可以得到：

我们令：

根据公式(8)(9)则有：

将公式(13)，带入公式(11)，可以得到：

这样我们就得到了输出层L相关的权值参数ω的偏导数计算公式！

接下来，同理可得输出层L相关的阈值θ的偏导数计算公式为：

而根据公式(3)可以得到：

将公式(16)带入公式(15)可以得到：

这就是输出层L相关的阈值θ的偏导数计算公式！

## 3、根据L层，求前一层参数的偏导函数

由公式(3)可知，一个权值参数ω只影响一个L-1层的神经元，因此有：

根据公式(3)可以得到：

将公式(19)带入公式(18)可以得到：

根据公式(12)可以得到：

将公式(21)带入公式(20)可以得到：

同理，我们可以得到：

根据公式(3)可以得到：

将公式(24)带入公式(23)可以得到：

这样我们就得到了L-1层神经元相关参数的计算公式！

下面，我们还需要推导一下之间的关系，根据公式(2)可以得到：

同样根据公式(2)可以得到：

将公式(27)带入公式(26)可以得到：

由公式(3)可知，一个权值参数ω只影响一个L-1层的神经元，但这个L-1层神经元影响了所有L层的神经元。因此，根据“多元复合函数求导法则”有：

根据公式(12)可以得到：

将公式(27)带入公式(26)可以得到：

根据公式(3)可以得到：

将公式(32)带入到公式(31)可以得到：

将公式(33)带入公式(28)可以得到：

**这样我们就得到了反向传播，逐层推导的通用公式：**

在这里，ω和z都是正向传播过程中，已经算好的常数，而 可以从L层开始逐层向前推导，直到第1层，第0层是输入层，不需要调整参数。而第L层的 可参考公式(13)。