

# 反向传播理解--从抽象到具体

神经网络从计算的角度看，数据是从底层输入，经过每一层，根据与该层之间的权重计算以一个中间结果，这个中间结果再经过一个非线性激活函数作用，得到该层的输出结果，然后把该层的输出结果传递给下一层，继续进行下一层的中间结果计算与非线性函数激活，直到最后一层，得到最终的结果，最后的结果或者用来表示分类结果，或者表示回归结果。

如果一个网络训练好了，所有层之间的权重都是确定好的，给定任意的一个输入，根据每一层之间的权重系数，从底层到顶层，计算出最终的结果，这一阶段叫做推理或者叫做测试。

反向传播发生是在学习过程中，每一次迭代根据计算出的结果与实际对应结果的误差，反向调整网络层之间的系数。

网络结构对于解决特定的问题是关键，但是最底层还是网络的反向传播过程，这是学习过程的基础，理解这部分，才能更清晰的理解一个网络结构，以及如何修改网络或者涉及全新的网络。

很久之前，写过一个从 BP 神经网络理解反向传播文章，那是从细节描述，反而理解起来不是很容易，尤其是跟同事讲或者带新人的时候，不太容易让他们清晰的理解反向传播的过程。这里反过来，从高层的抽象开始，逐步到细节，先从整体上有一个概念，然后逐步细化。

## 一个典型的网络

如下图，表示一个抽象的多层神经网络的结构。

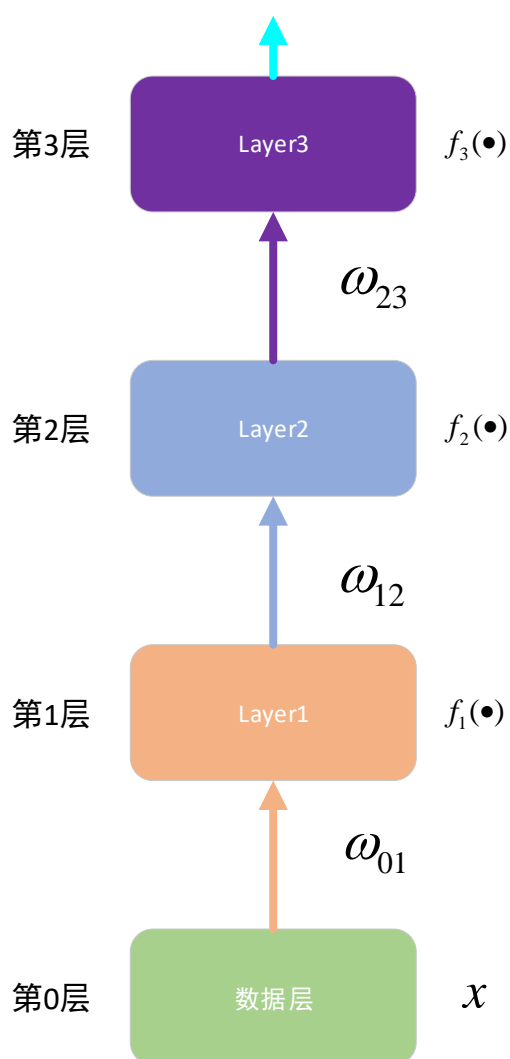


图 1 多层网络抽象结构

第 0 层是数据层，没有特别的，用  $x$  表示；

第 1 层是中间网络层，不管是卷积层，还是全连接层，还是 BN 层，与前一层之间的权重是  $\omega_{01}$ ，这一层的非线性激活函数是  $f_1(\bullet)$ ；

第 2 层也是中间网络层，不知道是个什么东西，与第一层之间的权重是  $\omega_{12}$ ，非线性激活函数是  $f_2(\bullet)$ ；

第 3 层是最后的输出层，与第 2 层之间的权重是  $\omega_{23}$ ，非线性激活函数是  $f_3(\bullet)$ 。

一般的网络结构都类似与这个样子，下面图描述了网络的前向计算过程中的数据传递。

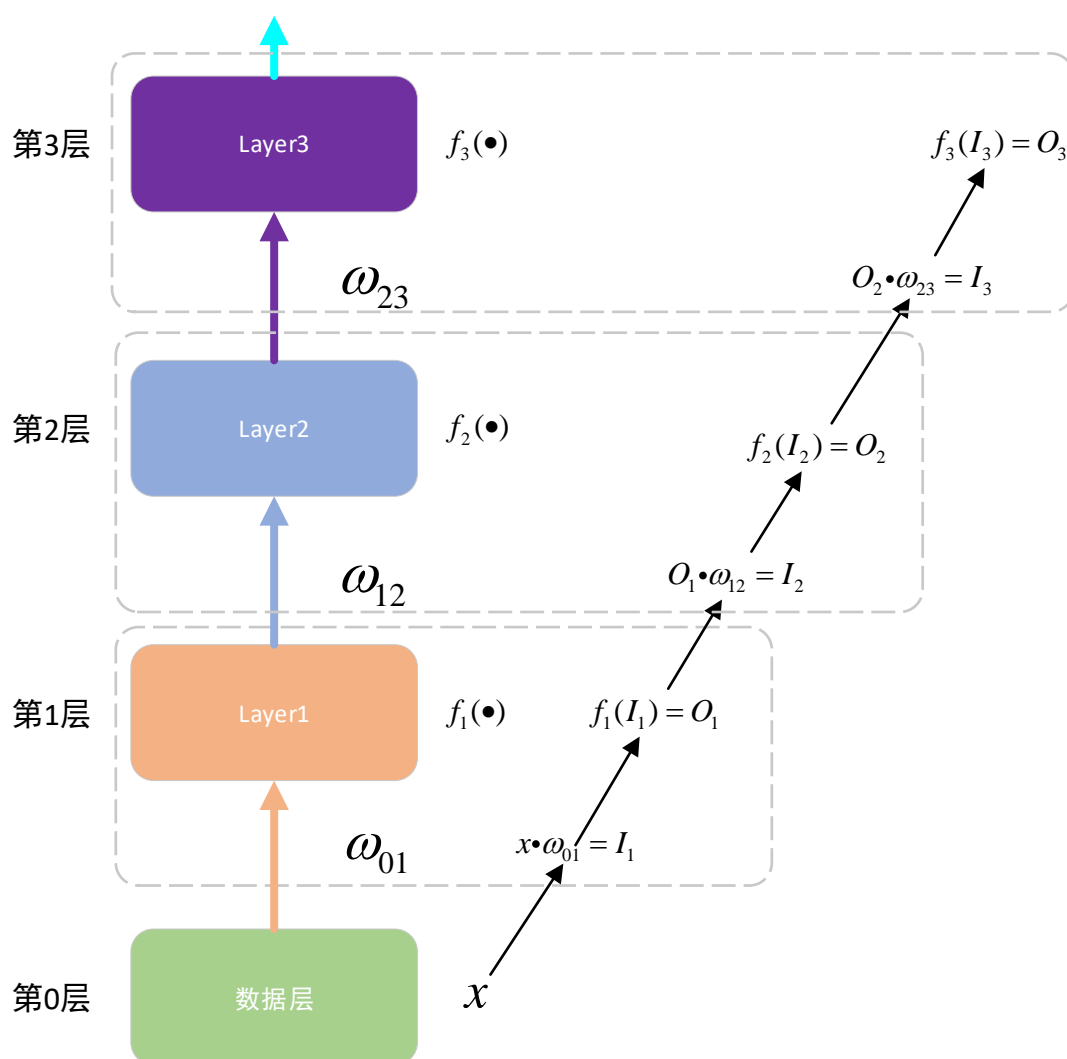


图 2 网络前向计算数据传递过程

从图中可以看出，第 1、2、3 层之间的计算过程。前向传递理解起来比较容易，不管这样抽象的表示，还是具体的网络实现。下面看看反向传递是怎么回事。

下图是反向传播时，从顶端的误差如何反向传递给下面的每一层。

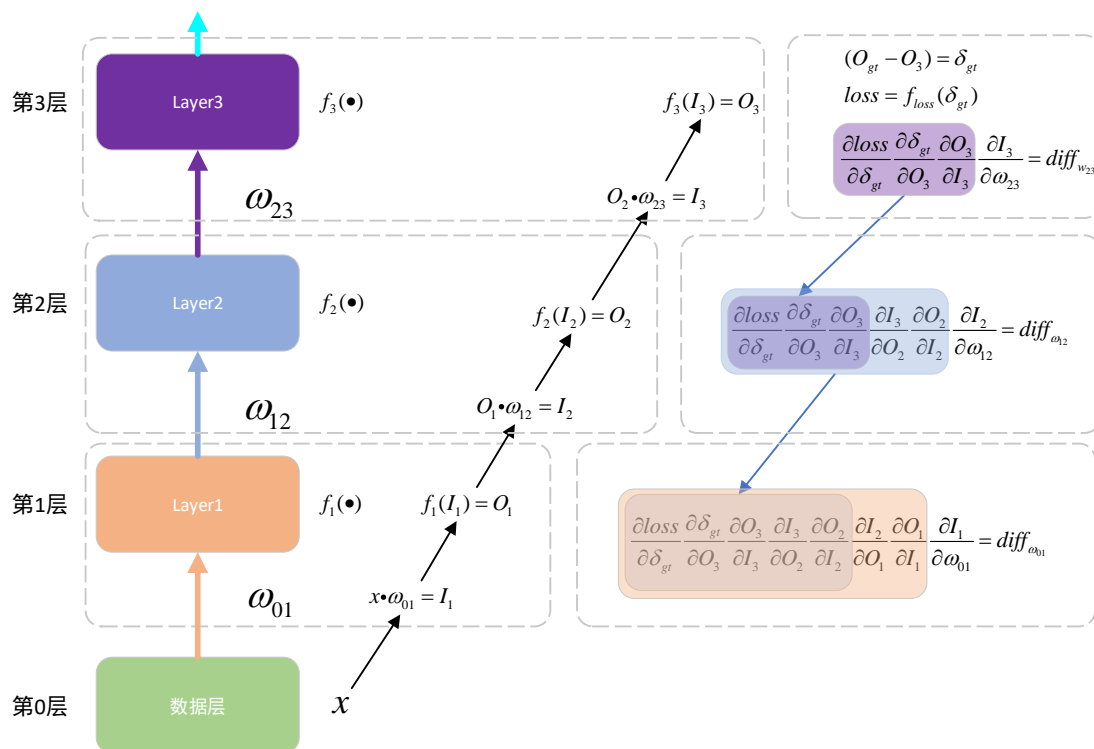


图 3 反向传播误差传递

第3层是最后一层，可以直接计算误差的损失，根据损失计算对权重的导数；但是到了内部的层，没有办法直接获取损失，只能通过上图中的链式规则计算对内层权重大导数。

从上面的图中发现，层次越深，这个导数计算越多，这哪能受得了，每一层的导数对输入的导数，在下面一层的计算中还会有用，所以，可以把这一部分直接传递给下面的层，这也就是误差回传的本质。

对上面的图进行修改能更清晰的理解。

