Softmax及交叉熵损失函数反向传播求导过程分析

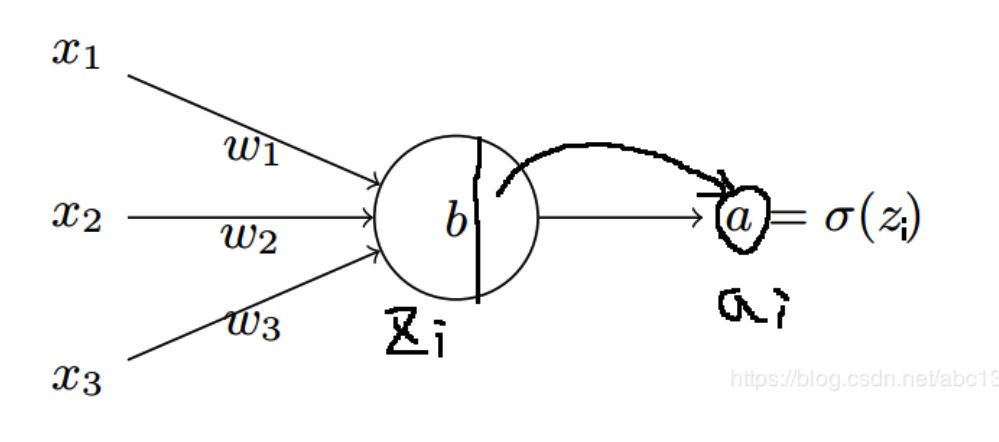
1. softmax函数

一般在神经网络中，softmax函数可以作为分类任务的输出层，其实可以认为softmax输出的是几个类别选择的概率，比如我有一个分类任务，要分为三个类，softmax函数可以根据他们相对的大小，输出三个类别选取的概率，并且概率和为1.

Softmax函数的公式是这种形式的：

其中代表的是第i个神经元的输出。其实就是在输出后面套一下这个函数，在推导之前，我们统一一下网络中的各个表示符号。

首先是神经元的输出，一个神经元如下图：



神经元的输出设为：

其中表示第i个神经元的第j个权重，b是偏置项，表示该网络的第i个神经元的输出。给这个输出加上一个softmax函数，那就变成了这样：

其中代表softmax的第i个输出值，就是套用了softmax函数。

1. 损失函数

损失函数可以有多种形式，这里采用的是交叉熵损失函数，原因如下：

* 1、这个求导结果比较简单，易于计算。
* 2、交叉熵解决某些损失函数学习缓慢的问题。

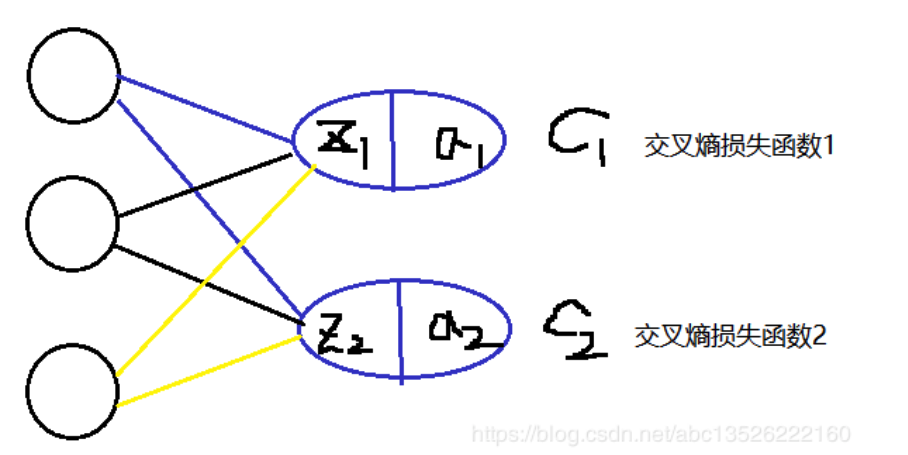
交叉熵的函数是这样的：

注意其中的一项为(一个节点):

其中表示真实的分类结果。

1. 具体的推导过程

假设是这样的模型：



下面，我们需要先明确一下，我们要求的是loss对于神经元的输出的梯度【注意求得不是对激活函数后的输出的梯度】，即：

根据复合函数求导法则，可得：

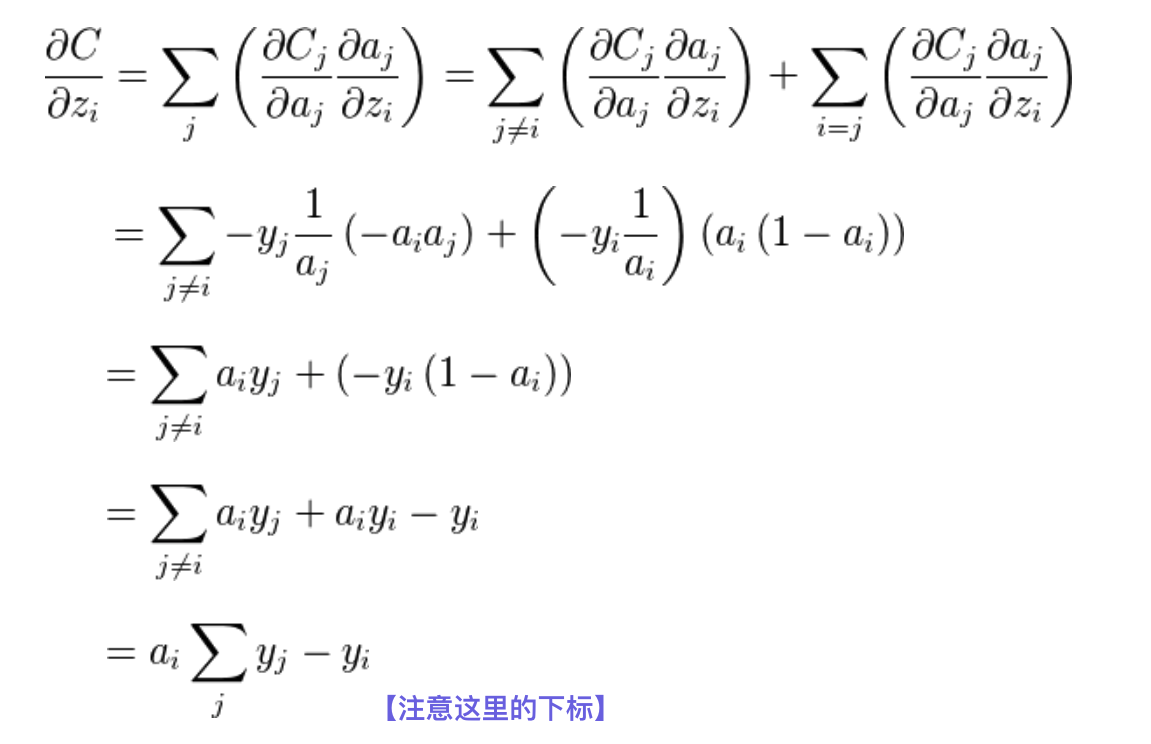
注意这里是不是，这里要看下softmax的公示了，因为softmax公式的特性，它的分母包含了所有的神经元的输出，所以，对于不等于i的其他输出里面，也包含着，所以所有的输出都要纳入到考虑范围内，并且后面的计算可以看到需要分为i=j和两种情况求导。

下面我们一个一个的推，首先求的值：

第二个就稍微复杂点，我们需要将它分为两种情况：

* 1、如果i==j：
* 2、如果:

接下来，我们只需要把上面的组合起来就行了：



最后的结果看起来简单多了。针对分类问题，我们给定的结果最终只会有一个类别是1，其他类别都是0，因此，对于分类问题，这个梯度等于：