**RNN之BPTT理论推导**

**一、RNN (循环神经网络)**

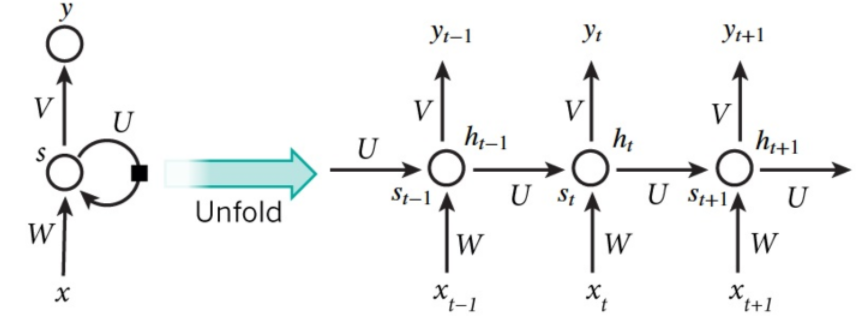
RNN是一种具有长时记忆能力的神经网络模型，被广泛用于序列标注问题。序列标注问题中，模型的输入是一段时间序列，记为



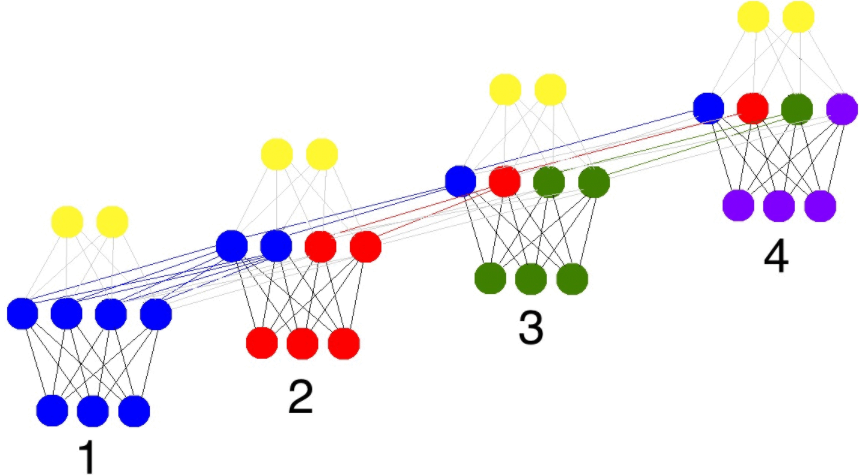
我们的目标是为输入序列的每个元素打上标签集合中的对应标签，记为



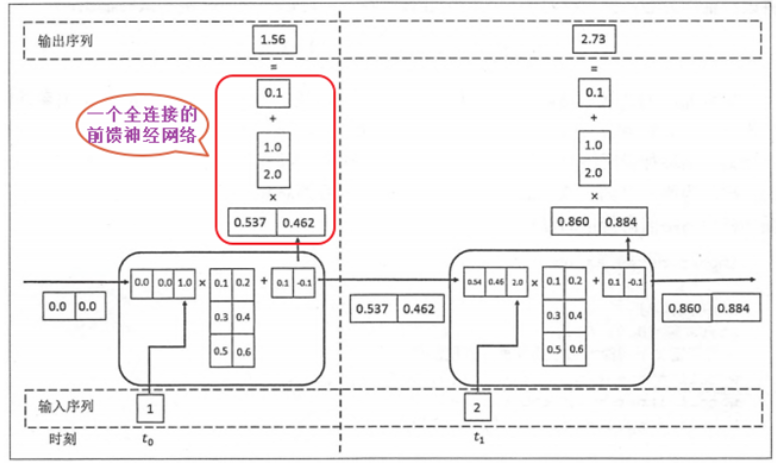
1. **RNN一般结构图**
2. 符号描述



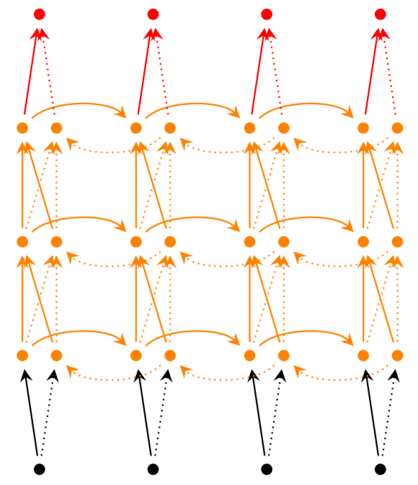
1. 节点描述



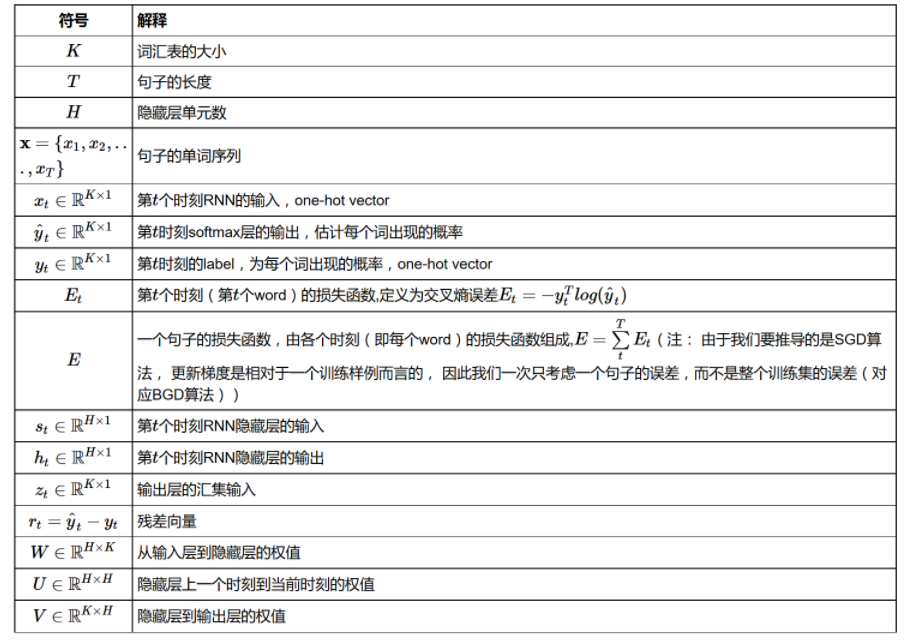
1. RNN前向传播的具体计算过程



1. 深度RNN（包含多个隐藏层）示意图



1. **符号说明**



1. **RNN的BPTT**

RNN中上述符号之间关系如下：



为RNN共享参数，BPTT目标是求



又因为



只需对每个时刻的损失函数求偏导再相加即可

1. **U、W、V、b、c偏导**

首先参考文献[1]定义prod运算符，对输入或输出X，Y，Z为任意形状张量的函数Y=f(X)和Z=g(Y)，通过链式法则，我们有



其中prod运算符将根据两个输入的形状，在必要的操作（如转置和互换输入位置）后对两个输入做乘法。

1. 对V求偏导

为标量，为权值矩阵，即标量对矩阵求偏导





为向量外积，对于推导细节见附录1。

1. 对c求偏导





1. 对U、W、b求偏导
2. W、b虽然参数共享，但是他们不仅对t时刻输出有贡献，同时对t+1时刻隐藏层的输入有贡献。所以，以U为例有



对t时刻的U、W、b求导，利用链式法则可得

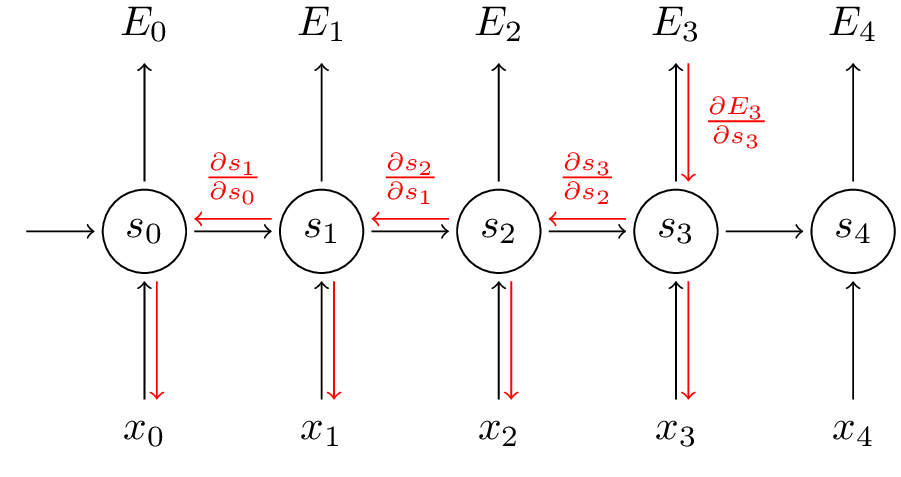






以上均需求出，根据如下传递关系





上图隐藏了，简化了误差传递路径。

有如下关系





其中为向量点积，即得到如下递推关系





通过可以推出，可推出如下







得到







参数更新











其中

1. **RNN梯度爆炸或消失**

以上推到可知

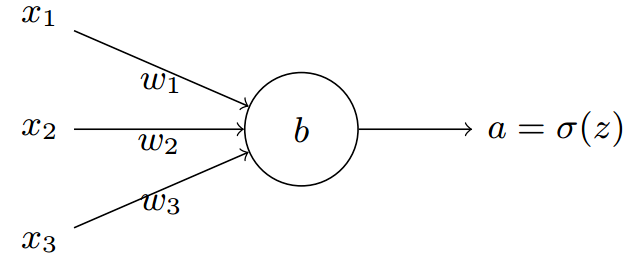


其中为各自矩阵模的上界，为，较大时，如果引起梯度爆炸，反之引起梯度消失。

# 附录1. softmax交叉熵损失函数求导

1，softmax函数

神经网络分类中经常用到softmax，如下神经元



其输出为



是第i个神经元的第j个权重，b是偏移量，表示该网络的第i个输出



代表softmax的第i个输出值，右侧就是套用了softmax函数。

1. 损失函数loss function

损失函数可以有很多形式，这里用的是交叉熵函数，主要是由于这个求导结果比较简单，易于计算，并且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

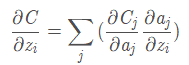


其中表示真实的分类结果

我们的目的是求loss对神经网络输出的导数即



根据复合函数求导法则



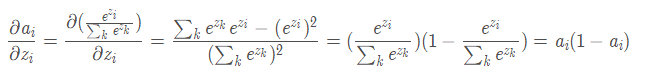
**注意：**因为softmax公式的特性，它的分母包含了所有神经元的输出，所以对于不等于i的其他输出里面，也包含，需要把所有纳入计算范围。

对于有

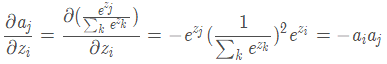


对于，分为两种情况

1. 



1. 



以上两种情况组合为



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，进一步化简为

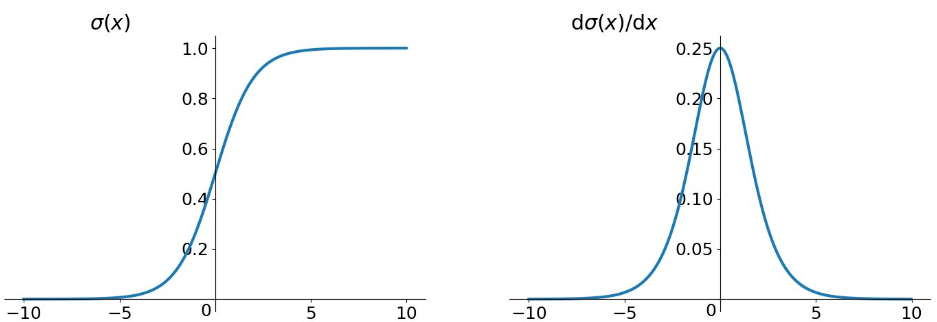


**附录2. 常用激活函数及其导数**

1. Sigmoid函数，表达式

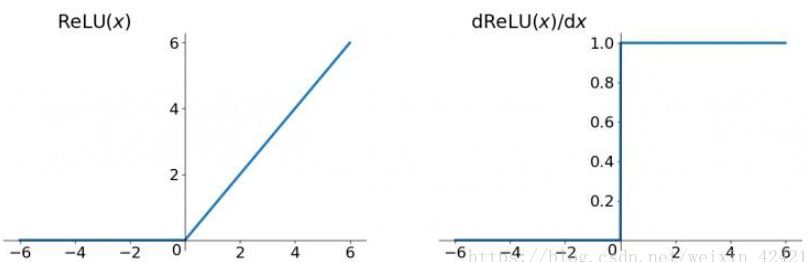
****

****



1. ReLu函数，表达式

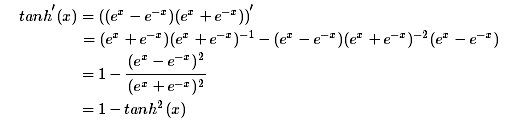


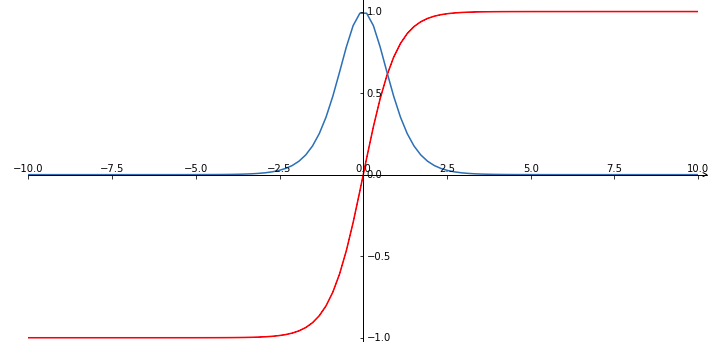


1. tanh函数，表达式



导数为





**参考文献**

1. A guide to recurrent neural networks and backpropagation ,Mikael Boden, Dallas Project Sics Technical Report T Sics, 2001
2. <http://zh.gluon.ai/chapter_deep-learning-basics/backprop.html>
3. <https://images2015.cnblogs.com/blog/583155/201608/583155-20160819153519750-1203124108.png>
4. <https://zybuluo.com/hanbingtao/note/541458>