**第七章读书报告**

09118223 吴亦珂

读书进度：第七章读完

1. **问题列表**

（我提出）P15页7.5中说neural language model不需要smothing是指什么吗？handle much longer histories是具体什么意思？

讨论结果：不需要smoothing是指不需要像n-gram 模型那样初始值设为一个非零值，从而避免某些项为0。handle much longer histories值可以考虑更多的前文内容，而不是前后几个词。

（我提出）P15页为什么神经网络学习中是初始化权重用一些小的随机数？以及为什么输入值均值为0，用有统一的方差效果较好？

讨论结果：防止某个值过大，权重过大，影响训练效果。

（别人提出）比起logistic regression等传统模型，神经网络模型的优势体现在哪里？  
讨论结果：书7.2中提到神经网络可以自动的学习出一种对输入的有用的表示形式。例如可以解决异或问题。7.5中提到神经网络用于语言模型的一些优势。总的来说，第一，神经网络模型比楚童模型复杂，因此可以解决更复杂的分类问题。拥有较强的学习能力。第二，神经网络具有更好的抽象性，不需要人为地指定相应的训练特征。相当于是一个黑盒。第三，神经网络具有较好的移植能力，便于迁移学习的使用。一个很复杂的神经网络可以通过预训练模型的方式为其他分类任务所直接使用。

（别人提出）基于神经网络的LM比起传统的基于贝叶斯的LM的优点是什么？为什么会有好的效果？

讨论结果：神经网络语言模型自带smoothing，可以考虑更多的上下文内容，并且更好地处理相似的词。同时还可以通过向量来表示语义之间的相似，不局限于训练的语料。

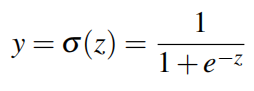
1. **读书收获**

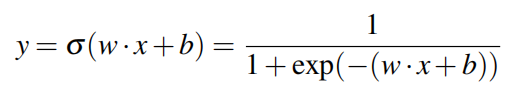
**7.1 Units**



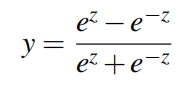
激活函数：

Sigmoid function：

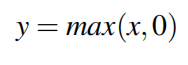




Tanh function:



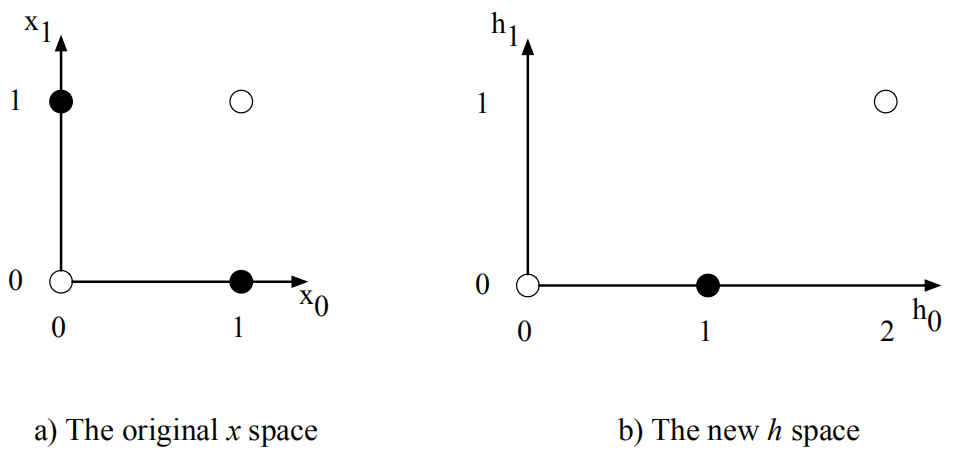
RELU function:



**7.2 The XOR problem**

需要使用多层神经网络才能解决异或的问题，因为单一神经元的判定边界是线性的。

多层神经网络一个很大的优势是将输入转化成有用的表示形式。

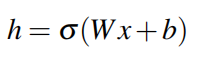


**7.3 Feed-Forward Neural Networks**

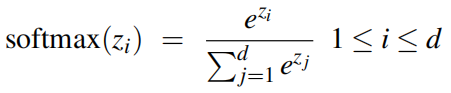
feedforward network是一个多层神经网络并且没有环，每一层的输出都会传到下一层。

有三种节点：输入层，隐藏层，输出层。

神经网络的核心是隐藏层。在标准的结构下，隐藏层是全连接的。

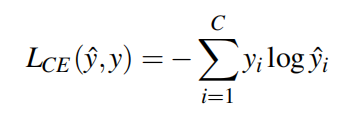


在输出层，通常使用softmax函数：

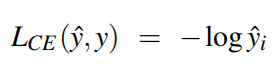


**7.4 Training Neural Nets**

损失函数：交叉熵损失函数

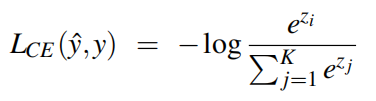


特别地，对于一个硬分类问题，只有一类为正确分类，则可以使用one hot编码，即只有正确的一维为1，其余均为0，这样一来，原本损失函数可以写成

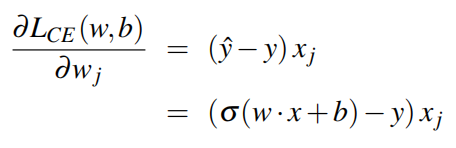


这种形式又被称作negative log likelihood loss

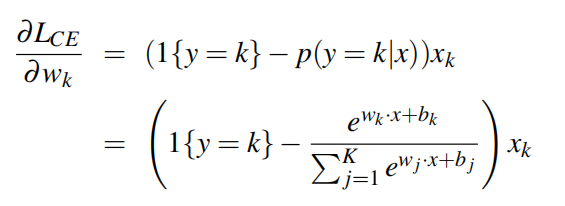
代入k个类分类的式子中，可以得到：

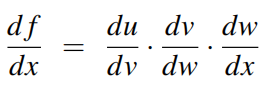


Logistics regression损失函数导数：



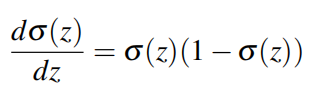
Softmax损失函数导数：



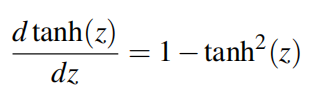
但是这只能求解一层神经网络的导数，如果要求解多层神经网络，可以采用error backpropagation or backprop的方法。这种方法其实就是使用链式法则来求解导数，例如：  


各种激活函数的导数：

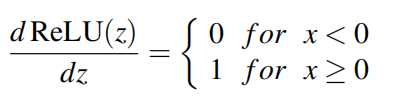
Sigmoid function:



Tanh:



RELU:



More details on learning:

对于逻辑回归，可以将权值以及偏差项设置为0。而神经网络则应该初始化为一些较小的值。

很多正则化方法防止过拟合的发生。可以采用dropout的方法，以及调整参数的方法。

**7.5 Neural Language Models**

相较于n-gram，神经网络的方法不需要smoothing，也可以考虑更多的上文内容。因此准确率也更高。但另一方面，神经网络训练起来更慢。

同时相较于n-gram，神经网络可以考虑到语义的相似性。