**导论大作业报告**

**-对于梯度消失问题采用放大梯度的手段的效果**

张润权 10174507143

1. 内容概述
   1. 使用sigmoid函数的深层神经网络在使用SGD训练时面临梯度消失的问题,然而,如果我们启发的认为梯度有用的是其方向信息,而其大小是不重要的,那么我们可以对过小的梯度采取某种放大手段使训练能正常进行.事实上,由于学习率,我们原本便并非在意梯度的原本大小.
2. 算法思路
   1. 作为放大程度的标准,采用如下的启发式规则:令学习率动态变化,对于每层,使得到的梯度乘上学习率后,其范数与该层权重向量的范数的比值为预先设定的定值,此定值可称为”相对学习率”.
3. 实验环境
   1. 数据集
      1. Mnist数据集,其扩展版本使用<http://neuralnetworksanddeeplearning.com/>的代码生成
   2. 网络
      1. 深层NN
   3. 框架
      1. 主要使用keras进行模型构建,使用numpy进行预实验
4. 实现思路
   1. 在预实验中,在基于numpy的NN上实现目标算法是容易的.
   2. 在keras中,需要继承抽象基类,自己实现一个和平台接口一致的标准优化器,这样的好处是仍可以使用框架的其它工具.
5. 实验过程介绍
   1. 预实验
      1. 实现
         1. 通过一本教材[1]上的源码,获得了numpy实现的NN,在其上修改实现目标算法
      2. 结果
         1. 经预实验验证,普通SGD确实几乎无法训练,在尝试的数个学习率中,均表现为验证集准确率始终为随机猜测,loss不下降
         2. 发现改进的SGD不仅能够正常训练,且收敛很快
   2. 在keras上的验证
      1. 首先重复了预实验的内容,结果一致.
      2. 在训练过程中,发现算法虽然能从无法训练变为可训练,但其最终准确率不高,甚至差于三层感知机,至此,实验的主要目标变为优化最终准确率,证明该优化方法可以收敛到一定程度上较好的结果
   3. 性能优化
      1. 使用多种常见方法进行了尝试,包括dropout,l2正则化,通过预处理扩展数据集,特殊的初始化,更换损失函数,更换激活函数等,某些方案取得了一定效果.
      2. 对于算法中的范数,首先使用了l2范数,后在实验中发现l1范数的效果一般更好
      3. 基于一篇论文[2]中描述的结果,尝试了数种已被证实可能取得好效果的网络结构(NN层数和神经元个数),但由于某些技术手段实现较繁,没能复现其结果
      4. 最终得到的最好结果为98.47%的正确率
6. 实验存在的问题
   1. 速度问题:在尝试过的范围内,方法速度并不比流行的优化器快,与adam对比,在扩增到250000的数据集下,两者仅在第一个epoch相差不大,之后表现不如adam快速和稳定.
   2. 难以训练到较好准确率,在达到90%左右准确率后,表现出难收敛,或者在准确率还较低时就过拟合的现象,很难训练到较好的准确率
   3. 存在数值上的问题,某些情况下会出现loss变为无意义的值(越界,除0等),更换sigmoid为tanh,或更换cross\_entropy为mse均会导致无法正常训练.
   4. 对于普通SGD是否完全不能训练存疑,因为实验只能说明尝试过的范围内效果不好,事实上,论文[2]中指出其正是使用了普通SGD进行训练的,但是即使普通SGD可以训练,本算法仍可以看作一种自动设置合适学习率的启发式,不失实践意义
7. 关于代码和细节
   1. 提交三份自己写的代码的notebook,包含了一部分实验内容,报告未提及的实现细节可以参照源码
   2. 为实验方便,很多代码和结果均没保存下来,提交的仅为部分,关于预实验未提交代码.
   3. expand\_mnist.py是来自[1]的开源代码
8. 主要参考文献
   1. [1] <http://neuralnetworksanddeeplearning.com/>

# [2] Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition

[Dan Claudiu Ciresan](https://arxiv.org/search/cs?searchtype=author&query=Ciresan%2C+D+C), [Ueli Meier](https://arxiv.org/search/cs?searchtype=author&query=Meier%2C+U), [Luca Maria Gambardella](https://arxiv.org/search/cs?searchtype=author&query=Gambardella%2C+L+M), [Juergen Schmidhuber](https://arxiv.org/search/cs?searchtype=author&query=Schmidhuber%2C+J)

Neural Computation, Volume 22, Number 12, December 2010