## Deep-Learning实验报告

## （反向传播参数对比）实验

日期：2017.7.3

学号：24320142202455

姓名：林金鹏

## 实验目标：

反向传播进行调参对比，查看不同参数会造成的结果

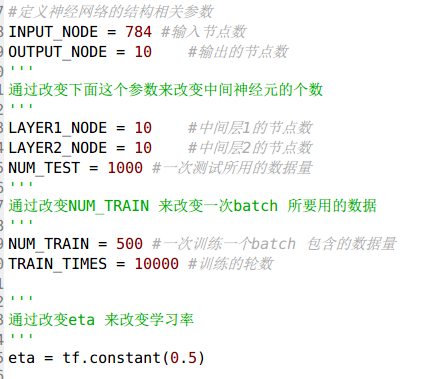
## 实验内容：

利用back propagation 反向传播进行调参对比

## 实验设计：

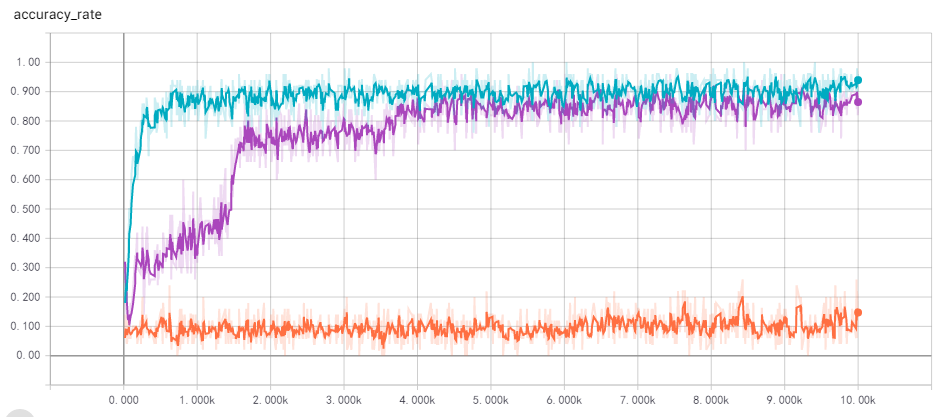
根据变量唯一的方式进行对比实验。本实验会对比的变量有 隐藏层的层数，神经元的个数，还有每次训练时的batch 的数据量。

在代码中，通过将可变的参数提取到开头，便于变量值的改变。

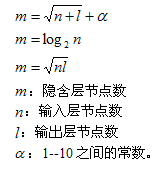


## 隐藏层神经元个数不同进行对比：

10个神经元，30个神经元，100个神经元

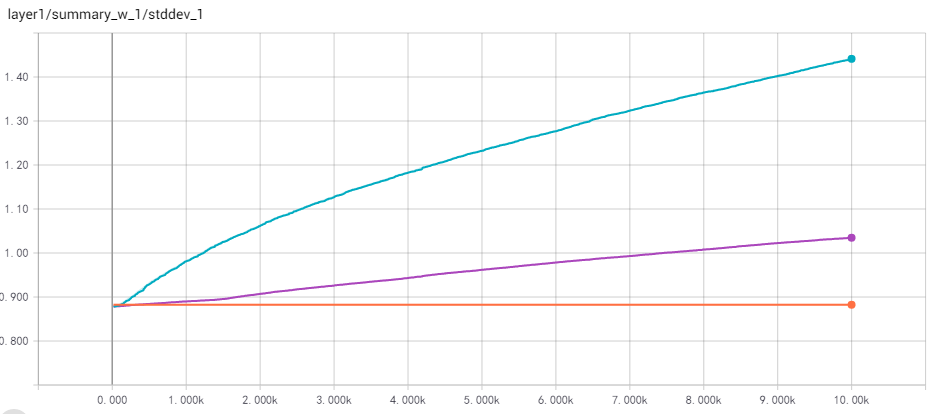


* 10个神经元准确率最高，同时收敛也是最快的。神经元过多会导致收敛变慢，甚至导致不收敛，可能是神经元过拟合。
* 一般有以下经验公式确定神经元的个数：

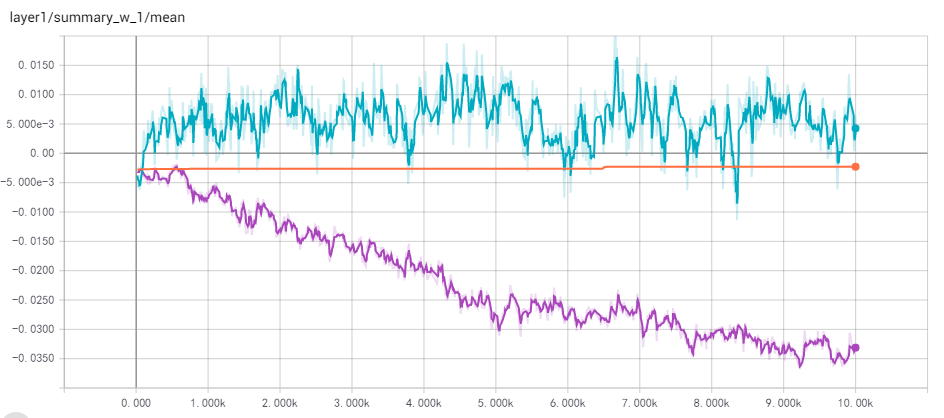


Log 784 约等于 10，大致符合第二个式子。

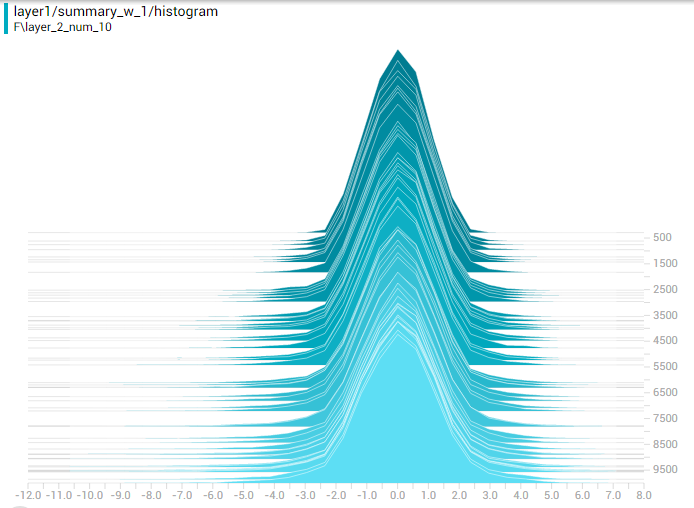
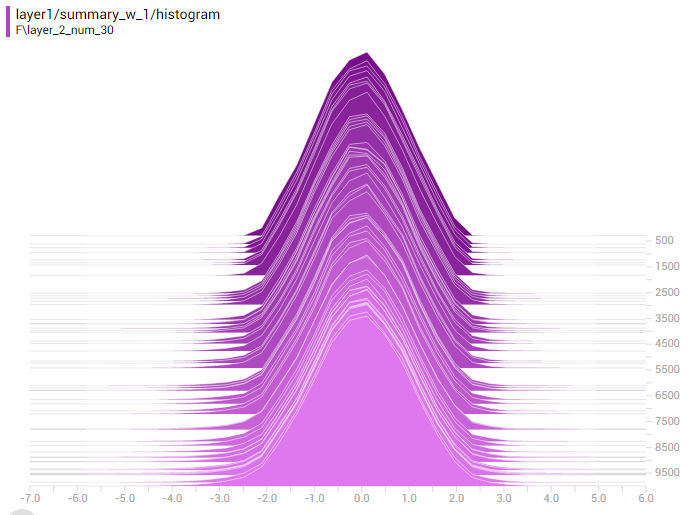
隐藏层的权值的统计对比

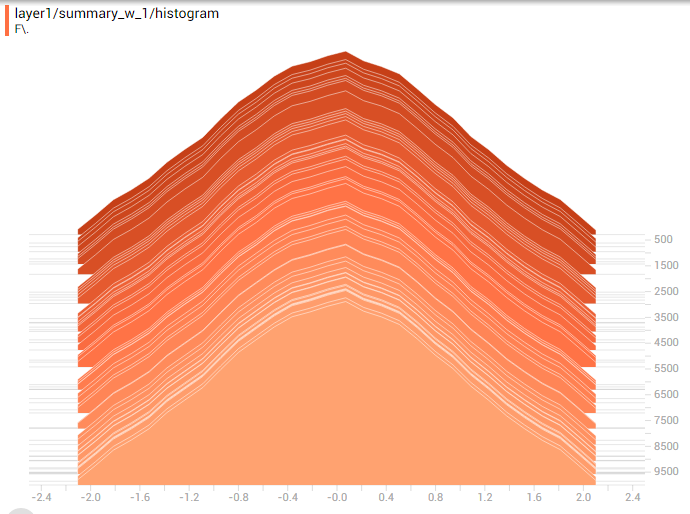


* 10个神经元的方差最大，说明权值的波动最大。权值间的分布比较分散。

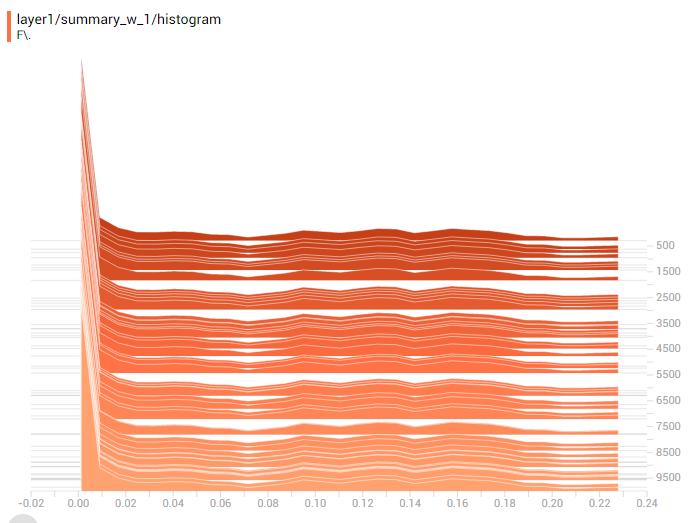


隐藏层权值的分布情况（初始值为正太分布）：



隐藏层权值的分布情况（初始值为全为0）：

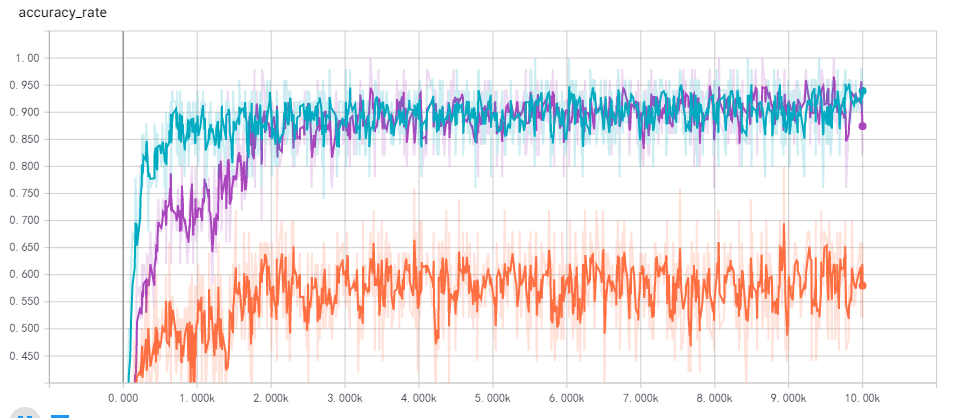


* 初始的时候权值是正太分布的，到了后面变化不大，依旧呈正太分布。当初始值全为0时，训练10000次后，权值的分布比较集中在0附近，而且准确率只有20%。所以初始权值的选择对于模型的训练是比较重要的，一般取较小的正太分布。

## 隐藏层数不同进行对比：

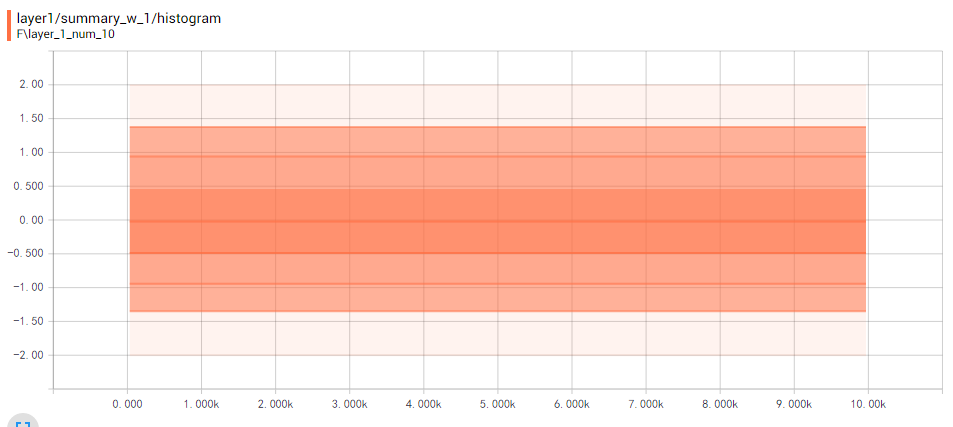
准确率的对比：

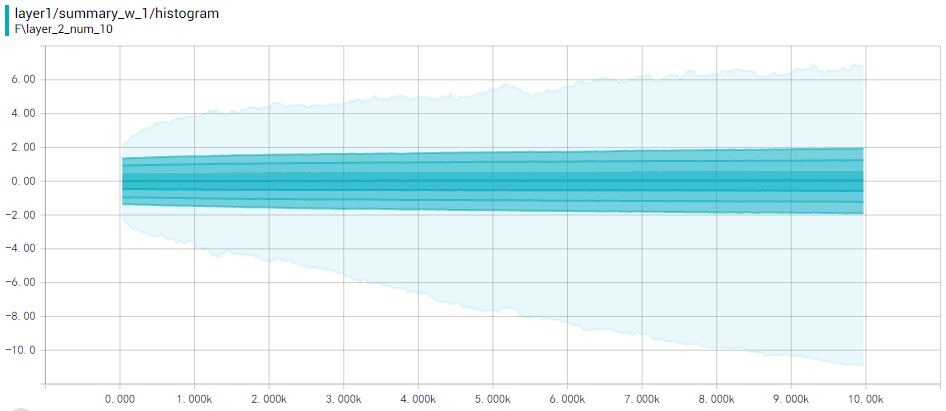
1个隐藏层，2个隐藏层，没有隐藏层

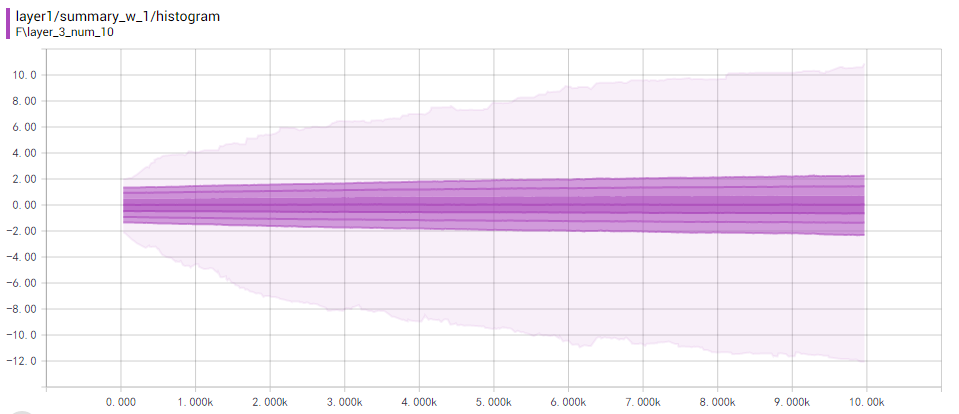


* 在2000次之前，1个隐藏层的收敛更快，2个隐藏层相对较慢，但达到一定次数之后，两者的准确率没有明显差别。较多的层数会让学习变得缓慢，但不一定让训练的效果变得更好。对于没有隐藏层来说，识别效果并不理想，最终只能达到55%左右。

权值分布情况对比：







* 效果比较好的权值的分布一般比较分散，波动比较大。

## 训练batch量不同进行对比(一层隐藏，10个神经元)：

准确率的对比：

Batch=10,Batch=50，Batch=500



* 当batch=50时,比batch=10的收敛更快，而且最后的准确度会稍好一点。适当的增加训练数据量可以提高模型的训练效果。但当batch=500的时候，10%的概率，相当于等概率随机猜。可能的原因是，代价函数（cost = (y-y\_)^2）的计算结果会随着数据集的增大而变大，反映在W和B 的变化上，会使W,B的更新过度，导致不能收敛。
* 通过调整学习率来解决参数更新过于明显的问题。将学习率又0.5变成0.01,可以得到虽然收敛的速度变慢，但是最终还是识别准确率能回到正常水平。

