**前言**

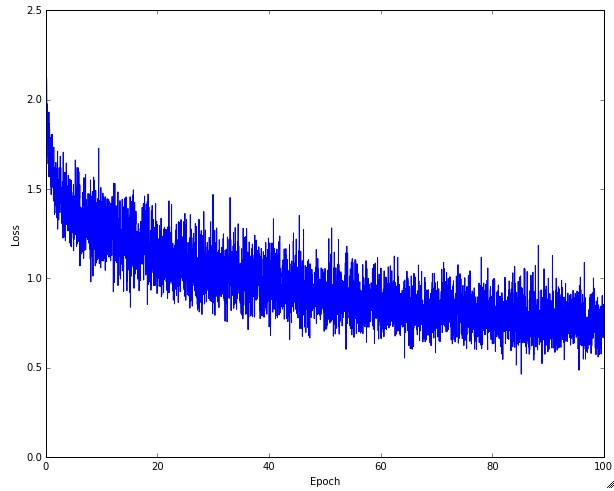
<https://www.cnblogs.com/liujshi/p/5646102.html>

在训练神经网络时，调参占了很大一部分工作比例，下面主要介绍在学习cs231n过程中做assignment1的调参经验。  
主要涉及的参数有隐藏层大小hidden\_size，学习率learn\_rate以及训练时的batch\_size.

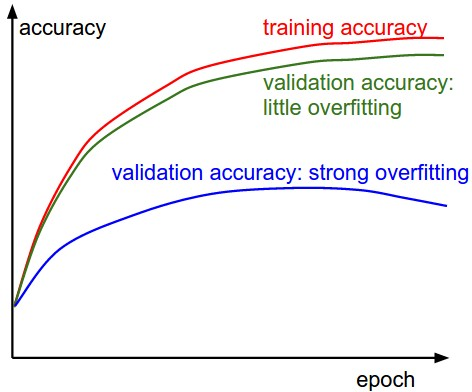
**理论部分**

首先介绍一下讲义上关于以上三个参数的可视化分析。

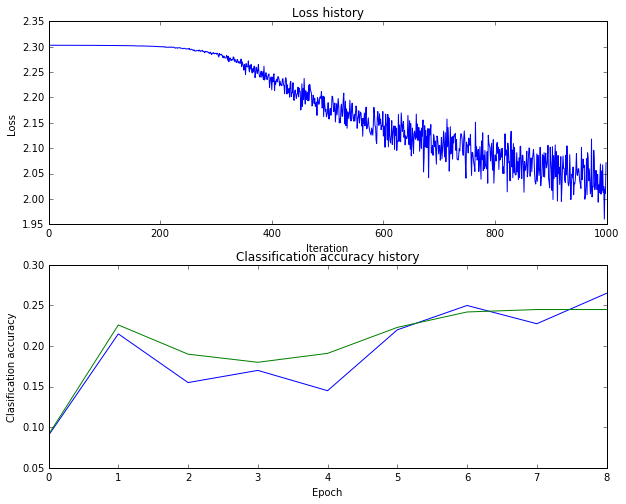
  
上图是learn\_rate对最终loss的影响，可以从图中看到低的learn\_rate曲线会趋近于线性（蓝线），而过高的learn\_rate曲线会趋近于指数（绿线）。  
高的学习率会使得曲线迅速下降，但是可能会在一个不好的结果下收敛。过高的学习率甚至会使得结果往差的方向发展（黄线）。比较合理的学习率应该体现出入红色线所示的形状。

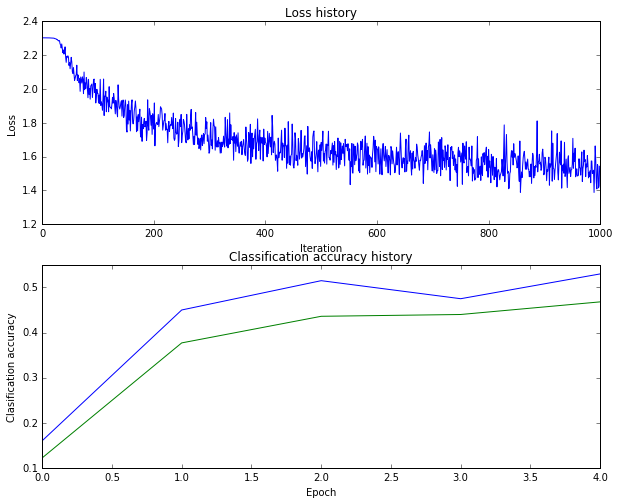


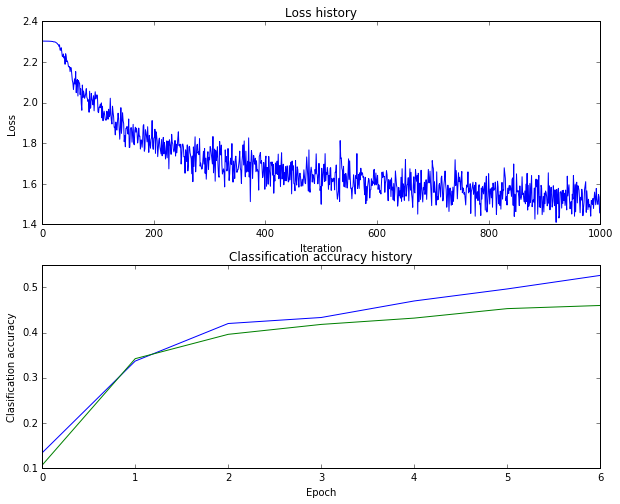
实际中学习曲线可能不是平滑的，而是如图上所示有很多的噪声，曲线的震幅反映了batch\_size选择的合理程度，一般batch\_size大的话曲线会变得更加平滑，也就是说训练的参数和实际的参数拟合度更高。  
极端情况是batch\_size的大小和dataset的大小一样，那么曲线很可能就是平滑的。

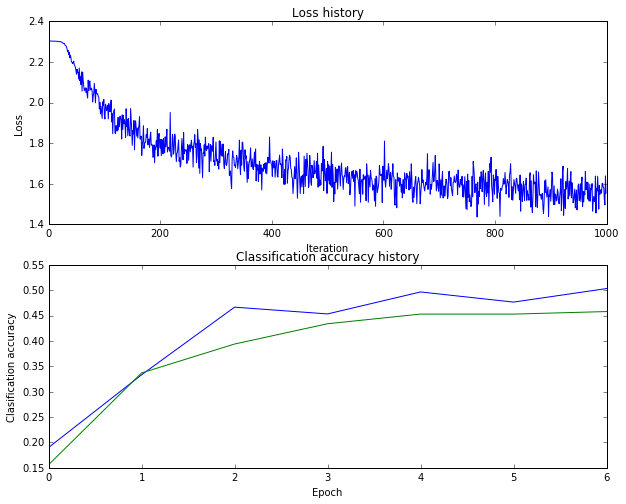
  
蓝色的曲线显示出验证准确率和训练准确率差别较大，说明可能出现过拟合现象，这种情况可以通过增加正则化项的惩罚系数，或者增加训练样本大小。  
绿色曲线显示出验证准确率和训练准确率差别较小，说明可能训练模型过小，可以适当增加网络的参数。

**实践部分**

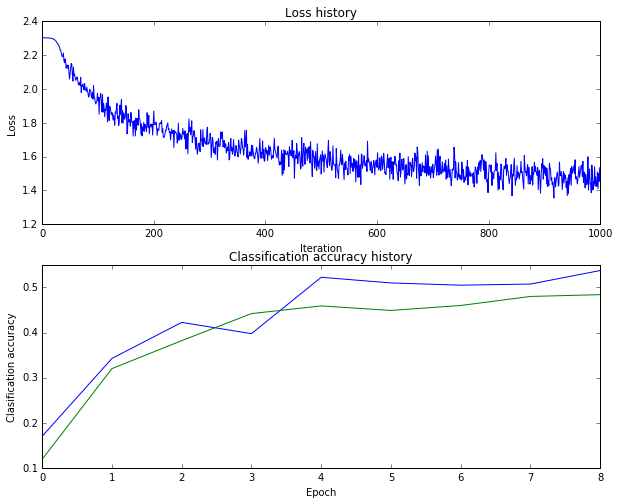
下面是根据以上的理论进行调参的过程。  
首先是原始的参数  
hidden\_size = 50 batch\_size=200 learning\_rate=1e-4 learning\_rate\_decay=0.95  
得到的曲线如下所示  


准确率为28%，从分析图像开始，上图的第二部分训练集合验证集曲线并没有一个合理的gap说明训练模型可能过小，所以我们需增加hidden\_size,或者增加batch\_size的大小。  
上图的第一部分的曲线形状也不太对，在最初的迭代过程中loss并没有下降，可能learn\_rate过低或者learning\_rate\_decay不明显。  
所以首先提高learn\_rate  
hidden\_size = 50 batch\_size=200 learning\_rate=1e-3 learning\_rate\_decay=0.9  
运行之后得到的结果是准确率到了47%  
  
这个修改出乎意料的有效，但是上图的曲线振幅较大，我们适当增加batch\_size来平滑准确率曲线。

hidden\_size = 50 batch\_size=300 learning\_rate=1e-3 learning\_rate\_decay=0.9  
运行之后得到的结果是46%  
  
出现了过拟合现象。而且图的上部分曲线后期震荡过大了，说明learning\_rate\_decay可能还是不够明显

hidden\_size =100 batch\_size=300 learning\_rate=1e-3 learning\_rate\_decay=0.8  
识别率依然是46%  
  
结果不明显，但是过拟合现象有所缓解。

最后

hidden\_size =100 batch\_size=400 learning\_rate=5e-4 learning\_rate\_decay=0.95  
识别率是48%  
  
应该还是有点过拟合