

作业1

本次实验利用numpy实现4层ann，并在MNIST数据集上测试效果。

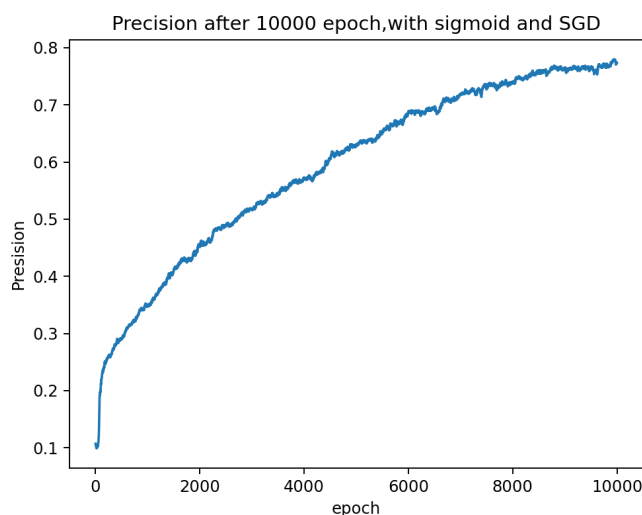
主要工作

首先尝试不同优化方法（SGD，ADAM，GD）的效果。接着尝试不同的激活函数（sigmoid，tanh和relu系），并用precision-epoch的曲线可视化展示结果。

SGD

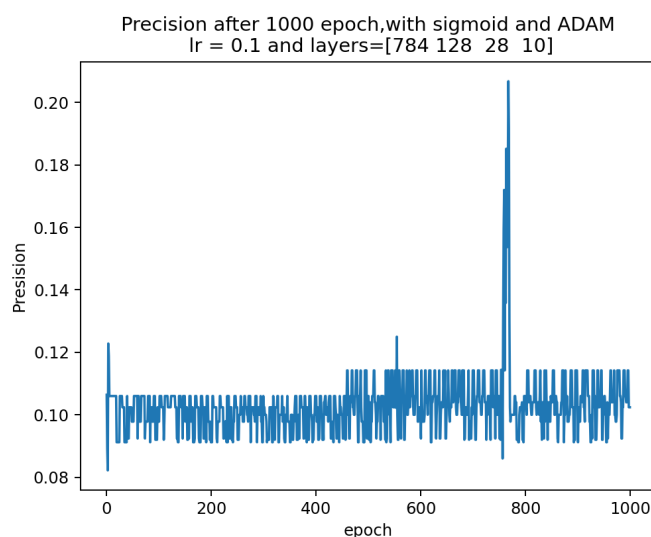
下面是SGD经过10000个epoch的准确率--训练轮数曲线。

发现SGD收敛较慢，且稳定性不好，换用别的激活函数后SGD的效果相差比较大。猜测可能是SGD 更加简单，只采用了一个与参数更新相关的常数，因此它在处理复杂模型时可能会比较慢。



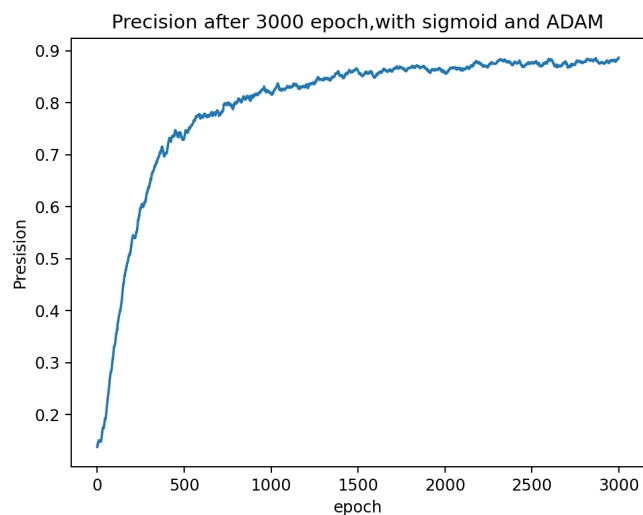
ADAM

在跑ADAM的时候，我先是遇到了如下图的问题：准确度完全没有上升，而是在一个很低水平震荡。



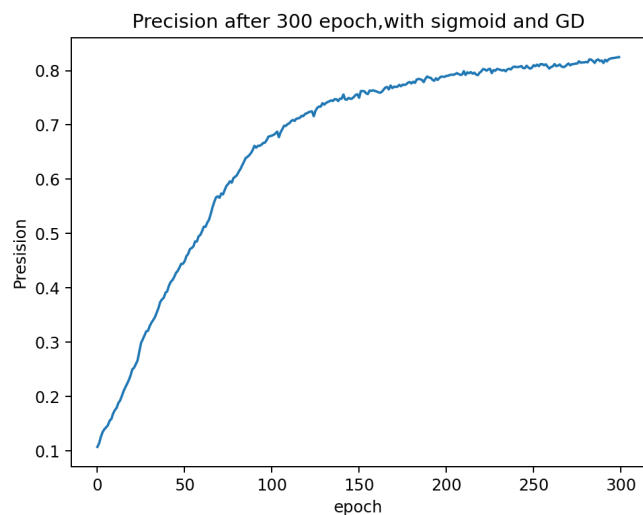
在思考后，我确定代码写的没有问题，经过调试后发现是**学习率太大**的原因，在上面的SGD中取了lr=0.1，而用ADAM至少需要0.05以下的学习率才能得到好的结果。

下面是用ADAM, $lr=0.05$ 得到的结果, 可以看到: 相比SGD更快, 且最终效果更好。



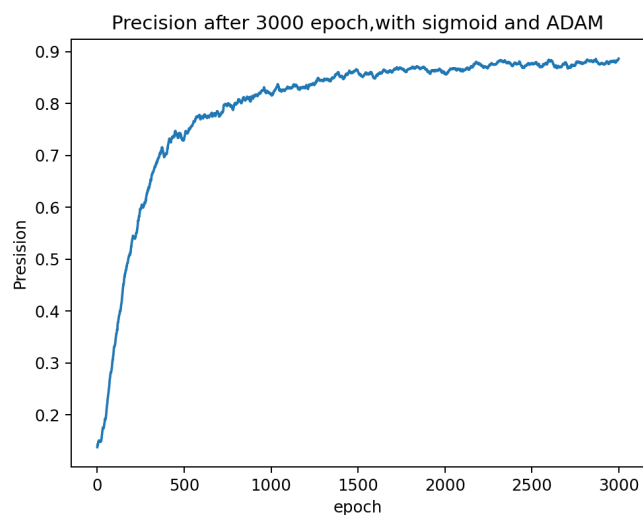
GD

GD一开始也遇到了学习率太大的问题, 通过测试取 $lr=0.001$ 效果很棒。我只训练了300轮, 就达到了SGD10000轮左右的正确率。

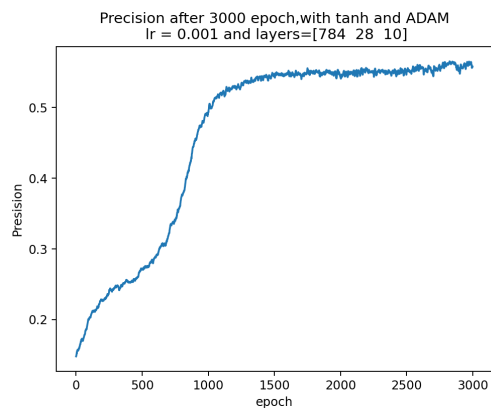
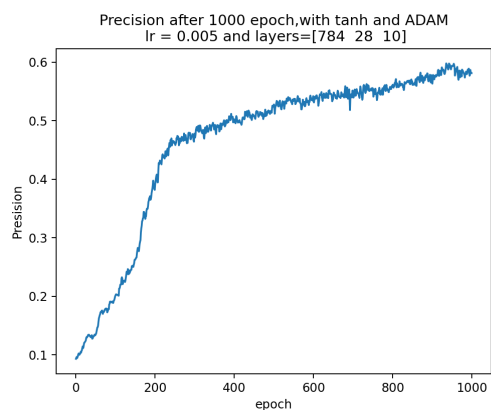


接下来尝试不同的激活函数:

sigmoid: 表现较好

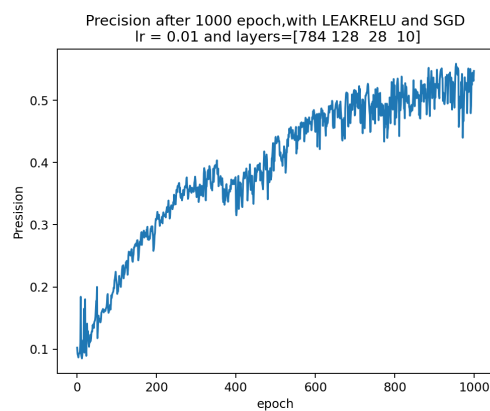
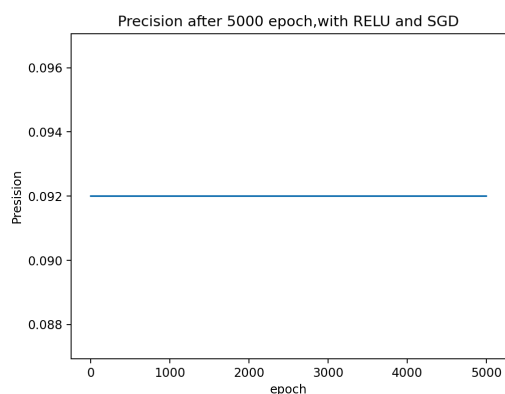


tanh: 效果较差



RELU&LEAKRELU:

效果不好，经常出现卡在一个值不动的情况（如下图1），通过调参后能又提升，但依然震动较大且效果不好（如下图2）。



遇到的问题&思考

1. ADAM和GD下loss完全不下降 \Rightarrow 调低学习率解决

2. RELU卡在一个值不动:

查阅资料发现: relu 函数在 0 附近的表現不稳定，容易出现梯度消失或梯度爆炸的问题，这可能会导致神经网络的训练困难或无法收敛。

尝试方法: 尝试了LeakRELU, 表现略有改进。另外有更多的函数 (如SELU())发明来解决这个问题。