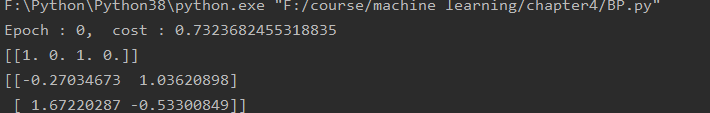
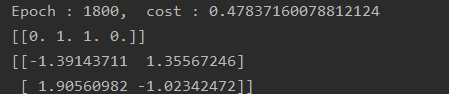
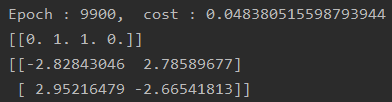
第一次训练时的损失，预测标签以及输入层到隐含层的权重参数如下



到第1800次训练时，此时预测标签已经正确

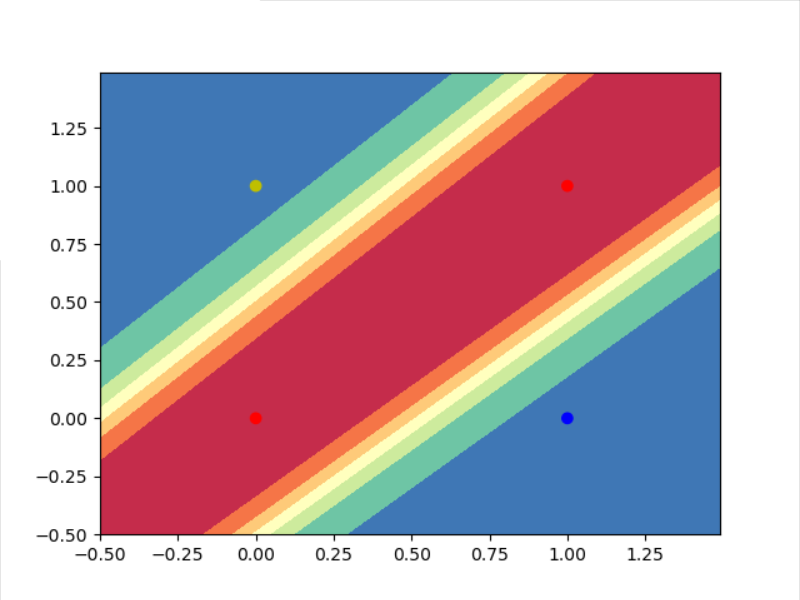


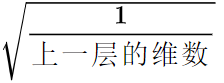
最后损失可以降到0.05以下



此后损失和权重基本变化不大

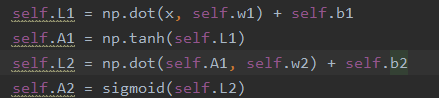
可视化结果如下：



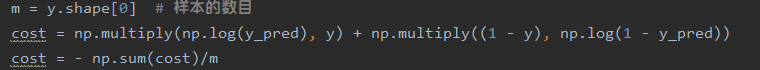
初始化参数时采用的是Xavier Initialization，即在随机化权重之后乘

因为还是随机化权重参数，所以每次运行结果会有差异

前向传播中第一层和第二层的激活函数分别用的是tanh和sigmoid函数



损失函数使用交叉熵损失函数



也可以搭建pytorch神经网络模型来实现

