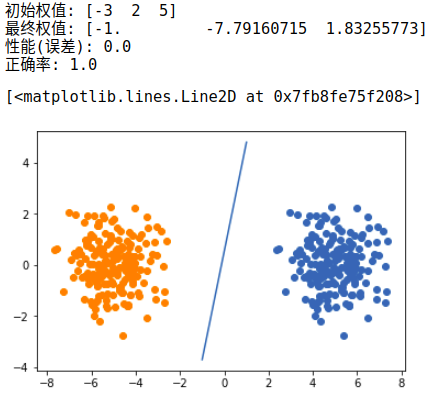
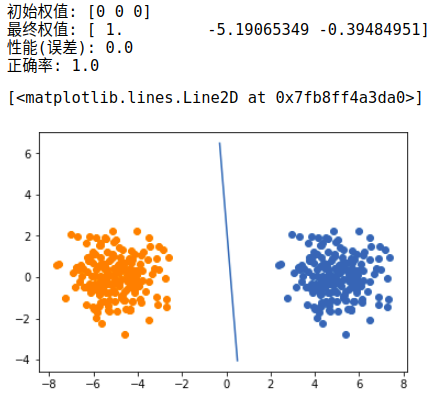
PLA算法编程作业实验报告

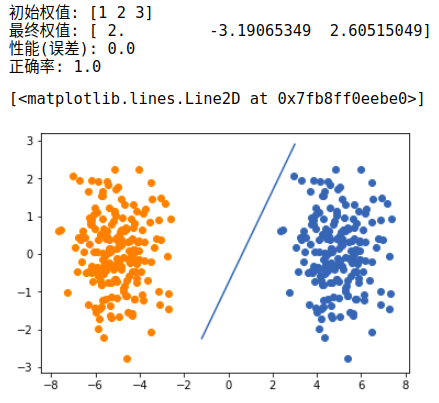
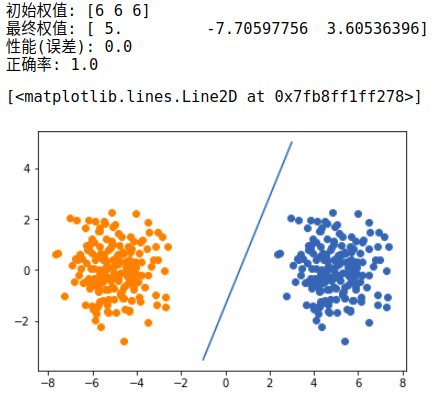
曾文正 U201715853 自动化校际交流1701

一、实验结果

1.生成的数据集为均值为[-5,0],[5,0]的向量各200个。

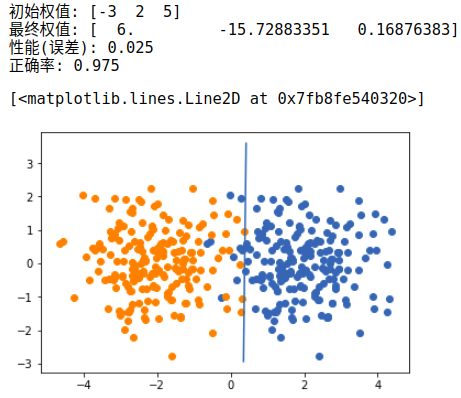
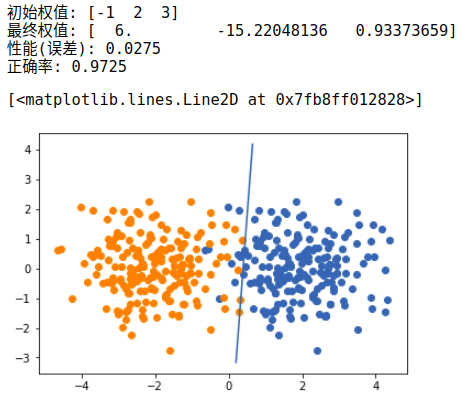
分别设置初始权值为[-3,2,5],[0,0,0],[6,6,6],[1,2,3],最终结果正确率均为100%

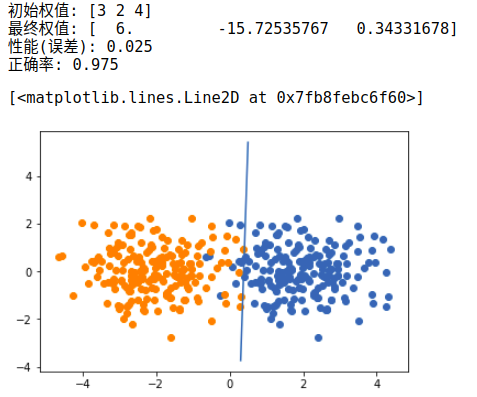
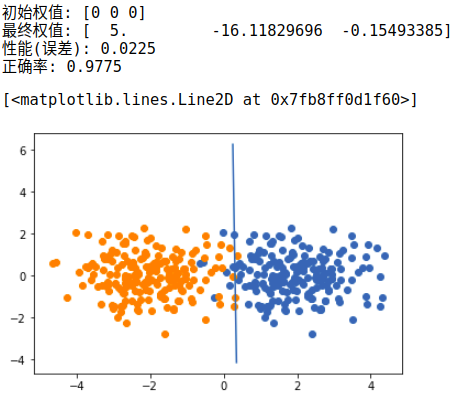




2.生成的数据集为均值为[-2,0],[2,0]的向量各200个。

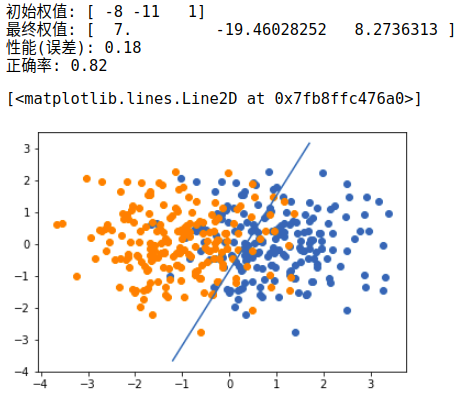
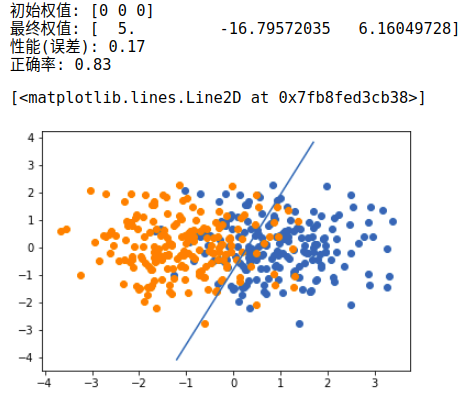
分别设置初始权值为[-1,2,3],[-3,2,5],[0,0,0],[3,2,4]，最终结果正确率略有差异，但都在97.5%左右。

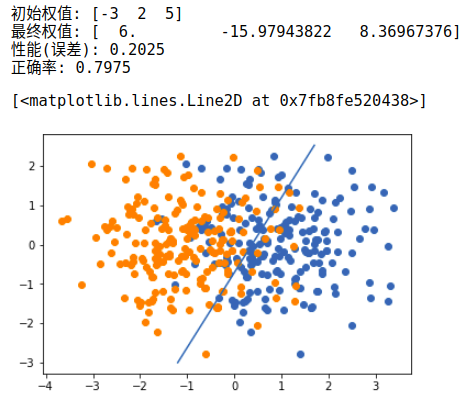
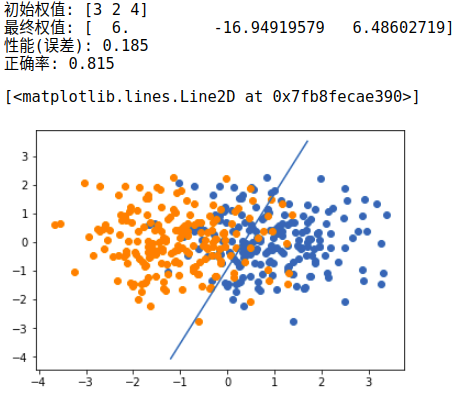




3.生成的数据集为均值为[-1,0],[1,0]的向量各200个。

分别设置初始权值为[0,0,0],[-8,-11,1],[3,2,4],[-3,2,5]，最终结果正确率略有差异，但都在79%-83%之间。





二、试验结果分析与总结

1.改变初始权值会对分类结果产生一定影响，如何选择初始权值仍然是这个算法做不到的。我开始自己编写不同权值，发现并不会对结果有太大影响（可能是因为我打乱了样本顺序，如果不打乱依然会有较大波动），只是在一定范围内波动，后来我还利用np.random.rand()函数随机产生初始权值，对结果的影响仍然不大。

2.样本送入算法的顺序会影响最终分类结果，尤其体现在线性不可分的情况。我将生成的样本顺序打乱，发现这样在样本均值为[1,0],[-1,0]时分类效果好了很多。对于不打乱样本顺序时，我得到的正确率甚至会不足50%，且对于不同权值，不同迭代次数，结果的波动很大。当我打乱样本顺序后，正确率得到明显提高，并且迭代次数和初始权值并不会对结果产生太大的影响。但如何选择顺序最好，这个算法并不能解决这个问题。

3.迭代次数也会影响最终分类结果。开始我没打乱样本顺序时影响很大，打乱后在一定次数后影响较小。只能说对于线性可分样本，会趋于正确分类。对于线性不可分，结果难以预测。

**对于PLA实现方法的总结**：

1. 将w0和入w，使得X，w变为增广矩阵。
2. 根据判断条件（内积）来决定是否对现有w进行调整（向量的合成）。
3. 将得到的权值（3维）转化为分界线（2维），权值是分界线的法向量。
4. 打乱样本顺序会得到更好的结果。