Lecture 5编程作业

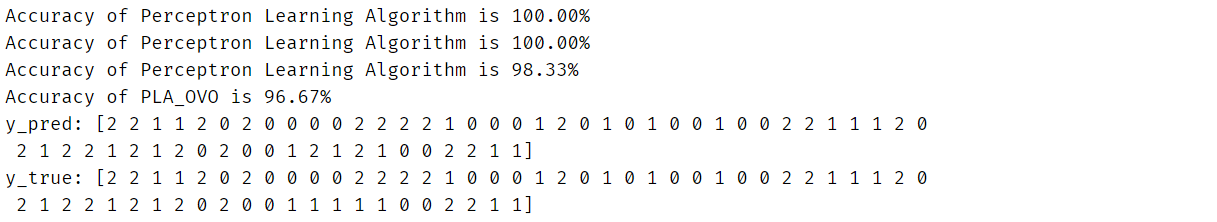
李星毅 U201712072 自实1701

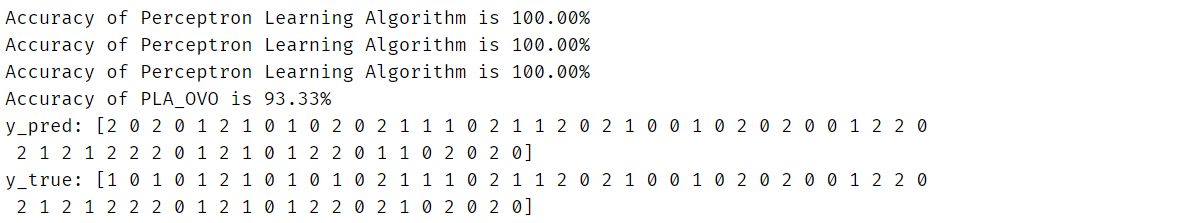
给定IRIS数据集做如下题目实验，该数据集有三类目标，每个类别有50个样本，每个样本有四维特征。实验时每个类别随机选30个样本进行训练，另外20个样本用于测试。

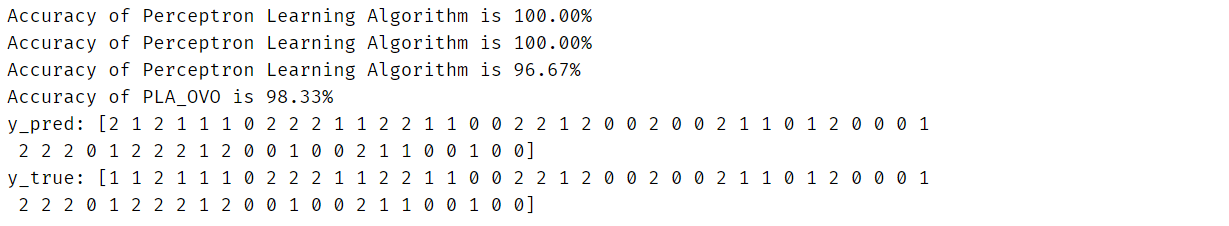
1，以感知器算法为基础分类算法，编写一个OVO多类分类器算法，对上述数据集进行实验，分析结果。

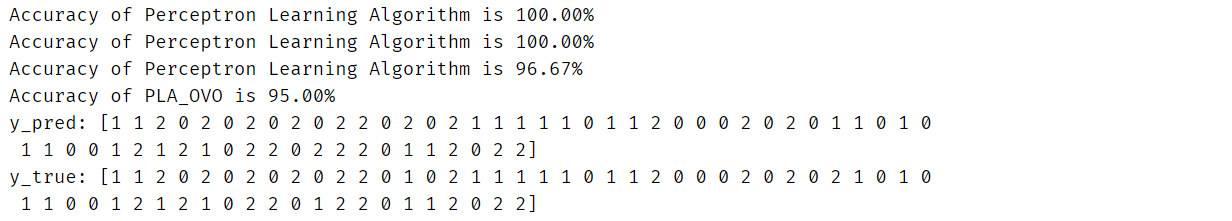
解：记setosa为第0类，versicolor为第1类，virginica为第2类。

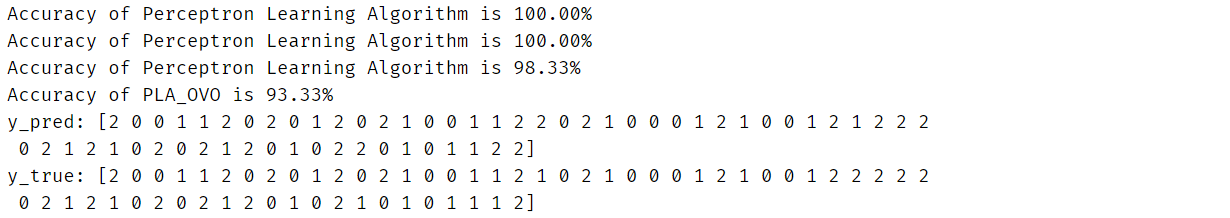
（1）结果展示：其中从上到下依次是第0、1类，第0、2类，第1、2类和OVO三分类感知器的准确率，y\_pred是OVO三分类感知器的预测结果，y\_true是正确答案。











（2）总结：

I. 数据预处理：对数据进行标准化，这一步可充分发挥算法性能，提高精度。在对训练集标准化的同时，使用训练集的均值和方差对测试集进行标准化。需要指出的是，这里进行的标准化，是分别对每一个分类器所使用的训练集和测试集进行标准化，而不是对整个iris数据集直接进行标准化。

II．改进感知器算法：不一定像原始感知器一样设置学习率为1，可以设置为其他合理的学习率。

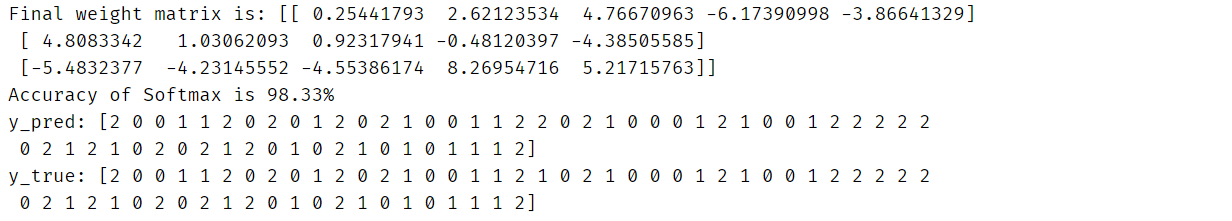
III．改进感知器算法：打乱所有错误点，随机选择错误点对权重进行修正。

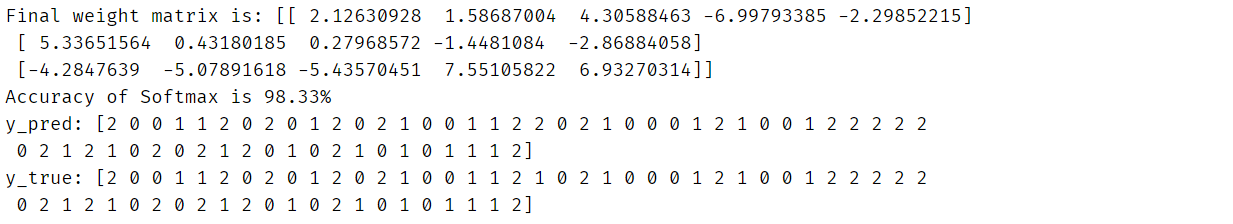
IV．结果分析：可以看到，第0、1类，第0、2类的分类准确率极高，可轻易分开，第1、2类重叠的程度更高，因此不是很容易分开，而OVO三分类感知器的准确率与三个单独的感知器的准确率有关，但不是完全成正比，也就是说泛化能力也很重要；PLA OVO算法不稳定，准确率的高低跟初始权重等许多因素有关。

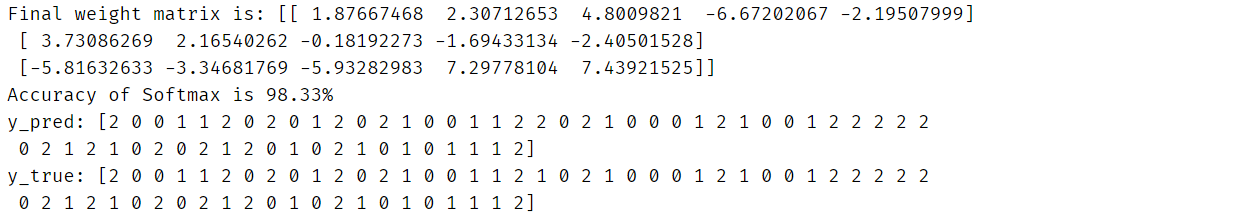
2，编写Softmax算法实现多类别分类，对上述数据集进行实验，分析结果。

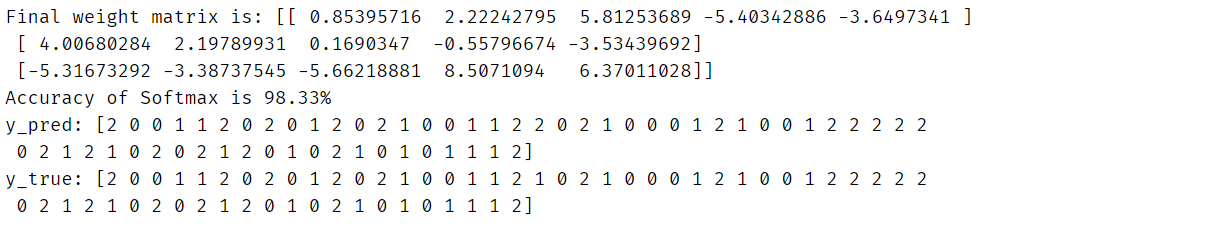
解：

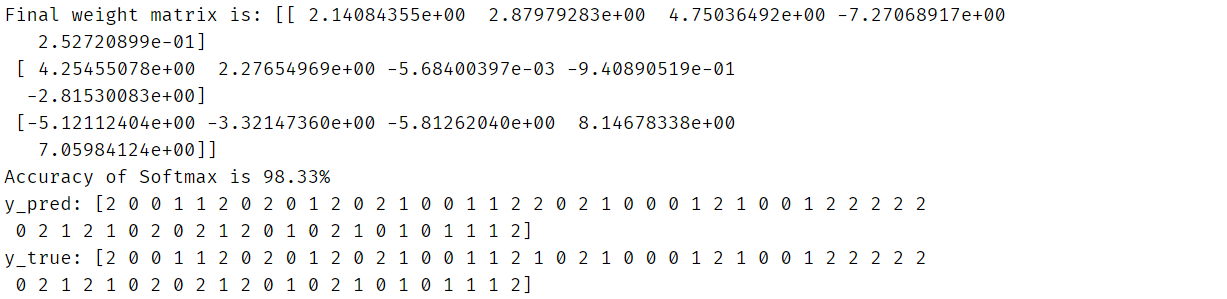
1. 结果展示：











1. 总结：

I. 数据预处理：对数据进行标准化，这一步可充分发挥算法性能，提高精度。在对训练集标准化的同时，使用训练集的均值和方差对测试集进行标准化。需要指出的是，这里进行的标准化，对整个iris数据集直接进行标准化。

II．结果分析：可以看到，Softmax算法是比PLA OVO要稳定很多的，并且预测精度也非常高；Softmax Regression算法是Logistic Regression算法的推广，用于多分类问题。与OVO策略不同的是，Softmax Regression是直接进行执行多分类。而OVO是把多个二分类器组合到一起完成多分类任务。