计算机与信息工程学院实验报告（七）

姓名：王赫 学号：1828070097 专业：数据科学与大数据技术(明德计划) 年级：2018级

课程：机器学习与数据挖掘 主讲教师：罗慧敏 辅导教师：\_\_\_\_\_\_\_

实验时间：2020年12月 9日 下午17时至18时，实验地点：606

实验题目： 编程实现AdaBoost算法

实验目的： 掌握AdaBoost算法的求解过程

实验环境（硬件和软件） Anaconda/Jupyter notebook/Pycharm

实验内容：

编码实现AdaBoost算法，以决策树为基学习器，训练一个AdaBoost集成模型，对测试样本进行判别。

实验数据记录：

**一、**已经给定部分代码，补充完整的代码，需要补充代码的地方已经用红色字体标注，包括：

**（1）#补充定义弱分类器的代码**

**（2）#补充定义、训练AdaBoost分类器的代码**

**（3）#补充评估模型在测试集上的精度的代码**

**二、**将补充完整的代码提交，并提交实验结果；（**也可以自己重写这部分的代码提交**）

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import AdaBoostClassifier

from sklearn.datasets import load\_iris

from sklearn.metrics import accuracy\_score

from sklearn.model\_selection import train\_test\_split

import numpy as np

data = np.loadtxt('data/wine.data',delimiter=',')

X = data[:,1:]

y = data[:,0:1]

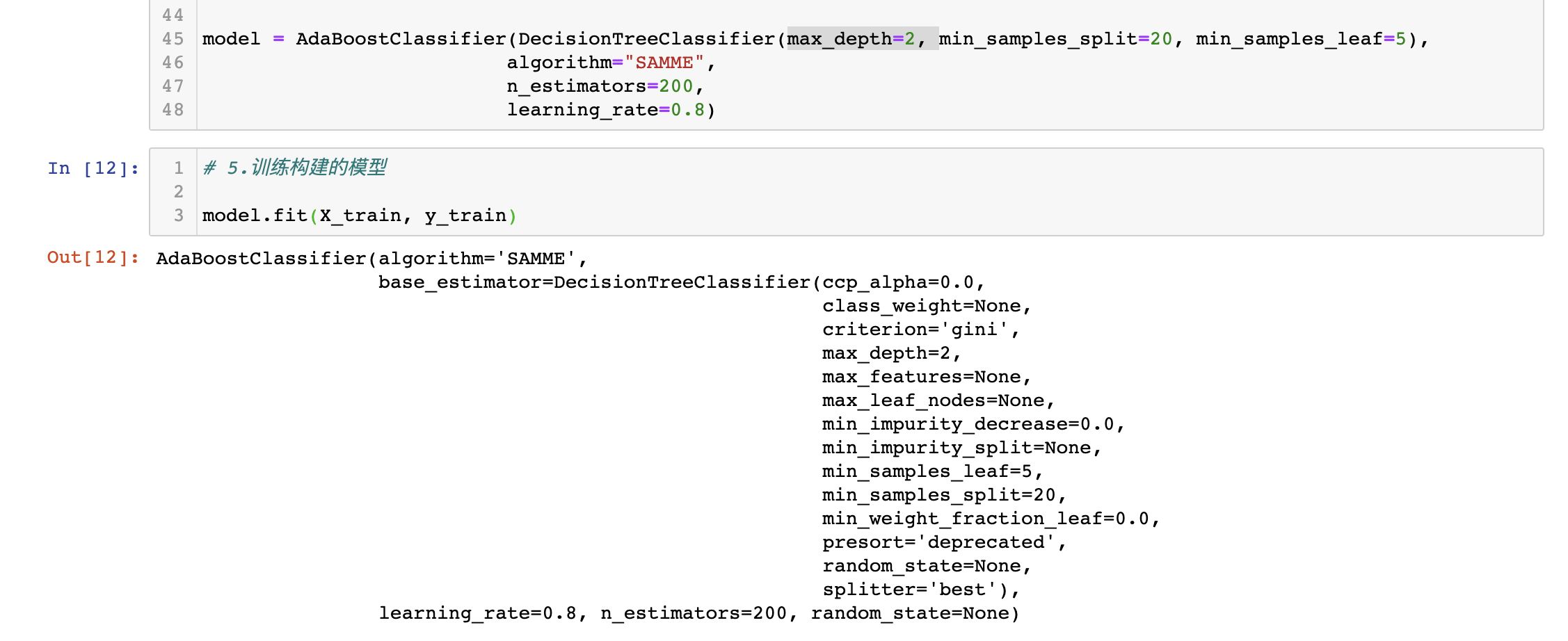
X\_train,X\_test,y\_train,y\_test = train\_test\_split(X,y.ravel(),train\_size=0.8,random\_state=0)

**#补充定义弱分类器的代码**

**model = AdaBoostClassifier(DecisionTreeClassifier(max\_depth=2, min\_samples\_split=20, min\_samples\_leaf=5), algorithm="SAMME", n\_estimators=200, learning\_rate=0.8)**

**#补充定义、训练AdaBoost分类器的代码**

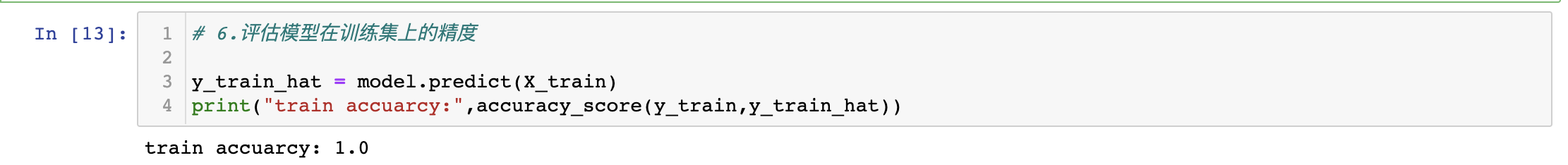
**model.fit(X\_train, y\_train)**

****

#评估模型在训练集上的精度

y\_train\_hat = model.predict(X\_train)

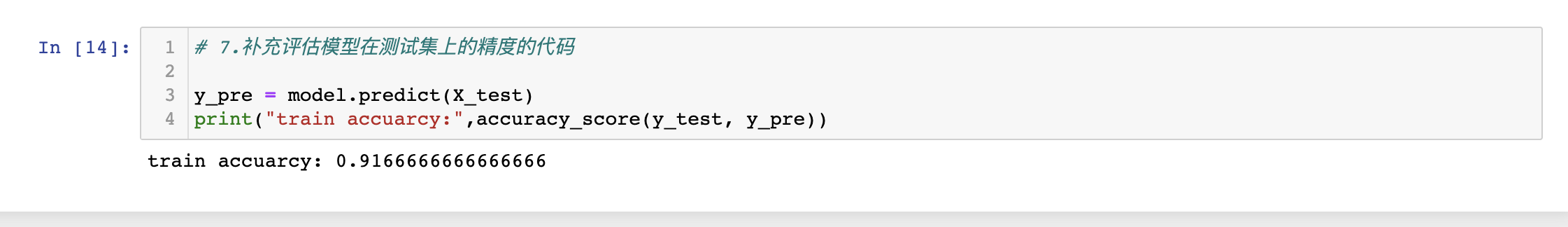
print("train accuarcy:",accuracy\_score(y\_train,y\_train\_hat))



**#补充评估模型在测试集上的精度的代码**

**y\_pre = model.predict(X\_test)**

**print("train accuarcy:",accuracy\_score(y\_test, y\_pre))**



问题讨论：

在进行本次实验时，我们需要注意AdaBoostClassfier弱分类器的构造方式，其中就有base\_estimatorobject基学习器参数，这里我们采用的基学习器是决策树模型。

但是这里就会遇到一些小问题，那就是参数的选择。在构建决策树需要如何选择合适的参数？在构建分类器需要如何选择合适的参数？

这里可以利用前面讲的交叉验证的方法选择合适的参数。