计算机与信息工程学院实验报告（一）

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：\_\_\_\_\_\_\_

实验题目： 作业1

实验目的： 对数几率回归西瓜数据集3.0

10折交叉验证法和留一法估计出对率回归的错误率

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

实验内容：

课后习题：

1、对数几率回归西瓜数据集3.0

2、10折交叉验证法和留一法估计出对率回归的错误率

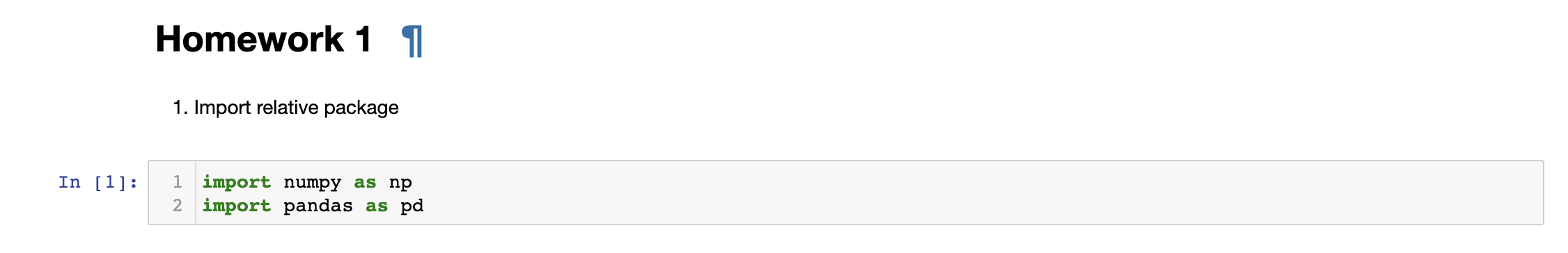
实验步骤：

1. 数学公式推导
2. 编程实现
3. 数据记录
4. 实验总结

实验数据记录：

题目1:

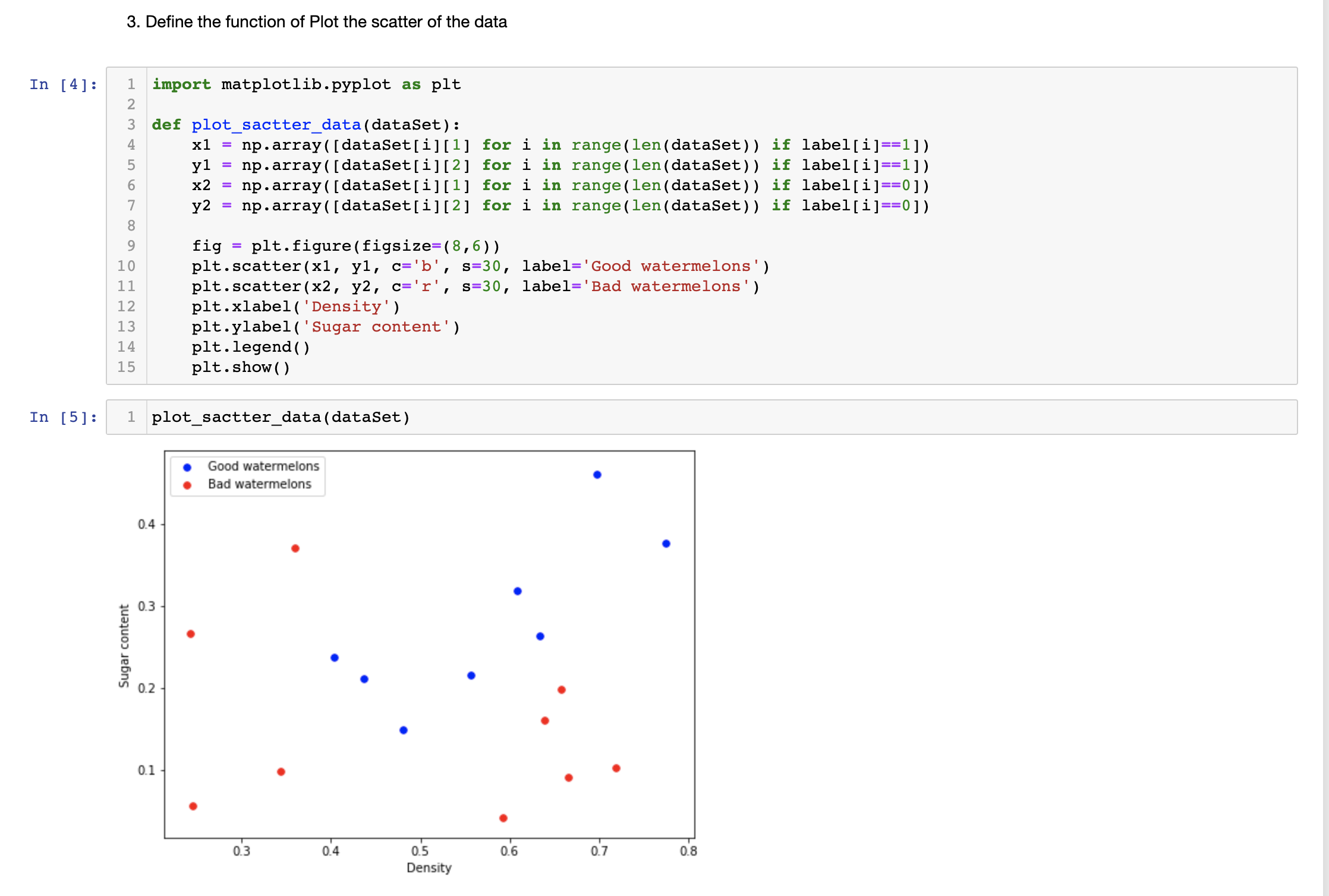
1. 导入依赖



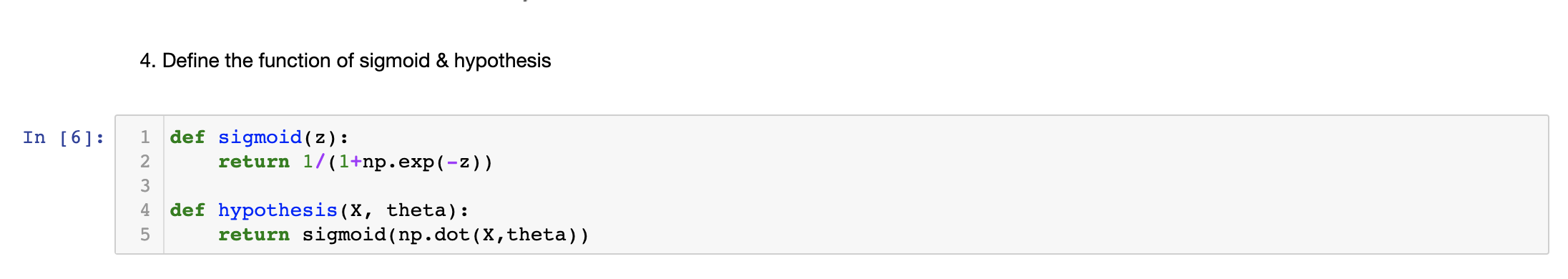
1. 西瓜数据集获取函数



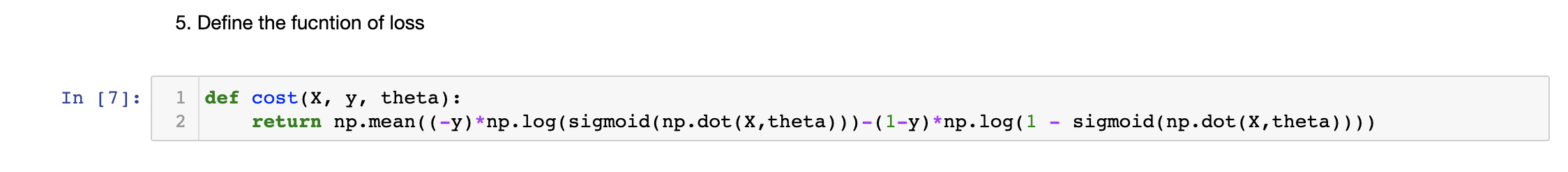
1. 绘制散点图



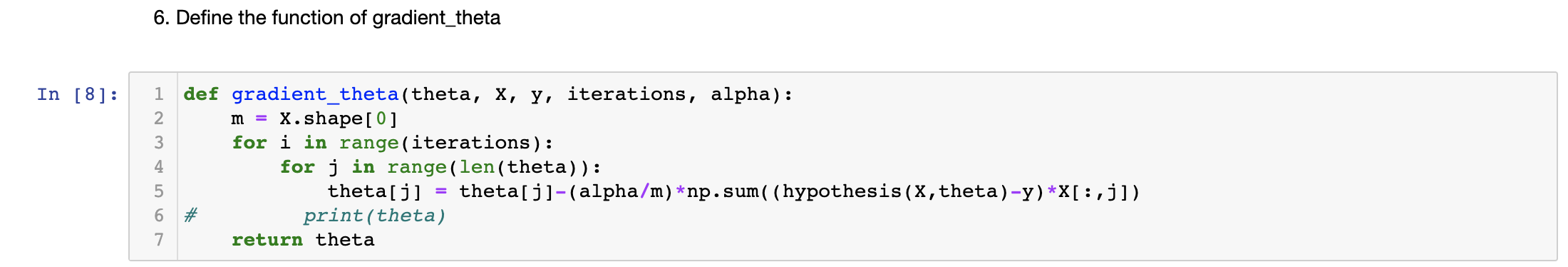
1. 定义sigmoid函数及其导数



1. 定义损失函数



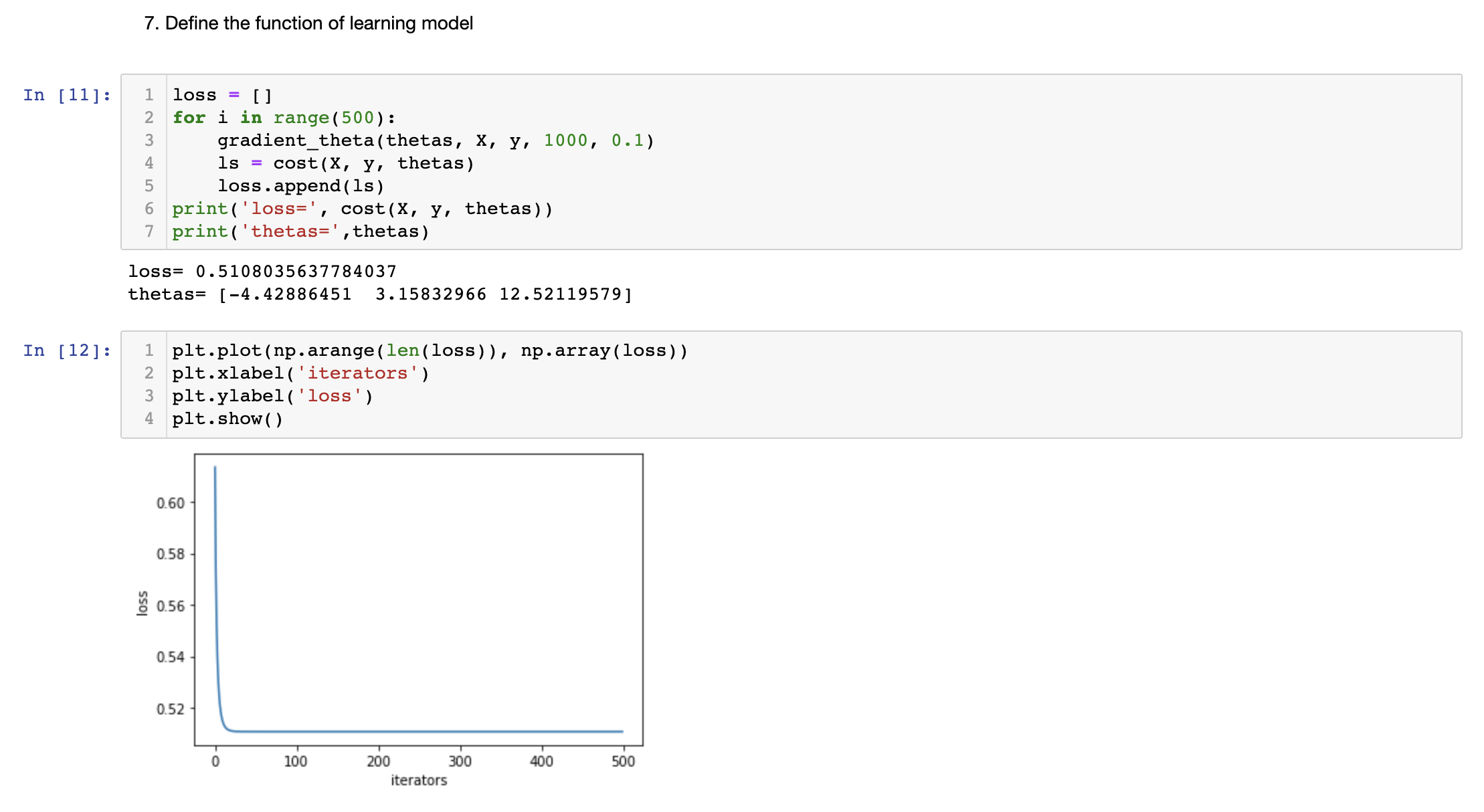
1. 定义梯度下降算法计算函数



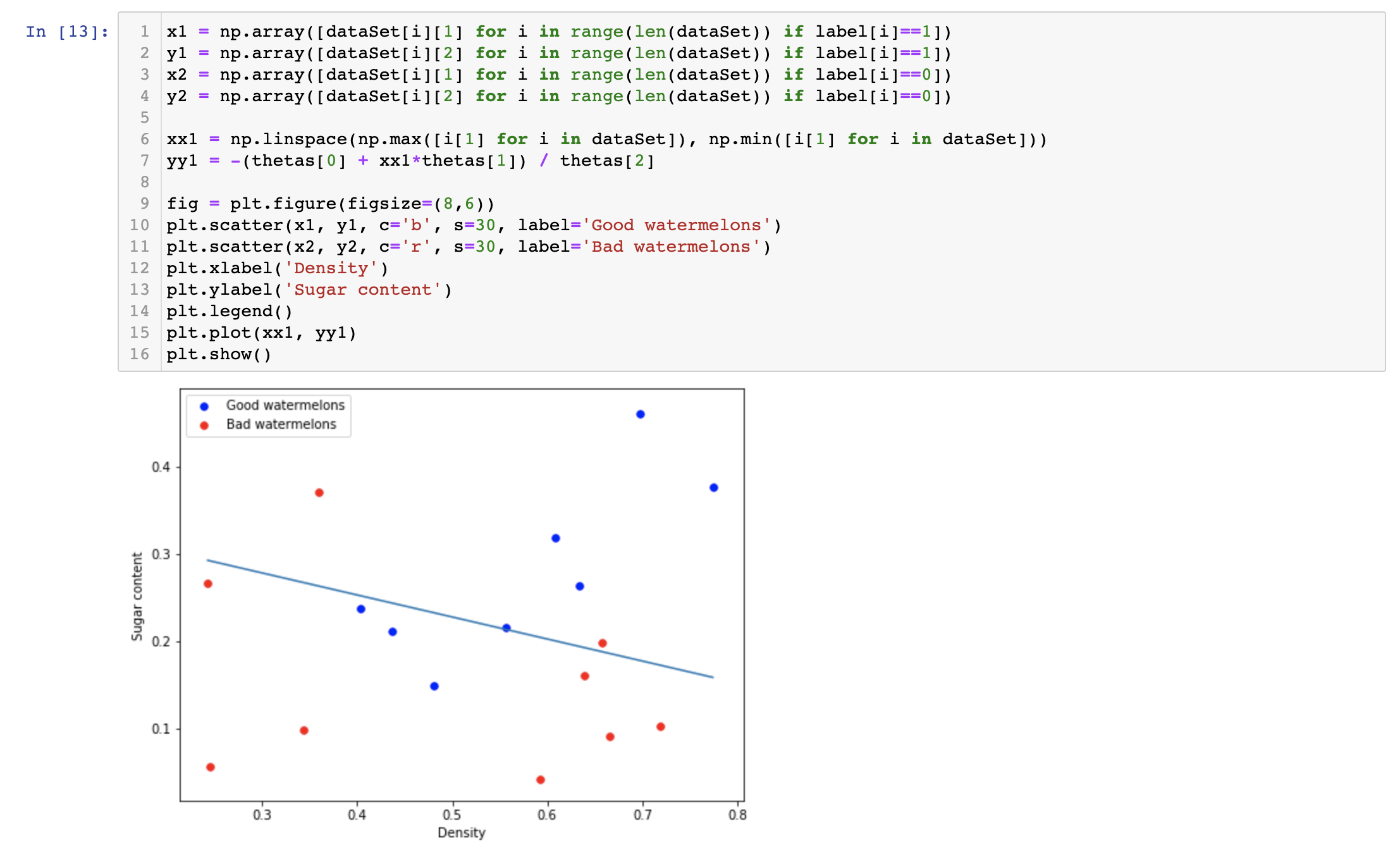
1. 初始化变量



1. 计算损失函数并绘制损失函数变化图



1. 绘制分类曲线图



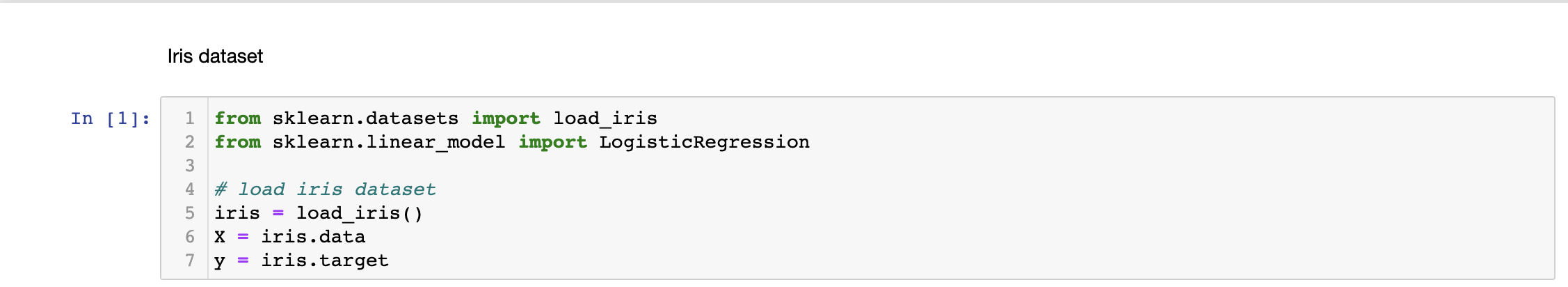
1. 计算准确率



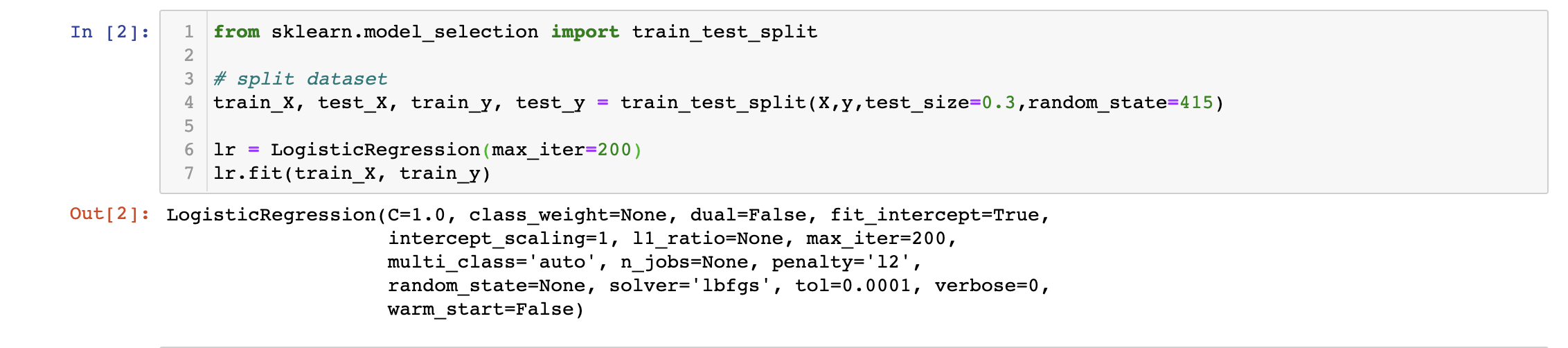
题目2:

选用UCI中的iris和wine数据集进行10折交叉验证法和留一法估计出对率回归的错误率的实验：

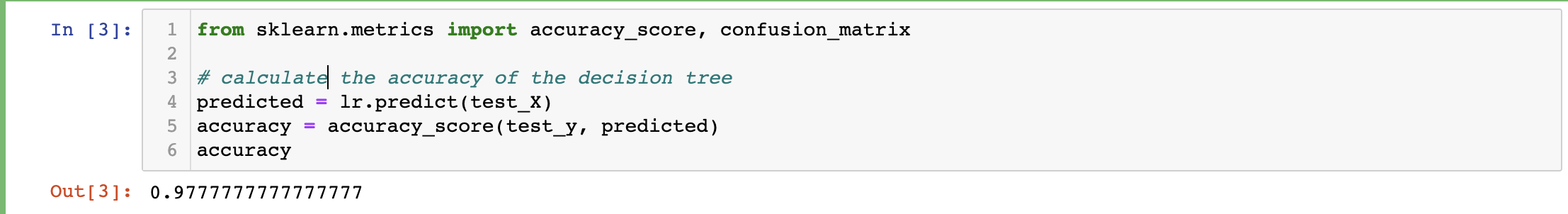
1. iris dataset
   1. load dataset



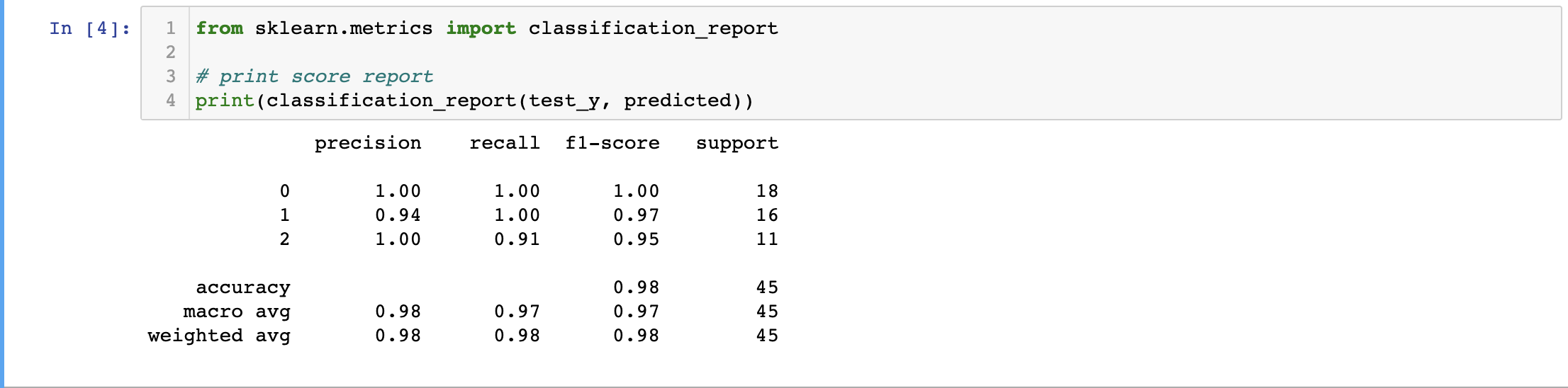
* 1. split dataset and create model to fit



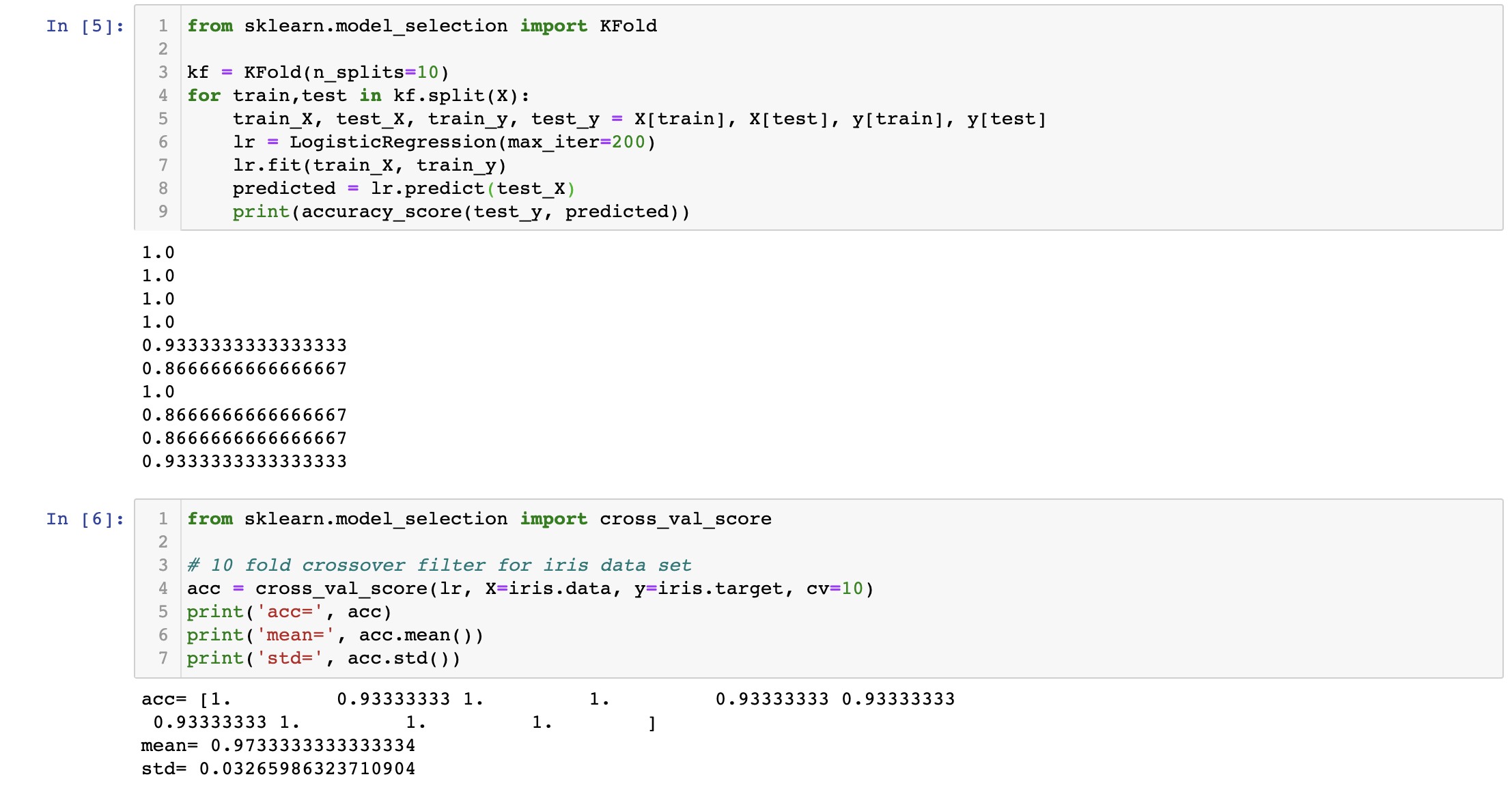
* 1. calculate the accuracy of Linear model



* 1. print score report



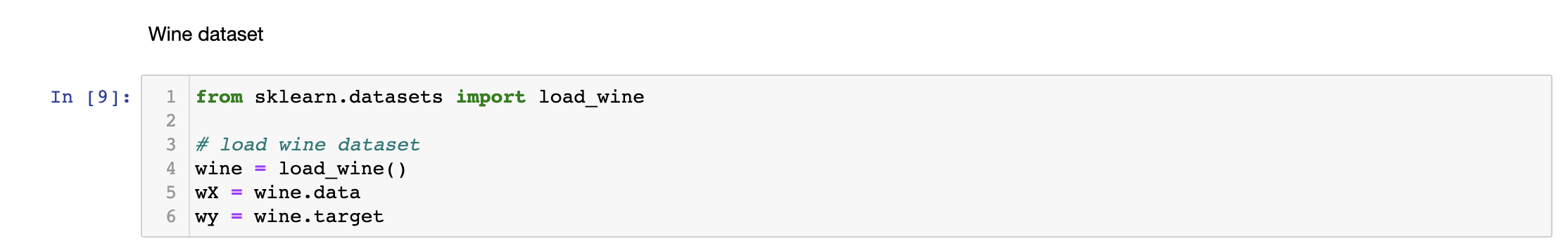
* 1. 10 fold crossover filter for iris data set



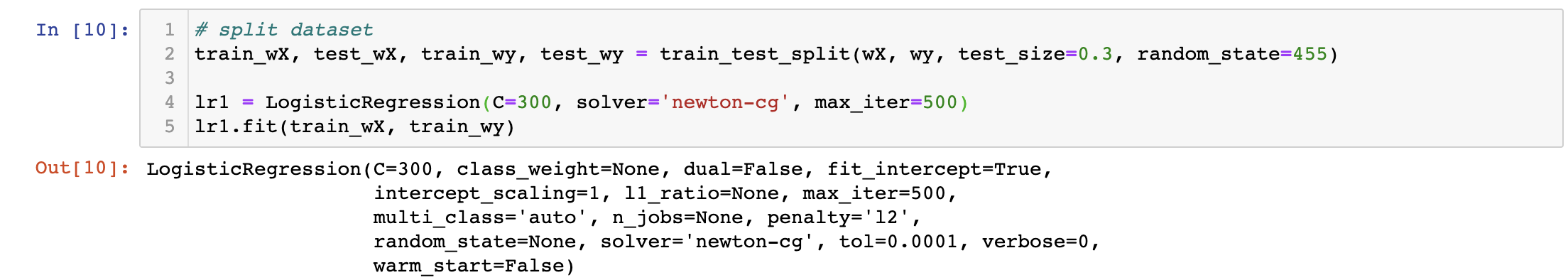
* 1. leave one method filter



1. wine dataset
   1. load dataset



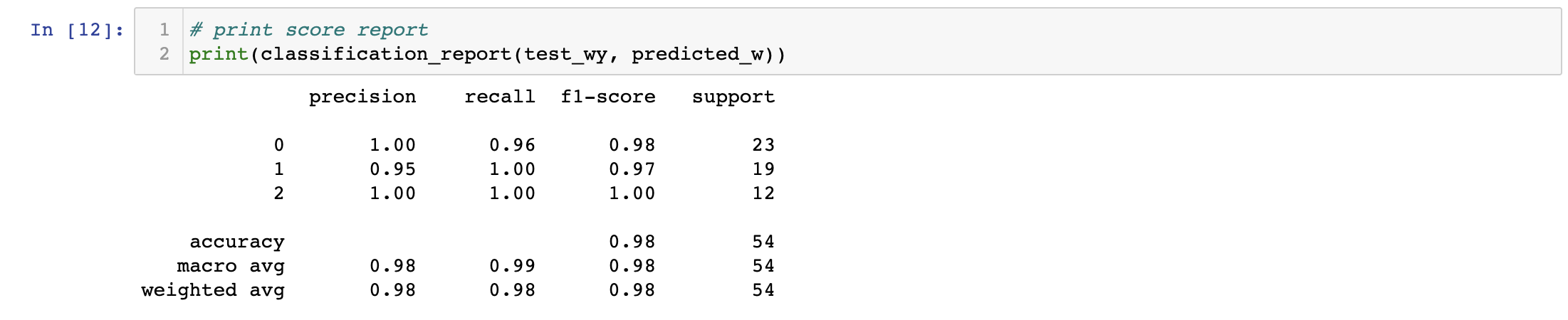
* 1. split dataset and create model to fit



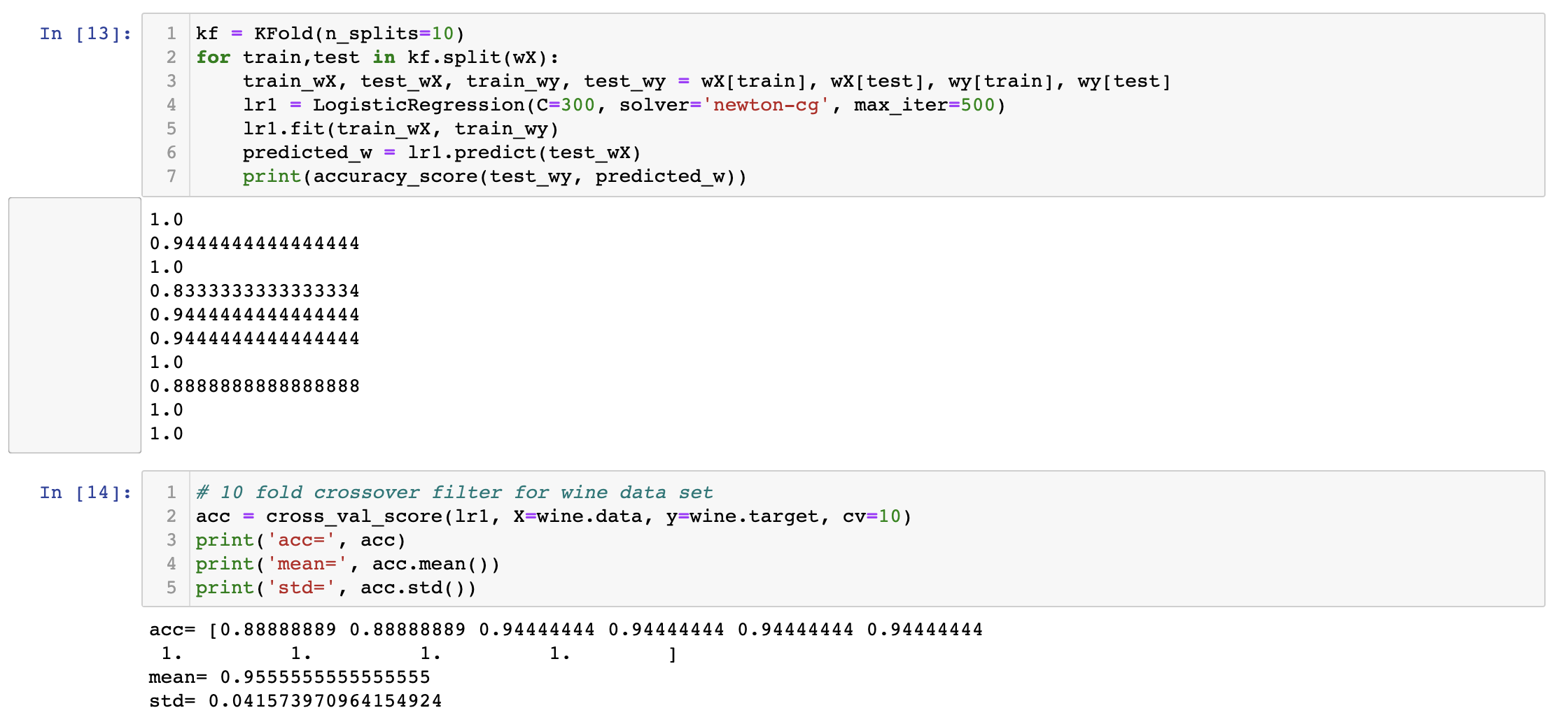
* 1. calculate the accuracy of Linear model



* 1. print score report



* 1. 10 fold crossover filter for iris data set



* 1. leave one method filter

