**作业文档命名方式：学号\_姓名\_课堂作业5**

1. **实验目的**
2. 掌握机器学习建模分析
3. 掌握回归分析、分类分析、聚类分析、降维等
4. 了解各分类器之间的差异
5. **实验环境**

操作系统：Windows 11

应用软件：Python 3.10.8, Anaconda Jupyter Notebook

1. **实验内容与结果**

（题目、源程序、运行结果）

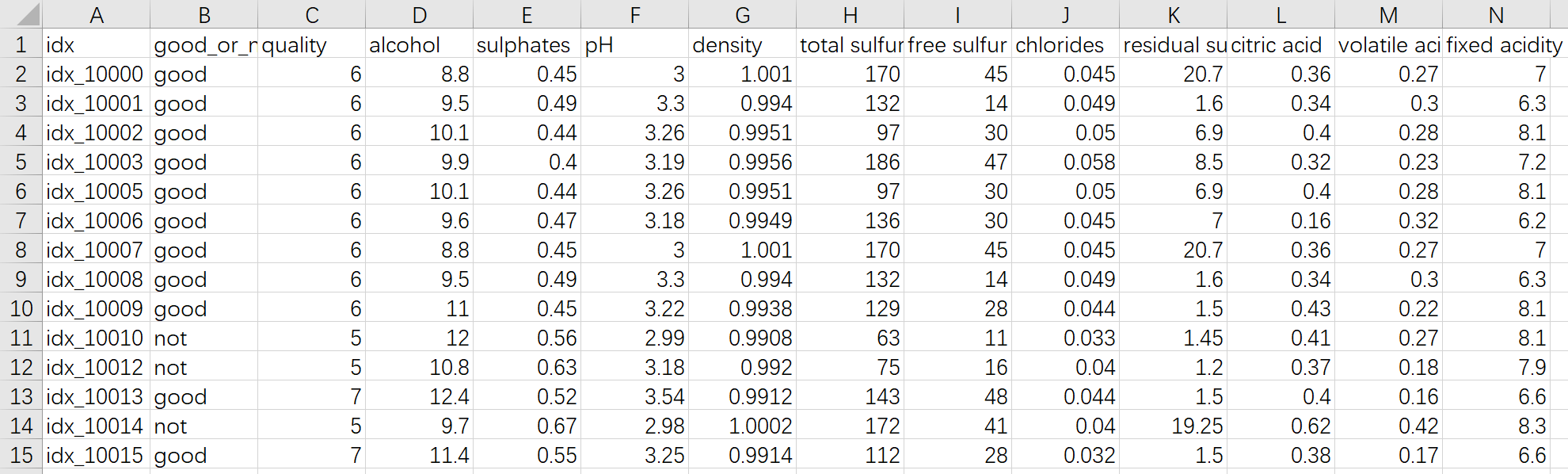
1.使用scikit-learn建立SVM模型为葡萄酒数据集构造分类器（分类结果为’good’或‘not’ ） (40分)[“不可使用quantity”列]

2.评估分类器在此数据集上的分类性能(20分)

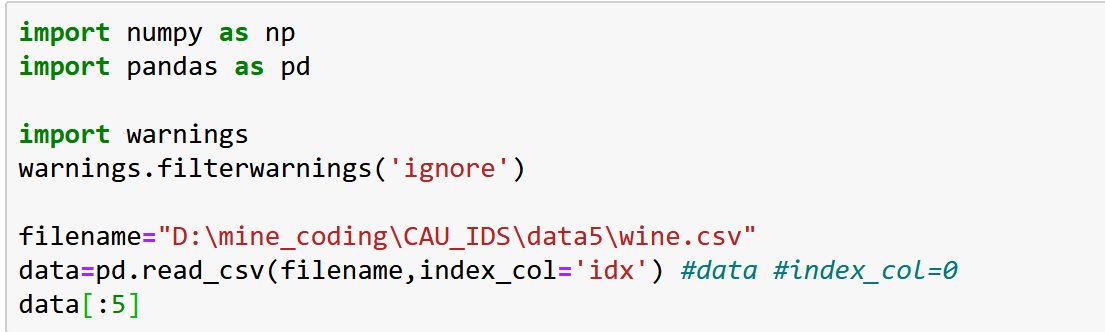
使用SVM多分类的时候需要用到Kappa系数

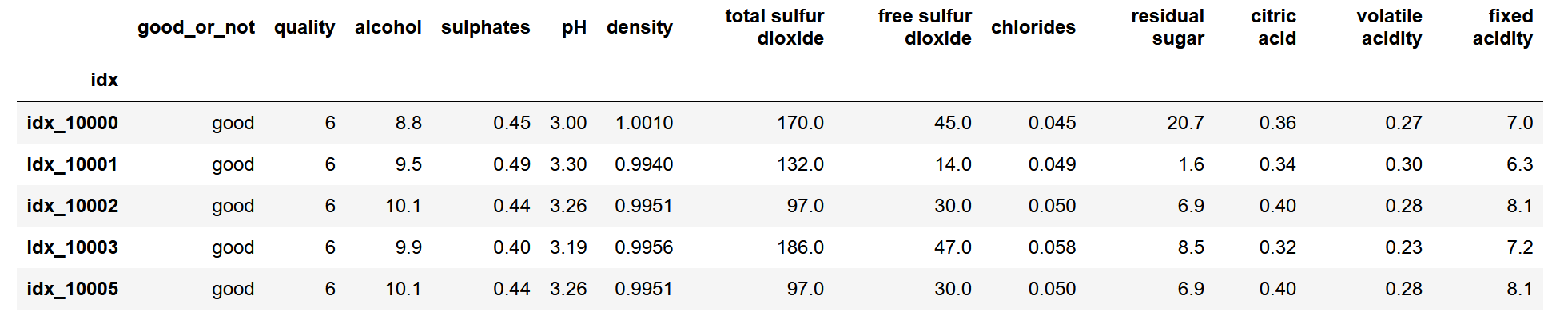
\* 需要划分训练集和测试集

算法：支持向量机（SVM， Support Vector Machine）



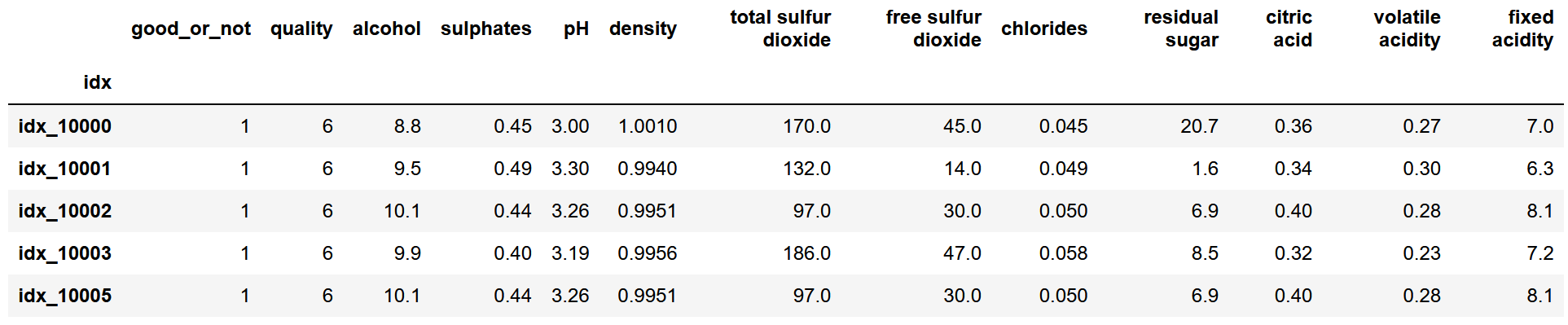
读入wine.csv数据：

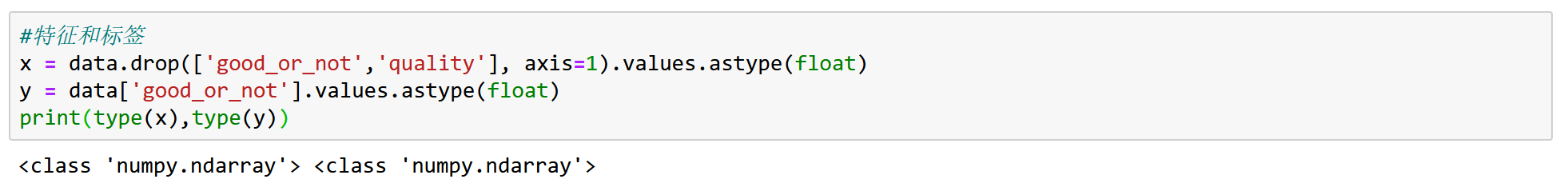




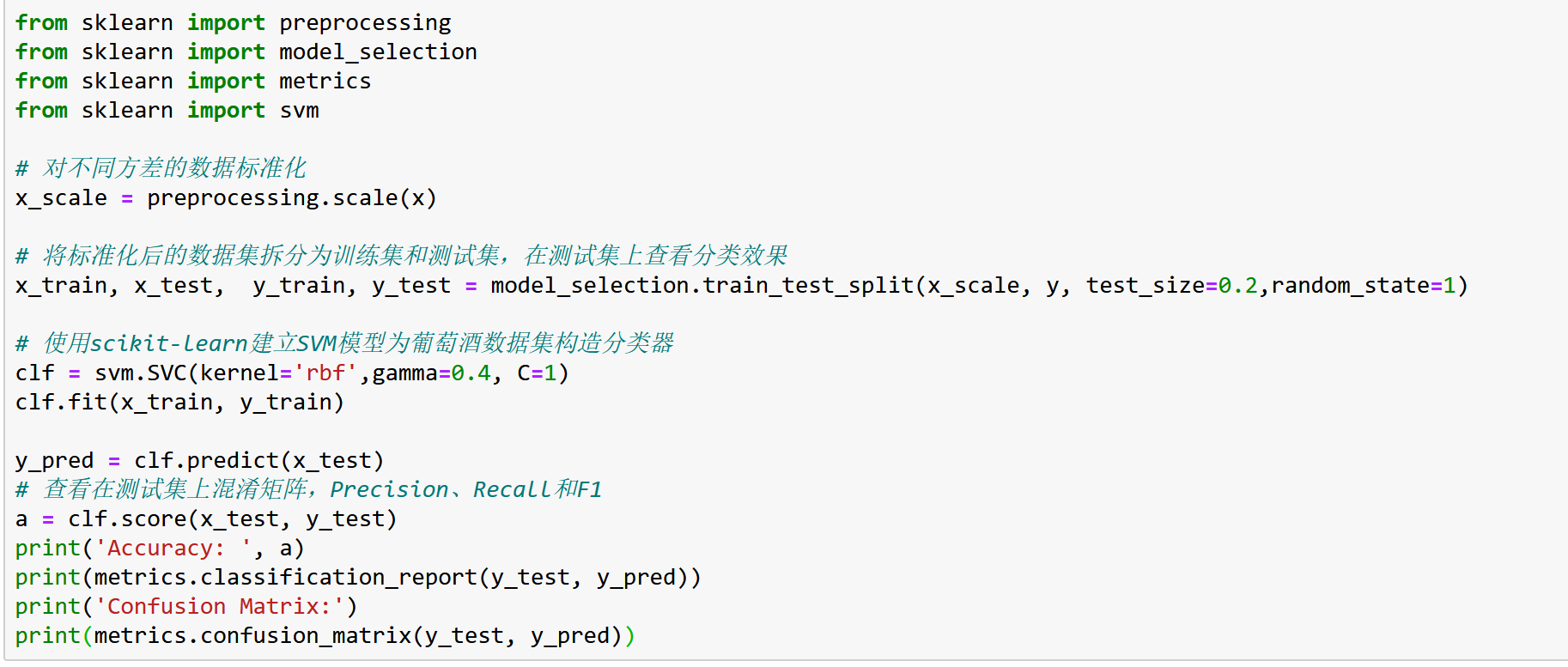
对数据进行预处理，将所有数据换为数值型数据：

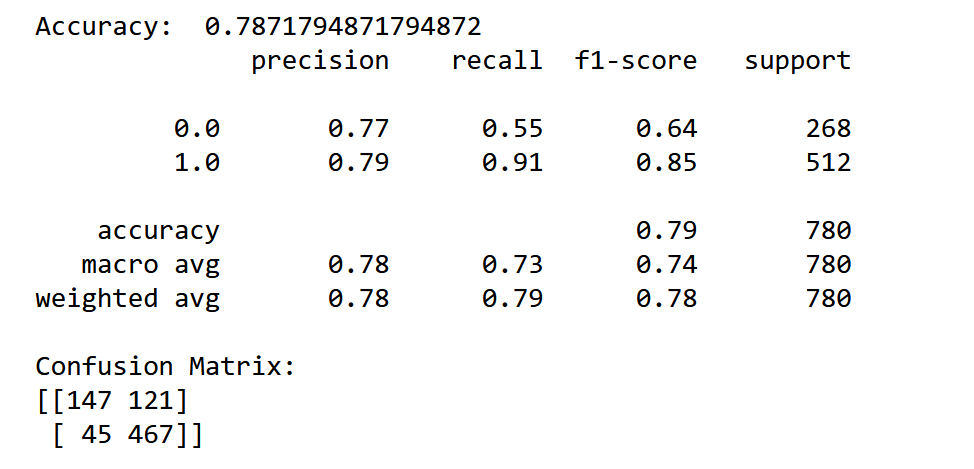






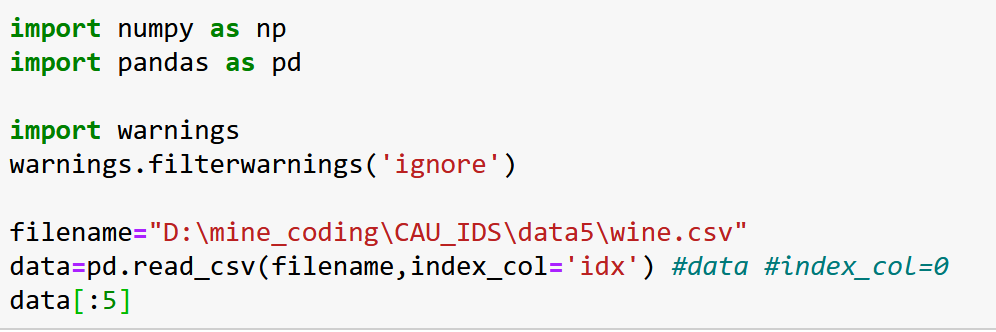
处理数据并获取结果：

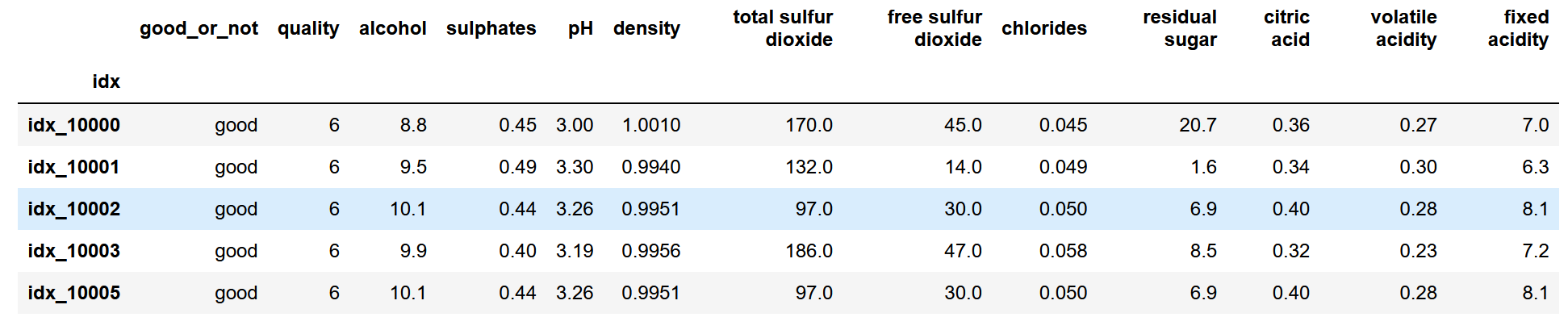




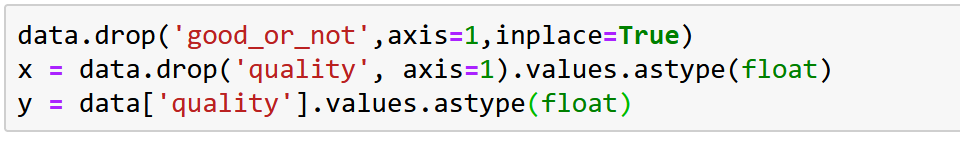
3.使用scikit-learn建立随机森林回归模型预测葡萄酒质量（1-10之间）(40分)[“不可使用good\_or\_not”列]

读入wine.csv数据：





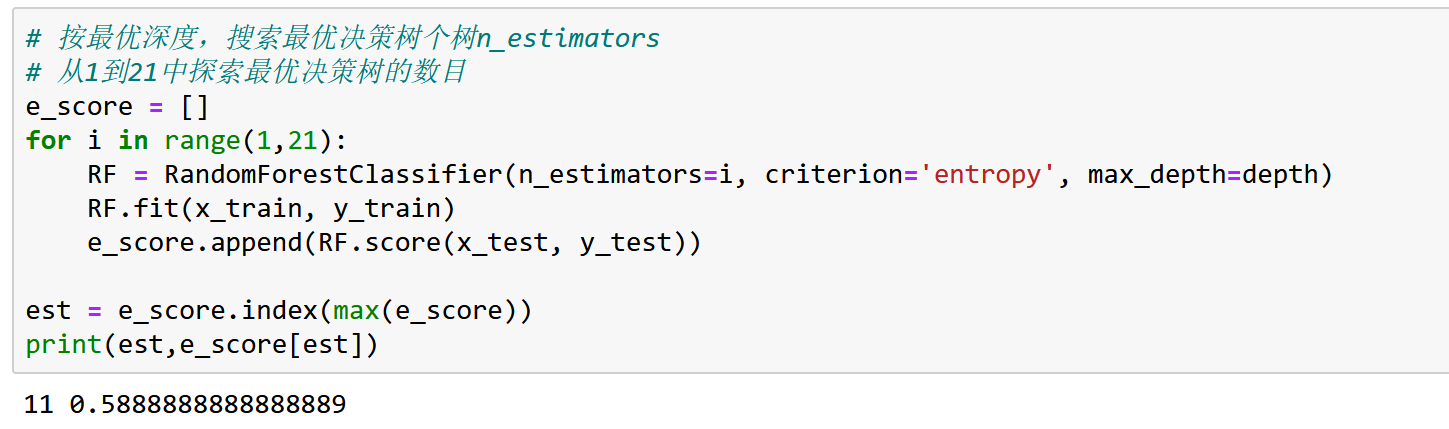
对数据进行预处理，将所有数据换为数值型数据：



使用随机森林算法训练集成分类器，建立决策树：









1. **实验总结**

收获：

（1）SVM模型：在机器学习中，支持向量机是在分类与回归分析中分析数据的监督式学习模型与相关的学习算法。给定一组训练实例，每个训练实例被标记为属于两个类别中的一个或另一个，SVM训练算法创建一个将新的实例分配给两个类别之一的模型，使其成为非概率二元线性分类器。SVM模型是将实例表示为空间中的点，这样映射就使得单独类别的实例被尽可能宽的明显的间隔分开。然后，将新的实例映射到同一空间，并基于它们落在间隔的哪一侧来预测所属类别。

（2）随机森林算法：在机器学习中，随机森林是一个包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定。通过随机的方式建立一个森林.