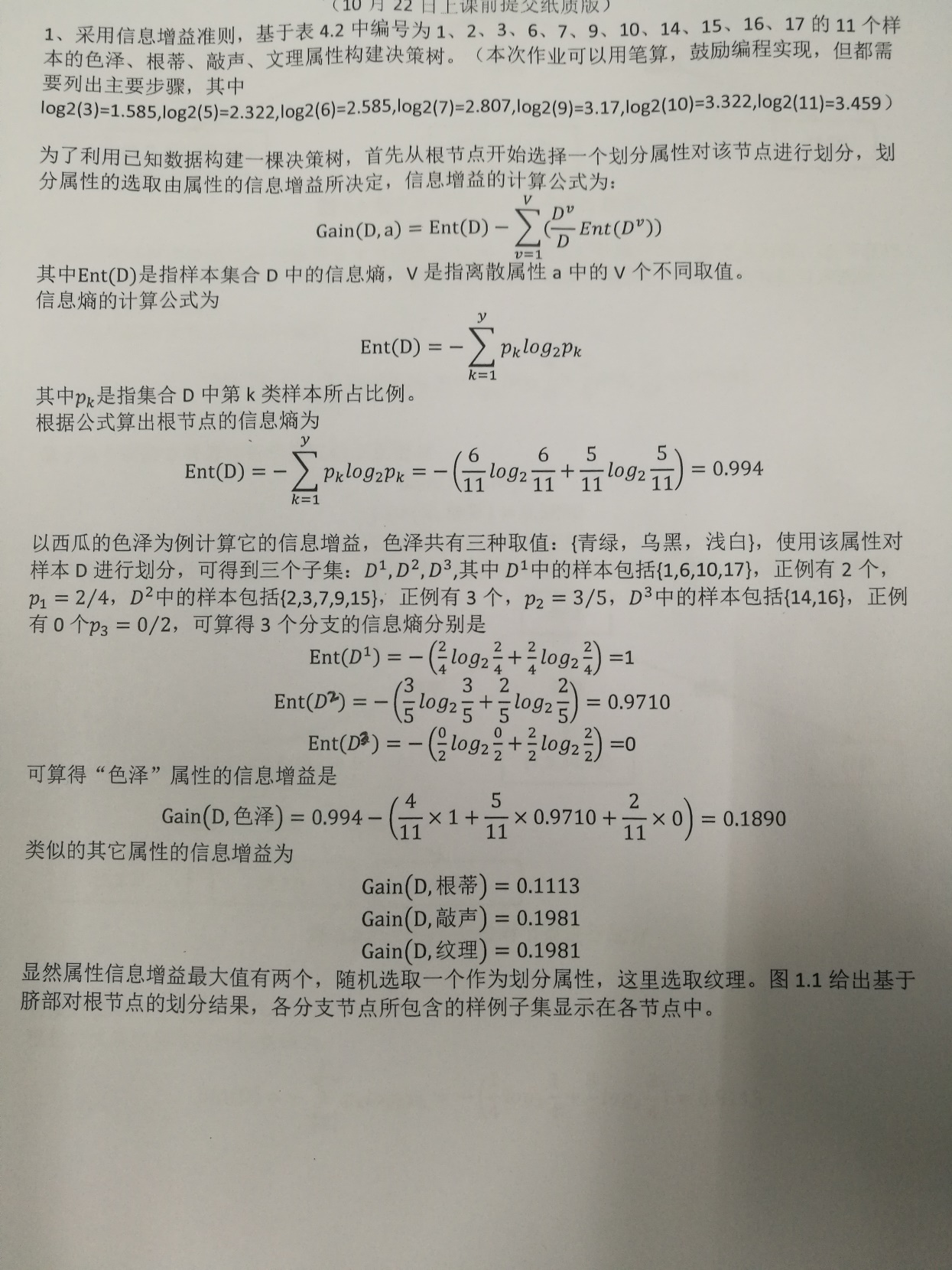
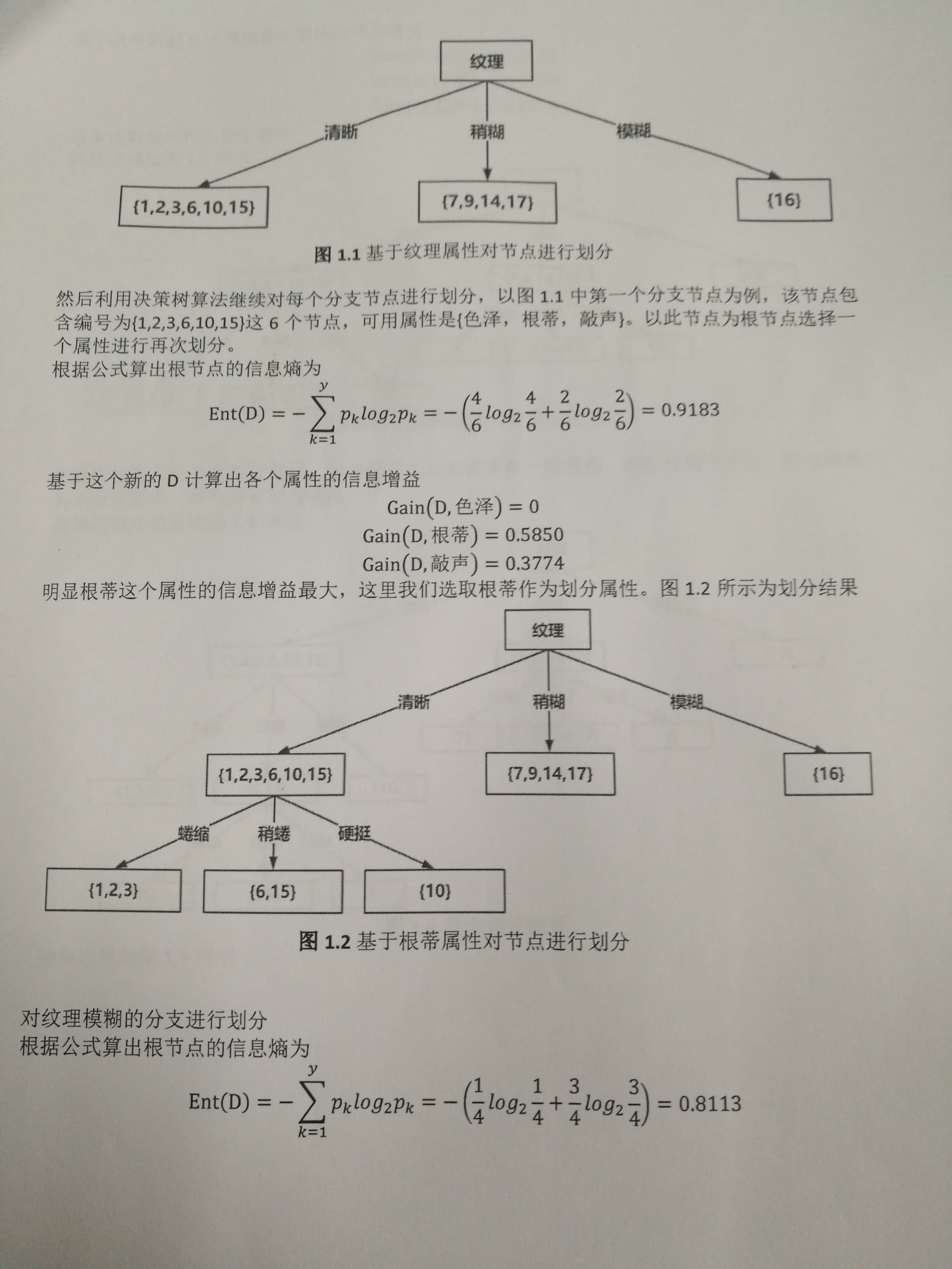
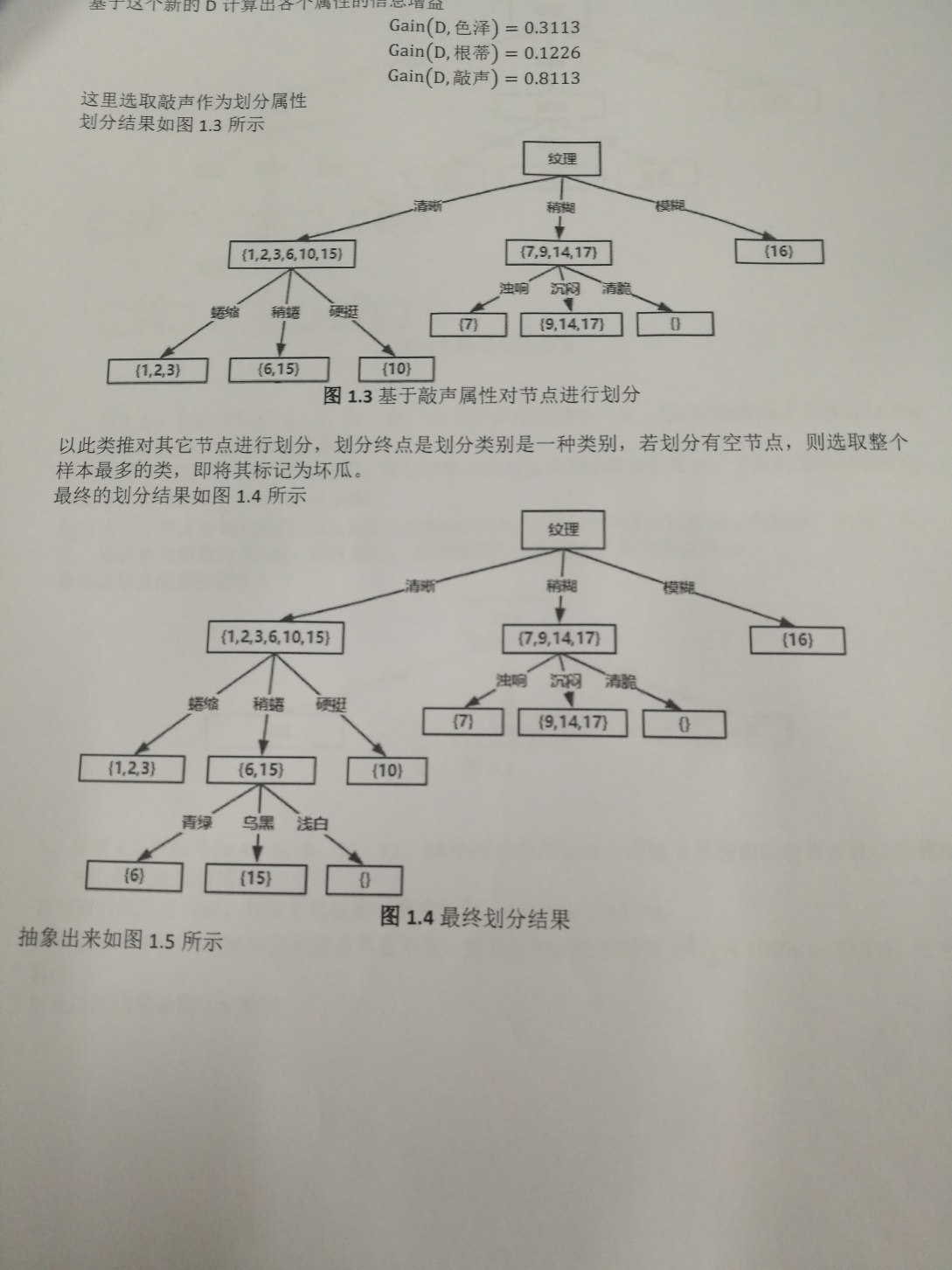
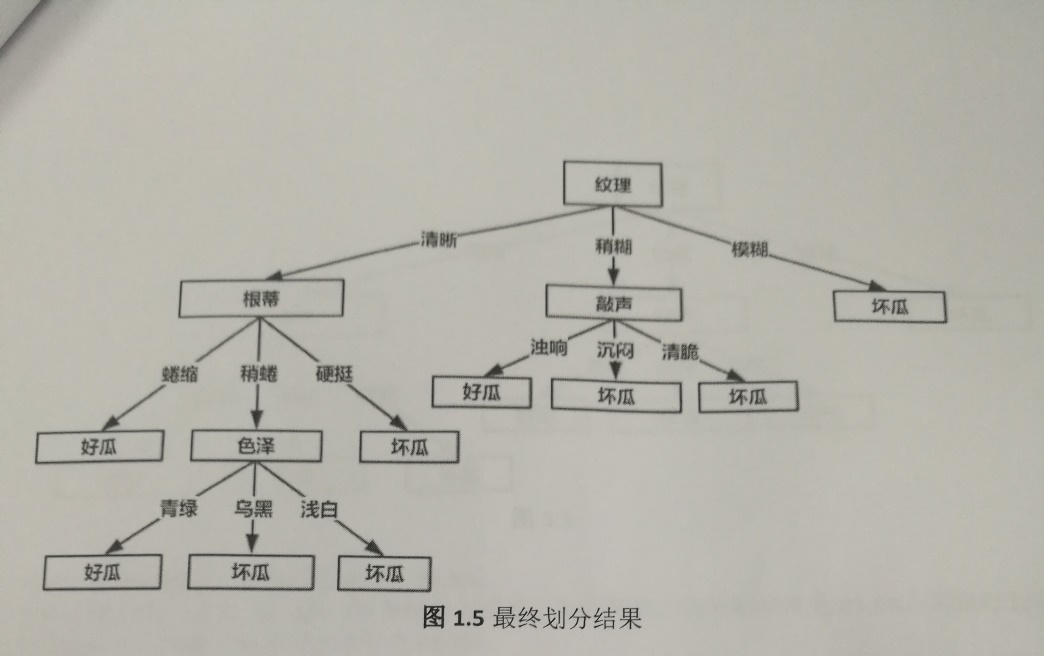
作业二

（10月22日上课前提交纸质版）

1. 采用信息增益准则，基于表4.2中编号为1、2、3、6、7、9、10、14、15、16、17的11个样本的色泽、根蒂、敲声、文理属性构建决策树。（本次作业可以用笔算，鼓励编程实现，但都需要列出主要步骤，其中log2(3)=1.585,log2(5)=2.322,log2(6)=2.585,log2(7)=2.807,log2(9)=3.17,log2(10)=3.322,log2(11)=3.459）(40分) 

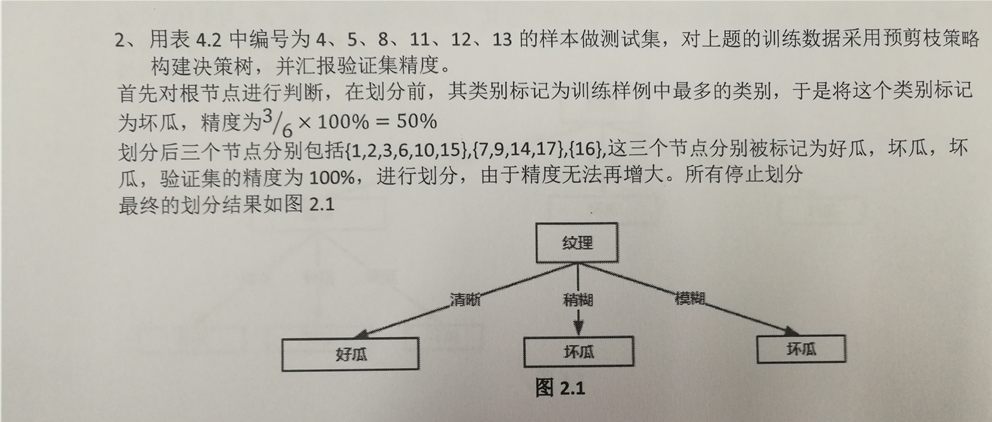






注：此题也可用敲声做第一次划分属性，结果会不一样

1. 用表4.2中编号为4、5、8、11、12、13的样本做测试集，对上题的训练数据采用预剪枝策略构建决策树，并汇报验证集精度。(35分)



1. 用表4.2中编号为4、5、8、11、12、13的样本做测试集，对题1所构建的决策树进行后剪枝，并汇报验证集精度。(25分)

