**6.2 试编程实现拉普拉斯修正的朴素贝叶斯分类器，并以p.84-西瓜数据集3.0为训练集，对p.151 "测1" 样本进行判别。**

**①问题描述：**

本问题要求使用西瓜数据集训练带有拉普拉斯修正的朴素贝叶斯分类器，并使用给定的样本进行测试，从而验证该训练出的分类器的性能。相较于普通朴素贝叶斯分类器，带有拉普拉斯修正的朴素贝叶斯分类器能够避免因训练集样本不充分而导致概率估值为零的问题，并且在训练集变大时，修正过程所引入的先验的影响也会逐渐变得可忽略，使得估值渐趋向于实际概率值。该问题是典型的二分类问题。

**②实现过程：**

a.使用Pandas库将WaterMelon3.0.csv文件中的数据集载入，并将第一列序号数据抹去，之后将数据转变为list类型。

b.根据数据集中最后一列的数据，将数据集划分为好瓜和坏瓜两类。

c.分别根据好瓜和坏瓜数量使用拉普拉斯修正算出各自的类先验概率。

d.在好瓜的分类数据内，依照测试数据各个属性的顺序，依次寻找与其某一属性值相同的数据的个数，并使用拉普拉斯修正计算出各个属性值对应的条件概率，并依次乘上之前求得的的先验概率，得到带拉普拉斯修正的朴素贝叶斯分类器的计算好瓜预测概率值。其中，针对离散属性值，使用概率密度函数对条件概率进行计算，设置计算均值和方差的函数进行概率密度函数的运算，最终得到相应的条件概率。

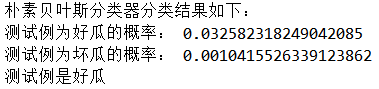
e.在坏瓜的分类数据内，依照测试数据各个属性的顺序，依次寻找与其某一属性值相同的数据的个数，并使用拉普拉斯修正计算出各个属性值对应的条件概率，并依次乘上之前求得的的先验概率，得到带拉普拉斯修正的朴素贝叶斯分类器的计算坏瓜预测概率值。其中，针对离散属性值，使用概率密度函数对条件概率进行计算，设置计算均值和方差的函数进行概率密度函数的运算，最终得到相应的条件概率。

f.分别打印出该分类器将测试例预测为好瓜和坏瓜的概率。

g.比较预测为好瓜和坏瓜的概率值，将该测试例标记为概率较高的那一类，并打印出其类别。

**③实现结果：**

Spyder的Console窗口打印结果：



**④结果分析：**

经训练集训练得到的贝叶斯分类器将该测试例预测为好瓜的概率为0.03258，将该测试例预测为坏瓜的概率为0.00104，因此该测试例被分类为好瓜。本问题中所提供的训练数据较少，因此进行拉普拉斯修正时引入的先验的影响仍然较大，估计值并不非常接近于实际概率值，但是能够避免其他属性携带的信息被训练集中未出现的属性值“抹去”。所提供的测试例较少，无法根据测试集对该分类器的分类准确率进行计算分析。