**第3章 朴素贝叶斯**

**一、简述朴素贝叶斯的原理、实现的关键操作点（不超过300字，逻辑要通，要有小标题，要排版清晰，请不要网络大幅摘抄）**

1. **朴素贝叶斯原理**

朴素贝叶斯的过程主要分为两个阶段。第一阶段，对实验样本进行分类，分别计算不同条件下其概率。第二阶段，输入测试样本，计算不同的类条件改率，比较其概率大小，从而完成对测试样本的分类。

实现样本的分类，需要通过计算条件概率而得到，计算条件概率的方法称为贝叶斯准则。条件概率的计算方法：

1. **关键操作点**
2. 计算先验概率：统计训练集中每个类别的出现频率，作为先验概率。
3. 计算条件概率：对于每个特征，计算其在每个类别下的条件概率。
4. 特征独立性假设：假设特征之间相互独立，进行对于类条件概率的计算。
5. 后验概率计算：利用贝叶斯定理，结合先验概率和条件概率，计算样本属于每个类别的后验概率。
6. 分类决策：选择后验概率最大的类别作为预测结果。
7. **实现要点**
8. 数据预处理：对离散特征进行编码，对连续特征进行离散化或假设其服从某种分布（如高斯分布）。
9. 概率平滑处理：使用拉普拉斯平滑处理零概率问题，避免因未出现的特征导致概率为零，将离散的数据进行相应的平滑操作，对于概率的计算和。
10. 模型训练：统计训练数据中的概率分布，构建概率表。
11. 预测：对新样本，根据概率表计算其属于各类别的概率，选择最大概率对应的类别。

**二、利用朴素贝叶斯解决问题（预测或分类）。**

利用朴素贝叶斯对鸢尾花数据进行分类。

要求：

1）针对当前数据、模型，**详叙数据预处理、模型参数调整过程及效果比对**；

2）代码要列出，重点代码加注释说明，特别是自己调试过程中的自我理解；

3）运行结果要截图，结果要文字说明；

4）注意排版

**数据预处理部分：**

def open\_wwcsv():  
 readbook = xlrd.open\_workbook('./iris.xls') # 打开csv文件  
 sheet = readbook.sheet\_by\_index(0) # 获取表格  
 return sheet

**特征数据提取、并且对其进行标签化：**

def invalue(sheet,nrows) :  
 iris = []  
 y = []  
 # v1,v2,v3,v4 , name  
 for i in range(1, nrows):  
 v1 = sheet.cell(rowx=i, colx=1).value  
 v2 = sheet.cell(rowx=i, colx=2).value  
 v3 = sheet.cell(rowx=i, colx=3).value  
 v4 = sheet.cell(rowx=i, colx=4).value  
 name = sheet.cell(rowx=i, colx=5).value  
 iris.append([v1, v2, v3, v4,name])  
 y.append(name)  
 return iris, y

**三、[附加题].利用朴素贝叶斯解决自己遇到的问题。**

[解题要求同题2]