**朴素贝叶斯分类器实验指导书**

1. **实验目的**
   1. 掌握贝叶斯判别定理；
   2. 熟悉如何通过Python编程实现贝叶斯分类器设计。
2. **实验环境**

系统：windows

软件：Python

示例代码版本：seaborn==0.11.1

Python 3.8.3

numpy==1.19.5

1. **背景知识**

1. 贝叶斯定理

http://latex.codecogs.com/gif.latex?P(A|B)表示事件B已经发生的前提下，事件A发生的概率，即事件B发生下事件A的条件概率，其基本求解公式为：

http://latex.codecogs.com/gif.latex?P(A|B)=\frac%7bP(AB)%7d%7bP(B)%7d

贝叶斯定理打通了从P(A|B)获得P(B|A)的道路，即：

http://latex.codecogs.com/gif.latex?P(B|A)=\frac%7bP(A|B)P(B)%7d%7bP(A)%7d

2. 朴素贝叶斯分类器

朴素贝叶斯分类是一种简单的分类算法，其基本思想是：对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大，就认为此待分类项属于哪个类别。

朴素贝叶斯分类器的基本流程如下：

（1）设http://latex.codecogs.com/gif.latex?x=\%7ba_1,a_2,...,a_m\%7d为一个待分类项，而每个*a*为*x*的一个特征属性值；

（2）有类别集合http://latex.codecogs.com/gif.latex?C=\%7by_1,y_2,...,y_n\%7d；

（3）计算http://latex.codecogs.com/gif.latex?P(y_1|x),P(y_2|x),...,P(y_n|x)；

（4）如果http://latex.codecogs.com/gif.latex?P(y_k|x)=max\%7bP(y_1|x),P(y_2|x),...,P(y_n|x)\%7d，则http://latex.codecogs.com/gif.latex?x%20\in%20y_k。

其中的关键就是如何计算上述第3步中的各个后验概率。我们可以这么做：

（1）找到一个已知类别的样本集合，这个集合叫做训练样本集；

（2）统计得到在各类别下各个特征属性值的条件概率估计，即：，其中;

（3）根据贝叶斯定理有如下推导：

http://latex.codecogs.com/gif.latex?P(y_i|x)=\frac%7bP(x|y_i)P(y_i)%7d%7bP(x)%7d

因为分母对于所有类别为常数，因为我们只要将分子最大化即可。又假设各特征属性是条件独立的，则有：http://latex.codecogs.com/gif.latex?P(x|y_i)P(y_i)=P(a_1|y_i)P(a_2|y_i)...P(a_m|y_i)P(y_i)=P(y_i)\prod%5em_%7bj=1%7dP(a_j|y_i)

1. **实验内容**

以西瓜数据集3.0为训练集，试编程实现拉普拉斯修正的朴素贝叶斯分类器，并对样本

['青绿', '蜷缩', '浊响', '清晰', '凹陷', '硬滑', '0.697','0.460']

['青绿', '硬挺', '清脆', '清晰', '平坦', '软粘', '0.243','0.267']

['乌黑', '稍缩', '沉闷', '稍糊', '稍凹', '硬滑', '0.666','0.091']

['乌黑', '蜷缩', '沉闷', '清晰', '凹陷', '硬滑', '0.774','0.376']

['青绿', '蜷缩', '浊响', '清晰', '凹陷', '硬滑', '0.697','0.460']

['浅白','蜷缩','浊响','模糊','平坦','软粘',’0.343’,’0.099’]

进行判别；

1. **实验步骤**

基于西瓜数据集3.0

1. 拉普拉斯修正朴素贝叶斯分类器的实现（如.ipynb提供样例）,并探究拉普拉斯修正系数对结果的影响。
2. Sklearn贝叶斯分类器的实现（需同学们自行完成，可参考课件中的截图），相关实现可参考理论课ppt相关截图，或<https://www.kaggle.com/code/prashant111/naive-bayes-classifier-in-python>
3. 对比两者结果，对输出不同的样例进行思考 对比结果为什么不同

对比方向：（多项式和高斯方法） （纯离散模型，离散加连续混合模型）

六、实验报告

实验报告要求：

1. 使用深圳大学实验报告模板

2. 要有运行结果截图

3. 建议附上代码，并以注释方式记录写代码的思考过程

4. 截止日期2024年11月12日11:59pm, [发送电子版至TA邮箱iamironmancc@qq.com，格式：2024](mailto:发送电子版至TA邮箱jecdlinjw@126.com，格式：2022)机器学习第二次实验报告-姓名。