实验名称 – 基于Naïve Bayes的分类算法实现

目的：

通过实现Naïve Bayes算法，理解Naïve Bayes算法的基本原理和基本的程序实现方法。

带类标的数据：titanic.dat（详见相应的文件），该文件为文本文件，以@开头的行为注释行，剩下每一行为一个带类标的训练样例，由空格隔开的属性值构成，最后一个属性值为类标，取值为1.0或者-1.0，分别表示两个不同的类别。

要求：

1. 对各个属性的值进行离散化，离散化成两个区间（即把各个属性的取值变成布尔类型）。要求以信息增益作为标准，对每个属性选择信息增益最大的区间划分点（也叫做阈值点）；
2. 对给定的数据集随机划分，70%作为训练数据，30%作为测试数据；
3. 实现Naïve Bayes算法，给出测试数据集中每个测试样例的预测类标，同时输出每个测试样例属于每个类别的后验概率值。
4. 统计算法在测试集上类别预测的准确率（预测类别正确的测试样例的个数/测试样例的个数）。

程序设计语言

C、C++、Java或Python