计算机学院实验报告

|  |  |  |
| --- | --- | --- |
| 实验题目：朴素贝叶斯算法 | | 学号：201600181073 |
| 日期： | 班级：智能16 | 姓名： 唐超 |
| **Email：[1984386166@qq.com](mailto:1984386166@qq.com)** | | |
| 实验目的：  通过本实验掌握朴素贝叶斯算法原理，了解贝叶斯算法如何应用。 | | |
| 实验环境介绍：  Python3.7 | | |
| 解决问题的主要思路：  朴素贝叶斯算法是有监督的学习算法，解决的是分类问题，如客户是否流失、是否值得投资、信用等级评定等多分类问题。该算法的优点在于简单易懂、学习效率高、在某些领域的分类问题中能够与决策树、神经网络相媲美。但由于该算法以自变量之间的独立（条件特征独立）性和连续变量的正态性假设为前提，就会导致算法精度在某种程度上受影响。  http://demo.cjudge.net:8081/userfiles/image/1529748940249039753839.png  把P(A)称为”先验概率”（Prior probability），即在B事件发生之前，我们对A事件概率的一个判断。  P(A|B)称为”后验概率”（Posterior probability），即在B事件发生之后，我们对A事件概率的重新评估。  P(B|A)/P(B)称为”可能性函数”（Likelyhood），这是一个调整因子，使得预估概率更接近真实概率。  用所给的可训练数据得出P(A)和P(B|A), P(B)视情况是可以舍去的，最后我们可以用P(A)\*P(B|A)来近似P(A|B)，毕竟我们想要的不是P(A|B)准确的绝对大小，而是B属于A下不同类别概率之间的相对大小来判断B的所属。 | | |
| **实验步骤及结果：**  创建实验样本并加载数据：      构建词表：      构造文本向量矩阵：        训练函数：      返回的p0v代表32个词每一个词属于非侮辱类的概率，p1v代表32个词每一个词属于侮辱类的概率。pab是先验概率。因为先验概率是0.5，所以实际上p0v也可以说是非侮辱类这个类别出现各个词的概率，p1v可类似说成是侮辱类这个类别出现各个词的概率。 | | |