# 基于概率论的分类方法：朴素贝叶斯

朴素贝叶斯应该是我接触的第一个分类算法。研究生刚联系好导师就被扔了一本张学工的《模式识别》，第一个介绍的算法就是朴素贝叶斯。由于当时太傻，玩心也重，只用了java做了实现，理论大概也是张无忌学武功，学到最后全忘了。今天重新总结下：

### 1.朴素贝叶斯分类的原理

学过概率论应该都知道，有两个概念，先验概率和后验概率。基于贝叶斯的方法和基于似然估计的方法最大的区别应该就在于是否基于先验概率的计算。先验概率是根据以往经验和分析得到的概率。后验概率类似于条件概率，。

联合概率：

全概率公式：

根据贝叶斯定理：

其中，样本X包含n个特征，我们假设各个特征之间是独立的，这也就是所谓“朴素”的原因。

因此，上上式可表示为

由于对于不同的类别，上式中分母部分均一致，因此只需比较分子的大小，分子越大说明是该类别的可能性越大，因此选择该类别正确的可能性越大。

### 2.朴素贝叶斯分类的优缺点

优点：在数据较少的情况下依然有效，可以处理多类别问题

缺点：对于输入数据的形式有要求

适用数据类型：标称型数据