朴素贝叶斯原理

朴素贝叶斯(naive Bayes)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。

贝叶斯方法把计算 “具有某特征的条件下属于某类” 的概率转换成计算 “属于某类的条件下具有某特征” 的概率，属于有监督学习。

算法优缺点

\*\*优点：\*\*数据较少的情况下仍然有效，可以处理多类别问题

\*\*缺点：\*\*输入的数据有条件限制，精度不高

贝叶斯理论

朴素贝叶斯作为贝叶斯决策理论的一部分，首先先了解一下贝叶斯理论的概念。假设一个数据集，有两类组成。我们现在用p1(x,y)表示数据点(x,y)属于类别1的概率，用p2(x,y)表示数据点(x,y)属于类别2的概率，那么对于一个新数据点(x,y)，可以用下面的规则来判断它的类别：

如果p1(x,y) > p2(x,y)，那么类别为1

如果p1(x,y) < p2(x,y)，那么类别为2

也就是说，我们会选择高概率对应的类别 。这就是贝叶斯决策理论的核心思想，即选择具有最高概率的决策。

