一、 朴素贝叶斯算法简介

贝叶斯定理由英国数学家贝叶斯（Thomas Bayes 1702-1761）发展，用来描述两个条件概率之间的关系。通常，事件 A 在事件 B(发生)的条件下的概率，与事件 B 在事件 A 的条件下的概率是不一样的；然而，这两者是有确定的关系，贝叶斯法则就是这种关系的陈述。贝叶斯法则是关于随机事件 A 和 B 的条件概率和边缘概率的。$P(A|B)$ 表示事件 B 已经发生的前提下，事件 A 发生的概率，叫做事件 B 发生下事件 A 的条件概率。其基本求解公式为：

$$P(A|B) = \frac{P(AB)}{P(B)}$$

贝叶斯公式为：

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

朴素贝叶斯分类算法的基础就是贝叶斯定理。它的思路非常简单清晰，对于给出的任何待分类项，求解该项出现的条件下各个类别出现的概率，哪个概率最大，就认为这个待分类项属于哪个类别。

二、 朴素贝叶斯算法流程

1. 设$x=\{a_1, a_2, \cdots a_m\}$为一个待分类项，而每个 a 为 x 的一个特征属性；
2. 有类别集合$C = \{y_1, y_2, \cdots y_n\}$；
3. 计算$P(y_1|x), P(y_2|x), \dots, P(y_n|x)$；
4. 如果$P(y_k|x) = max\{P(y_1|x), P(y_2|x), \dots, P(y_n|x)\}$，则$x \in y_k$。

由此可知，在这个流程中，最重要的步骤就是如何计算第三步中的条件概率。实际操作中，一般是按下面的步骤来进行操作的：

1. 找到一个已知分类的待分类项集合，这个集合叫做训练样本集；

2. 统计得到在各类别下各特征属性的条件概率估计，也就是 $P(a_1|y_1), P(a_2|y_1), \dots, P(a_m|y_1); P(a_2|y_2), \dots, P(a_m|y_2); \dots; P(a_1|y_n), P(a_2|y_n), \dots, P(a_m|y_n)$。

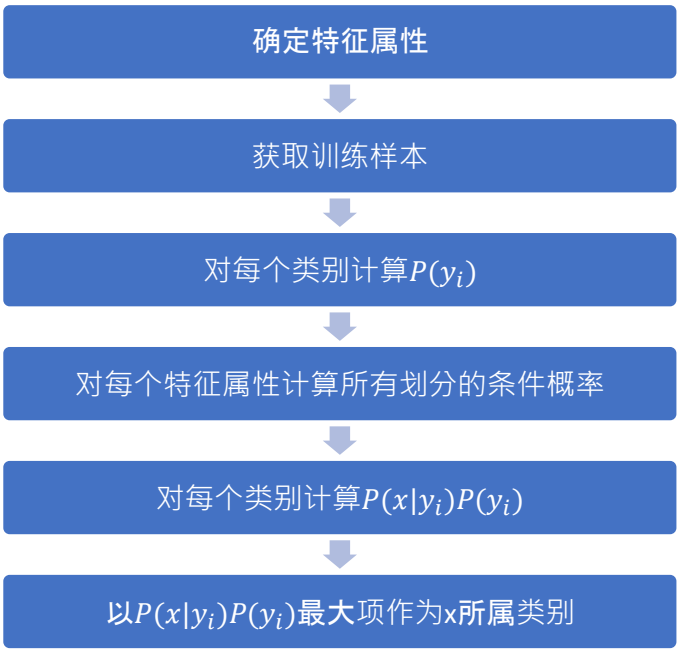
3. 若各个特征属性是条件独立的，那么按照贝叶斯公式可知

$$P(y_i|x) = \frac{P(x|y_i)P(y_i)}{P(x)}$$

4. 由于分母对于所有的类别是常数，因此只需要分子最大化。又因为每个特征属性都是条件独立的，因此有

$$P(x|y_i)P(y_i) = P(a_1|y_i)\ P(a_2|y_i) \dots P(a_m|y_i) = P(y_i)\prod_{j=1}^{m} P(a_j|y_i)$$

因此，由上面的步骤我们可以得到朴素贝叶斯分类的一个流程图：



在这个流程中，首先要为朴素贝叶斯分类做必要的准备，也就是根据具体情况确定特征属性，并且对每个特征属性进行适当划分，形成训练样本集合。这个阶段中，输入是所有的待分类数据，输出是特征属性和训练样本。这一阶段是整个朴素贝叶斯分类中唯一需要人工完成的阶段，其质量对整个过程将有重要影响，

分类器的质量很大程度上由特征属性、特征属性划分及训练样本质量决定。

形成训练样本集合后，就需要生成分类器。主要工作是计算每个类别在训练样本中的出现频率及每个特征属性划分对每个类别的条件概率估计，并将结果记录。其输入是特征属性和训练样本，输出是分类器。

最终，使用分类器对待分类项进行分类，其输入是分类器和待分类项，输出是待分类项与类别的映射关系。

三、 朴素贝叶斯分类算法应用示例

训练数据如下表，要求学习一个朴素贝叶斯分类器，并且确定 x=$(2,S)^T$的类标记 y。表中的$X^{(1)}$和$X^{(2)}$为特征，取值的集合分别为$A_1=\{1,2,3\}$，$A_2=\{S,M,L\}$,Y 为类标记，Y∈C={1，-1}。

训练数据

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

计算过程：

根据贝叶斯公式可以求出：

$P(Y=1)=\frac{9}{15}, P(Y=-1)=\frac{6}{15}$,

$P(X^{(1)}|Y=1)=\frac{2}{9}, P(X^{(1)}=2|Y=1)=\frac{3}{9}, P(X^{(1)}=3|Y=1)=\frac{4}{9}$

$P(X^{(2)}=S|Y=1)=\frac{1}{9}, P(X^{(2)}=M|Y=1)=\frac{4}{9}, P(X^{(2)}=L|Y=1)=\frac{4}{9}$

$P(X^{(1)}=1|Y=-1)=\frac{3}{6}, P(X^{(1)}=2|Y=-1)=\frac{2}{6}, P(X^{(1)}=3|Y=-1)=\frac{1}{6}$

对于给定的$x=(2,S)^T$,计算：

$$P(Y=1)P\left(X^{(1)}=2\middle|Y=1\right)P\left(X^{(2)}=S\middle|Y=1\right)=\frac{9}{15}\times\frac{3}{9}\times\frac{1}{9}=\frac{1}{45}$$

$$P(Y=-1)P\left(X^{(1)}=2|Y=-1\right)P\left(X^{(2)}=S\middle|Y=-1\right)=\frac{6}{15}\times\frac{2}{6}\times\frac{3}{6}=\frac{1}{15}$$

由于$P(Y=-1)P\left(X^{(1)}=2|Y=-1\right)P\left(X^{(2)}=S\middle|Y=-1\right)$最大，所以 y=-1

四、 朴素贝叶斯分类算法优缺点总结

- 朴素贝叶斯模型有稳定的分类效率；
- 对小规模的数据表现很好，能处理多分类任务，适合增量式训练；
- 对缺失数据不太敏感，比较简单，常用于文本分类。

朴素贝叶斯的主要缺点有：

- 朴素贝叶斯模型给定输出类别的情况下,假设属性之间相互独立 ,这个假设在实际应用中往往是不成立的，在属性个数比较多或者属性之间相关性较大时，分类效果不好。而在属性相关性较小时，朴素贝叶斯性能最为良好。
- 需要知道先验概率，且先验概率很多时候取决于假设，假设的模型可以有很多种，因此在某些时候会由于假设的先验模型的原因导致预测效果不佳。
- 由于我们是通过先验和数据来决定后验的概率从而决定分类，所以分类决策存在一定的错误率。
- 对输入数据的表达形式很敏感。