朴素贝叶斯分类

朴素贝叶斯分类算法的核心是贝叶斯概率公式

贝叶斯概率公式　　　有监督的分类

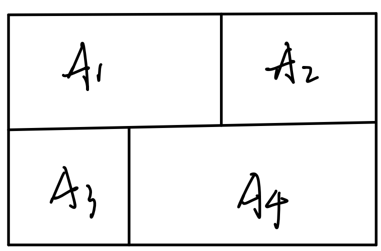
1. 完备事件组

$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$，且 $A_i \cap A_j = \emptyset$，$1 \le i \ne j \le n$

则称 $A_1, A_2, \cdots A_n$ 为一个完备事件组

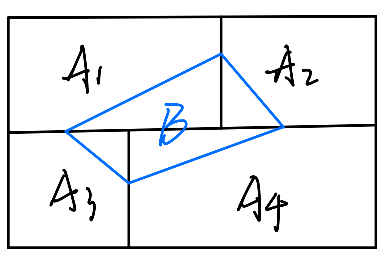
也就是说 $A_1, A_2, \cdots A_n$ 的并集是完整的事件空间

且两两之间相互独立



假设有一事件B，只有在 A_i 事件发生后，B事件才发生

并且 A_i 事件发生后，B事件发生的概率可能不相等



先验概率：由简单分析可获得的概率

条件概率：事件Ai发生的前提下，B发生的概率

联合概率：事件Ai和B同时发生的概率

全概率公式：

$$p(B)=\sum_{i=1}^{n}p(A_i,B)=\sum_{i=1}^{n}p(A_i)\cdot p(B|A_i) \quad (1)$$

2. 贝叶斯概率公式

第一个场景：B事件发生后，Ai事件发生的概率

$$p(A_i|B)=\frac{p(A_i,B)}{p(B)}=\frac{p(B|A_i)\cdot p(A_i)}{\sum_{i=1}^{n}p(A_i)\cdot p(B|A_i)} \quad (2)$$

上式即为贝叶斯概率公式。

作用：已知事件B发生了，去探求是某个原因Ai导致这一结果发生的概率 $p(A_i|B)$

3. 朴素贝叶斯分类算法

3.1 单一属性的朴素贝叶斯分类算法

☆ 3.2 多属性的朴素贝叶斯分类算法

假设有一个容量为n，包含m个特征的数据集

$D=\{(x_1^{(1)},x_2^{(1)},\cdots,x_m^{(1)},y_1),(x_1^{(2)},x_2^{(2)},\cdots,x_m^{(2)},y_2),\cdots,$
$(x_1^{(n)},x_2^{(n)},\cdots,x_m^{(n)},y_n)\}$

其中y是分类标签，$y_i\in(C_1,C_2,\cdots,C_k)$

通过对D进行统计分析，得到每个类别的

先验概率：$P(Y=C_k),\ k=1,2,\cdots,K$

以及在每个类别下不同特征的条件概率：

$P(X_j=x_j\mid Y=C_k)$（在C_k类中第j维特征为x_j的概率）

朴素贝叶斯算法有一个前提，即各个特征间独立

表示为 $P(X=x\mid Y=C_k)=P(X_1=x_1,X_2=x_2,\cdots,X_m=x_m\mid C_k)$

$$=\prod_{j=1}^{m}P(X_j=x_j\mid C_k) \quad (3)$$

利用贝叶斯概率公式，可得

$$P(Y=C_k\mid X=x)=\frac{P(X=x,Y=C_k)}{P(X=x)}=\frac{P(X=x\mid Y=C_k)\cdot P(Y=C_k)}{\sum_{k=1}^{K}P(Y=C_k)P(X=x\mid Y=C_k)}$$

$$=\frac{\prod_{j=1}^{m}P(X_j=x_j\mid Y=C_k)\cdot P(Y=C_k)}{\sum_{k=1}^{K}P(Y=C_k)\cdot\prod_{j=1}^{m}P(X_j=x_j\mid Y=C_k)} \quad (4)$$

比较各个概率对应的大小，选择概率最大的一个类别作为分类结果

注意：在计算过程中无需计算分母，因为分母不变

全概率公式

$$P(B)=\sum_{i=1}^{n}P(A_i,B)=\sum_{i=1}^{n}P(A_i)\cdot P(B|A_i)$$

贝叶斯概率公式

$$P(A_i|B)=\frac{P(A_i,B)}{P(B)}=\frac{P(A_i)\cdot P(B|A_i)}{P(B)}$$

$$=\frac{P(A_i)\cdot P(B|A_i)}{\sum_{i=1}^{n}P(A_i)\cdot P(B|A_i)}$$

朴素：各特征之间相互独立