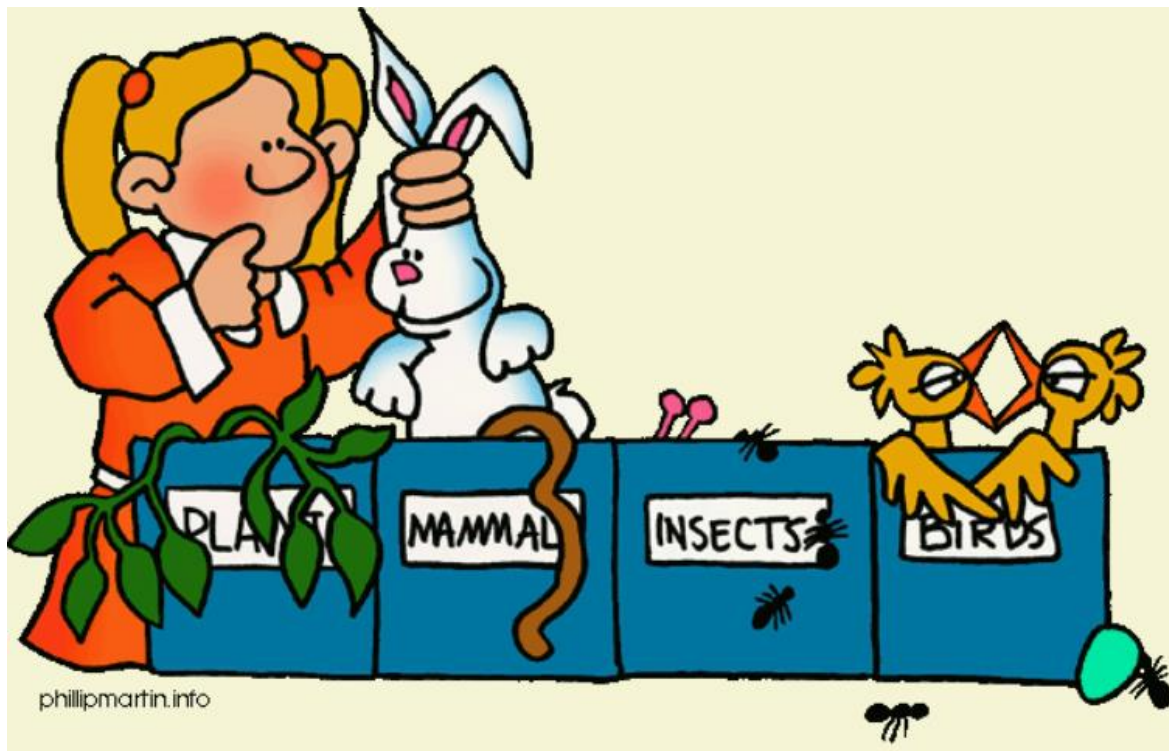


第六章：朴素贝叶斯 Ch6: Naïve Bayes



贝叶斯



Thomas Bayes 贝叶斯 (约1701-1761)，英国数学家。贝叶斯在数学方面主要研究概率论。他首先将**归纳推理**法用于概率论基础理论，并创立了**贝叶斯统计理论**，对于统计决策函数、统计推断、统计的估算等做出了贡献。他死后，理查德·普莱斯(Richard Price)于1763年将他的著作An essay towards solving a problem in the doctrine of chances寄给了英国皇家学会，对于现代概率论和数理统计产生了重要的影响。

THE ROYAL SOCIETY
PUBLISHING

All Journals ▾

Twitter Facebook WeChat

Home Content ▾ Information for ▾ About us ▾ Sign up ▾

✓ You have access

View PDF

Tools

Share

Cite this article ▾

Section

Abstract

Footnotes

Letter

LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S

Thomas Bayes

Published: 01 January 1763 | <https://doi.org/10.1098/rstl.1763.0053>

Abstract

Dear Sir, I Now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved.

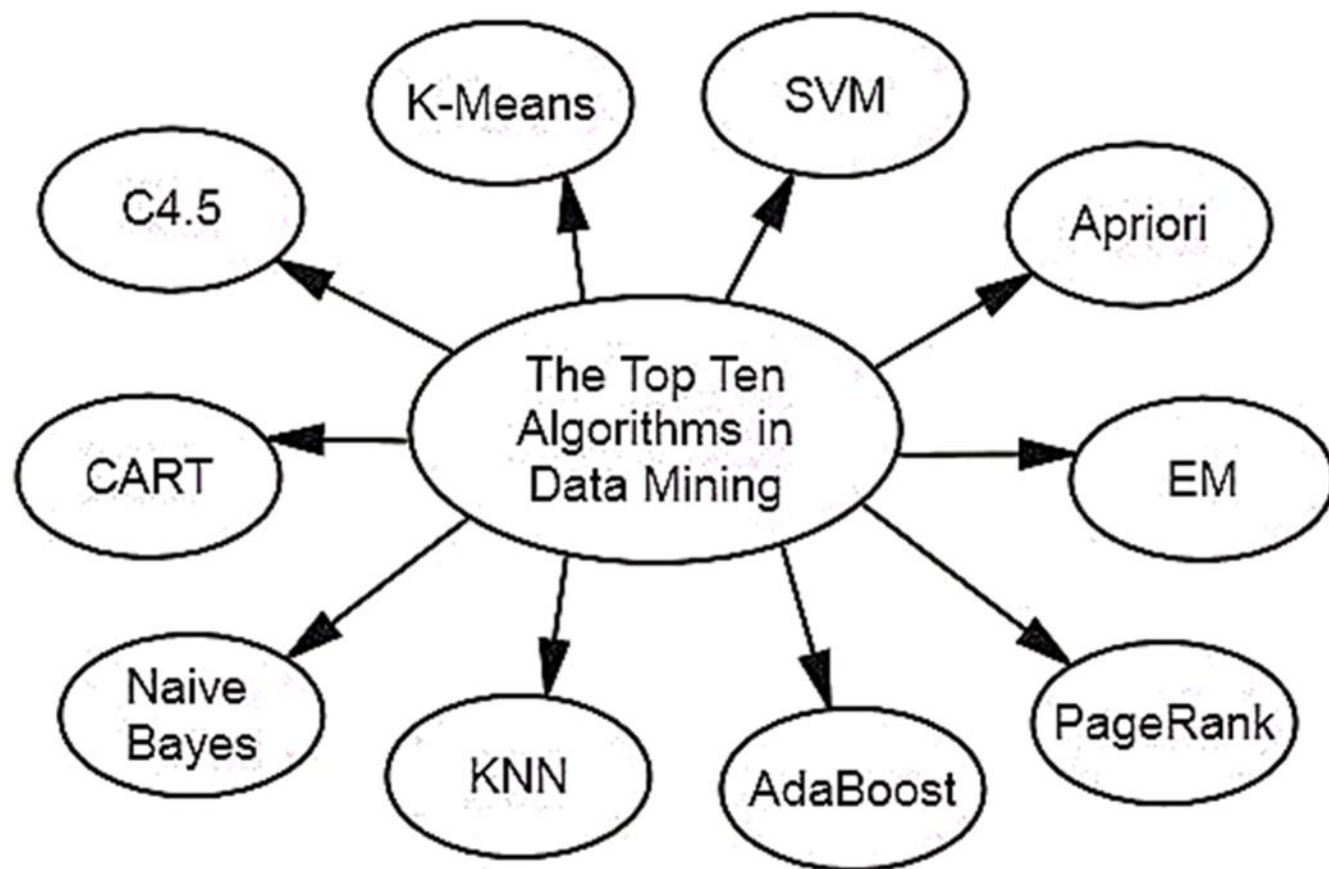
贝叶斯



“从特殊推论一般、从样本推论全体”
第一人。

该1763年发表的论文中，首先提出了贝叶斯定理。在这篇论文中，他为了解决一个“**逆概率**”问题，而提出了**贝叶斯定理**。最初提出贝叶斯定理是为了解决“逆概率”问题，然而后来，贝叶斯定理席卷了概率论，并将应用延伸到各个领域，比如癌症的检测、垃圾邮件的过滤。可以说，所有需要作出概率预测的地方都可以见到贝叶斯定理的影子，特别地，贝叶斯是**数据挖掘**的核心算法之一、也是**机器学习**的核心方法之一。

贝叶斯



贝叶斯



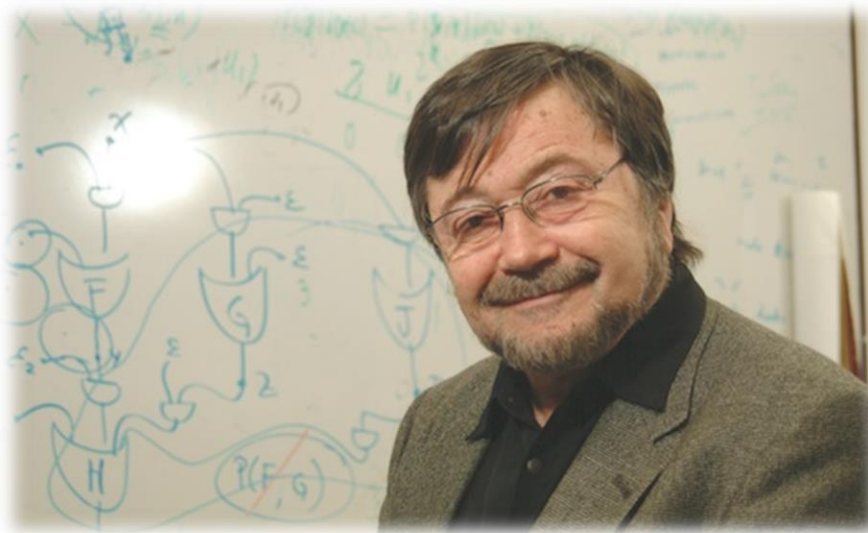
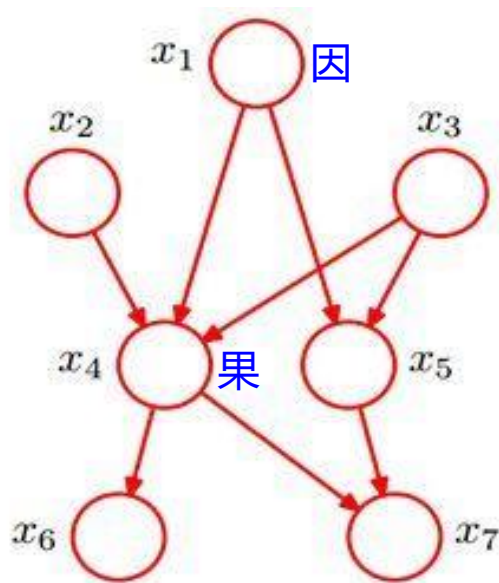
贝叶斯决策就是在不完全情报下，对**部分未知的状态用主观概率估计**，然后用贝叶斯公式对发生概率进行修正，最后再利用期望值和修正概率做出最优决策。

贝叶斯决策理论方法是统计模型决策中的一个基本方法，其基本思想是：

- 1、已知**类条件概率**密度参数表达式和**先验概率**。
- 2、利用贝叶斯公式转换成**后验概率**。
- 3、根据后验概率大小进行**决策分类**。

贝叶斯网络

贝叶斯网络 (Bayesian network), 又称信念网络 (Belief Network), 或有向无环图模型 (directed acyclic graphical model), 是一种**概率图模型**, 于1985年由Judea Pearl首先提出。它是一种模拟人类推理过程中因果关系的不确定性处理模型, 其网络拓扑结构是一个有向无环图 (DAG)。



2011年图灵奖获得者Judea Pearl教授

贝叶斯网络之父

贝叶斯网络的应用

最早的PathFinder系统，该系统是淋巴疾病诊断的医学系统，它可以诊断60多种疾病，涉及100多种症状；后来发展起来的Internist-I系统，也是一种医学诊断系统，但它可以诊断多达600多种常见的疾病。

1995年，微软推出了第一个基于贝叶斯网的专家系统，一个用于幼儿保健的网站OnParent (www.onparenting.msn.com), 使父母们可以自行诊断。

- 故障诊断 (diagnose)
- 专家系统 (expert system)
- 规划 (planning)
- 学习 (learning)
- 分类 (classifying)

贝叶斯网络的应用

总库 1.08万		中文	学术期刊 5517	学位论文 4950	会议 226	报纸 3	年鉴	图书 3	专利	标准 0
科技		社科	检索范围: 总库 主题: 贝叶斯网络 主题定制 检索历史							
主题		排序: 相关度↓ 发表时间 被引 下载 综合								
主要主题		批量下载 导出与分析								
次要主题		全选 已选: 0 清除								
<input type="checkbox"/> 贝叶斯网络(2870)		题名 作者 来源 发表时间 数据库								
<input type="checkbox"/> 故障诊断(544)		<input type="checkbox"/> 1 基于故障树和贝叶斯网络集成的重大工程弃渣场风险诊断与预测 网络首发 李玉龙; 侯相宇 系统管理学报 2022-05-10 17:24 期刊								
<input type="checkbox"/> 动态贝叶斯网络(256)		<input type="checkbox"/> 2 基于贝叶斯网络的态势评估方法研究 黄克英; 任磊; 贺治飞 山西电子技术 2022-04-15 期刊								
<input type="checkbox"/> 风险评估(237)		<input type="checkbox"/> 3 基于证据理论层次分析法云贝叶斯网络的预警雷达毁伤效果评估 网络首发 邓力源; 杨萍; 刘卫东; 汪江鹏 兵工学报 2022-04-12 18:51 期刊								
<input type="checkbox"/> 数据挖掘(228)		<input type="checkbox"/> 4 基于HAZOP分析和贝叶斯网络的压缩机故障定位方法研究 邢建文; 梁熙; 李伟 自动化仪表 2022-03-20 期刊								
<input type="checkbox"/> 网络结构(165)		<input type="checkbox"/> 5 基于贝叶斯网络的老年人公交出行行为研究 网络首发 刘建荣; 刘志伟 武汉理工大学学报(交通科学与工程版) 2022-03-07 19:26 期刊								
<input type="checkbox"/> 可靠性分析(150)		<input type="checkbox"/> 6 基于贝叶斯网络的HAZOP-LOPA煤矿安全风险评价方法应用研究 张景钢; 王清焱; 赵淑枫 矿业安全与环保 2022-03-01 16:23 期刊								
<input type="checkbox"/> 结构学习(144)		<input type="checkbox"/> 7 基于贝叶斯网络的地铁驾驶员行车作业人因风险分析方法 张智贤; 李健; 刘志钢; 朱琳 智能计算机与应用 2022-03-01 期刊								
<input type="checkbox"/> 模型研究(138)		<input type="checkbox"/> 8 基于贝叶斯网络的尾矿库溃坝风险动态评价 网络首发 崔旭阳; 叶义成; 胡南燕; 周栋; 谭文侃 化工矿物与加工 2022-02-28 12:41 期刊								
<input type="checkbox"/> 故障诊断方法(127)		<input type="checkbox"/> 9 基于贝叶斯网络的南水北调中线工程大型河渠交叉建筑物风险评估 郭政宇; 袁源; 陈宁; 郭治君; 尹传森 居业 2022-02-20 期刊								
学科										
<input type="checkbox"/> 自动化技术(2882)										
<input type="checkbox"/> 计算机软件及计算机... (1757)										
<input type="checkbox"/> 电力工业(798)										
<input type="checkbox"/> 数学(663)										
<input type="checkbox"/> 互联网技术(633)										
<input type="checkbox"/> 公路与水路运输(604)										
<input type="checkbox"/> 安全科学与灾害防治(553)										

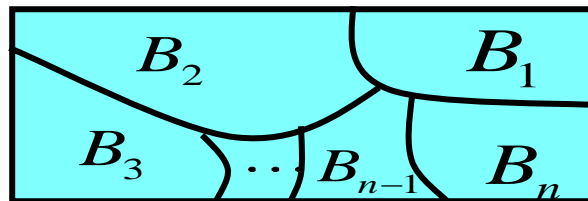
概率(回顾)

样本空间的划分

定义：设 Ω 为试验 E 的样本空间， B_1, B_2, \dots, B_n 为 E 的一组事件，
若：

$$1^0 \quad B_i B_j = \emptyset, i, j = 1, 2, \dots, n;$$

$$2^0 \quad B_1 \cup B_2 \cup \dots \cup B_n = \Omega$$



则称 B_1, B_2, B_n 为样本空间 Ω 的一个划分。

概率(回顾)

概率有限可加性

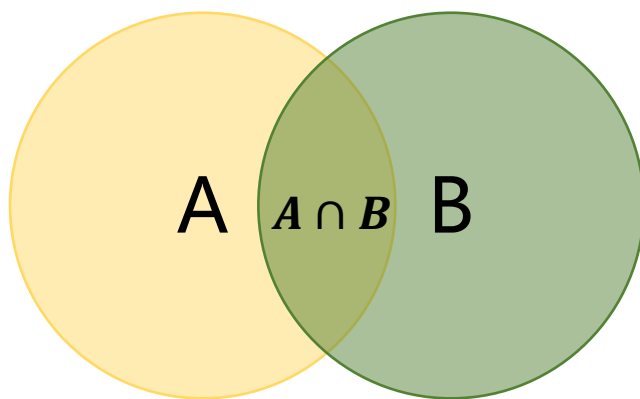
设 A_1, A_2, \dots, A_n 两两互不相容, 则

$$P\left(\bigcup_i^n A_i\right) = \sum_i^n P(A_i)$$

概率(回顾)

条件概率

指在事件 B 发生的情况下，事件 A 发生的概率，用 $P(A|B)$ 来表示。



可以很清楚地看到在事件 B 发生的情况下，事件 A 发生的概率就是 $P(A \cap B)$ 除以 $P(B)$ ，即如下： $P(A|B) = \frac{P(A \cap B)}{P(B)}$ 或者， $P(A|B) = \frac{P(AB)}{P(B)}$

也就是： $P(A \cap B) = P(A|B) P(B)$ 或者， $P(AB) = P(A|B) P(B)$

乘法定理

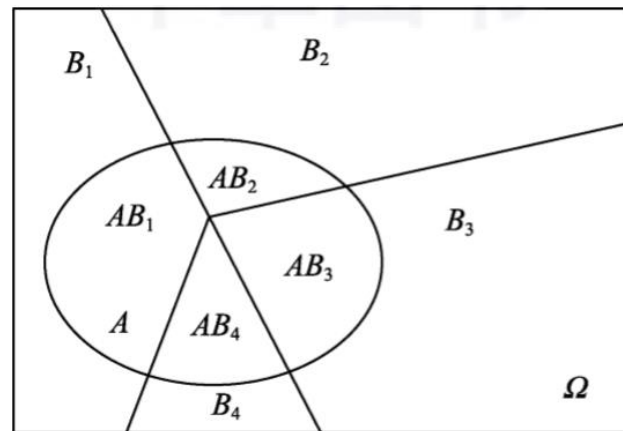
概率(回顾)

全概率公式

定义 设 Ω 为试验 E 的样本空间, A 为 E 的事件,

B_1, B_2, \dots, B_n 为 Ω 的一个划分, 且 $P(B_i) > 0$
($i = 1, 2, \dots, n$), 则

$$\begin{aligned} P(A) &= P(A | B_1)P(B_1) + P(A | B_2)P(B_2) \\ &\quad + \dots + P(A | B_n)P(B_n) \\ &= \sum_{i=1}^n P(B_i)P(A | B_i) \end{aligned}$$



事件 A 的概率 $P(A)$ 应该处于最小的 $P(A|B_i)$ 和最大的 $P(A|B_j)$ 之间, 它不是所有条件概率 $P(A|B_k)$ 的算术平均, 因为事件被使用的机会权重 (即 $P(B_i)$) 各不相同, 因此全概率 $P(A)$ 就是各条件概率 $P(A|B_k)$ 以 $P(B_k)$ 为权重的加权平均值。

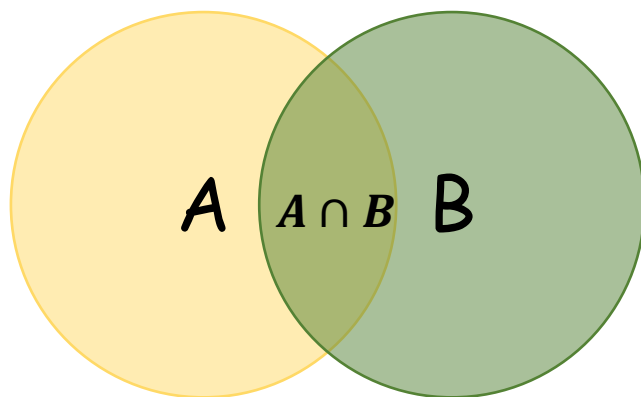
概率(回顾)

条件独立公式

如果 $P(A)=P(A|B)$ ，说明事件 B 的发生与否对事件 A 是否发生毫无影响。这时，我们就称 A 和 B 这两个事件独立，并且由条件概率的定义式进行转换可以得到：

$$\begin{aligned}P(A|B) &= \frac{P(AB)}{P(B)} \Rightarrow P(AB) = P(A|B)P(B) \\ &= P(A)P(B)\end{aligned}$$

贝叶斯定理



同理，在事件 A 发生的情况下，事件 B 发生的概率就是 $P(A \cap B)$ 除以 $P(A)$ ，即如下：

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

也就是： $P(A \cap B) = P(B|A) P(A)$

最后得到： $P(B|A) P(A) = P(A|B) P(B)$ 或者， $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

贝叶斯定理

后验概率 = 先验概率 × 调整因子

先验概率 + 数据 = 后验概率

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

似然函数

先验概率

可能性函数
(调整因子)

后验概率

B事件发生之后，我们对A事件概率的重新评估。

贝叶斯定理

“逆概率思维”

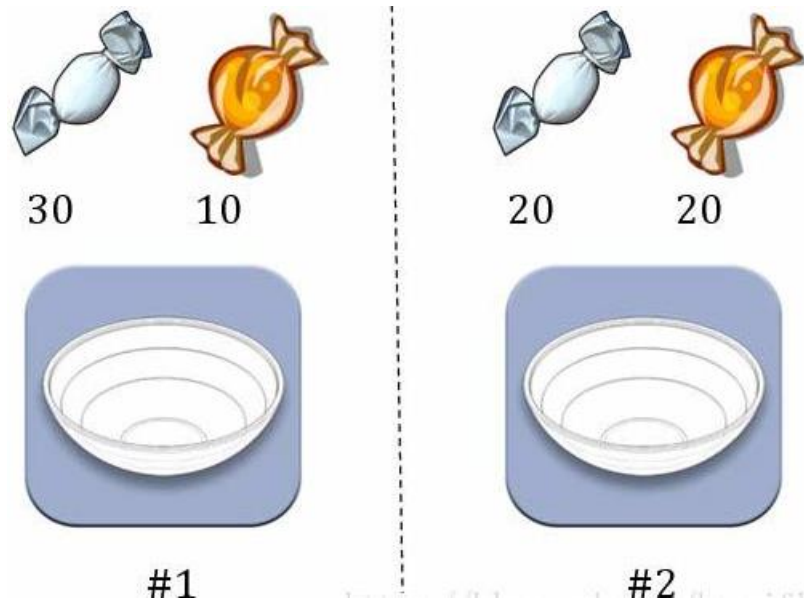
$$P(B_i | A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A | B_i)}{P(A)}$$
$$= \frac{P(B_i)P(A | B_i)}{P(B_1)P(A | B_1) + P(B_2)P(A | B_2) + \cdots + P(B_n)P(A | B_n)}$$

$$P(B_i | A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A | B_i)}{\sum_j P(B_j)P(A | B_j)}$$

条件概率 $P(B_i | A)$ ，就是在观察到结果事件 A 已经发生的情况下，推断结果事件 A 是由原因 B_i 造成的概率的大小。

贝叶斯定理

两个一模一样的碗，#1碗有30颗水果糖和10颗巧克力糖，#2碗有水果糖和巧克力糖各20颗。现在随机选择一个碗，从中摸出一颗糖，发现是水果糖。请问这颗水果糖来自#1碗的概率有多大？



贝叶斯定理

我们假定， H_1 表示#1碗， H_2 表示2#碗。由于这两个碗是一样的，所以 $P(H_1)=P(H_2)$ ，也就是说，在取出水果糖之前，这两个碗被选中的概率相同。因此， $P(H_1)=0.5$ ，我们把这个概率就叫做“先验概率”，即没有做实验之前，来自#1碗的概率是0.5。再假定， E 表示水果糖，所以问题就变成了在已知 E 的情况下，来自#1碗的概率有多大，即求 $P(H_1|E)$ 。我们把这个概率叫做“后验概率”，即在 E 事件发生之后，对 $P(H_1)$ 的修正。根据条件概率公式，得到：

解：根据条件概率公式，得到：
$$P(H_1|E) = P(H_1) \frac{P(E|H_1)}{P(E)}$$

已知， $P(H_1)$ 等于0.5， $P(E|H_1)$ 为一号碗中取出水果糖的概率，等于0.75，那么求出 $P(E)$ 就可以得到答案。根据全概率公式，

$$P(E) = P(E|H_1)P(H_1) + P(E|H_2)P(H_2) = 0.75 \times 0.5 + 0.5 \times 0.5 = 0.625$$

来自一号碗的概率是：
$$P(H_1|E) = 0.5 \times \frac{0.75}{0.625} = 0.6$$

也就是说，取出水果糖之后， H_1 事件的可能性得到了增强。

贝叶斯定理

一般来说，先验概率一般来说就是我们对于数据所在领域的历史经验，但是这个经验常常是难以量化或者是模型化，于是贝叶斯分类就是一个大胆的假设先验分布的模型，一般来说有正态分布、beta分布、泊松分布等等，但这些假设一般没有特定的依据，因此一直被认为很荒谬。但现代的某些特定领域的很多实际应用中，贝叶斯理论很好用，因此也流行了起来。

- ◆ 朴素贝叶斯分类器
- ◆ 判别函数和决策面
- ◆ 贝叶斯参数估计

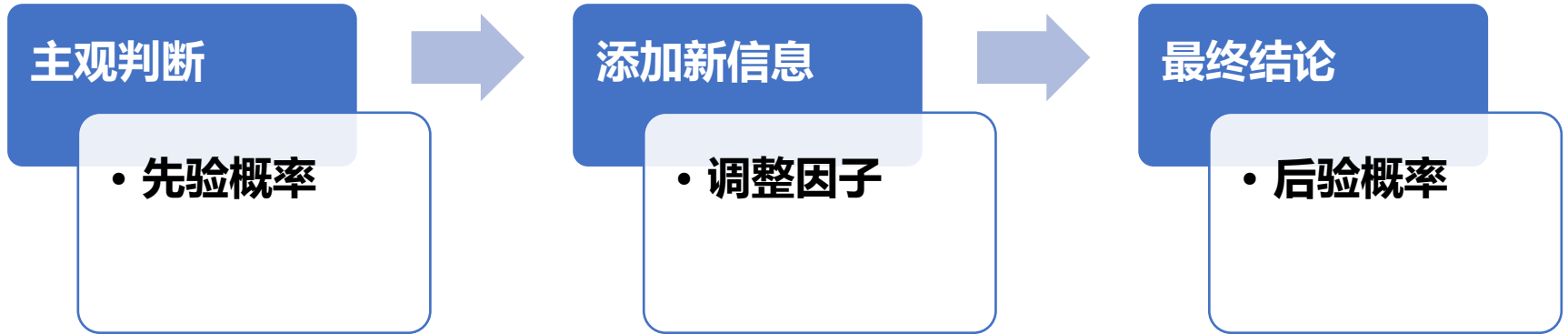
贝叶斯思维：例子



小哥哥给姐姐买朵花吧

主观判断

贝叶斯思维



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

↓ ↓ ↓

先验概率 可能性函数
 (调整因子)

后验概率

朴素贝叶斯

朴素贝叶斯 (Naïve Bayes) 法是基于**贝叶斯定理**与**特征条件独立假设**的分类方法。对于给定的训练数据集，首先基于**特征条件独立假设**学习**输入输出的联合概率分布**；然后基于此模型，对给定的输入利用贝叶斯定理求出**后验概率最大的输出**。朴素贝叶斯法实现简单，学习与预测的效率都很高，是种常用的方法。

本章叙述朴素贝叶斯法，包括**朴素贝叶斯法的学习与分类**、朴素贝叶斯法的**参数估计算法**。

基本方法

设输入空间 $\mathcal{X} \subseteq \mathbf{R}^n$ 为 n 维向量的集合, 输出空间为类标记集合 $\mathcal{Y} = \{c_1, c_2, \dots, c_K\}$ 。输入为特征向量 $x \in \mathcal{X}$, 输出为类标记 (class label) $y \in \mathcal{Y}$ 。 X 是定义在输入空间 \mathcal{X} 上的随机向量, Y 是定义在输出空间 \mathcal{Y} 上的随机变量。 $P(X, Y)$ 是 X 和 Y 的联合概率分布。训练数据集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

由 $P(X, Y)$ 独立同分布产生。

朴素贝叶斯通过训练数据集学习联合概率分布 $P(X, Y)$,

先验概率分布: $P(Y = c_k), \quad k = 1, 2, \dots, K$

条件概率分布:

$$P(X = x | Y = c_k) = P(X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)} | Y = c_k), \quad k = 1, 2, \dots, K$$

基本方法

条件概率分布 $P(X=x|Y=C_k)$ 有指数级数量的参数，其估计实际是不可行的事实上，假设有 $x^{(j)}$ 可取值有 S_j 个 $j=1, 2, \dots, n$ ， Y 取值有 K 个，

那么参数个数 $K \prod_{j=1}^n S_j$

朴素贝叶斯法对条件概率分布作了条件独立性的假设。由于这是一个较强的假设，朴素贝叶斯法也由此得名。具体地，条件独立性假设是

$$\begin{aligned} P(X = x|Y = c_k) &= P(X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)}|Y = c_k) \\ &= \prod_{j=1}^n P(X^{(j)} = x^{(j)}|Y = c_k) \end{aligned}$$

朴素贝叶斯法实际上学习到生成数据的机制，所以属于**生成模型**。条件独立假设等于是说用于分类的特征在类确定的条件下都是条件独立的。这一假设使朴素贝叶斯法变得简单，但有时会牺牲一定的分类准确率。

生成模型vs判别模型

生成模型：就是生成“数据分布”的模型。依据统计理论： $p(y|x) = \frac{p(x, y)}{p(x)}$ ，

通过拟合 $p(x, y)$ ，并计算 $p(x)$ ，从而间接的拟合出 $p(y|x)$

- 朴素贝叶斯 (Naïve)，隐马尔科夫模型 (HMM)，贝叶斯网络 (Bayesian Networks) 和隐含狄利克雷分布 (Latent Dirichlet Allocation)、混合高斯模型

判别模型：就是判别“数据输出值”的模型。用一个模型或函数直接拟合 $p(y|x)$ ，即 $f(x) = p(y|x)$

- 线性回归 (Linear Regression)，逻辑回归 (Logistic Regression)，支持向量机 (SVM)，传统神经网络 (Traditional Neural Networks)，线性判别分析 (Linear Discriminative Analysis)，条件随机场 (Conditional Random Field)

基本方法

- 条件独立性假设: $P(X = x | Y = c_k) = P(X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)} | Y = c_k)$

$$= \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_k)$$

- “朴素” 贝叶斯名字由来, 牺牲分类准确性。

- 贝叶斯定理: $P(Y = c_k | X = x) = \frac{P(X = x | Y = c_k)P(Y = c_k)}{\sum_k P(X = x | Y = c_k)P(Y = c_k)}$

$$P(B_i | A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A | B_i)}{\sum_j P(B_j)P(A | B_j)}$$

基本方法

- 条件独立性假设: $P(X = x | Y = c_k) = P(X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)} | Y = c_k)$

$$= \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_k)$$

- 贝叶斯定理: $P(Y = c_k | X = x) = \frac{P(X = x | Y = c_k)P(Y = c_k)}{\sum_k P(X = x | Y = c_k)P(Y = c_k)}$

- 代入上式: $P(Y = c_k | X = x) = \frac{P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}{\sum_k P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}$

朴素贝叶斯法分类的基本公式

基本方法

朴素贝叶斯法分类的基本公式：

$$P(Y = c_k | X = x) = \frac{P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}{\sum_k P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}$$

朴素贝叶斯分类器可表示为：

$$y = f(x) = \arg \max_{c_k} \frac{P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)}{\sum_k P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)} \rightarrow \text{定值}$$

等价于：

$$y = \arg \max_{c_k} P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k)$$

后验概率最大化的含义

- 假设选择0-1损失函数: $f(X)$ 为决策函数

- 期望风险函数: $L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$

- 取条件期望: $R_{\text{exp}}(f) = E[L(Y, f(X))]$

$$R_{\text{exp}}(f) = E_X \sum_{k=1}^K [L(c_k, f(X))] P(c_k | X)$$

后验概率最大化的含义

- 只需对 $X=x$ 逐个极小化，得：

$$\begin{aligned} f(x) &= \arg \min_{y \in \mathcal{Y}} \sum_{k=1}^K L(c_k, y) P(c_k | X = x) \\ &= \arg \min_{y \in \mathcal{Y}} \sum_{k=1}^K P(y \neq c_k | X = x) \\ &= \arg \min_{y \in \mathcal{Y}} (1 - P(y = c_k | X = x)) \\ &= \arg \max_{y \in \mathcal{Y}} P(y = c_k | X = x) \end{aligned}$$

- 推导出后验概率最大化准则：

$$f(x) = \arg \max_{c_k} P(c_k | X = x)$$

将实例分到后验概率最大的类中，等价于期望风险最小化

朴素贝叶斯法的参数估

- 应用极大似然估计法估计相应的概率
- 先验概率 $P(Y=c_k)$ 的极大似然估计是： $P(Y=c_k) = \frac{\sum_{i=1}^N I(y_i = c_k)}{N}$, $k=1,2,\dots,K$

先验概率

样本空间中各类样本所占的比例，可通过各类样本出现的频率估计（大数定理）

- 设第 j 个特征 $x(j)$ 可能取值的集合为： $\{a_{j1}, a_{j2}, \dots, a_{js_j}\}$
- 条件概率的极大似然估计：

$$P(X^{(j)} = a_{jl} | Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl}, y_i = c_k)}{\sum_{i=1}^N I(y_i = c_k)}$$

$$j=1,2,\dots,n; \quad l=1,2,\dots,S_j; \quad k=1,2,\dots,K$$

朴素贝叶斯法的参数估

学习与分类算法Naïve Bayes Algorithm:

输入: 训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})^T$

$x_i^{(j)}$ 是第*i*个样本的第*j*个特征, $x_i^{(j)} \in \{a_{j1}, a_{j2}, \dots, a_{jS_j}\}$

a_{jl} 是第*j*个特征可能取的第*l*个值, $j=1, 2, \dots, n$; $l=1, 2, \dots, S_j$ $y_i \in \{c_1, c_2, \dots, c_K\}$

输出: x 的分类

朴素贝叶斯法的参数估

步骤:

1、计算先验概率和条件概率

$$P(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k)}{N}, \quad k = 1, 2, \dots, K$$
$$P(X^{(j)} = a_{jl} | Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl}, y_i = c_k)}{\sum_{i=1}^N I(y_i = c_k)}$$
$$j = 1, 2, \dots, n; \quad l = 1, 2, \dots, S_j; \quad k = 1, 2, \dots, K$$

2、对于给定的实例, $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})^T$, 计算

$$P(Y = c_k) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_k), \quad k = 1, 2, \dots, K$$

3、确定x的类别

$$y = \arg \max_{c_k} P(Y = c_k) \prod_{j=1}^n P(X^{(j)} = x^{(j)} | Y = c_k)$$

例子

例 4.1 试由表的类标记 y 。表中 $X^{(1)}$, $X^{(2)}$ 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, $y \in C=\{1, -1\}$ 。

4.1 的训练数据学习一个朴素贝叶斯分类器并确定 $x = (2,S)^T$

表 4.1 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

解: 根据算法4.1, 由表4.1, 容易计算下列概率:

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

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

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

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

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

例子

对于给定的 $x=(2, S)^T$ 计算:

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

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

因为 $P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1)$ 最大, 所以 $y= -1$ 。

贝叶斯估计

极大似然估计可能会出现所要估计的**概率值为0**，这时会影响到后验概率的计算结果，分类产生偏差。解决这一问题的方法是采用**贝叶斯估计**。

条件概率的贝叶斯估计：

$$P_{\lambda}(X^{(j)} = a_{jl} | Y = c_k) = \frac{\sum_{i=1}^N I(x_i^{(j)} = a_{jl}, y_i = c_k) + \lambda}{\sum_{i=1}^N I(y_i = c_k) + S_j \lambda} \quad (4.10)$$

式中 $\lambda \geq 0$ 。等价于在**随机变量各个取值的频数上赋予一个正数 $\lambda > 0$** 。

$\lambda = 0$ ，为极大似然估计； $\lambda = 1$ ，为拉普拉斯平滑 (Laplacian smoothing) 。

对任何 $l=1, 2, \dots, S_j$, $k=1, 2, \dots, K$, 有：

$$\sum_{l=1}^{S_j} P(X^{(j)} = a_{jl} | Y = c_k) = 1$$

先验概率的贝叶斯估计：

$$P_{\lambda}(Y = c_k) = \frac{\sum_{i=1}^N I(y_i = c_k) + \lambda}{N + K \lambda} \quad (4.11)$$

例子1

例4.2：问题同例4.1，按照拉普拉斯平滑估计概率，即取 $\lambda=1$ 。

解： $A_1=\{1,2,3\}$, $A_2=\{S,M,L\}$, $C=\{1,-1\}$ 。按式(4.10)和式(4.11)计算下列概率：

$$P(Y=1) = \frac{10}{17}, \quad P(Y=-1) = \frac{7}{17}$$

$$P(X^{(1)}=1|Y=1) = \frac{3}{12}, \quad P(X^{(1)}=2|Y=1) = \frac{4}{12}, \quad P(X^{(1)}=3|Y=1) = \frac{5}{12}$$

$$P(X^{(2)}=S|Y=1) = \frac{2}{12}, \quad P(X^{(2)}=M|Y=1) = \frac{5}{12}, \quad P(X^{(2)}=L|Y=1) = \frac{5}{12}$$

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

$$P(X^{(2)}=S|Y=-1) = \frac{4}{9}, \quad P(X^{(2)}=M|Y=-1) = \frac{3}{9}, \quad P(X^{(2)}=L|Y=-1) = \frac{2}{9}$$

对于给定的 $x=(2,S)^T$ ，计算：

$$P(Y=1)P(X^{(1)}=2|Y=1)P(X^{(2)}=S|Y=1) = \frac{10}{17} \cdot \frac{4}{12} \cdot \frac{2}{12} = \frac{5}{153} = 0.0327$$

$$P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1) = \frac{7}{17} \cdot \frac{3}{9} \cdot \frac{4}{9} = \frac{28}{459} = 0.0610$$

由于 $P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1)$ 最大，所以 $y=-1$ 。 ■

例子2

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

欲测试: $\langle \text{Outlook}=\text{sunny}, \text{Temperature}=\text{cool}, \text{Humidity}=\text{high}, \text{Wind}=\text{strong} \rangle$

$$c(x) = \arg \max_{c \in \{\text{yes}, \text{no}\}} P(c)P(\text{sunny} | c)P(\text{cool} | c)P(\text{high} | c)P(\text{strong} | c)$$

例子2

$$P(yes)=(9+1)/(14+2)=10/16$$

$$P(no)=(5+1)/(14+2)=6/16$$

$$P(sunny | yes)=(2+1)/(9+3)=3/12$$

$$P(sunny | no)=(3+1)/(5+3)=4/8$$

$$P(cool | yes)=(3+1)/(9+3)=4/12$$

$$P(cool | no)=(1+1)/(5+3)=2/8$$

$$P(high | yes)=(3+1)/(9+2)=4/11$$

$$P(high | no)=(4+1)/(5+2)=5/7$$

$$P(strong| yes)= (3+1)/(9+2)=4/11$$

$$P(strong| no)=(3+1)/(5+2)=4/7$$

$$P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes)=0.0069$$

$$P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no)=0.0191$$

课后作业

最近某车行做了一项顾客买车意愿调查，结果如下表所示，请你通过贝叶斯估计（拉普拉斯平滑因子 $\lambda=1$ ）判断小明是否会买一辆红色国产SUV汽车？请书写完整计算过程。

调查编号	颜色	种类	产地	是否购买
1	红色	跑车	国产	是
2	红色	跑车	国产	否
3	红色	跑车	国产	是
4	黄的	跑车	国产	否
5	黄的	跑车	进口	是
6	黄的	SUV	进口	否
7	黄的	SUV	进口	是
8	黄的	SUV	国产	否
9	红色	SUV	进口	否
10	红色	跑车	进口	是