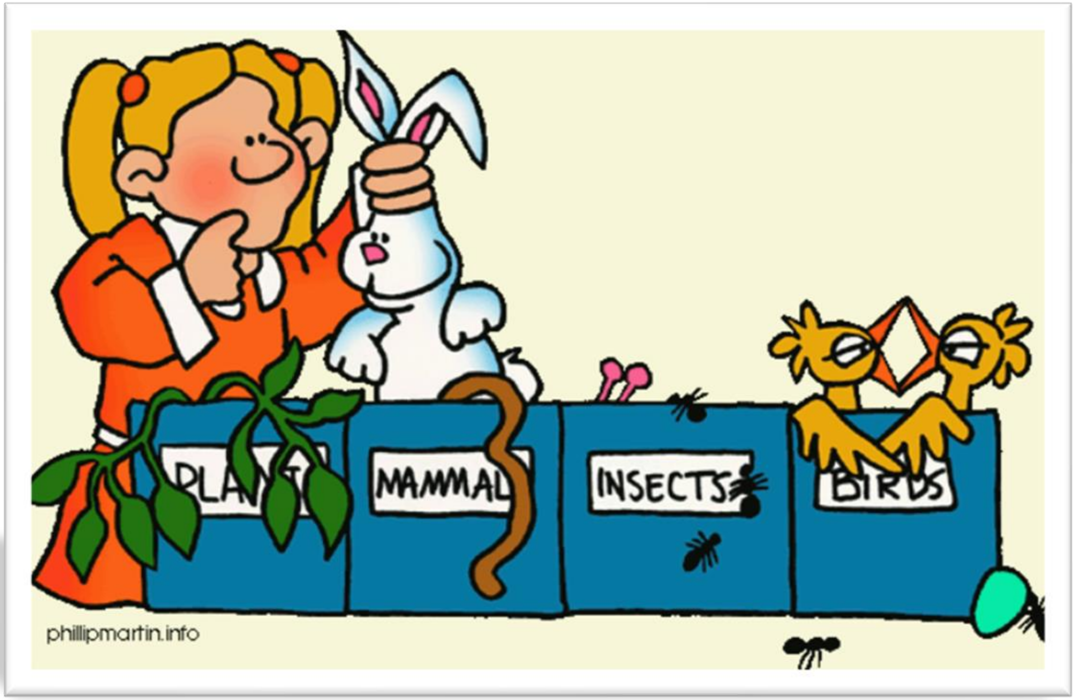
第六章：朴素贝叶斯
Ch6: Naïve Bayes



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。遗著An essay towards solving a problem in the doctrine of chances对现代概率论和数理统计产生了重要的影响。

THE ROYAL SOCIETY
PUBLISHING

All Journals

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

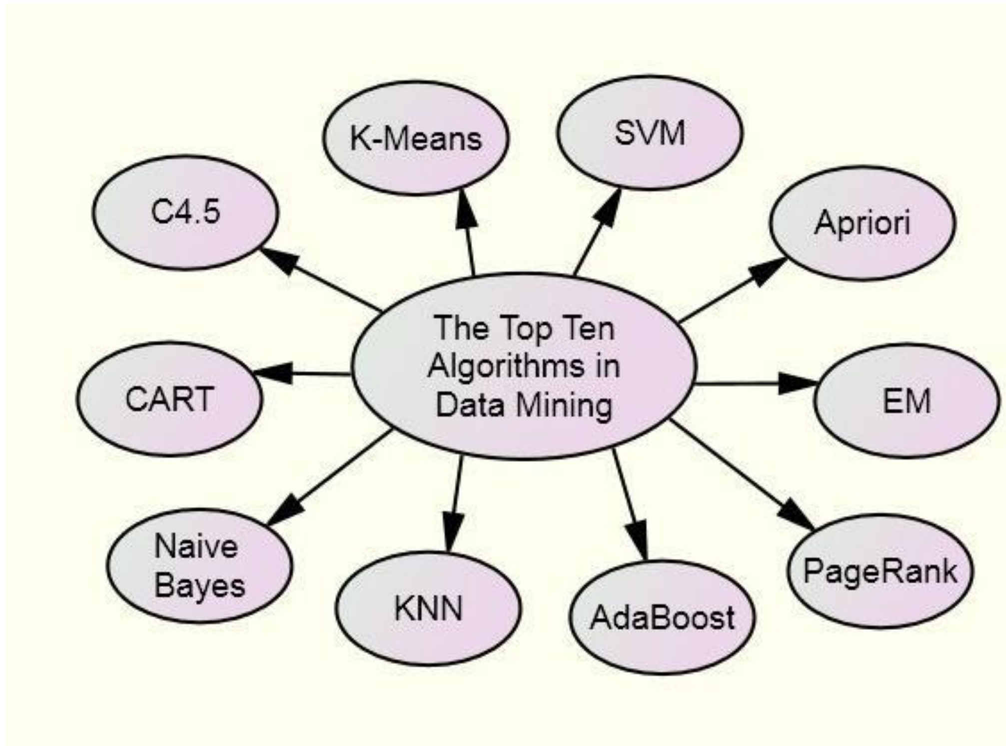
An essay towards solving a problem in the doctrine of chances

“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问题领域，比如癌症的检测、垃圾邮件的过滤。

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



The Top Ten Algorithms in Data Mining
K-Means
SVM
C4.5
Apriori
CART
EM
Naive Bayes
KNN
AdaBoost
PageRank

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：

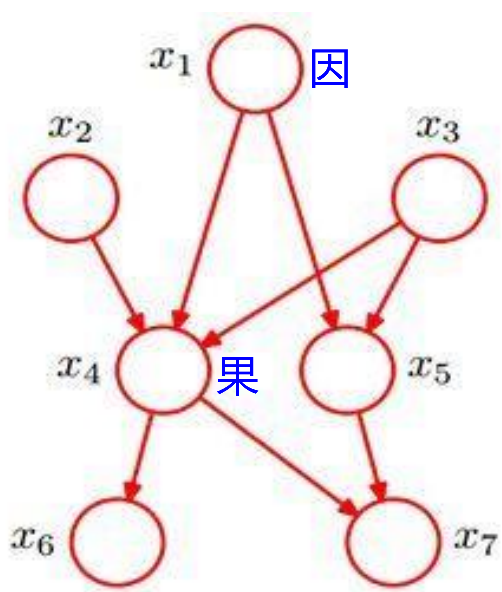
1、已知类**条件概率**密度参数表达式和**先验概率**。

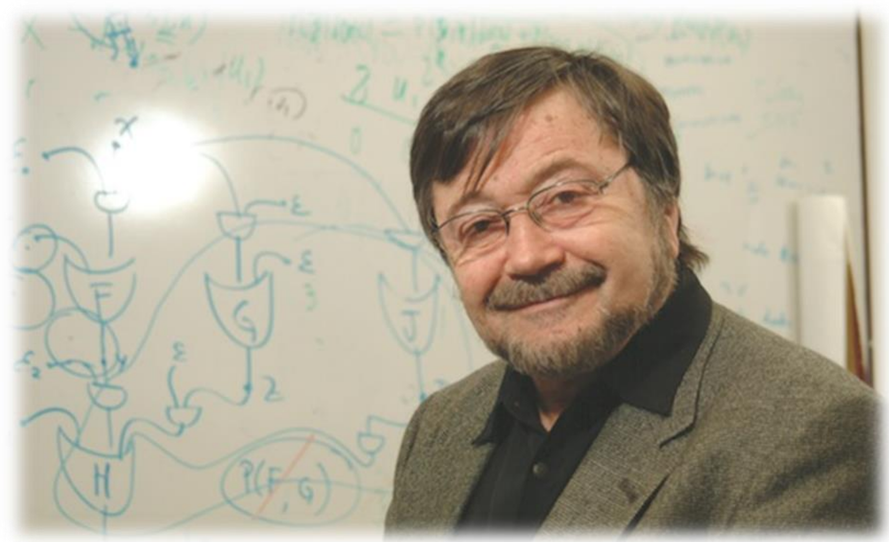
2、利用贝叶斯公式转换成**后验概率**。

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，其网络拓朴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

优点

模型简洁：贝叶斯网络是一种有向无环图，其简洁性使得它在表示概率关系方面非常有效

可解释性强：由于贝叶斯网络的结构清晰，每个节点表示一个随机变量，每条边表示一个条件依赖关系，因此贝叶斯网络具有较强的可解释性

可扩展性好：贝叶斯网络可以轻松地扩展到包含更多变量的模型，因此它在处理复杂问题方面具有很大的优势

可计算性强：贝叶斯网络的结构使得许多问题可以通过简单的计算得到解决，例如计算条件概率、最大后验概率等

缺点

数据需求：贝叶斯网络需要大量的数据来估计参数，因此在实际应用中需要收集大量的数据。

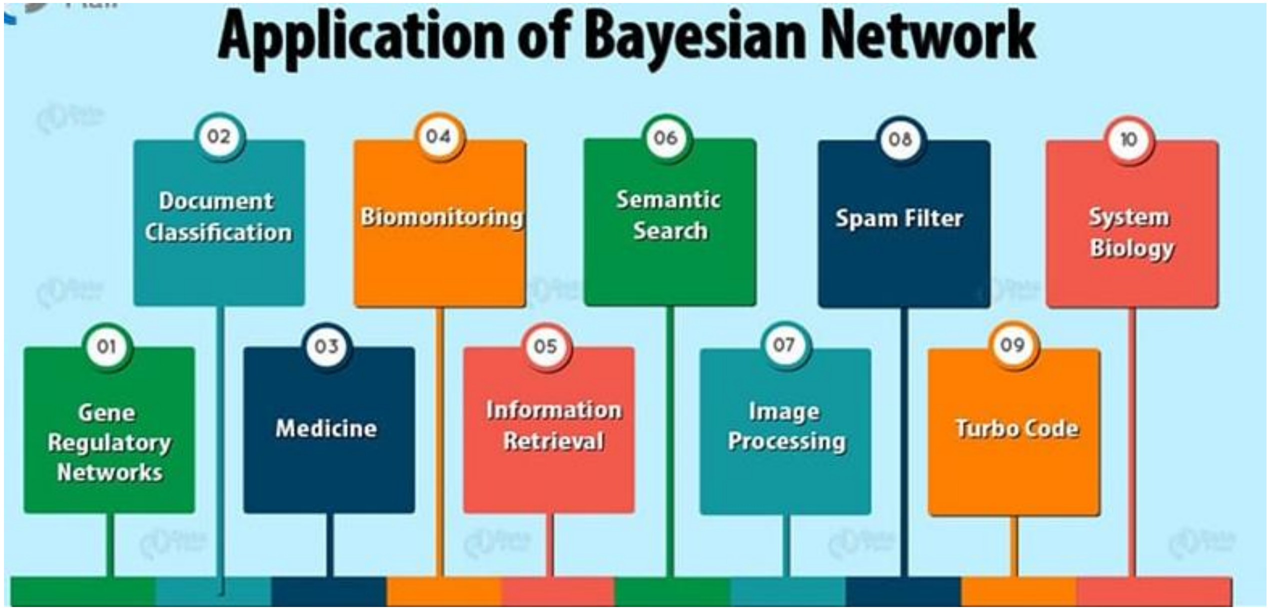
模型选择：贝叶斯网络需要选择合适的模型结构，这可能需要大量的试错和调整

计算复杂性：在某些情况下，贝叶斯网络的计算复杂性可能较高，需要使用高效的算法来解决

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。



https://data-flair.training/blogs/bayesian-network-applications

贝叶斯网络的应用

Find articles with these terms

bayesian network

Advanced search

121,903 results

Set search alert

Refine by:

Subscribed journals

Years

2025 (15)

2024 (10,669)

2023 (15,063)

Show more

Article type

Review articles (9,801)

Research articles (93,864)

Encyclopedia (989)

Book chapters (5,483)

Download selected articles   Export   sorted by releva

Research article • Full text access

1 Fuzzy Bayesian network analysis for quantifying risk reduction rate of Hierarchy of Controls

Journal of Loss Prevention in the Process Industries, Available online 14 May 2024

Mi-Jeong Lee, Sejong Bae, ... Jong Bae Baek

View PDF   Abstract   Extracts   Figures   Export

Research article • *Open access*

2 Being Bayesian about learning Bayesian networks from ordinal data

International Journal of Approximate Reasoning, July 2024

Marco Grzegorczyk

View PDF   Abstract   Extracts   Figures   Export

Get a personalized search experience

Recommendations, reading history, search & journals alerts, and more registration benefits.

Personalize

Suggested topic

Bayesian Network

in Mathematics

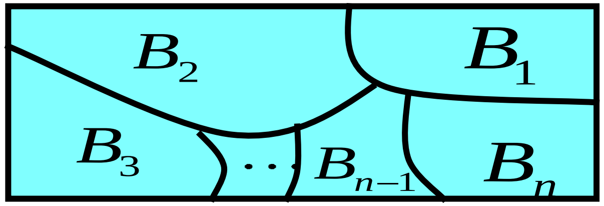
概率(回顾)

样本空间的划分

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

1^0 $B_i \cap B_j = \emptyset, i, j = 1,2,\cdots,n;$

2^0 $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$



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

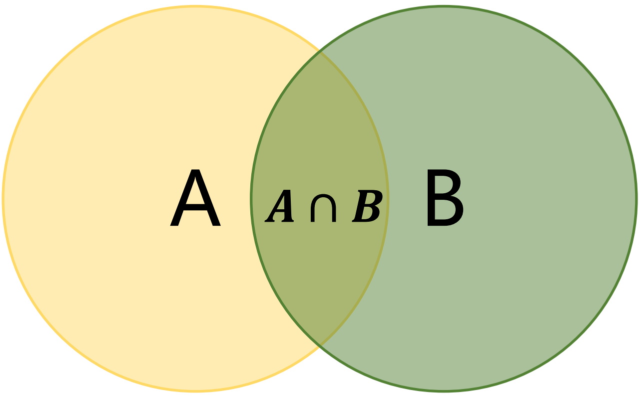
设$A_1, A_2, \dots, A_n$两两互不相容，则

$$P\left(\bigcup_{i}^{n} A_i\right) = \sum_{i}^{n} P(A_i)$$

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。

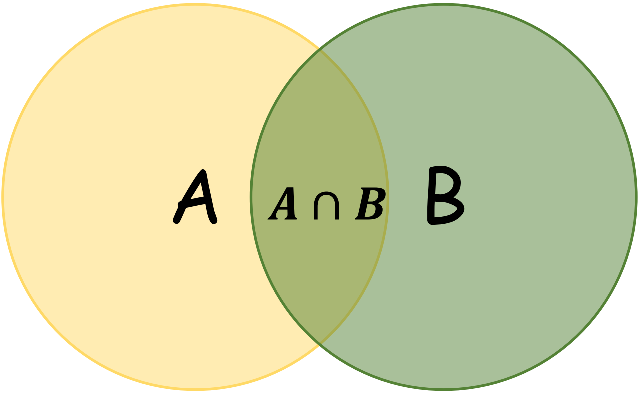


可以很清楚地看到在事件B发生的情况下，事件A发生的概率就是$P(A\cap B)$除以$P(B)$，即如下：$P(A|B)=\frac{P(A\cap B)}{P(B)}$ 或者，$P(A|B)=\frac{P(AB)}{P(B)}$

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)



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：$P(A\cap B)=P(AB)=P(B|A)\,P(A)$

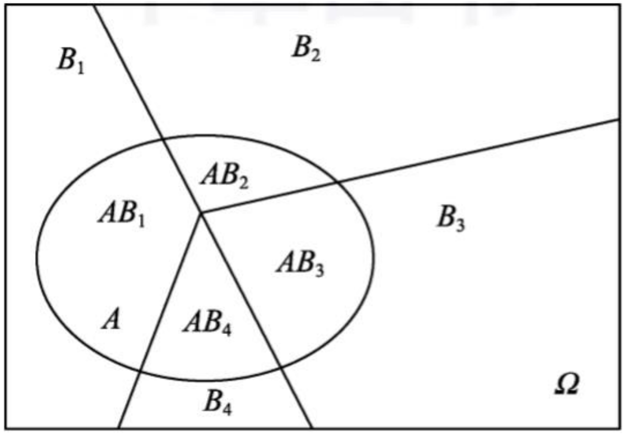
最后得到：$P(B|A)\,P(A)=P(A|B)\,P(B)$ 或者，$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$

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

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n) = \sum_{i=1}^{n} P(B_i)P(A \mid B_i)$$



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：

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

$$=P(A)P(B)$$

贝叶斯定理

“逆概率思维”

$$P(B_i \mid A) = \frac{P(AB_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{P(A)}$$

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

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

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

后验概率＝先验概率×调整因子

先验概率＋数据＝后验概率

似然函数

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

先验概率

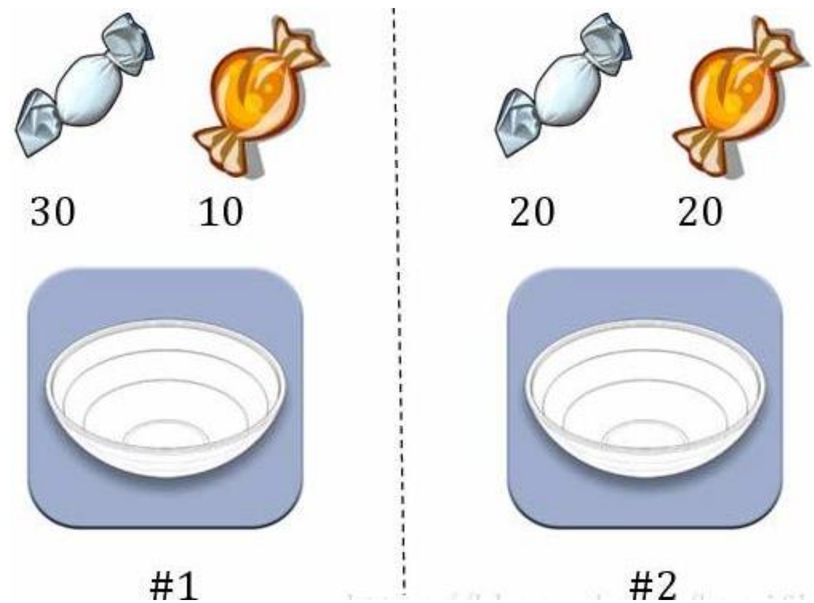
可能性函数（调整因子）

后验概率

*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

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，H1表示#1碗，H2表示2#碗。由于这两个碗是一样的，所以*P*(H1)=*P*(H2)，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*(H1)=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*(H1|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*(H1)的修正。根据条件概率公式，得到：

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

已知，*P*(H1)=0.5，*P*(E|H1)=0.75，根据全概率公式，

$$P(E)=P(E|H_1)P(H_1)+P(E|H_2)P(H_2)=0.75\times0.5+0.5\times0.5=0.625$$

E来自#1碗的概率是：$P(H1|E)=0.5\times\dfrac{0.75}{0.625}=0.6$

也就是说，取出水果糖之后，H1事件的可能性得到了增强。

贝叶斯思维：例子

主观判断

贝叶斯思维

主观判断
- 先验概率

添加新信息
- 调整因子

最终结论
- 后验概率

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

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$\mathcal{X} \subseteq \mathbb{R}^n$为$n$维向量的集合，输出空间为类标记集合$\mathcal{Y} = \{c_1, c_2, \cdots, c_K\}$。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

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

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：$P(Y = c_k)$, $k = 1, 2, \cdots, K$

条件概率分布：

$$P(X = x \mid Y = c_k) = P(X^{(1)} = x^{(1)}, \cdots, X^{(n)} = x^{(n)} \mid Y = c_k), k = 1, 2, \cdots, K$$

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，假设想$x^{(j)}$可取值有S_j个 $j=1, 2, \ldots, n$, 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

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

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），条件随机场（Conditional Random Field）

基本方法

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

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

- 贝叶斯定理： $P(Y=c_k\mid X=x)=\dfrac{\boxed{P(X=x\mid Y=c_k)}P(Y=c_k)}{\sum_k P(X=x\mid Y=c_k)P(Y=c_k)}$

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

朴素贝叶斯法分类的基本公式

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

基本方法

朴素贝叶斯法分类的基本公式：

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

朴素贝叶斯分类器可表示为：

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

等价于：$y=f(x)=\arg\max_{c_k} P(Y=c_k)\prod_{j=1}^{n} P(X^{(j)}=x^{(j)} \mid Y=c_k)$

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 $L(Y, f(X)) = \begin{cases} 1, & Y \neq f(X) \\ 0, & Y = f(X) \end{cases}$

- 期望风险函数： $R_{\exp}(f) = E[L(Y, f(X))]$

- 取条件期望： $R_{\exp}(f) = E_X \sum_{k=1}^{K} [L(c_k, f(X))] P(c_k \mid X)$

后验概率最大化的含义

- 只需对X=x逐个极小化，得：$$\begin{aligned} f(x) &= \arg\min_{y\in Y} \sum_{k=1}^{K} L(c_k, y) P(c_k \mid X = x) \\ &= \arg\min_{y\in Y} \sum_{k=1}^{K} P(y \neq c_k \mid X = x) \\ &= \arg\min_{y\in\mathcal{Y}} (1 - P(y = c_k \mid X = x)) \\ &= \arg\max_{y\in\mathcal{Y}} P(y = c_k \mid X = x) \end{aligned}$$

- 推导出后验概率最大化准则：$f(x) = \arg\max_{c_k} P(c_k \mid X = x)$

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：$P(Y=c_k)=\frac{\sum_{i=1}^{N} I(y_i=c_k)}{N}, k=1,2,\cdots,K$

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：$\{a_{j1}, a_{j2}, \cdots, a_{jS_j}\}$

- 条件概率的极大似然估计：$P(X^{(j)}=a_{ji} \mid Y=c_k)=\frac{\sum_{i=1}^{N} I(x_i^{(j)}=a_{ji}, y_i=c_k)}{\sum_{i=1}^{N} I(y_i=c_k)}$

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

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

其中 $x_i=(x_i^{(1)},x_i^{(2)},\cdots,x_i^{(n)})^{\mathrm{T}}$ $y_i\in\{c_1,c_2,\cdots,c_K\}$

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

a_{jl} 是第j个特征可能取的第*l*个值, *j*=1,2,…,*n*; l=1,2,…, S_j,

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：

$$P(Y=c_k)=\frac{\sum_{i=1}^{N} I(y_i=c_k)}{N}, k=1,2,\cdots,K$$

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

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

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

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

3、确定x的类别： $y=\arg\max_{c_k} P(Y=c_k)\prod_{j=1}^{n} P(X^{(j)}=x^{(j)} \mid Y=c_k)$

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:

$$P(Y=1)P(X^{(1)}=2|Y=1)P(X^{(2)}=S|Y=1)=\frac{9}{15}\bullet\frac{3}{9}\bullet\frac{1}{9}=\frac{1}{45}$$
$$P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1)=\frac{6}{15}\bullet\frac{2}{6}\bullet\frac{3}{6}=\frac{1}{15}$$

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λ≥0。等价于在**随机变量各个取值的频数上赋予一个正数**λ>0。

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

对任何l=1, 2,…, S_j，k=1, 2,…, K, 有：

$$P_\lambda(X^{(j)}=a_{jl}|Y=c_k)>0$$

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

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λ=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：

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

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

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

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

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

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： $< Outlook = sunny, Temperature = cool, Humidity = high, Wind = strong >$

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

例子2

$$P(yes) = (9+1)/(14+2) = 10/16$$
$$P(sumny \mid yes) = (2+1)/(9+3) = 3/12$$
$$P(cool \mid yes) = (3+1)/(9+3) = 4/12$$
$$P(high \mid yes) = (3+1)/(9+2) = 4/11$$
$$P(strong|yes) = (3+1)/(9+2) = 4/11$$

$$P(no) = (5+1)/(14+2) = 6/16$$
$$P(sumny \mid no) = (3+1)/(5+3) = 4/8$$
$$P(cool \mid no) = (1+1)/(5+3) = 2/8$$
$$P(high \mid no) = (4+1)/(5+2) = 5/7$$
$$P(strong \mid no) = (3+1)/(5+2) = 4/7$$

$$P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes) = 0.0069$$
$$P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no) = 0.0191$$

作业4

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）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