第2节 贝叶斯分类方法

——贝叶斯基本原理

一、贝叶斯分类原理

一个例子：

某汽车公司下属有两个汽车制造厂，如下表：

厂名	产品份额	次品率
甲厂	40%	1%
乙厂	60%	2%

问：从公司生产的汽车中随机抽取一辆为次品，更可能是哪个厂生产的？

条件概率：

在参数θ 存在的条件下，事件 x 发生的概率称为在θ 存在的条件下 x 发生的条件概率，并记为：

$$p(x \mid \theta) = \frac{p(x,\theta)}{p(\theta)}$$

取：

参数 θ_1 = 甲厂出产

参数 θ_2 = 乙厂出产

事件 x = 出现次品

例子问题应求：$p(\theta_1 | x)$和$p(\theta_2 | x)$，与条件概率不同。

通过比较上述两个概率的大小，就可以知道次品最可能是什么厂生产的。这就是贝叶斯分类的原理。

1. Bayes概率

1) 先验概率：

根据历史资料或主观判断所确定的参数θ分布概率 $\pi(\theta)$，称为先验概率，或者主观概率。

本例中：

$$\pi(\theta_1) = 0.4$$
$$\pi(\theta_2) = 0.6$$

称为汽车出厂先验概率

2) 似然函数，联合概率密度：

在参数θ存在的条件下，样本数据 x 的概率分布情况，标记为 $L(\theta')$，或者 $p(x \mid \theta')$。

当样本 $x=(x_1, x_2, \ldots, x_n)$的任意两个分量间两两独立时，可以按下式计算：

$$p(x \mid \theta') = \prod_{i=1}^{n} p(x_i \mid \theta')$$

本例中，

$$p(x \mid \theta_1') = 0.01$$
$$p(x \mid \theta_2') = 0.02$$

为通过抽样调查后，获得的新的关于汽车出厂样本的信息。

3) 联合概率：

综合先验分布 $\pi(\theta)$ 和联合概率密度 $p(x|\theta')$，可得到样本x 和参数θ的联合分布概率，记为：

$$h(x,\theta) = p(x|\theta') \cdot \pi(\theta)$$

本例中，

$$h(x,\theta_1) = p(x|\theta_1') \cdot \pi(\theta_1) = 0.4 \times 0.01 = 0.004$$

$$h(x,\theta_2) = p(x|\theta_2') \cdot \pi(\theta_2) = 0.6 \times 0.02 = 0.012$$

4) 边缘概率：

把联合分布概率对变量θ 积分，则得到关于事件 x 的边缘概率 $m(x)$，也称为全概率。

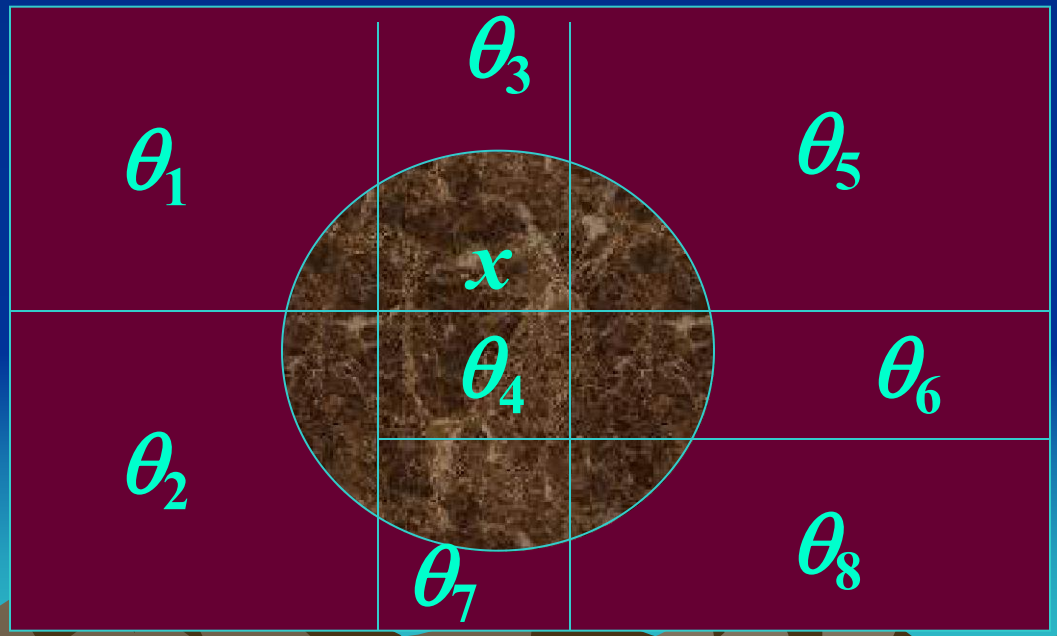
$$m(x)=\int_D h(x,\theta)d\theta=\int_D p(x\mid\theta)\cdot\pi(\theta)d\theta$$

本例中，

$$m(x)=h(x,\theta_1)+h(x,\theta_2)$$

$$=p(x\mid\theta_1')\cdot\pi(\theta_1)+p(x\mid\theta_2')\cdot\pi(\theta_2)=0.016$$

因此，全概率(边缘概率)是“从原因推结果”，即每个原因都对结果的发生有一定的作用，从而在求取结果的时候需要对全部的子原因引起的结果进行累加和。



每个θ_i都是产生x结果的原因

5) 后验概率：

利用Bayes公式，结合采样等方法获取了新的附加信息，对先验概率进行修正后得到的更符合实际的概率，也称为逆概率，标记为 $\pi(\theta \mid x)$。

本例中，次品汽车出厂后验概率 $\pi(\theta_1 \mid x)$和 $\pi(\theta_2 \mid x)$ 可以采用贝叶斯定理计算得出。

2. 贝叶斯(Bayes)公式

$$\pi(\theta \mid x) = \frac{p(x \mid \theta) \cdot \pi(\theta)}{\sum_{D} p(x \mid \theta) \cdot \pi(\theta)}$$

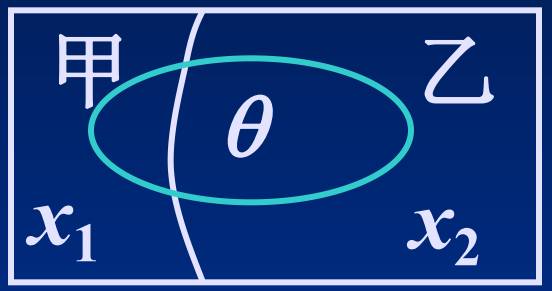
该公式于1763年由贝叶斯(Bayes)给出。它是在观察到事件 x 已发生的条件下，寻找导致 x 发生的参数 θ 的概率分布。

贝叶斯定理的意义在于：

我们很容易直接得出 $p(x|\theta)$，即在参数θ出现的条件下，事件 x 出现的条件概率。 但是，$p(\theta|x)$ 则很难根据条件概率公式直接得出。贝叶斯定理则为我们提供了根据先验分布 $\pi(\theta)$和已观察到的样本数据 $p(x|\theta)$来推断参数θ的分布 $p(\theta|x)$的方法。

对数据量大的问题十分适用，在云计算和大数据时代再次成为研究热点。

根据贝叶斯公式，可得：



$$\pi(\theta_1 \mid x) = \frac{h(x,\theta_1)}{m(x)}$$

$$= \frac{\pi(\theta_1)\cdot p\left(x \mid \theta_1'\right)}{\pi(\theta_1)\cdot p\left(x \mid \theta_1'\right) + \pi(\theta_2)\cdot p\left(x \mid \theta_2'\right)}$$

$$= \frac{0.4\times 0.01}{0.4\times 0.01 + 0.6\times 0.02} = 0.25$$

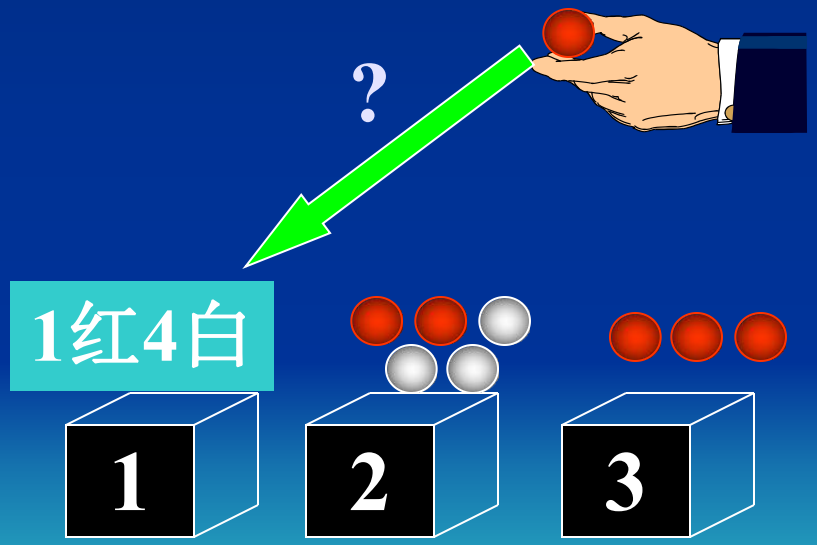
$$\pi(\theta_2 \mid x) = \frac{h(x,\theta_2)}{m(x)}$$

$$= \frac{\pi(\theta_2)\cdot p\left(x \mid \theta_2'\right)}{\pi(\theta_1)\cdot p\left(x \mid \theta_1'\right) + \pi(\theta_2)\cdot p\left(x \mid \theta_2'\right)}$$

$$= \frac{0.6\times 0.02}{0.4\times 0.01 + 0.6\times 0.02} = 0.75$$

由于 $\pi(\theta_1 \mid x) < \pi(\theta_2 \mid x)$，所以次品汽车更有可能出自乙厂。

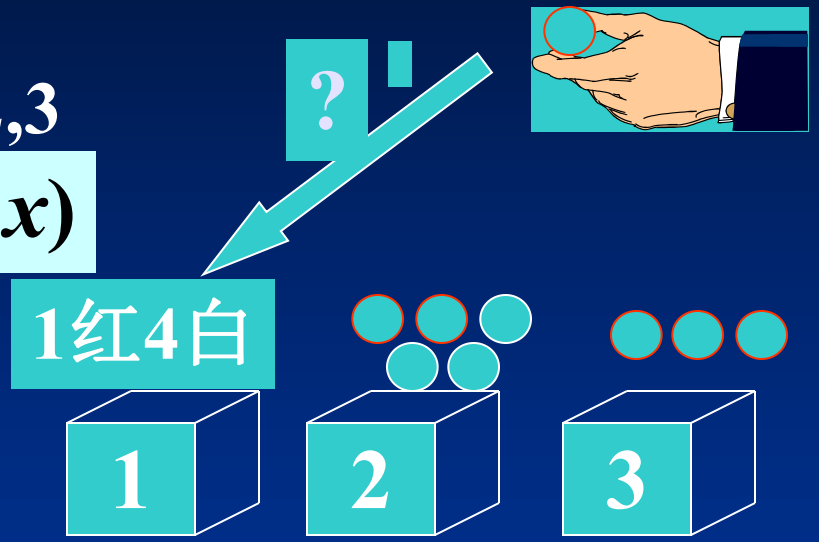
又例如：有三个箱子，分别编号为1,2,3，1号箱装有1个红球4个白球，2号箱装有2红球3白球，3号箱装有3红球. 某人从三箱中任取一箱，从中任意摸出一球，发现是红球,求该球是取自1号箱的概率 .



记 θ_i={球取自i号箱}, i =1,2,3

x ={取得红球}

求$\pi(\theta_1|x)$



由题可知：

$\pi(\theta_1)=1/3$ $=\pi(\theta_2)$ $=\pi(\theta_3)$

$p(x|\theta_1)=1/5$ $p(x|\theta_2)=2/5$ $p(x|\theta_3)=1$

$$\pi(\theta_1 \mid x)=\frac{h(\theta_1,x)}{m(x)}$$

$$=\frac{\pi(\theta_1)\cdot p(x\mid\theta_1)}{\pi(\theta_1)\cdot p(x\mid\theta_1)+\pi(\theta_2)\cdot p(x\mid\theta_2)+\pi(\theta_3)\cdot p(x\mid\theta_3)}$$

=0.125

3. Bayes方法

(1) 用概率表示形式

学习或其他形式的推理都是用概率规则来实现。

(2)Bayes学习结果

表示为随机变量的概率分布，即对不同可能性的信任度。

(3)Bayes先验概率假设

如果随机变量的先验分布未知，假设其服从均匀分布，即参数在其变化范围内，取各个值的机会是相同的。

4. 连续函数的Bayes定理

假定随机向量 x，θ 的联合分布密度是 $h(x,\theta)$，边缘概率为 $m(x)$。

一般假设 x 是观测向量，θ 是未知参数向量。通过对观测向量 x 采样，来获取对未知参数向量 θ 的估计。

连续型的Bayes定理表示为:

$$\pi(\theta \mid x) = \frac{h(x,\theta)}{m(x)} = \frac{\pi(\theta)\cdot p(x \mid \theta)}{\int \pi(\theta)\cdot p(x \mid \theta)d\theta}$$

其中，$\pi(\theta)$是 θ 的先验分布。

5. Bayes方法对未知参数向量θ的估计

1) 将未知参数θ 看成随机向量；(这是与传统估计方法的最大区别)

2) 根据对参数θ 的知识，假设其先验分布$\pi(\theta)$；

3) 根据贝叶斯公式及观察样本 x，计算其后验分布密度，对未知参数θ 进行估计。

二、Bayes问题求解过程

1. 先验分布的求取

求先验概率分布是Bayes求解问题的第一步，也是较关键的一步。

设 θ 为待估计参数，事件 $x = (x_1, x_2, \ldots\ldots, x_n)$ 为观测数据。$\pi(\theta)$ 是参数 θ 的先验分布。

先验分布的选取原则：

- ✓共轭分布
- ✓杰弗莱原则
- ✓最大熵原则

1) 共轭分布族

如果先验分布采取共轭分布族的函数，则根据贝叶斯公式计算得到的后验分布将会有与先验分布相同的函数形式。

共轭分布 —— 后验分布与先验分布属于同一分布类型的分布。

共轭先验分布的优势：

$$\pi(\theta \mid x) = \frac{p(x \mid \theta) \cdot \pi(\theta)}{m(x)}$$

从贝叶斯公式可以看出，由于$p(x \mid \theta)$、$m(x)$项都没有参数θ，可以被当做正则化常数。

此时，θ的后验分布$\pi(\theta \mid x)$正比于其先验分布$\pi(\theta)$，如果先验分布是共轭函数族里的函数，则可以很快地补出所需的常数项，从而计算出后验分布。

2) 最大熵原则

熵的定义：

设随机变量 x 是离散的，它可取 $a_1, a_2, \ldots a_i, \ldots,$ 个值，且 $p(x = a_i) = p_i, i = 1, 2, \ldots$ 则

$$H(x) = -\sum_i p_i \ln p_i$$

称为 x 的熵。

对于连续型随机变量 x，它的概率密度函数为 $p(x)$。若积分：

$$H(x) = -\int p(x) \ln p(x) dx$$

有意义，则称其为连续随机变量的熵。

熵是信息论中描述事物不确定性的程度的一个概念。如果一个随机变量只取与两个不同的值，
比较下面两种情况：

(1) $p(x = a_1) = 0.98$ $p(x = a_2) = 0.02$
(2) $p(x = a_1) = 0.45$ $p(x = a_2) = 0.55$

很明显，(1)的不确定性要比(2)的不确定性小得多，而且从直觉上也可以看得出当取的两个值得概率相等时，不确定性达到最大。

最大熵原则：

在没有先验知识的应用中，未知参数θ的先验分布应取其变化范围内熵最大的分布。

设随机变量θ只取有限个值$a_1, a_2, \ldots, a_n$，其相应的先验概率记为π_1，π_2，…，π_n，θ的熵$H(\theta)$最大的充分必要条件是：

$$\pi_1 = \pi_2 = \ldots\ldots = \pi_n = \frac{1}{n}$$

3. 杰弗莱原则

杰弗莱原则较好地解决了Bayes方法中的一个矛盾：即若对参数θ选用均匀分布，其函数$g(\theta)$往往不服从均匀分布。

(1) 有些场合要求参数θ与其函数 $g(\theta)$的先验分布有相同的函数形式；

(2) 杰弗莱原则给出满足上述要求的具体方法，以求取适合要求的先验分布。

$N(\theta,\delta^2)$

2. 学习机制和问题求解步骤

1) Bayes公式的学习机制

以正态分布为例，分析以Bayes公式求后验分布的学习机制。

设$x=\{x_1, x_2, \ldots, x_n\}$是正态分布随机变量$N(\theta, \sigma_1^2)$的一个样本，其中$\sigma_1^2$是方差，它是一个已知的量，$\theta$未知，求它的估计量　。

该正态分布样本x的似然函数为：

$$p(x|\theta)=\left(\frac{1}{\sqrt{2\pi}\sigma_1}\right)^n \cdot \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}$$

求解过程：

(1) 取另一个正态分布$N(\mu_0, \sigma_0^2)$作为该待估计参数θ的先验分布，则有先验分布为：

$$\pi(\theta)=\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)\cdot\exp\left\{-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right\},-\infty<\theta<+\infty$$

由此可以写出样本x与参数θ的联合概率密度：

$$h(x,\theta)=k_1\cdot\exp\left\{-\frac{1}{2}\left[\frac{n\theta^2-2n\theta\cdot\bar{x}+\sum_{i=1}^{n}x_i^2}{\sigma_1^2}+\frac{\theta^2-2\mu_0\theta+\mu_0^2}{\sigma_0^2}\right]\right\}$$

取

$$A=\frac{\pi}{\sigma_1^2}+\frac{1}{\sigma_0^2},B=\frac{\pi\cdot x}{\sigma_1^2}+\frac{\mu_0}{\sigma_0^2},C=\frac{\pi}{\sigma_1^2}\cdot\sum_{i=1}^{n}x_i^2+\frac{\mu_0}{\sigma_0^2}$$

则可得到：

$$h(x,\theta)=k_1\cdot\exp\left\{-\frac{1}{2}\left[A\cdot\theta^2-2\theta\cdot B+C\right]\right\}$$

(2) 计算样本 x 的边缘分布：

$$m(x)=\int_{-\infty}^{+\infty}h(x,\theta)d\theta$$

$$=\int_{-\infty}^{+\infty}k_1\cdot\exp\left\{-\frac{1}{2}\left[A\cdot\theta^2-2\theta\cdot B+C\right]\right\}d\theta$$

$$=k_1\cdot\exp\left\{-\frac{1}{2}\left(C-B^2/A\right)\right\}\cdot\left(\frac{2\pi}{\mathrm{A}}\right)^{\frac{1}{2}}$$

(3) 用Bayes公式计算参数 θ 的后验分布 $p(\theta \mid x)$：

$$\pi(\theta|x) = \frac{p(x|\theta)\cdot\pi(\theta)}{m(x)}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma_1}\cdot\frac{1}{\sqrt{2\pi}\sigma_0}\cdot exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i-\theta)^2-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right\}}{k_1\cdot exp\left\{-\frac{1}{2}(C-B^2/A)\right\}\cdot\left(\frac{2\pi}{A}\right)^{\frac{1}{2}}}$$

(4) 样本 x 均值的后验分布仍为正态分布：

$$\pi(\theta \mid \bar{x}) = \left(\frac{2\pi}{A}\right)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{(\theta - B/A)^2}{2/A}\right\} \propto N(a_1, d_1^{\,2})$$

其中：$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$　　（样本均值）

$$a_1 = \left(\frac{1}{\sigma_0^{\,2}}\mu_0 + \frac{n}{\sigma_1^{\,2}}\bar{x}\right) / \left(\frac{1}{\sigma_0^{\,2}} + \frac{n}{\sigma_1^{\,2}}\right)$$

$$d_1^{\,2} = \left(\frac{1}{\sigma_0^{\,2}} + \frac{n}{\sigma_1^{\,2}}\right)^{-1}$$

(4) 用后验分布 $\pi(\theta \mid \bar{x})$ 的数学期望a_1作为参数θ的估计值，得到：

$$\bar{\theta} = a_1 = \left(\frac{1}{\sigma_0^{\,2}}\mu_0 + \frac{n}{\sigma_1^{\,2}}\bar{x}\right)\cdot d_1^{\,2}$$

以上可见：得到θ的估计值 $\bar{\theta}$ 是先验分布中的期望μ_0 和样本均值 $\bar{x}$ 的加权平均。

$\bar{\theta}$ 是由先验分布的方差σ_0^2的倒数 $\frac{1}{\sigma_0^{\,2}}$ 对先验分布期望进行加权，样本分布的方差σ_1^2的 n 倍倒数 $\frac{n}{\sigma_1^{\,2}}$ 对样本均值 $\bar{x}$ 进行加权后形成的。当n越大时，则 $\bar{x}$ 在后验中的作用越大，而先验分布 μ_0 的影响越小。

2) Bayes 方法的问题求解步骤：

① 定义随机变量

将未知参数看成随机变量θ， $h(x_1, x_2, \ldots, x_n, \theta)$ 做为样本$x_1, x_2, \ldots, x_n$ 在 θ 条件下的联合分布密度，则样本$x_1, x_2, \ldots, x_n$ 的条件分布函数记为：

$$p(x_1, x_2, \ldots, x_n \mid \theta) \text{ 或 } p(x \mid \theta)$$

② 确定先验分布密度 $\pi(\theta)$，采用共轭先验分布的方法。如果对先验分布没有任何知识，则采用无信息先验分布的Bayes假设。

③ 用Bayes 定理计算后验分布函数$\pi(\theta|x)$。

④ 利用计算所得的后验分布值对所求的问题进行推断。