

模式识别

第三章 概率密度函数的估计

郭园方

北京航空航天大学计算机学院

贝叶斯决策的实际使用

* 计算贝叶斯后验概率进行决策如何实现?

$$P(\omega_i | x) = \frac{p(x | \omega_i) P(\omega_i)}{\sum_{j=1}^2 p(x | \omega_j) P(\omega_j)}$$

* 实现中有问题吗?能直接计算吗?

贝叶斯决策的实际使用

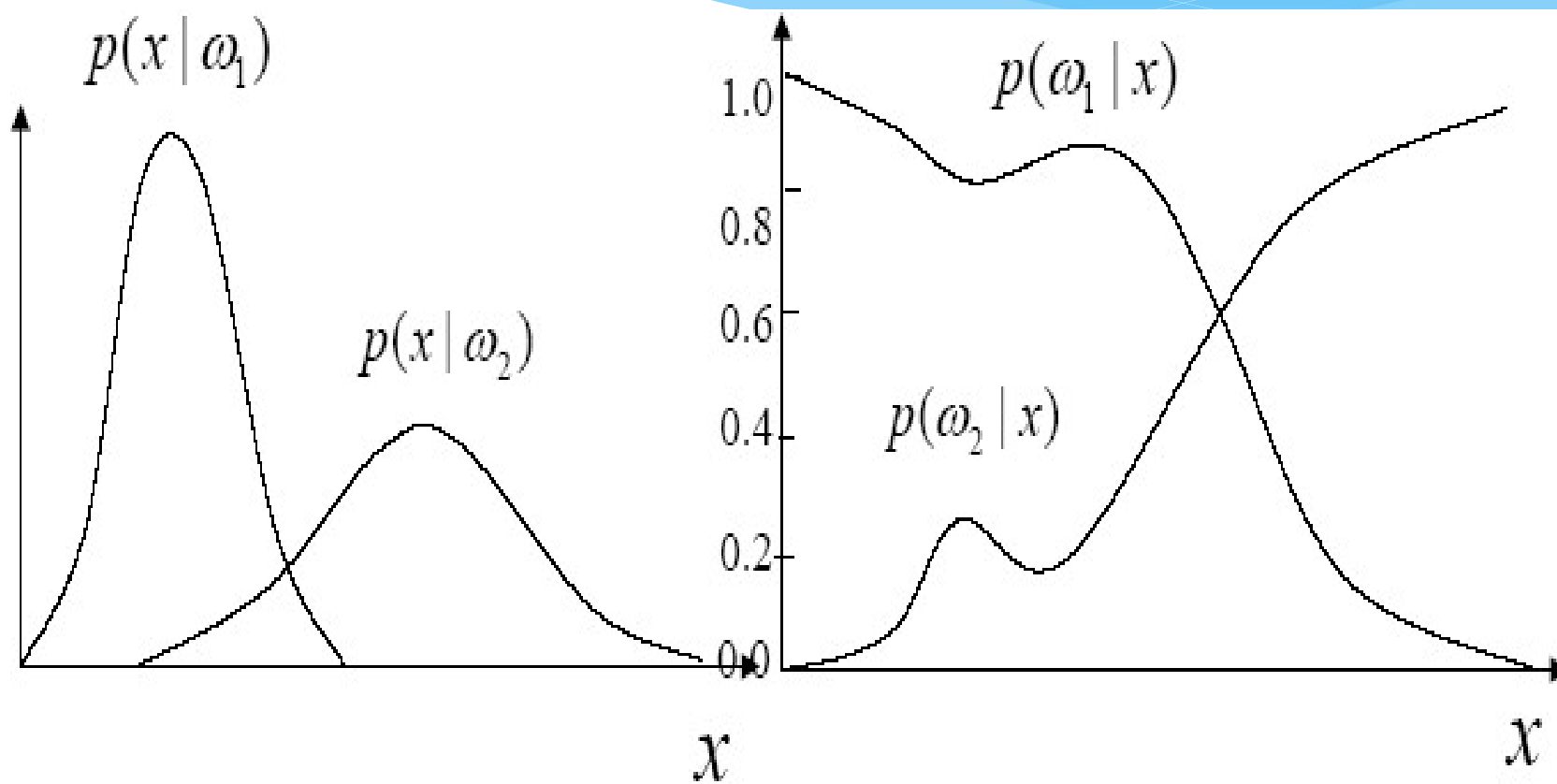
- * 以第二章医生诊病为例：

$$P(\omega_1), P(\omega_2)$$

$$P(x|\omega_1), P(x|\omega_2)$$

- * 先验概率未知,医生可大致估计
- * 类条件概率密度呢?
- * 只能通过已知的一些数据去估计

基于样本的两步贝叶斯决策



类条件概率密度函数

后验概率

基于样本的两步贝叶斯决策

- * 利用样本集估计 $P(\omega_i)$ 和 $p(x|\omega_i)$
- * 得到 $\hat{P}(\omega_i)$ 和 $\hat{p}(x|\omega_i)$
- * 然后将估计量带入贝叶斯决策规则中
- * 得到决策结果

基于样本的两步贝叶斯决策

- * 首先通过训练样本估计概率密度函数
 - * 先验概率估计 - 训练样本分布情况/根据领域知识认定
 - * 类条件概率密度估计难，本章重点
- * 统计决策进行类别判定
- * 注意！与上一章的贝叶斯决策有区别，什么区别？

基于样本的两步贝叶斯决策

- * 训练样本有限，难以涵盖所有情况！
- * 当训练样本无限多，基于样本的两步贝叶斯决策可无限趋近第二章中的理论贝叶斯决策

当 $N \rightarrow \infty$

$$\begin{aligned}\hat{P}(\omega_i) &\rightarrow P(\omega_i) \\ \hat{p}(x|\omega_i) &\rightarrow p(x|\omega_i)\end{aligned}$$

概率密度估计方法

- * 由训练样本集估计总体概率密度的方法可分为：
 - * 监督参数估计：样本所属类别及类条件总体概率密度函数形式已知，表征概率密度函数的某些参数未知。
 - * 非监督参数估计：已知总体概率密度函数形式但未知样本所属类别，要求推断出概率密度函数的某些参数。
 - * 非参数估计：已知样本所属类别，未知总体概率密度函数的形式，直接推断概率密度函数本身。

参数估计的基本概念

- * 统计量: 样本中包含着总体的信息, 针对不同要求构造出样本的某种函数, 这种函数在统计学中称统计量。
- * 参数空间: 假设总体概率密度函数形式已知, 未知分布中的参数 θ , θ 的全部可容许值组成的集合称为参数空间, 记为 Θ

参数估计的基本概念

- * 点估计：点估计问题就是要构造一个统计量 $d(x_1, x_2, \dots, x_N)$ 作为参数 θ 的估计 $\hat{\theta}$ 。
- * 估计量： $\hat{\theta}$ 被称为 θ 的估计量。
- * 估计值：如果 $x_1^i, x_2^i, \dots, x_N^i$ 是属于类别 ω_i 的样本的观察值，代入统计量 d 就得到对于第 i 类的 $\hat{\theta}$ 的具体数值，这一数值被称为 θ 的估计值。

参数估计的基本概念

- * 区间估计: 估计出区间 (d_1, d_2) 作为 θ 可能的取值范围, 这个区间叫置信区间, 这类问题称为区间估计。

概率密度估计的评估

- * 如何评估概率密度估计的好坏?
- * 单次抽样得到的估计值与真实值的偏差?
- * 基于平均和方差进行评估较为公平!
- * 常用标准:
 - * 无偏性
 - * 有效性
 - * 一致性

概率密度估计的评估

- * 无偏性： θ 的估计量 $\hat{\theta}(x_1, x_2, \dots, x_N)$ 的数学期望是 θ
- * 渐进无偏： N 趋于无穷时估计具有无偏性
- * 有效性：一种估计比另一种的方差小，此种估计更有效

- * 对于任意正数 ε ，有

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$$

则 $\hat{\theta}$ 是 θ 的一致估计

最大似然估计

基本假设

- * **Maximum Likelihood Estimation.**
- * 参数 θ 是确定（非随机）的而未知的量。
- * 按类别把样本集分开， \mathcal{R}_j 类中的每个样本都是独立地从概率密度为 $p(x|\omega_j)$ 的总体中独立地抽取出来的 - 独立同分布。

最大似然估计

基本假设

- * 类条件概率密度 $p(x|\omega_j)$ 为已知分布，参数向量未知

$$p(x|\omega_j) \rightarrow p(x|\omega_j, \theta_j)$$

- * 假设 \mathcal{R}_i 中不包含关于 $\theta_j (j \neq i)$ 的信息，即不同类别的参数在函数上是独立的。

最大似然估计

似然函数的定义

- * 在一类中独立抽取样本集来估计未知参数
- * 假设某类样本集中包含有 N 个样本 $\mathcal{R} = \{x_1, x_2, \dots, x_N\}$
- * 因样本独立抽取，样本出现在样本集中的概率为

$$P(\mathcal{R}|\theta) = p(x_1, x_2, \dots, x_N|\theta) = \prod_{k=1}^N p(x_k|\theta)$$

最大似然估计

似然函数的定义

* $P(\mathcal{R}|\theta)$ 也被称为相对于样本集 \mathcal{R} 的 θ 的似然函数(likelihood function)

* 似然函数的定义

$$\begin{aligned} l(\theta) &= p(x_1, x_2, \dots, x_N | \theta) \\ &= p(x_1 | \theta) p(x_2 | \theta) \dots p(x_N | \theta) \end{aligned}$$

最大似然估计

似然函数的定义

- * 似然函数示例
- * 假设 $N = 1$ ， x 为一维，均值为 6，方差为 1 的正态分布
- * 最可能出现的样本 $x'_1 = 6$
- * 此时，似然函数为
$$l(\theta) = p(x'_1 | 6, 1) = \max_x p(x | 6, 1)$$

最大似然估计

最大似然估计量

* 令 $l(\theta)$ 为样本集 \mathcal{R} 的似然函数

$$\mathcal{R} = \{x_1, x_2, \dots, x_N\}$$

* 如果

$$\hat{\theta} = d(\mathcal{R}) = d(x_1, x_2, \dots, x_N)$$

$\hat{\theta}$ 是参数空间中能使似然函数最大化的 θ

则 $\hat{\theta}$ 是 θ 的最大似然估计量

最大似然估计

最大似然估计的求解

- * 假设第*i*类样本的类条件概率密度:

$$p(x^i|\omega_i) = p(x^i|\omega_i, \theta^i) = p(x^i|\theta^i)$$

- * 属于*i*类的训练样本为 $X^i = (x_1, x_2, \dots, x_N)$
- * 求 θ^i 的最大似然估计即为将 $p(x^i|\theta^i)$ 看成 θ^i 的函数, 求使它最大时 θ^i 的值

最大似然估计

最大似然估计的求解

- * 由于训练样本独立从总体样本集中抽取的

$$P(X^i | \omega_i, \theta^i) = P(X^i | \theta^i) = \prod_{k=1}^N p(x_k | \theta^i)$$

- * 对N个训练样本出现概率的乘积取对数：

$$\log \prod_{k=1}^N p(x_k | \theta^i) = \sum_{k=1}^N \log p(x_k | \theta^i)$$

最大似然估计

最大似然估计的求解

* 对 θ^i 求导，令其等于0

$$\begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \ddots \\ \frac{\partial}{\partial \theta_p} \end{bmatrix} \sum_{k=1}^N \log P(X_k | \theta^i) = 0$$

最大似然估计

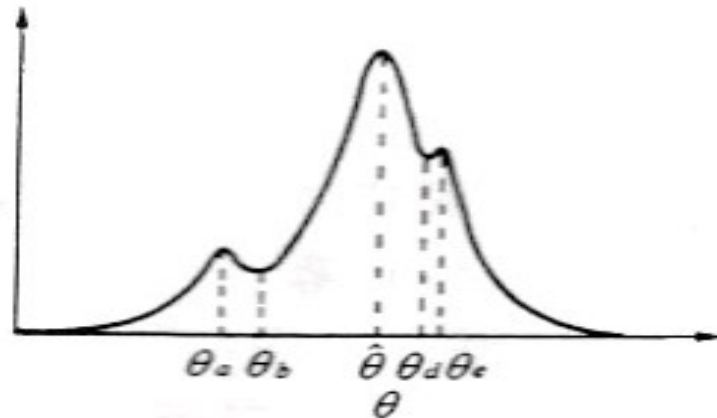
最大似然估计的求解

$$\left\{ \begin{array}{l} \sum_{k=1}^N \frac{\partial}{\partial \theta_1} \log P(X_k | \theta^i) = 0 \\ \dots\dots\dots \\ \dots\dots\dots \\ \sum_{k=1}^N \frac{\partial}{\partial \theta_p} \log P(X_k | \theta^i) = 0 \end{array} \right.$$

最大似然估计

最大似然估计的求解

- * 利用上式可以求出 θ^i 的估值 $\hat{\theta}$
- * 则 $\theta^i = \hat{\theta}$
- * 注意：有时上式是多解的
- * 例：此图有5个解,只有一个解最大



最大似然估计

多维正态分布的最大似然估计

- * 情况1: Σ 已知, μ 未知, 估计 μ
- * $P(X^i|\theta^i)$ 服从正态分布
- * 待估计参数为 $\theta^i = \theta_1 = \mu$
- * 求解方程

$$\sum_{k=1}^N \frac{\partial}{\partial \mu} \log p(x_k|\mu) = 0$$

最大似然估计

多维正态分布的最大似然估计

* 正态分布时

$$p(x_k|\mu) \\ = -\frac{1}{2}\log[(2\pi)^n|\Sigma|] - \frac{1}{2}(x_k - \mu)^T\Sigma^{-1}(x_k - \mu)$$

* 求偏导

最大似然估计

多维正态分布的最大似然估计

* 所以

$$\Sigma^{-1} \sum_{k=1}^N (x_k - N\mu) = 0$$

$$\mu = \hat{\mu} = \frac{1}{N} \sum_{k=1}^N x_k$$

* 说明未知均值的最大似然估计正好是训练样本的算术平均

最大似然估计

多维正态分布的最大似然估计

- * 情况2: Σ, μ 均未知
- * 一维时, $n=1$ 对于每个训练样本只有一个特征的简单情况:

$$\theta_1 = \mu_1, \theta_2 = \sigma_1^2$$

$$\therefore \log P(X_k | \theta^i) = -\frac{1}{2} \log 2\pi \theta_2 - \frac{1}{2\theta_2} (X_k - \theta_1)^2$$

最大似然估计

多维正态分布的最大似然估计

$$\text{代入} \sum_{k=1}^N \frac{\partial}{\partial \theta_1} \log P(X_k | \theta^i) = \sum_{k=1}^N \frac{1}{\theta_2} (X_k - \theta_1) = 0$$

$$\sum_{k=1}^N \frac{\partial}{\partial \theta_2} \log P(X_k | \theta^i) = \sum_{k=1}^N \left[-\frac{1}{2\theta_2} + \frac{(X_k - \theta_1)^2}{2\theta_2^2} \right] = 0$$

$$\therefore \hat{\theta}_1 = \hat{\mu}_1 = \frac{1}{N} \sum_{k=1}^N X_k \quad \text{训练样本的算术平均}$$

$$\therefore \hat{\theta}_2 = \hat{\sigma}_1^2 = \frac{1}{N} \sum_{k=1}^N \left(X_k - \hat{\mu} \right)^2 \quad \text{样本方差}$$

最大似然估计

多维正态分布的最大似然估计

- * 一维的结论:
 - * 正态总体均值的最大似然估计即为训练样本的算术平均
 - * 正态总体方差的最大似然估计与样本的方差不同(一个是算术平均, 一个是期望值), 当 N 较大的时候, 二者的差别不大。
- * 注意, 这两个也正是人们经常使用的对于均值和方差的估计!

最大似然估计

多维正态分布的最大似然估计

- * 多维情况：N个特征
- * 与一维时的情况类似
- * 估计值：

$$\hat{\theta}_1 = \hat{\mu} = \frac{1}{N} \sum_{k=1}^N X_k$$

$$\hat{\theta}_2 = \hat{\Sigma} = \frac{1}{N} \sum_{k=1}^N \left(X_k - \hat{\mu} \right) \left(X_k - \hat{\mu} \right)^T$$

最大似然估计

多维正态分布的最大似然估计

- * 结论:
- * μ 的估计即为训练样本的算术平均
- * 估计的协方差矩阵是如下矩阵的算数平均

$$\left(X_k - \hat{\mu}\right)\left(X_k - \hat{\mu}\right)^T$$

最大似然估计

多维正态分布的最大似然估计

- * 这两个估计是否为无偏的估计量?
- * θ_1 为无偏估计量
- * θ_2 并不是无偏的
- * 小作业：无偏的 θ_2 的估计量是什么?

贝叶斯估计和贝叶斯学习

- * **Bayesian Estimation**

- * 贝叶斯估计实质：贝叶斯决策来决策参数的取值
- * 与最大似然估计同为概率密度估计中的主要参数估计方法
- * 结果多数情况下与最大似然估计相同
- * 区别：
 - * 最大似然估计把待估计的参数当作未知但固定的量
 - * 贝叶斯估计把待估计的参数也看为随机变量

贝叶斯估计和贝叶斯学习

- * **Bayesian Learning**
- * 把贝叶斯估计的原理用于直接从数据对概率密度函数进行迭代估计

贝叶斯估计和贝叶斯学习

最小风险贝叶斯决策回顾

* 回顾最小风险贝叶斯决策

$$\begin{aligned} R(a_i|x) &= E[\lambda(a_i, \omega_j)] \\ &= \sum_{j=1}^c \lambda(a_i, \omega_j) P(\omega_j|x), i = 1, 2, \dots, a \end{aligned}$$

$$R(a_k|x) = \min_{i=1, \dots, a} R(a_i|x)$$

* 则 a_k 就是最小风险贝叶斯决策

贝叶斯估计和贝叶斯学习

贝叶斯估计量

- * $R(\hat{\theta}|x)$ 为给定 x 条件下估计量 $\hat{\theta}$ 的期望损失, 称为条件风险
- * 定义: 如果 θ 的估计量 $\hat{\theta}$ 使得条件风险最小, 则称 $\hat{\theta}$ 是 θ 的贝叶斯估计量

贝叶斯估计和贝叶斯学习

损失函数

- * 决策分类时我们需要事先定义决策风险表即损失表
- * 估计连续随机变量时我们需要定义损失函数
- * 损失函数有许多种，最常见的损失函数为平方误差损失函数

$$\lambda(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

贝叶斯估计和贝叶斯学习

贝叶斯估计

* 定理:

* 如果损失函数为二次函数, 即

$$\lambda(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

* 则 θ 的贝叶斯估计量 $\hat{\theta}$ 为在给定 x 时 θ 的条件期望, 如下

$$\hat{\theta} = E[\theta|x] = \int_{\Theta} \theta p(\theta|x) d\theta$$

贝叶斯估计和贝叶斯学习

参数估计问题

- * 假设有一个样本集 \mathcal{R} ，要求我们找出估计量 $\hat{\theta}$ 用来估计 \mathcal{R} 所属的总体分布的某个真实参数 θ 使带来的贝叶斯风险最小，这就是贝叶斯估计

贝叶斯估计和贝叶斯学习

贝叶斯估计

- * 最大似然估计是把待估的参数看作固定的未知量，而贝叶斯估计则是把待估的参数作为具有某种先验分布的随机变量
- * 通过对第 i 类训练样本 X^i 的观察，使得概率密度分布 $P(X^i|\theta)$ 转化为后验概率 $P(\theta|X^i)$ ，再求贝叶斯估计

贝叶斯估计和贝叶斯学习

贝叶斯估计具体步骤

- * 确定 θ 的先验分布 $P(\theta)$,待估参数为随机变量
- * 用第 i 类样本 $X^i=(X_1, X_2, \dots, X_N)$ 求出样本的联合概率密度分布 $P(X^i|\theta)$, 它是 θ 的函数
- * 利用贝叶斯公式,求 θ 的后验概率

$$P(\theta | X^i) = \frac{P(X^i | \theta) \cdot P(\theta)}{\int_{\theta} P(X^i | \theta) P(\theta) d\theta}$$

- * 最后

$$\text{求贝叶斯估计 } \hat{\theta} = \int_{\theta} \theta P(\theta | X^i) d\theta \text{ (证明略)}$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

- * 一维正态分布:
- * 已知 σ^2 ,估计 μ
- * 假设概率密度服从正态分布
- * $P(X|\mu) = N(\mu, \sigma^2)$
- * $P(\mu) = N(\mu_0, \sigma_0^2)$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

* 第*i*类样本 $X^i=(X_1, X_2, \dots, X_N)$

* 后验概率:

$$P(\mu | X^i) = \frac{P(X^i | \mu) \cdot P(\mu)}{\int P(X^i | \mu) P(\mu) d\mu}$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

* 因为N个样本是独立抽取的：

$$\therefore P(\mu | X^i) = a \prod_{k=1}^N P(X_k | \mu) \cdot P(\mu)$$

其中a为比例因子，只与x有关，与μ无关

$$a = \frac{1}{\int P(X^i | \mu) P(\mu) d\mu}$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

* 由

$$P(X|\mu) = N(\mu, \sigma^2)$$

$$P(\mu) = N(\mu_0, \sigma_0^2)$$

$$\begin{aligned}\therefore P(\mu | X^i) &= a \prod_{k=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{X_k - \mu}{\sigma}\right)^2\right\} \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right] \\ &= a' \exp\left\{-\frac{1}{2}\left[\sum_{k=1}^N \left(\frac{X_k - \mu}{\sigma}\right)^2 + \left(\frac{\mu - \mu_0}{\sigma_0}\right)^2\right]\right\} \\ &= a'' \exp\left\{-\frac{1}{2}\left[\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{1}{\sigma^2} \sum_{k=1}^N X_k + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right\}\end{aligned}$$

其中 a', a'' 包含了所有与 μ 无关的因子

贝叶斯估计和贝叶斯学习

正态分布的均值估计

- * $P(\mu|X^i)$ 是 μ 的二次函数的指数函数
- * $P(\mu|X^i)$ 仍然是一个正态函数, $P(\mu|X^i) = N(\mu_N, \sigma_N^2)$
- * 另外后验概率可以直接写成正态形式:

$$P(\mu | X^i) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_N}{\sigma_N}\right)^2\right]$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

* 比较以上两个式子,对应的系数应该相等

$$\begin{cases} \frac{1}{\sigma_N^2} = \frac{N}{\sigma^2} + \frac{1}{\sigma_0^2} \\ \frac{\mu_N}{\sigma_N^2} = \frac{1}{\sigma^2} \sum_{k=1}^N X_k + \frac{\mu_0}{\sigma_0^2} \end{cases}$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

* 求解，可得：

$$\mu_N = \frac{\sigma_0^2}{N\sigma_0^2 + \sigma^2} \sum_{k=1}^N X_k + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0$$

$$\sigma_N^2 = \frac{\sigma_0^2 \sigma^2}{N\sigma_0^2 + \sigma^2}$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

* 将 μ_N 和 σ_N^2 代入 $P(\mu|X^i)$ 可以得到后验概率，再用

$$\hat{\theta} = \int_{\theta} \theta P(\theta | X^i) d\theta, \text{ 求 } \mu \text{ 的估计}$$

$$\hat{\mu} = \int \mu P(\mu | X^i) d\mu = \mu_N \text{ (因为正态分布)}$$

$$\hat{\mu}_N = \mu_N = \frac{\sigma_0^2}{N\sigma_0^2 + \sigma^2} \sum_{k=1}^N X_k + \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0$$

贝叶斯估计和贝叶斯学习

正态分布的均值估计

$$P(\mu_N) = N(\mu_0, \sigma_0) = N(0, 1)$$

$$\therefore \hat{\mu}_N = \frac{1}{N+1} \sum_{k=1}^N X_k$$

* 与最大似然估计结果相似，分母不同

贝叶斯估计和贝叶斯学习

贝叶斯学习

- * 前面学习了两种参数估计的方法. 最终目的是估计总体分布

$$p(x|\mathcal{R})$$

$$\mathcal{R} \rightarrow X^i$$

贝叶斯估计和贝叶斯学习

贝叶斯估计的另一计算方式

- * 确定 θ 的先验分布 $P(\theta)$,待估参数为随机变量
- * 用第 i 类样本 $X^i=(X_1, X_2, \dots, X_N)$ 求出样本的联合概率密度分布 $P(X^i|\theta)$, 它是 θ 的函数
- * 利用贝叶斯公式,求 θ 的后验概率

$$P(\theta | X^i) = \frac{P(X^i | \theta) \cdot P(\theta)}{\int_{\theta} P(X^i | \theta) P(\theta) d\theta}$$

- * 最后

$$\text{求贝叶斯估计 } \hat{\theta} = \int_{\theta} \theta P(\theta | X^i) d\theta \text{ (证明略)}$$

贝叶斯估计和贝叶斯学习

贝叶斯估计的另一计算方式

- * 我们在第三步后可以直接通过联合密度求类条件概率密度

$$p(x|X^i) \int p(x, \theta|X^i) d\theta = \int p(x|\theta) p(\theta|X^i) d\theta$$

贝叶斯估计和贝叶斯学习

非监督参数估计

- * 在未知样本类别的条件下的参数估计称为非监督参数估计.
- * 几个基本假设:
 - * 样本来自类数为 c 的各类中, 但不知道每个样本究竟来自哪一类
 - * 每类的先验概率 $P(\omega_j), j = 1, 2, \dots, c$ 已知
 - * 类条件概率密度的形式 $p(x|\omega_j, \theta_j)$ 已知
 - * 未知的只是 c 个参数向量 $\theta_1, \theta_2, \dots, \theta_c$ 的值

贝叶斯估计和贝叶斯学习

非监督参数估计

* 似然函数：为样本集的联合密度 $l(\theta) = p(\mathcal{R}|\theta)$

* 假设每个样本的类别未知，但样本从混合密度为

$$p(x|\theta) = \sum_{j=1}^c p(x|\omega_j, \theta_j) P(\omega_j)$$

* 的总体中独立被抽出来，被观察函数的似然函数定义为

$$l(\theta) = p(\mathcal{R}|\theta) = \prod_{k=1}^N p(x_k|\theta)$$

贝叶斯估计和贝叶斯学习

非监督参数估计

* 对数似然函数为：

$$H(\theta) = \ln[l(\theta)] = \sum_{k=1}^N \ln p(x_k | \theta)$$

* 最大似然估计 $\hat{\theta}$ 就是使似然函数最大的 θ

* 满足

$$l(\hat{\theta}) = \max_{\theta \in \Theta} \prod_{k=1}^N p(x_k | \theta)$$

$$H(\hat{\theta}) = \max_{\theta \in \Theta} \sum_{k=1}^N \ln[p(x_k | \theta)]$$

贝叶斯估计和贝叶斯学习

非监督参数估计

* 如何求得估计量?

$$\nabla_{\theta_i} H(\theta) = \sum_{k=1}^N \frac{1}{p(x_k | \theta)} \nabla_{\theta_i} \left[\sum_{j=1}^c p(x_k | \omega_j, \theta_j) P(\omega_j) \right] = 0$$

贝叶斯估计和贝叶斯学习

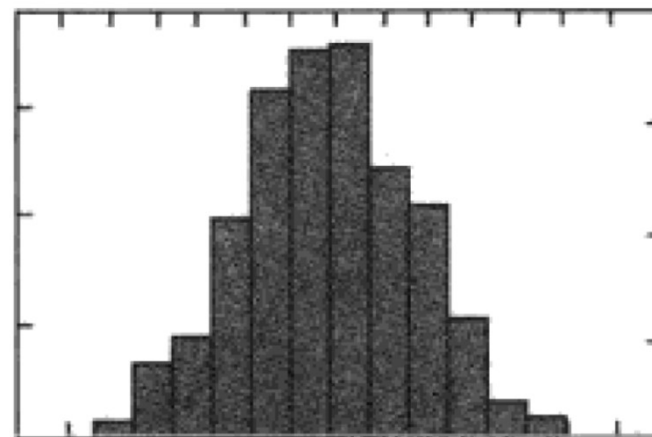
非参数设计

- * 多数情况下我们对样本的分布没有充分的了解，无法给出密度函数形式，部分样本情况难以用简单的函数来描述。
- * 此时，非参数设计 - 不对密度函数进行假设，直接利用样本估计整个函数，登场
- * 只有数值形式，无封闭函数形式。
- * 从所有可能的函数中进行一种选择

贝叶斯估计和贝叶斯学习

非参数设计

- * 几种常见的非参数设计
 - * 直方图方法
 - * K近邻估计方法
 - * Parzen窗法



贝叶斯估计和贝叶斯学习

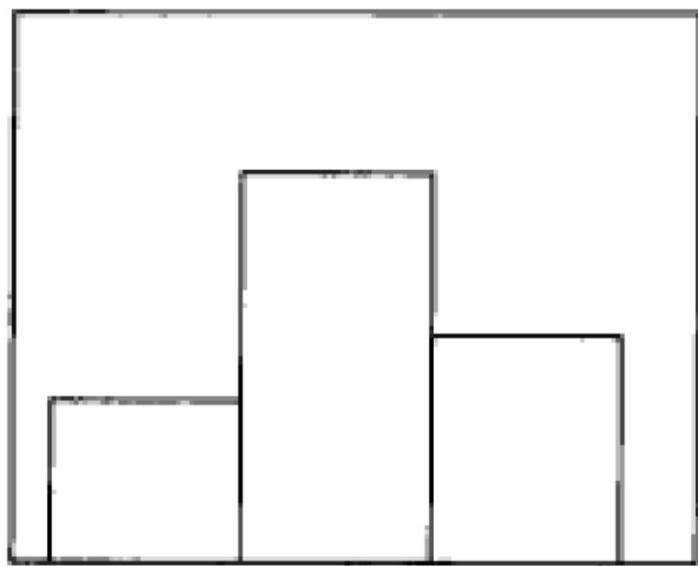
非参数设计－直方图估计

- * 直方图方法为日常人们常用之方法
- * 过程：
 - * 将样本 x 的每个分量在其取值范围内分割成 k 个等间隔小窗。如果 x 是 d 维向量，则这种分割会得到 k^d 个小舱，每个小舱体积为 V
 - * 统计落入每个小舱内的样本数目 q_i
 - * 把每个小舱的概率密度看为常数，并用 $q_i/(NV)$ 作为其估计值， N 为样本总数。

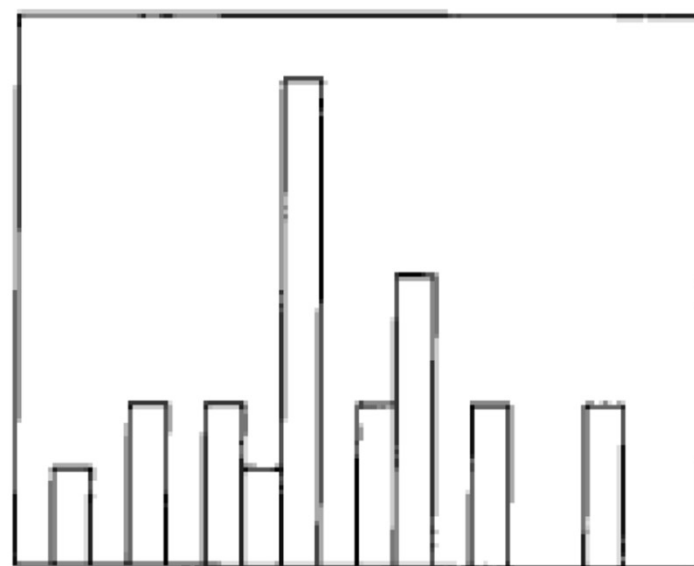
贝叶斯估计和贝叶斯学习

非参数设计－直方图估计

* 小窗的选择与估计效果息息相关



(a) 小窗过宽



(b) 小窗过窄

请大家自由组合，
每3位同学一组，
为做大作业分组

大作业1

- * 1. 学习总体分布的非参数估计
- * 2. 阅读文章
 - * **Density Estimation based on Parzen Window**
 - * **K-nearest Algorithm**
- * 3. 实现**PARZEN**窗估计和**K**近邻算法(非分类)
- * 4. 以**PPT**形式口头报告，包括两种算法的原理实现

谢谢