第二章 贝叶斯决策理论

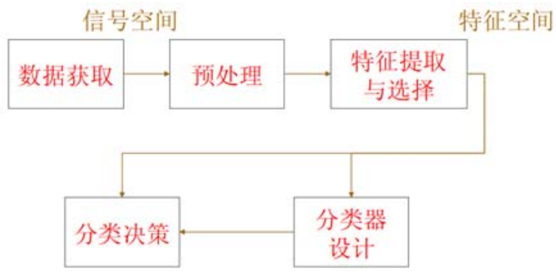
2009.09.22

本章主要内容

- 引言
- 基于最小错误率的Bayes决策
- 基于最小风险的Bayes决策
- 基于判别函数的分类器设计
- 正态分布的最小错误率Bayes决策
- 讨论

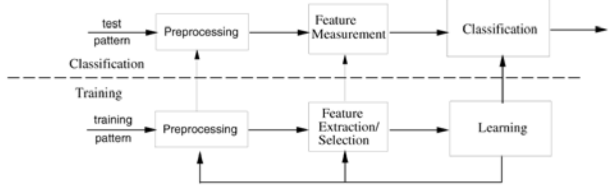
引言

- 模式识别是一种分类(classification)问题，即根据识别对象所呈现的观察值，将其分到某个类别中去。



引言

- 统计决策理论是处理模式分类问题的基本理论之一，对模式分析和分类器(classifier)的设计起指导作用。贝叶斯(Bayes)决策理论是统计模式识别中的一个基本方法，我们先讨论这一决策理论，然后讨论涉及统计判别方法的一些基本问题。



引言

- 客观现象或事物的发生和发展，按照"可预见性"可分两类情况 — 确定性和随机性。
  - 随机性事物的结果无法预知，但具有统计规律。
  - 随机性事物的特征观察值是随机变量。
- 特征的观察值总含有某种误差，其具有一定的随机性；而且同类的不同对象的某个特征的值通常也是按某种规律散布的。

➔ 模式类别和判决结果的随机性

➔ 用概率统计的理论和方法来解决识别问题是合理的。

引言

- 随机模型是用来描述自然界中不确定现象的数学模型。
- 统计模式识别的要点：将模式的特征量考虑为符合某种统计规律（概率密度/分布函数）的随机量。而任一个样本是取自总体中的一个个体。
- 需要解决三个问题：
  - **判别问题**：已知若干总体分布，当给出一个个体样本时，要确定这个样本属于哪个总体？
  - **训练问题**：已知一些个体样本，分别属于某些总体，要确定这些总体的分布规律（或参数）。
  - **误判率问题**：研究运用上述模型所造成的误判率的计算。

概念和名词约定

- 样本sample：待研究对象的个体，包括性质已知或未知的个体（统计学中有不同的约定）。
- 类别class：将所研究的样本性质离散化成有限的类别，认为同一类的样本在该性质上是不可区分的。
  - 类别用 ω_i (i=1，2，…，c，共c类)表示；如两个类别用 ω_1， ω_2表示，也可用{-1，1}表示。
- 已知样本：类别情况已知的样本。
- 未知样本：类别情况未知的样本。
- 样本集：若干样本的集合，分已知样本集和未知样本集。

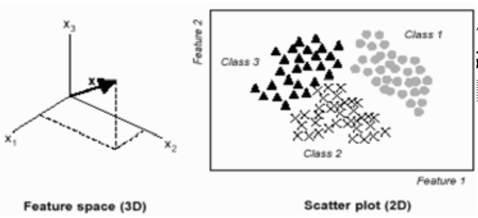
概念和名词约定

- 特征features：样本的任何可区分的且可观测的方面（属性）。
  - 包括定量特征和定性特征，通常最后转化为定量特征。
- 特征向量feature vectors：样本的所有特征组成的d维向量。
  - 是样本在数学上的表达，因此也称为样本。

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

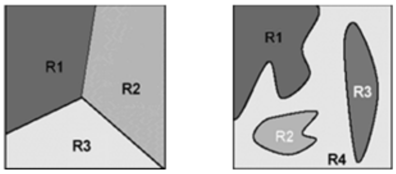
概念和名词约定

- 特征空间feature space：d维特征向量的所有可能取值范围构成的d维特征空间。
  - 每一个样本（特征向量）是该空间中的一个点，一个类别是该空间中的一个区域。
  - 例 … 位）
    在 … 位）
    在 … ，y
    值 … 量的
    特



概念和名词约定

- 分类器classifier：能够将每个样本都分到某个类别中去（或者拒绝）的计算机算法。
  - 是从特征空间到决策空间的映射。
- Decision region：分类器将特征空间划分为若干区域（决策域）。
- Decision boundary：不同类别区域之间的边界称作分类边界、决策边界或分类面，决策面。



贝叶斯定理

Reverend Thomas Bayes

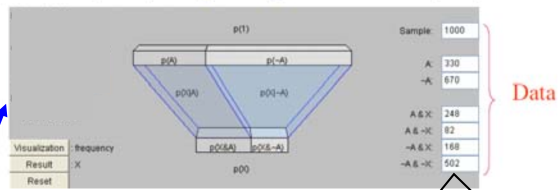
1702-1761

Bayes set out his theory of probability in *Essay towards solving a problem in the doctrine of chances* published in the *Philosophical Transactions of the Royal Society of London* in 1764. The paper was sent to the Royal Society by Richard Price, a friend of Bayes', who wrote:-

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... In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

贝叶斯定理

Two Classes(A, ~A) , Single Binary-Valued Feature (X,~X)



$$P(A|X) = \frac{P(X|A)P(A)}{P(X)}$$
$$= \frac{P(X|A)P(A)}{P(X,A)+P(X,\sim A)}$$
$$= \frac{P(X|A)P(A)}{P(X|A)P(A)+P(X|\sim A)P(\sim A)}$$
$$= \frac{0.75\times 0.33}{0.75\times 0.33+0.25\times 0.67} = 0.596$$

$$P(X|A) = \frac{P(X,A)}{P(A)} = \frac{0.248}{0.330} = 0.7515$$

$$P(X|\sim A) = \frac{P(X,\sim A)}{P(\sim A)} = \frac{0.168}{0.670} = 0.2507$$

贝叶斯决策理论概述

- 贝叶斯决策理论是解决模式分类问题的一种基本统计途径。
- 对问题的**要求/条件**
  - 决策问题可以用概率的形式来描述；
  - 所有有关的概率结构均已知。
- **出发点**是利用概率的不同分类决策和相应的决策代价之间的定量折中。
- 对于同一个问题，采用不同的决策标准将得到不同意义下“最优”的决策。其中最具代表性的是：
  - 最小错误率
  - 最小风险

贝叶斯决策理论概述

- 例1：医生根据病人血液中白细胞的浓度来判断病人是否患有血液病。
  - 一个人的白细胞浓度是3100，医生应该做出怎样的判断？ （两类别的识别问题）
  - 根据医学知识和以往的经验医生知道：
    - 一般人群中，患病的人数比例为0.5%。
    - 患病的人白细胞的浓度服从均值2000，方差1000的正态分布；未患病的人白细胞的浓度服从均值7000，方差3000的正态分布；

贝叶斯决策理论概述

- 数学表示
  - 用Ω表示“类别”这一随机变量，类别ω_1和ω_2分别表示“患病” 和“未患病”。
  - 用x表示“白细胞浓度值”这一随机变量。
  - 决策空间Θ={ω_1， ω_2}。

贝叶斯决策理论概述

- 先验概率 (*priori* probabilities / prior)
  - 根据大量统计数据确定某个类别事物出现的比例。
  - 例1中的两个类别的先验概率分布分别是:

$$P(\Omega=\omega_1)=0.5\%$$
$$P(\Omega=\omega_2)=99.5\%$$
$$P(\Omega=\omega_1)+P(\Omega=\omega_2)=1 \qquad \text{(排他性、穷举性)}$$

  - “先验” ——
    - 没有获得观测数据（病人白细胞浓度）之前类别的分布。
    - 只针对ω_1和ω_2出现的可能性，不考虑其他任何因素（如白细胞浓度）。

贝叶斯决策理论概述

- 先验概率 (*priori* probabilities / prior)
  - 仅依据先验信息的判决规则：

$$\text{Decide} \begin{cases} \omega_1, & \text{if } P(\omega_1) > P(\omega_2) \\ \omega_2, & \text{otherwise} \end{cases}$$

  - 判决的误差概率

$$P(error)=\min\{P(\omega_1),\ P(\omega_1)\}$$

贝叶斯决策理论概述

- 类条件概率(Class-conditional Probabilities)
  - 可利用白细胞浓度值x（连续随机变量）来帮助判决；
  - x的分布取决于类别状态（患病或未患病），用类条件概率密度函数来表示：

$$p(x\,|\,\omega_1)\sim N(2000,1000);$$
$$p(x\,|\,\omega_2)\sim N(7000,3000);$$

    - $p(x\,|\,\omega_i)$ 是指在类别ω_i下，在一个连续的函数空间中观测到x的可能性。
    - $p(x\,|\,\omega_1)$ 和 $p(x\,|\,\omega_2)$ 间的区别表示了血液病人和非血液病人之间白细胞浓度值的区别。

贝叶斯决策理论概述

- 类条件概率(Class-conditional Probabilities)
  - 同一类事物的各个属性都有一定的变化范围，在其变化范围内的分布概率用一种函数形式表示，即类条件概率密度函数。
  - 这种分布概率只针对同一类别事物，与其他类别的事物无关。
  - 用条件概率形式表示，以强调是同一类别事物的内部。
    - 例，用x表示某一个学生的身高，则男生身高的概率密度表示成p(x|男生)，女生身高表示成p(x|女生)，两者之间没有任何关系。

贝叶斯决策理论概述

- 后验概率(posteriori probabilities/posterior)
  - 问题：已知先验概率 $P(\omega_i)$ 和类条件概率密度函数 $p(x|\omega_i)$，i=1,2；对于一个样本 x=3100，判定 x∈ ω_1 或 x∈ ω_2？
    - 计算在观测样本x下，其类别状态是 ω_i (i=1,2)的概率：$P(\omega_i|x)$ 。
  - 后验概率是一个具体事物属于某种类别的概率。
    - 一个样本只可能属于两个类别之一，即有约束

$$P(\omega_1|x)+P(\omega_2|x)=1$$

    - 区别 $P(\omega_i|x)$ 和 $P(\omega_i)$ 。

贝叶斯决策理论概述

- 后验概率(posteriori probabilities/posterior)

$$P(\omega_i|x)=\frac{P(\omega_i,x)}{p(x)}$$

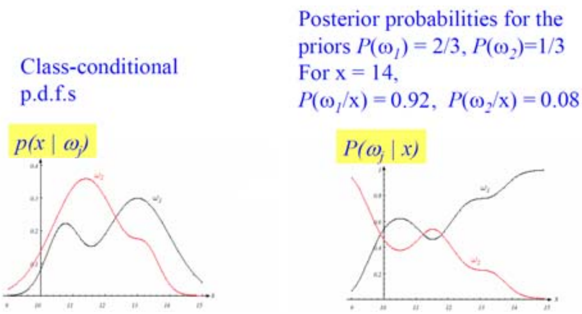
$$=\frac{P(\omega_i)p(x|\omega_i)}{\sum_j P(\omega_j)p(x|\omega_j)}$$

  - $p(x|\omega_i)$ 是 ω_i 关于x的似然(likelihood)函数，表明了在其他条件都相等的情况下，使得 $p(x|\omega_i)$ 较大的 ω_i 更有可能是真实的类别。

$$posterior=\frac{prior\times likelihood}{evidence}$$

贝叶斯决策理论概述

- 后验概率(posteriori probabilities/posterior)



贝叶斯决策理论概述

- 后验概率(posteriori probabilities/posterior)
  - 实质上，贝叶斯定理是通过观察样本x，把类别状态的先验概率 $P(\omega_i)$ 转化成后验概率 $P(\omega_i|x)$ 。
  - 贝叶斯定理的必要性：计算概率需要有大量的数据，而对于某一特定的事件（如白细胞浓度值x=3100）要搜集大量的样本是很困难的。
  - 贝叶斯定理综合了先验概率（类别出现的可能性）和类条件概率（类别符合观测样本的可能性）两方面因素。

基于最小错误率的贝叶斯决策

- 为什么会有错分类，在何种情况下会出现错分类？
  - 当某一特征向量值X只为某一类物体所特有，即

$$P(\omega_k|x)=\begin{cases}1, & k=i\\ 0, & k\neq i\end{cases}$$

  对其作出决策是容易的，也不会发生错误。
  - 问题在于出现模棱两可的情况，即不同类别在特征空间的分布有重叠。此时，任何决策都存在误判的可能性。

基于最小错误率的贝叶斯决策

- 目标：最小化决策的**平均错误率** $P(error)$
- $P(error)$：在特征向量观测值的整个可能取值范围内的错误率的均值。

$$
\begin{aligned}
P(error) &= \int_{-\infty}^{+\infty} P(error, x)dx \\
&= \int_{-\infty}^{+\infty} P(error \mid x) p(x) dx \\
&= E(P(error \mid x)).
\end{aligned}
$$

  - 平均错误率是条件错误率的数学期望。

基于最小错误率的贝叶斯决策

- 计算**条件错误率**（以两类别为例）

$$
P(error \mid x) = \begin{cases} P(\omega_2 \mid x), & \text{if assign } x \in \omega_1 \\ P(\omega_1 \mid x), & \text{if assign } x \in \omega_2 \end{cases}
$$

  - 因为 $P(error \mid x) \ge 0, \quad p(x) \ge 0, \quad \forall x$

➔ $\min P(error) \Leftrightarrow \min P(error \mid x), \text{ for all } x$

基于最小错误率的贝叶斯决策

- **判决规则**
  - 最大化后验概率准则

$$
\text{assign } x \in \begin{cases} \omega_1, & \text{if } P(\omega_1 \mid x) > P(\omega_2 \mid x) \\ \omega_2, & \text{otherwise} \end{cases}
$$

➔ $P(error \mid x) = \min\{P(\omega_1 \mid x), P(\omega_2 \mid x)\}, \quad \forall\ x$

  - 最小错误率的贝叶斯决策是一致最优决策。

基于最小错误率的贝叶斯决策

- 判决规则的**等价形式**
  - 比较大小不需计算 $p(x)$ ，即

$$
\text{assign } x \in \begin{cases} \omega_1, & \text{if } p(x \mid \omega_1)P(\omega_1) > p(x \mid \omega_2)P(\omega_2) \\ \omega_2, & \text{otherwise} \end{cases}
$$

    - 如果对于某个x，有 $p(x \mid \omega_1) = p(x \mid \omega_2)$ ，判决取决于先验概率；
    - 如果 $P(\omega_1) = P(\omega_2)$ ，判决取决于似然概率。

基于最小错误率的贝叶斯决策

- 判决规则的**等价形式**
  - 最大化似然比准则

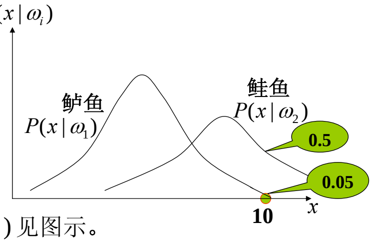
$$
\text{assign } x \in \begin{cases} \omega_1, & \text{if } l(x) = \dfrac{p(x \mid \omega_1)}{p(x \mid \omega_2)} > \dfrac{P(\omega_2)}{P(\omega_1)} \\ \omega_2, & \text{otherwise} \end{cases}
$$

  - 对似然比取负对数

$$
\text{assign } x \in \begin{cases} \omega_1, & \text{if } h(x) = -\ln p(x \mid \omega_1) + \ln p(x \mid \omega_2) > \ln \dfrac{P(\omega_2)}{P(\omega_1)} \\ \omega_2, & \text{otherwise} \end{cases}
$$

基于最小错误率的贝叶斯决策

- 例解：两类鱼的自动分类问题，鲈鱼(ω $_1$)和鲑鱼(ω $_2$)，用鱼长度的观察值(x)为特征。
  - 根据统计结果：
    - $P(\omega_1) = 1/3$;
    - $P(\omega_2) = 2/3$;



    - $p(x \mid \omega_1)$, $p(x \mid \omega_2)$ 见图示。
  - 如何将一条长为10的鱼分类？

基于最小错误率的贝叶斯决策

- 解法一：

$$P(\omega_1 \mid x=10)=\frac{p(x=10\mid\omega_1)P(\omega_1)}{p(x=10)}$$
$$=\frac{p(x=10\mid\omega_1)P(\omega_1)}{p(x=10\mid\omega_1)P(\omega_1)+p(x=10\mid\omega_2)P(\omega_2)}$$
$$=\frac{0.05\times1/3}{0.05\times1/3+0.50\times2/3}=0.048;$$

$$P(\omega_2\mid x=10)=1-P(\omega_1\mid x=10)=0.952;$$
$$\Rightarrow P(\omega_1\mid x=10)<P(\omega_2\mid x=10);$$
$\Rightarrow x=10\in\omega_2$， 即是鲑鱼。

基于最小错误率的贝叶斯决策

- 解法二（用似然比）：

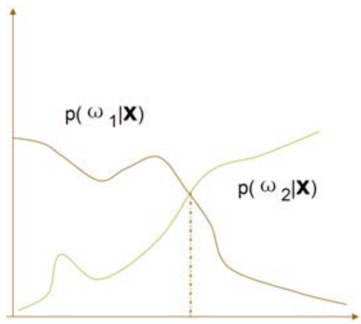
$$l_{12}(x=10)=\frac{p(x=10\mid\omega_1)}{p(x=10\mid\omega_2)}=\frac{0.05}{0.50}=0.1;$$

判决阀值 $\theta_{12}=\dfrac{P(\omega_2)}{P(\omega_1)}=\dfrac{2/3}{1/3}=2;$

$\because\quad l_{12}(x=10)<\theta_{12}$ ；

$\therefore\quad x=10\in\omega_2$ ，即是鲑鱼。

基于最小错误率的贝叶斯决策



后验概率

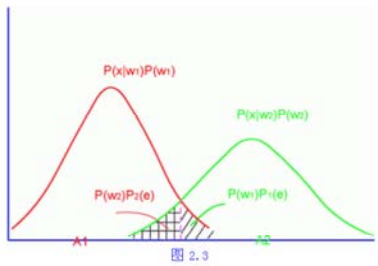
基于最小错误率的贝叶斯决策

- 再看决策的**错误率**
  - 设t为类别的分界面，则在特征向量x是一维时，t为x轴上的一点。两个决策区域:
    - R_1~(-∞，t)：决策为ω 1，$P(error\mid x)=P(\omega_2\mid x);$
    - R_2~(t，+∞)：决策为ω 2，$P(error\mid x)=P(\omega_1\mid x);$

$$P(error)=\int_{R_1}P(\omega_2\mid x)p(x)dx+\int_{R_2}P(\omega_1\mid x)p(x)dx$$
$$=\int_{R_1}P(x\mid\omega_2)P(\omega_2)dx+\int_{R_2}P(x\mid\omega_1)P(\omega_1)dx$$
$$=P(\omega_2)\int_{R_1}P(x\mid\omega_2)dx+P(\omega_1)\int_{R_2}P(x\mid\omega_1)dx$$

基于最小错误率的贝叶斯决策

- 再看决策的错误率



错误率为图中两个划线部分之和，
对应的错误率区域面积为最小。

基于最小错误率的贝叶斯决策

- 推广
  - 允许使用多于一个的特征，即用特征向量；
  - 允许多于两种的类别状态；
  - 允许除了判定类别以外的其他行为（如拒绝）；
  - 可以引入比误差概率更一般的损失函数。

(1) assign $x\in\omega_i$ if $P(\omega_i\mid x)=\max_{j=1,\cdots,c}P(\omega_j\mid x);$

(2) assign $x\in\omega_i$ if $p(x\mid\omega_i)P(\omega_i)=\max_{j=1,\cdots,c}p(x\mid\omega_j)P(\omega_j);$

$$P(error)=1-P(correct)=1-\sum_{j=1}^{c}\int_{R_j}p(x\mid\omega_j)P(\omega_j)dx$$

基于最小风险的贝叶斯决策

- 决策的风险：
  - 以医生根据白细胞浓度判断一个人是否患血液病为例：
    - 没病被判为有病，还可以做进一步检查，（一般情况下）损失不大；
    - 有病被判为无病， 损失严重。
  - 做决策要考虑决策可能引起的损失。
- 最小风险的Bayes决策正是考虑各种不同的错误造成的损失不同而提出的一种决策规则。

基于最小风险的贝叶斯决策

- 几个定义
  - **状态空间**Ω：由c个可能的状态（ c类）组成

$$\Omega=\{\omega_1,\omega_2,\cdots,\omega_c\}$$

  - **决策空间**$\mathcal{A}$**：** 由所有可能采取的决策组成

$$\mathcal{A}=\{\alpha_1,\alpha_2,\cdots,\alpha_k\}$$

  - **损失函数** $\lambda(\alpha_i,\omega_j),\quad i=1,\cdots,k,\quad j=1,\cdots,c$ ：
    表示对真实状态为ω_j的样本，采取决策α_i 时所造成的损失。
    - 常用表格形式描述损失函数（决策表）。

基于最小风险的贝叶斯决策

损失 决策 \ 状态	自然状态					
	ω_1	ω_2	…	ω_j	…	ω_c
α_1	$\lambda(\alpha_1,\omega_1)$	$\lambda(\alpha_1,\omega_2)$	…	$\lambda(\alpha_1,\omega_j)$	…	$\lambda(\alpha_1,\omega_c)$
α_2	$\lambda(\alpha_2,\omega_1)$	$\lambda(\alpha_2,\omega_2)$	…	$\lambda(\alpha_2,\omega_j)$	…	$\lambda(\alpha_2,\omega_c)$
⋮	⋮	⋮	…	⋮	⋮	⋮
α_i	$\lambda(\alpha_i,\omega_1)$	$\lambda(\alpha_i,\omega_2)$	…	$\lambda(\alpha_i,\omega_j)$	…	$\lambda(\alpha_i,\omega_c)$
⋮	⋮	⋮	…	⋮	⋮	⋮
α_a	$\lambda(\alpha_a,\omega_1)$	$\lambda(\alpha_a,\omega_2)$	…	$\lambda(\alpha_a,\omega_j)$	…	$\lambda(\alpha_a,\omega_c)$

决策表

基于最小风险的贝叶斯决策

- **条件期望损失**：对于特定的观察样本x，决策α_i造成的损失对x实际所属类别的各种可能的平均，也叫做条件风险：

$$R(\alpha_i\mid x)=E\left[\lambda(\alpha_i,\omega_j)\right]$$
$$=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j\mid x)$$

- **期望风险**：对所有x取值所作的决策α(x)所带来的平均风险，即条件风险对x的数学期望。

$$R(\alpha)=E[R(\alpha(\mathbf{x})\mid\mathbf{x})]=\int R(\alpha(\mathbf{x})\mid\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

基于最小风险的贝叶斯决策

- 目标：决策带来的损失的平均值——（平均）风险最小。
- 决策规则

$$Decide\quad \alpha_k,\quad \text{if}\quad R(\alpha_k\mid x)=\min_{j=1,\cdots,a}R(\alpha_i\mid x)$$

  - 通过保证对于每个观测值下的条件风险最小，使得决策的数学期望——平均风险最小。基于最小风险的贝叶斯决策是一致最优决策。