

第二章 贝叶斯决策理论

2011. 9. 8

本章主要内容

- 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

引言



概述

- 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具有随机性；用概率统计的理论和方法来解决识别问题是合理的。

概述

- 统计模式识别的要点：将模式的特征向量考虑为符合某种统计规律（概率密度/分布函数）的随机向量；而任一样本是取自总体中的一个个体。
- 需要解决的问题：
 - 判别问题：已知若干总体分布，当给出一个个体样本时，如何确定这个样本属于哪个总体？
 - 训练问题：已知一些分属于某些总体的个体样本，如何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。

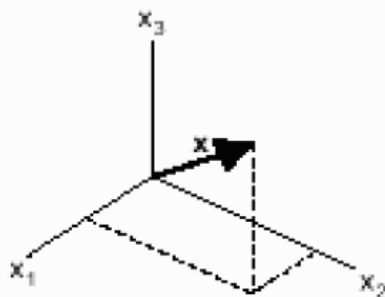
概念和名词约定

- 样本（sample）：待研究对象的个体，包括性质已知或未知的个体；
- 类别（class）：将所研究样本的性质离散化成有限的类别，同一类的样本在该性质上不可区分； ω_i ($i=1,2,\dots,c$) 表示第 i 个类别；
- 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，分为已知样本集和未知样本集。

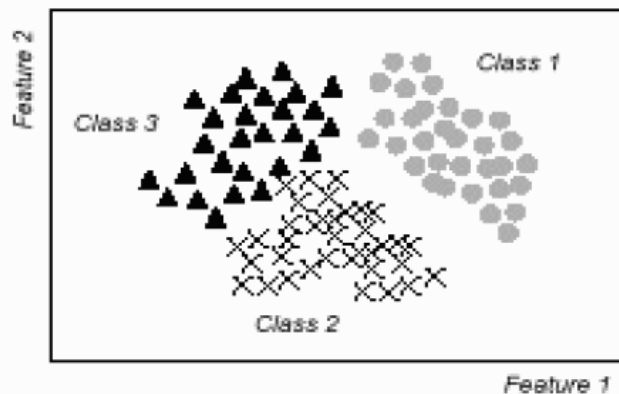
概念和名词约定

- ❑ 特征 (features): 样本的任何可区分且可观测的属性，包括定量特征和定性特征；
- ❑ 特征向量 (feature vector)：样本的所有特征组成的 d 维向量，是样本在数学上的表达；
- ❑ 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。

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



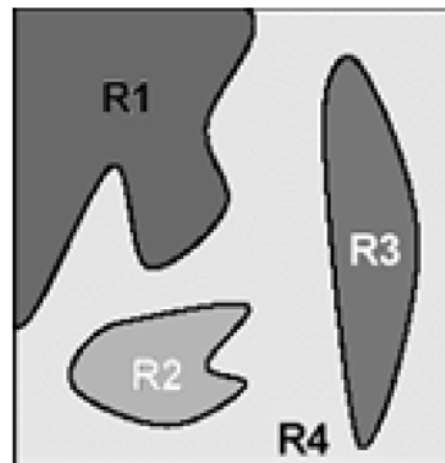
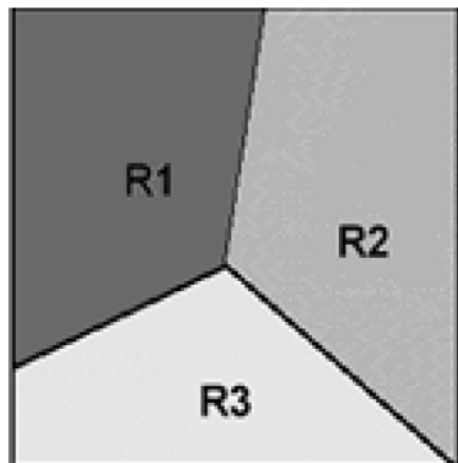
Feature space (3D)



Scatter plot (2D)

概念和名词约定

- 分类器（classifier）：将每个样本分到某个类别中（或拒绝）的计算机算法；是从特征空间到决策空间的映射；
 - 分类器将特征空间划分为若干决策域（decision region）；不同类别区域之间的边界称作分类/决策边界、分类/决策面（decision boundary）。



贝叶斯决策理论

贝叶斯定理

Reverend Thomas Bayes

1702-1761

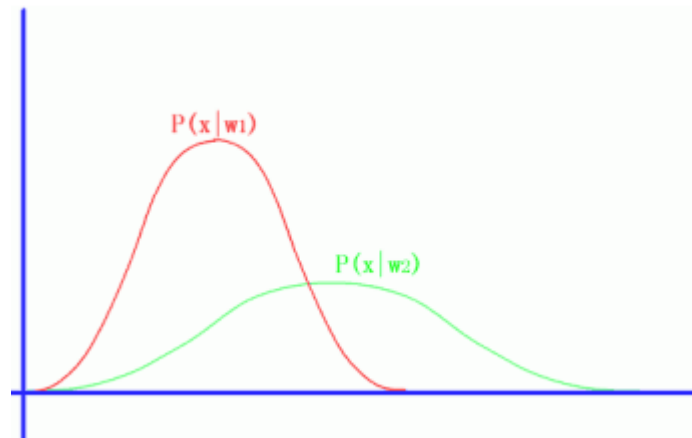


$$P(x|y) = \frac{P(y|x)P(x)}{\sum_{x \in \mathcal{X}} P(y|x)P(x)}$$

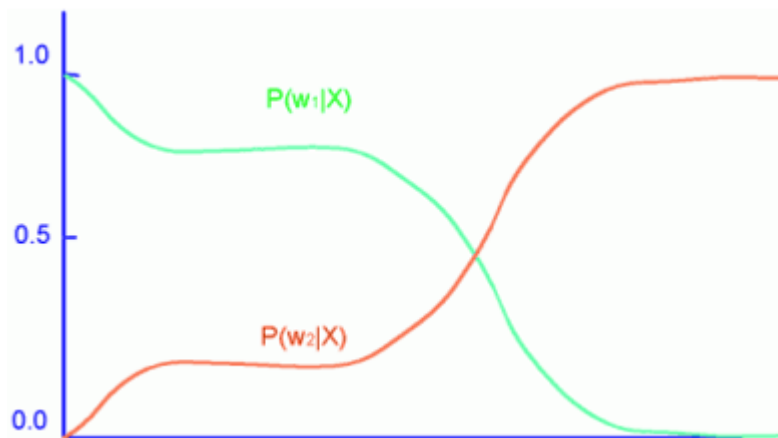
Bayes set out his theory of probability in *Essay towards solving a problem in the doctrine of chances* published in the *Philosophical Transactions of the Royal Society of London* in 1764. The paper was sent to the Royal Society by Richard Price, a friend of Bayes', who wrote:-
I now send you an essay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

贝叶斯决策理论概述

- **先验概率：**通过大量统计数据确定的类别出现的可能性；未获得观测样本前类别的分布；
- **类条件概率密度函数：**每一类别中样本特征向量的分布情况；
- **后验概率：**某模式类别在一定条件（特定的特征向量）下出现的概率。



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



基于最小错误率的贝叶斯决策

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

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

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

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

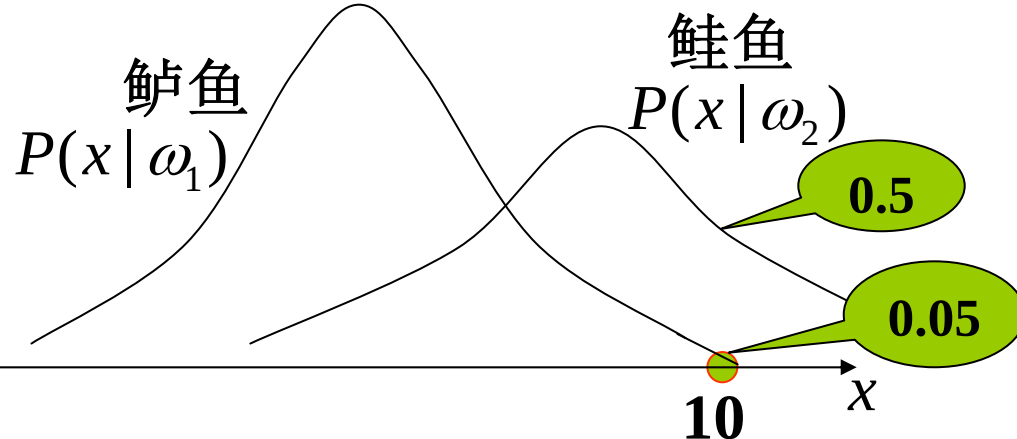
基于最小错误率的贝叶斯决策

- 例：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： $P(x|\omega_i)$

$$P(\omega_1) = 1/3;$$

$$P(\omega_2) = 2/3;$$



- 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

$$\begin{aligned}
 P(\omega_1 | x=10) &= \frac{p(x=10 | \omega_1)P(\omega_1)}{p(x=10)} \\
 &= \frac{p(x=10 | \omega_1)P(\omega_1)}{p(x=10 | \omega_1)P(\omega_1) + p(x=10 | \omega_2)P(\omega_2)} \\
 &= \frac{0.05 \times 1/3}{0.05 \times 1/3 + 0.50 \times 2/3} = 0.048;
 \end{aligned}$$

$$P(\omega_2 | x=10) = 0.952;$$

$$\Rightarrow P(\omega_1 | x=10) < P(\omega_2 | x=10);$$

$$\Rightarrow x=10 \in \omega_2, \text{ 即是鲑鱼。}$$

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

$$\text{判决阈值 } \theta_{12} = \frac{P(\omega_2)}{P(\omega_1)} = \frac{2/3}{1/3} = 2;$$

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

$$\therefore x=10 \in \omega_2, \text{ 即是鲑鱼。}$$

基于最小错误率的贝叶斯决策

$$\begin{aligned}
 P(\text{error}) &= \int_{R_1} P(\omega_2 | x) p(x) dx + \int_{R_2} P(\omega_1 | x) p(x) dx \\
 &= \int_{R_1} P(x | \omega_2) P(\omega_2) dx + \int_{R_2} P(x | \omega_1) P(\omega_1) dx \\
 &= P(\omega_2) \int_{R_1} P(x | \omega_2) dx + P(\omega_1) \int_{R_2} P(x | \omega_1) dx
 \end{aligned}$$

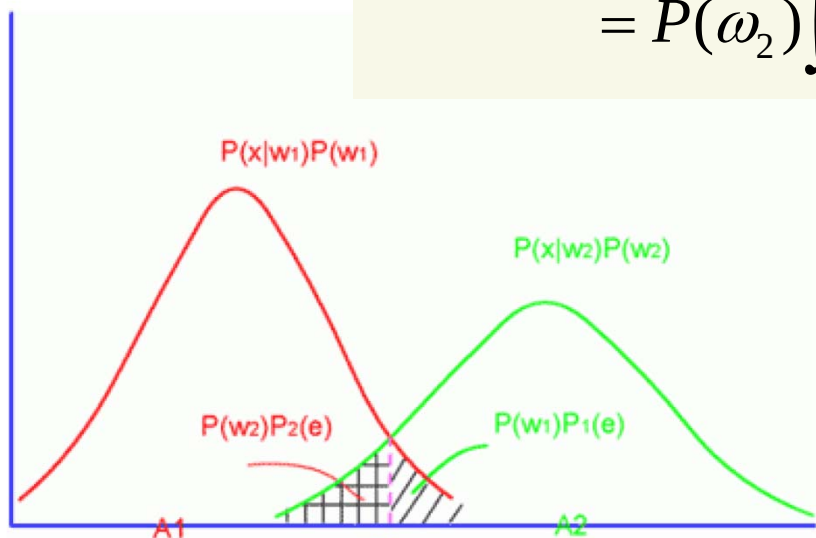


图 2.3

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。

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

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

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

基于最小风险的贝叶斯决策

- 状态空间 Ω : $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$
- 决策空间 $\mathcal{A}$: $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_a\}$
- 损失函数 $\lambda(\alpha_i, \omega_j)$, $i = 1, \dots, k$, $j = 1, \dots, c$: 对真实状态为 ω_j 的样本采取决策 α_i 时所造成的损失。

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	$\lambda(\alpha_2, \omega_1)$	$\lambda(\alpha_2, \omega_2)$	...	$\lambda(\alpha_2, \omega_j)$	...	$\lambda(\alpha_2, \omega_c)$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$
α_i	$\lambda(\alpha_i, \omega_1)$	$\lambda(\alpha_i, \omega_2)$	...	$\lambda(\alpha_i, \omega_j)$	...	$\lambda(\alpha_i, \omega_c)$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$
α_a	$\lambda(\alpha_a, \omega_1)$	$\lambda(\alpha_a, \omega_2)$	...	$\lambda(\alpha_a, \omega_j)$	...	$\lambda(\alpha_a, \omega_c)$

基于最小风险的贝叶斯决策

- 条件风险: $R(\alpha_i | \mathbf{x}) = E[\lambda(\alpha_i, \omega_j)] = \sum_{j=1}^c \lambda(\alpha_i, \omega_j) P(\omega_j | \mathbf{x});$
- 期望风险: $R(\alpha) = E[R(\alpha(\mathbf{x}) | \mathbf{x})] = \int R(\alpha(\mathbf{x}) | \mathbf{x}) p(\mathbf{x}) d\mathbf{x};$
- 决策规则: Decide α_k , if $R(\alpha_k | x) = \min_{j=1, \dots, a} R(\alpha_j | x);$
- 计算步骤: 计算后验概率 → 计算每个决策的条件风险 → 依据最小条件风险决策

基于最小风险的贝叶斯决策

□ 两类别问题

- 行为 α_1 : deciding ω_1 ;
- 行为 α_2 : deciding ω_2 ;
- 损失 $\lambda_{ij} = \lambda(\alpha_i, \omega_j)$, $i, j = 1, 2$.

$$R(\alpha_1 | \mathbf{x}) = \lambda_{11}P(\omega_1 | \mathbf{x}) + \lambda_{12}P(\omega_2 | \mathbf{x});$$

$$R(\alpha_2 | \mathbf{x}) = \lambda_{21}P(\omega_1 | \mathbf{x}) + \lambda_{22}P(\omega_2 | \mathbf{x});$$

$$\text{Decide } \begin{cases} \omega_1, & \text{if } (\lambda_{21} - \lambda_{11})P(\omega_1 | \mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2 | \mathbf{x}) \\ \omega_2, & \text{otherwise} \end{cases};$$

$$\text{Decide } \begin{cases} \omega_1, & \text{if } \frac{P(\mathbf{x} | \omega_1)}{P(\mathbf{x} | \omega_2)} > \frac{(\lambda_{12} - \lambda_{22}) P(\omega_2)}{(\lambda_{21} - \lambda_{11}) P(\omega_1)} \\ \omega_2, & \text{otherwise} \end{cases}.$$

基于最小风险的贝叶斯决策

□ 例：两类细胞识别问题：正常类 ω_1 和异常类 ω_2

$$P(\omega_1) = 0.9,$$

$$P(\omega_2) = 0.1;$$

$$p(\mathbf{x} | \omega_1) = 0.2,$$

$$p(\mathbf{x} | \omega_2) = 0.4;$$

$$\lambda_{11} = 0, \quad \lambda_{12} = 6,$$

$$\lambda_{21} = 1, \quad \lambda_{22} = 0;$$

$$P(\omega_1 | \mathbf{x}) = \frac{0.9 \times 0.2}{0.9 \times 0.2 + 0.1 \times 0.4} = 0.818,$$

$$P(\omega_2 | \mathbf{x}) = \frac{0.4 \times 0.1}{0.2 \times 0.9 + 0.4 \times 0.1} = 0.182;$$

$$R(\alpha_1 | \mathbf{x}) = \sum_{j=1}^2 \lambda_{1j} P(\omega_j | \mathbf{x}) = \lambda_{12} P(\omega_2 | \mathbf{x}) = 1.092,$$

$$R(\alpha_2 | \mathbf{x}) = \sum_{j=1}^2 \lambda_{2j} P(\omega_j | \mathbf{x}) = \lambda_{21} P(\omega_1 | \mathbf{x}) = 0.818;$$

$$\because R(\alpha_1 | \mathbf{x}) > R(\alpha_2 | \mathbf{x}), \quad \therefore \text{Decide } \alpha_2, \mathbf{x} \in \omega_2.$$

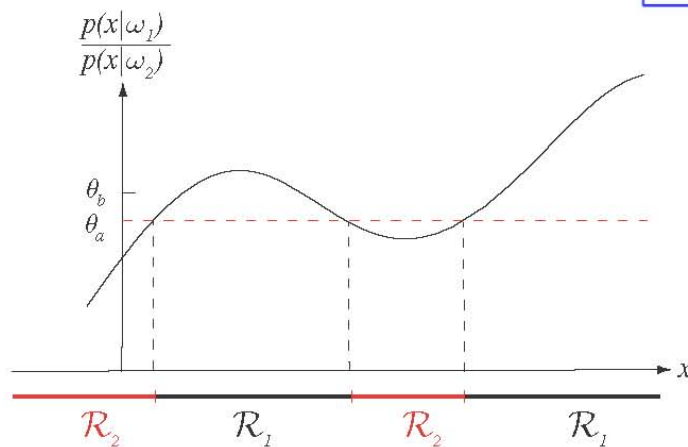
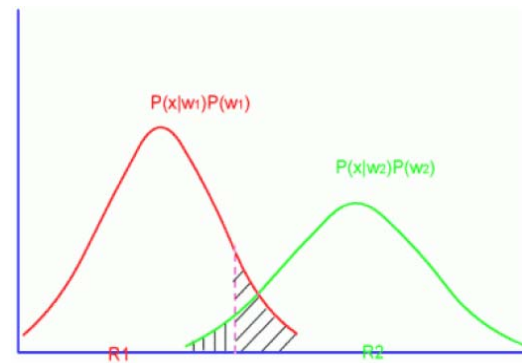
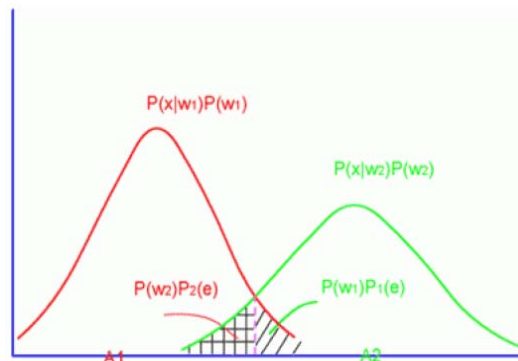
问题：依据最小风险决策
如何对细胞 $\mathbf{x}$ 进行分类？

两种决策方法之间的关系

$$\lambda(\alpha_i, \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}, \quad i, j = 1, \dots, c.$$

$$\begin{aligned} R(\alpha_i | \mathbf{x}) &= \sum_{j=1}^c \lambda(\alpha_i, \omega_j) P(\omega_j | \mathbf{x}) \\ &= \sum_{j=1, j \neq i}^c P(\omega_j | \mathbf{x}) = 1 - P(\omega_i | \mathbf{x}) \end{aligned}$$

$$\min_{\alpha} R(\alpha_i | \mathbf{x}) \Leftrightarrow \max_{\alpha} P(\omega_i | \mathbf{x})$$



$$\theta_a = P(\omega_2) / P(\omega_1)$$

$$\theta_b = \frac{P(\omega_2)(\lambda_{12} - \lambda_{22})}{P(\omega_1)(\lambda_{21} - \lambda_{11})}$$

两类错误率

w_1 : target is not present

w_2 : target is present

		Assigned	
		w_1	w_2
True	w_1	correct rejection	false alarm
	w_2	mis-detection	correct detection

Type – I Error (False positive): $P_1(e) = \int_{R_2} p(\mathbf{x} | \omega_1) d\mathbf{x}$,

Type – II Error (False negative): $P_2(e) = \int_{R_1} p(\mathbf{x} | \omega_2) d\mathbf{x}$

