

第 2 章 贝叶斯统计决策

贝叶斯统计决策方法基于贝叶斯决策理论，往往以某种概率的形式给出。本章首先介绍贝叶斯分类方法中的一般性的判决规则，并且抽象出随机模式的判决函数和决策面方程，给出两种分类器结构。

2.1 引言

在可以觉察到的客观世界中，存在着大量的物体和事件，他们在基本条件不变时，具有某种不确定性，每一次观测的结果没有重复性，这种模式就是随机模式。虽然随机模式样本测量值具有不确定性，但同类抽样实验的大量样本的观测值具有某种统计特性，这个统计特性是建立各种分类方法的基本依据。先看一下确定性模式判决函数的问题。

如下图所示：

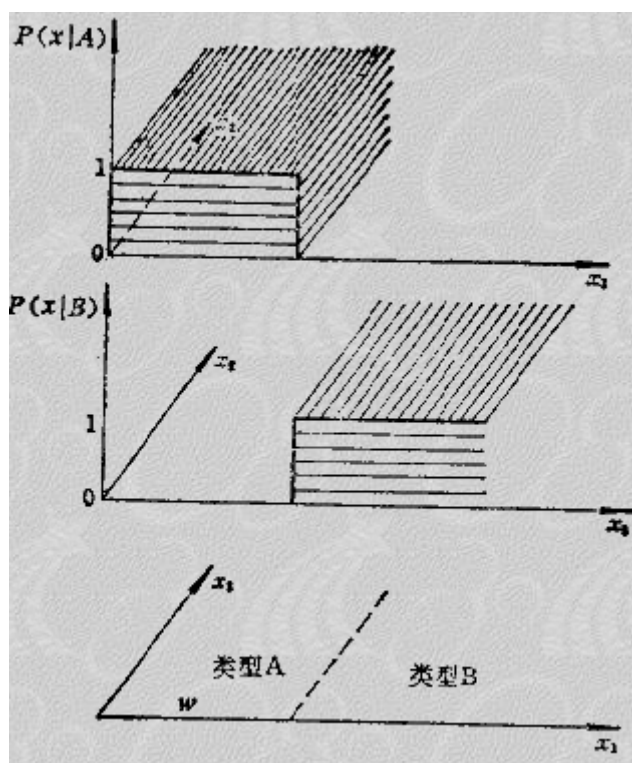


图 2-1 理想的概率分布

通过判决函数，特征空间被区分界面划分成两种类型的区域 A 和 B。由于模

式样本的观测值是确定性的，经常被正确分配到类型区域 A、B 之中。假如我们用概率的形式来表达，就是：在类型 A 的条件下观测模式样本 x ，则 x 位于区域 A 的概率为 1，而位于区域 B 的概率为 0。同样，在类型 B 的条件下观测模式样本 x ，情况正好相反， x 位于区域 A 的概率为 0，而位于区域 B 的概率为 1。这实际上是将概率的方法引入到确定模式，对于大多数实际情况，这是非常理想的概率分布。

许多实际情况，即使在类型 A 的条件下，模式样本 x 位于区域 A 的概率也往往小于 1，而位于区域 B 的概率也不为 0。对于类型 B 的条件也一样。这种交错分布的样本使分类发生错误，是模式随机性的一种表现。此时，分类方法就从确定性模式转到随机模式。

“如何使分类错误率尽可能小，是研究各种分类方法的中心议题。”

Bayes 决策理论是随机模式分类方法最重要的基础。下面是几个重要的概念：

1. 先验概率

先验概率是预先已知的或者可以估计的模式识别系统位于某种类型的概率。

若仍然用两个类型 A 和 B 为例，可用 $P(A)$ 和 $P(B)$ 表示各自的先验概率，此时满足：

$$P(A) + P(B) = 1 \quad (2-1)$$

推广到一般的 c 类问题中，用 w_1, w_2, \dots, w_c 表示类型，则各自的先验概率用 $P(w_1), P(w_2), \dots, P(w_c)$ 表示，且满足：

$$P(w_1) + P(w_2) + \dots + P(w_c) = 1 \quad (2-2)$$

其实，在处理实际问题时，有时不得不以先验概率的大小作为判决的依据。如：有一批木材，其中桦木占 70%，松木占 30%，A——桦木，B——松木，则 $P(A) = 0.7$ ， $P(B) = 0.3$ ，如果从中任取一块木材，而又要用先验概率做出判决，那就判为桦木。

先验概率不能作为判决的唯一依据，但当先验概率相当大时，它也能成为主要因素。

2. 类（条件）概率密度

它是系统位于某种类型条件下，模式样本 x 出现的概率密度分布函数，常用 $\rho(x|A), \rho(x|B)$ ，以及 $\rho(x|w_i)(i \in 1, 2, \dots, c)$ 来表示。

先验概率密度在分类方法中起至关重要的作用，它的函数形式及主要参数或者是已知的，或者是可通过大量抽样实验估计出来。

3. 后验概率

它是系统在某个具体的模式样本 x 条件下，位于某种类型的概率，常以 $P(A|x), P(B|x)$ ，以及 $P(w_i|x)(i \in 1, 2, \dots, c)$ 表示。

后验概率可以根据贝叶斯公式计算出来，可直接用作分类判决的依据。

例如：一个 2 类问题， w_1 表示诊断为无癌症， w_2 诊断为有癌症。 $P(w_1)$ 表示诊断正常的概率， $P(w_2)$ 表示某地区的人被诊断出患上癌症的概率，该值可以通过大量的统计得到， x 表示“试验反应呈阳性”。那么， $P(x|w_1)$ 表示诊断为无癌症且试验反应为阳性， $P(w_1|x)$ 表示试验为阳性，而且没有癌症。同样，可以有 w_2 的类概率密度和后验概率。

2.2 最小错误率判决规则

此规则的目标是，尽可能的减少错分类的情况，为了达成这一目标，我们需要一个决定测得的变量是属于哪一类的规则。这个规则会将输入空间划分为几个决策区域 R_k ，每个区域对应一个类别，如果变量落入了 R_k 我们就判定它属于 C_k 这一个类别。一个决策区域没必要是一个联通的个体，可以是由几个区域联合组成的。为了探究最佳的判决规则，我们首先来分析一个两类问题。在两类问题中若出现错误，应该是本该是属于 C_1 类别的变量却被判定为 C_2 ，或者是本该属于 C_2 类别的变量被判定为 C_1 ，这个错误的概率可以表示为：

$$P(\text{mistake}) = P(x \in R_1, C_2) + P(x \in R_2, C_1) \quad (2-3)$$

进一步改写为：

$$P(\text{mistake}) = \int_{R_1} p(x, C_2) dx + \int_{R_2} p(x, C_1) dx \quad (2-4)$$

为了使 $P(\text{mistake})$ 最小， x 应该被判定属于能够使上述积分达到较小值的类

别，由此推出一个规则，如果 $p(x, C_1) > p(x, C_2)$ 我们应将变量 x 归于类别 C_1 ，这样一来在 R_1 区域 $p(x, C_2)$ 的值较小，上述积分右边第一部分就达到最小了，第二部分的道理同上。从概率论的知识可知 $p(x, C_k) = p(C_k | x)p(x)$ 因此规则可以写为如果 $p(C_1 | x) > p(C_2 | x)$ （根据全概率公式 $p(x)$ 都是一样的），判断 x 为 C_1 类否则为 C_2 类。

以上一个例子为例，用 w_1 和 w_2 表示两种不同的类型，如 w_1 表示诊断正常， w_2 表示诊断出患有癌症。

用 $P(w_1)$ 和 $P(w_2)$ 分别表示先验概率。如： $P(w_1)$ 诊断正常的概率， $P(w_2)$ 表示某地人患癌症的概率，可通过大量的统计得到。

用 $\rho(x | w_1)$ 和 $\rho(x | w_2)$ 表示两个类概率密度。

样本 x 表示“试验反应阳性”，则 $\rho(x | w_1)$ 诊断为无癌症且试验反应为阳性， $P(w_1 | x)$ 试验为阳性且没有癌症。

根据全概率公式，模式样本 x 出现的全概率密度为：

$$\rho(x) = \rho(x | w_1) \cdot P(w_1) + \rho(x | w_2) \cdot P(w_2) \quad (2-5)$$

根据 Bayes 公式，在模式样本 x 出现的条件下，两个类型的后验概率为：

$$P(w_1 | x) = \frac{\rho(x | w_1) \cdot P(w_1)}{\rho(x)}, \quad P(w_2 | x) = \frac{\rho(x | w_2) \cdot P(w_2)}{\rho(x)} \quad (2-6)$$

此时，样本归属于“后验概率较高”的那种类型。

也就是：

$$\left. \begin{aligned} P(w_1 | x) &> P(w_2 | x), \text{ 则 } x \in w_1 \\ P(w_1 | x) &< P(w_2 | x), \text{ 则 } x \in w_2 \\ P(w_1 | x) &= P(w_2 | x), \text{ 则偶然决定 } x \in w_1, \text{ 或 } x \in w_2 \end{aligned} \right\} \quad (2-7)$$

根据 (2-6) 式，上述判决规则等价于：

$$\left. \begin{aligned} \rho(x | w_1) \cdot P(w_1) &> \rho(x | w_2) \cdot P(w_2), \text{ 则 } x \in w_1 \\ \rho(x | w_1) \cdot P(w_1) &< \rho(x | w_2) \cdot P(w_2), \text{ 则 } x \in w_2 \\ \rho(x | w_1) \cdot P(w_1) &= \rho(x | w_2) \cdot P(w_2), \text{ 则偶然决定 } x \in w_1, \text{ 或 } x \in w_2 \end{aligned} \right\} \quad (2-8)$$

下面我们以图形的方式来说明此规则的最佳性。

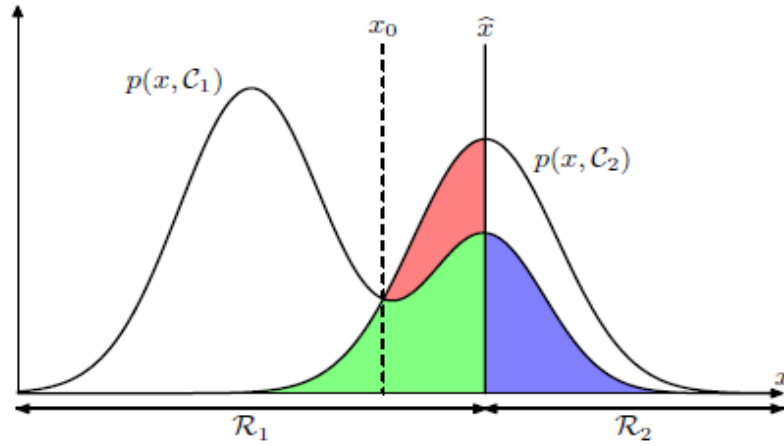


图 2-2 最佳门限

如上图所示，选定两个判定门限，第一个判定门限是 $x = \hat{x}$ ，该门限将空间划分为两个类型区域 R_1 , R_2 根据计算误判概率的公式（2-4）我们可以知道误判区域由，蓝色、绿色、红色三部分组成，对此先不做讨论。但我们选择第二个判定门限 $x = x_0$ 时，区域 R_1 的右边界会向左移动，进而以 x_0 为右边界，同时 R_2 会以 x_0 为左边界，这时误判区域由蓝色、绿色两部分组成，相比前边少了红色区域部分，也就是说误判率减小了。第一个门限是随意确定的门限，第二个门限是按照最小错误率判决规则选定的门限，故此图可以说明，最小错误率判决规则确实使误差率达到最小。

可以把上述两类问题导出的最小错误率判决规则一般化，推广到 c 类问题中，表达为：

若： $P(w_i | x) = \max_{j=1, \dots, c} \{P(w_j | x)\}$ ，则 $x \in w_i$ ，

等价于： $\rho(x | w_i) \cdot P(w_i) = \max_{j=1, \dots, c} \{\rho(x | w_j) \cdot P(w_j)\}$ ，则 $x \in w_i$

我们来论证此规则可以使错误率达到最小，对于这类判决问题判断结果只有两个，一个是判断正确，另一个是判断错误，这两种情况对应的概率是正确率 $P(\text{correct})$ 和误判率 $P(\text{mistake})$ ，且二者之间存在 $P(\text{correct}) + P(\text{mistake}) = 1$ 的关系，正确率的最大化自然代表着误判率的最小化， c 类问题中的误判率表达式很麻烦，但是正确率可以简单的表示如下：

$$P(\text{correct}) = \sum_{k=1}^c P(x \in R_k, C_k) = \sum_{k=1}^c \int_{R_k} p(x, C_k) dx \quad (2-9)$$

不难看出,按照我们的判断规则,对每个决策区域 R_k 选择的被积函数 $p(x, C_k)$ 都是最大的,所以最后的正确率会达到最大值,也就是说误判率达到最小。

例 1: 为了对癌症进行诊断,对一批人进行一次普查,各每个人打试验针,观察反应,然后进行统计,规律如下:

- (1) 这一批人中,每 1000 个人中有 5 个癌症病人;
- (2) 这一批人中,每 100 个正常人中有一个试验呈阳性反应;
- (3) 这一批人中,每 100 个癌症病人中有 95 人试验呈阳性反应。

问: 若某人(甲)呈阳性反应,甲是否正常?

解: 假定 x 表示实验反应为阳性,

(1) 人分为两类: w_1 —正常人, w_2 —癌症患者, $P(w_1) + P(w_2) = 1$

(2) 由已知条件计算概率值: 先验概率: $P(w_1) = 0.995$, $P(w_2) = 0.005$

类条件概率密度: $\rho(x | w_1) = 0.01$, $\rho(x | w_2) = 0.95$

(3) 决策过程

$$\begin{aligned} \rho(w_2 | x) &= \frac{\rho(x | w_2) \cdot P(w_2)}{\rho(x | w_1) \cdot P(w_1) + \rho(x | w_2) \cdot P(w_2)} \\ &= \frac{0.95 \times 0.005}{0.01 \times 0.995 + 0.95 \times 0.005} \\ &= 0.323 \end{aligned}$$

$$\rho(x | w_1) \cdot P(w_1) = 0.00995 \quad P(w_1 | x) = 1 - P(w_2 | x) = 1 - 0.323 = 0.677$$

$$\rho(x | w_2) \cdot P(w_2) = 0.00475 \quad P(w_1 | x) > P(w_2 | x)$$

$$\rho(x | w_1) \cdot P(w_1) > \rho(x | w_2) \cdot P(w_2)$$

由最小错误判决规则,可知: 甲 $\in w_1$

由于 $P(w_1)$ 比 $P(w_2)$ 大很多,所以先验概率起了较大作用。

2.3 最小风险判决规则

最小风险判决规则也是一种 Bayes 分类方法。最小错误率判决规则没有考虑错误判决带来的“风险”，或者说没有考虑某种判决带来的损失。

同一问题中，某种判决总会有一定的损失，特别是错误判决有风险。不同的错误判决有不同的风险，如上一节的例子中，判断细胞是否为癌细胞，可能有两种错误判决：

①正常细胞错判为癌细胞；

②癌细胞错判为正常细胞。

两种错误带来的风险不同。在①中，会给健康人带来不必要的精神负担，在②中，会使患者失去进一步检查、治疗的机会，造成严重后果。显然，第②种错误判决的风险大于第①种。

判决风险也可以理解为判决损失，即使在正确判决的情况下，一般也会付出某种代价，也会有损失。正是由于有判决风险的存在，最小错误率判决就不够了，必须引入最小风险判决规则。

假定有 c 类问题，用 $w_j (j=1,2,\dots,c)$ 表示类型，用 $\alpha_i (i=1,2,\dots,a)$ 表示可能做出的判决。实际应用中，判决数 a 和类型数 c 可能相等， $a=c$ ；也可能不等，即允许除 c 类的 c 个决策之外，可以采用其它决策，如“拒绝”决策，此时 $a=c+1$ 。

对于给定的模式样本 x ，令 $L(\alpha_i | w_j)$ 表示 $x \in w_j$ 而做出判决 α_i 带来的风险。若判决 α_i 一定，对 c 个不同类型的 w_j ，有 c 个不同的 $L(\alpha_i | w_j)$ 。

$L(\alpha_i | w_j)$ 的 c 个离散值随类型的性质变化，具有很大的随机性，可看成是随机变量。

另外，由于判决数目有 a 个，这样对于不同的判决和不同类型就有一个 $a \times c$ 维风险矩阵。

表 2.1 一般风险矩阵

类型 判决	w_1	w_2	w_c
α_1	$L(\alpha_1 w_1)$	$L(\alpha_1 w_2)$	$L(\alpha_1 w_c)$
α_2	$L(\alpha_2 w_1)$	$L(\alpha_2 w_2)$	$L(\alpha_2 w_c)$
.....
α_a	$L(\alpha_a w_1)$	$L(\alpha_a w_2)$	$L(\alpha_a w_c)$

假定某样本 x 的后验概率 $P(w_j | x)$ 已经确定，则有：

$P(w_1 | x) + P(w_2 | x) + \dots + P(w_c | x) = 1$ ，且 $P(w_j | x) \geq 0$ ， $j = 1, 2, \dots, c$ 对于每一种判决 α_i ，可求出随机变量 $L(\alpha_i | w_j)$ 的条件平均风险，也叫“条件平均损失”：

$$R(\alpha_i | x) = E[L(\alpha_i | w_j)] = \sum_{j=1}^c L(\alpha_i | w_j) \cdot P(w_j | x), \quad i = 1, 2, \dots, a \quad (2-10)$$

最小风险判决规则就是把样本 x 归属于“条件平均风险最小”的那一种判决。也就是说 $R(\alpha_i | x)$ 中的 i 应该具备的功能是，能够让 $R(\alpha_i | x)$ 满足如下等式：

$$\text{若 } R(\alpha_i | x) = \min_{k=1,2,\dots,a} \{R(\alpha_k | x)\}, \text{ 则 } x \in w_i \quad (2-11)$$

实施最小风险判决规则的步骤如下：

- (1) 在给定样本 x 条件下，计算各类后验概率 $P(w_j | x)$ ， $j = 1, 2, \dots, c$ 。
- (2) 按照(2-10)式求各种判决的条件平均风险 $R(\alpha_i | x)$ ， $i = 1, 2, \dots, a$ ，为此，需要知道风险矩阵。
- (3) 按照(2-11)式，比较各种判决的条件平均风险，把样本 x 归属于条件平均风险最小的那一种判决。

例 2：在例 1 的癌症诊断问题中，所有的化验结果可分为两类。

w_1 ——正常， w_2 ——癌症。

得到的判决也有两种 α_1 和 α_2 。

表 2.2 风险矩阵

判决 \ 类型	w_1	w_2
α_1	0.5	6
α_2	2	0.5

解：从风险矩阵中得到： $L(\alpha_1 | w_1) = 0.5$ ， $L(\alpha_1 | w_2) = 6$ ， $L(\alpha_2 | w_1) = 2$ ，
 $L(\alpha_2 | w_2) = 0.5$ 。

将例 1 中计算出的后验概率： $P(w_1 | x) = 0.677$ ， $P(w_2 | x) = 0.323$ 代入 (2-10) 式：

$$R(\alpha_1 = w_1 | x) = \sum_{j=1}^2 L(\alpha_1 | w_j) \cdot P(w_j | x) = 0.5 * 0.677 + 6 * 0.323 = 2.2765$$

$$R(\alpha_2 = w_2 | x) = \sum_{j=1}^2 L(\alpha_2 | w_j) \cdot P(w_j | x) = 2 * 0.671 + 0.5 * 0.323 = 1.5155$$

$$R(\alpha_2 = w_2 | x) < R(\alpha_1 = w_1 | x)$$

根据最小风险判决规则， $x \in w_2$ ，即试验人属于癌症病人，与例 1 的结论相反。

注意：实际工作中，列出合适的风险矩阵很不容易，要根据研究的具体问题，分析错误决策造成损失的严重程度，与有关专家共同商讨决定。

上面分析了两种决策规则，下面讨论它们之间的关系：

判决风险又叫判决损失， $L(\alpha_i | w_i)$ 又叫损失函数。现假设正确判决损失为 0，错误判决损失为 1，且判决数目与类型数目相等。即有 0—1 损失函数：

$$L(\alpha_i | w_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \quad (2-12)$$

可令 $L(\alpha_i | w_j) = 1 - \delta_{ij}$ ，其中 $\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ ，带入 (2-10) 式中可推导出：

$$\begin{aligned}
R(\alpha_i | w_j) &= \sum_{i=1}^c L(\alpha_i | w_j) \cdot P(w_j | x) \\
&= \sum_{i=1}^c (1 - \delta_{ij}) \cdot P(w_j | x) = \sum_{i=1}^c P(w_j | x) - \sum_{j=1}^c \delta_{ij} \cdot \rho(w_j | x) \\
&= 1 - \rho(w_i | x)
\end{aligned}$$

结果代入式 (2-11) 中, 得到:

若 $P(w_i | x) = \max_{k=1,2,\dots,c} \{P(w_k | x)\}$, 则 $x \in w_i$

这就是最小错误率判决规则。

结论: 在 0-1 损失函数情况下, 最小风险判决规则退化为最小错误率判决规则。也就是说, 最小错误率判决规则是最小风险判决规则的一个特例。

2.4 最大似然比判决规则

类概率密度 $\rho(x | w_i)$ 又称为“似然函数”, 两个类概率密度之比称为“似然比函数”。最大似然比判决规则也是一种 Bayes 分类方法。描述:

类型 w_i 分别与其它类型 $w_j (j=1,2,\dots,c, j \neq i)$ 的似然比均大于相应的门限值, 而其它类型 w_j 分别与 w_i 的似然比均小于相应的门限值, 则样本 $x \in w_i$ 。

(1) 由最小错误率判决规则引出最大似然比判决规则

若 $x \in w_1$, 最小错误率判决规则:

$$\rho(x | w_1) \cdot P(w_1) > \rho(x | w_2) \cdot P(w_2) \quad (2-13)$$

两边同时除以 $\rho(x | w_2) \cdot P(w_1)$ 有:

$$\frac{\rho(x | w_1)}{\rho(x | w_2)} > \frac{P(w_2)}{P(w_1)} \quad (2-14)$$

定义类型 w_1 与 w_2 的似然比为:

$$l_{12}(x) = \frac{\rho(x | w_1)}{\rho(x | w_2)} \quad (2-15)$$

则判决门限为:

$$\theta_{12} = \frac{P(w_2)}{P(w_1)} \quad (2-16)$$

一般先验概率已知， θ_{12} 也就已知了。

所以“最小错误率判决规则”就变为：

若： $l_{12}(x) > \theta_{12}$ ，则 $x \in w_1$

$$l_{12}(x) < \theta_{12}，\text{ 则 } x \in w_2 \quad (2-17)$$

$l_{12}(x) = \theta_{12}$ ，则偶然决定 $x \in w_1$ 或 $x \in w_2$

(2) 由最小风险判决规则引出最大似然比判决规则

若 $x \in w_1$ ，有 $R(\alpha_1 = w_1 | x) < R(\alpha_2 = w_2 | x)$ 代入

$$R(\alpha_i = w_i | x) = \sum_{j=1}^2 L(\alpha_i | w_j) \cdot P(w_j | x)，$$

有：

$$[L(\alpha_2 | w_1) - L(\alpha_1 | w_1)]P(w_1 | x) > [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)]P(w_2 | x)$$

即：

$$\frac{P(w_1 | x)}{P(w_2 | x)} > \frac{L(\alpha_1 | w_2) - L(\alpha_2 | w_2)}{L(\alpha_2 | w_1) - L(\alpha_1 | w_1)}$$

将 Bayes 公式： $\frac{P(w_1 | x)}{P(w_2 | x)} = \frac{\rho(x | w_1) \cdot P(w_1)}{\rho(x | w_2) \cdot P(w_2)}$ 代入上式可推导出：

$$\frac{\rho(x | w_1)}{\rho(x | w_2)} > \frac{L(\alpha_1 | w_2) - L(\alpha_2 | w_2)}{L(\alpha_2 | w_1) - L(\alpha_1 | w_1)} \cdot \frac{P(w_2)}{P(w_1)} \quad (2-18)$$

即： $l_{12}(x) > \theta_{12}$

$$\text{式中： } \theta_{12} = \frac{L(\alpha_1 | w_2) - L(\alpha_2 | w_2)}{L(\alpha_2 | w_1) - L(\alpha_1 | w_1)} \cdot \frac{P(w_2)}{P(w_1)} \quad (2-19)$$

为判决门限。

总结：最小风险判决引出的最大似然比判决与最小错误率判决引出的最大似然比判决的公式相同，只是判决门限 θ_{12} 的计算公式不同。

同样：在 (2-19) 中取 0—1 损失函数，即：

$$L(\alpha_1 | w_1) = 0, \quad L(\alpha_1 | w_2) = 1, \quad L(\alpha_2 | w_1) = 1, \quad L(\alpha_2 | w_2) = 0$$

则(2-19)退化为 (2-17)。也就是说在 0—1 损失函数情况下，最小风险判决退化为最小错误率判决。

将上述讨论进一步推广，假定有 c 个类型，分别用 w_1, w_2, \dots, w_c 表示，定义：

$$l_{ij}(x) = \frac{\rho(x | w_i)}{\rho(x | w_j)} \quad i, j = 1, 2, \dots, c, \quad \text{且 } i \neq j \quad (2-20)$$

由最小错误率判决规则导出，若对于任意的 $j, (j = 1, 2, \dots, c \text{ 且 } j \neq i)$ 有：

$$l_{ij}(x) > \theta_{ij}, \quad \text{则 } x \in w_i, \quad \theta_{ij} = \frac{P(w_j)}{P(w_i)} \quad (2-21)$$

由最小风险判决规则导出， θ_{ij} 定义为：

$$\theta_{ij} = \frac{[L(\alpha_i | w_j) - L(\alpha_j | w_j)] \cdot P(w_j)}{[L(\alpha_j | w_i) - L(\alpha_i | w_i)] \cdot P(w_i)} \quad (2-22)$$

同样在 0—1 损失函数的情况下，(2-22) 退化为 (2-21)。

似然函数的性质： $l_{ij}(x) = \frac{1}{l_{ji}(x)}$ ，因此，在 c 类问题中，若有一个 w_i 满足(2-21)

式，则不可能再有另外的类型 $w_j (i \neq j)$ 满足式 (2-21) 式。

例 3：对于前面的例 1、2 可以用上述办法求出。

对例 1 用似然比的方法求解

假定 x 表示实验反应为阳性人分为两类： w_1 —正常人， w_2 —癌症患者，

$$P(w_1) + P(w_2) = 1$$

由例 1 的解答过程中可以收集到的信息有：

1. 先验概率： $P(w_1) = 0.995$ ， $P(w_2) = 0.005$

2. 类条件概率密度： $\rho(x | w_1) = 0.01$ ， $\rho(x | w_2) = 0.95$

决策过程

$$l_{12}(x) = \frac{\rho(x|w_1)}{\rho(x|w_2)} = \frac{0.01}{0.95} = 0.0105$$

$$\theta_{12} = \frac{P(w_2)}{P(w_1)} = \frac{0.005}{0.995} = 0.005$$

$l_{12}(x) > \theta_{12}$ ，根据似然比的判定规则可知：甲 $\in w_1$ 与例 1 结果一致。

对例 2 用似然比的方法求解

所有的化验结果可分为两类。 w1——正常， w2——癌症。

得到的判决也有两种 α_1 和 α_2 。

表 2.3 风险矩阵

判决 \ 类型	w1	w2
α_1	0.5	6
α_2	2	0.5

由例 2 的解答过程可采集到的信息有：

1.风险： $L(\alpha_1 | w_1) = 0.5$ ， $L(\alpha_1 | w_2) = 6$ ， $L(\alpha_2 | w_1) = 2$ ， $L(\alpha_2 | w_2) = 0.5$ 。

2.先验概率： $P(w_1) = 0.995$ ， $P(w_2) = 0.005$

3.类条件概率密度： $\rho(x | w_1) = 0.01$ ， $\rho(x | w_2) = 0.95$

决策过程

$$l_{12}(x) = \frac{\rho(x|w_1)}{\rho(x|w_2)} = \frac{0.01}{0.95} = 0.0105$$

$$\theta_{12} = \frac{L(\alpha_1 | w_2) - L(\alpha_2 | w_2)}{L(\alpha_2 | w_1) - L(\alpha_1 | w_1)} \cdot \frac{P(w_2)}{P(w_1)} = \frac{6 - 0.5}{2 - 0.5} \times \frac{0.005}{0.995} = 0.0184$$

$l_{12}(x) < \theta_{12}$ ，根据似然比的判定规则可知甲 $\in w_2$ 与例 2 结果保持一致

结论：最大似然比判断规则，在很大程度上就是最小错误率判决规则和最小风险判决规则的一种变形，背景理论是一样的。

2.5 Neyman-Pearson 判决规则

在两类决策问题中，有犯两种错误分类的可能性，一种是在采取决策 w_1 时，其实际自然状态为 w_2 ；另一种是在采取决策 w_2 时，其实际自然状态为 w_1 。两种错误的概率分别为： $P(w_2) \cdot P_2(e)$ 和 $P(w_1) \cdot P_1(e)$ ，最小错误率 Bayes 决策是使这两种错误之和 $P(e)$ 最小。

$$P(e) = P(w_2) \cdot P_2(e) + P(w_1) \cdot P_1(e)$$

$$P_1(e) = \int_{R_2} \rho(x|w_1)dx, \quad P_2(e) = \int_{R_1} \rho(x|w_2)dx$$

在实际应用中，有时不知道先验概率，仅知道类概率密度，应如何确定判决门限呢？假定在处理过程中，先验概率保证不变，这时可以使用聂曼—皮尔逊 (Neyman—Pearson) 判决规则。

在两类问题中，两类的类概率密度曲线如下图所示，假定判决门限选为 t ，可能发生的两类分类错误与阴影区面积 ε_1 和 ε_2 成正比。

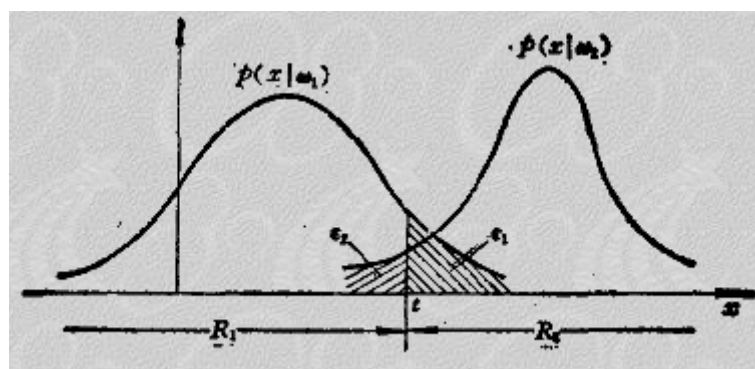


图 2-3 聂曼-皮尔逊判决示意图

聂曼—皮尔逊判决规则的基本思想是：在一种错误率不变的条件下，使另一种错误率最小。

这是具有实际意义的，例如，在细胞的化验中，由于把异常细胞错判为正常细胞的风险较大，可以要求这种错判的错误率不大于某个指定的常数作为前提条件，使正常细胞错判为异常细胞的错误率尽可能小，以此为原则来选择判决门限 t ，这就是聂曼—皮尔逊判决规则的基本思想。

从上图可以看出：

$$\varepsilon_1 = \int_{R_2} \rho(x | w_1) dx \quad (2-23)$$

$$\varepsilon_2 = \int_{R_1} \rho(x | w_2) dx \quad (2-24)$$

假定 ε_2 不变，为某个给定的正数，令：

$$\varepsilon = \varepsilon_1 + \mu \varepsilon_2 \quad (2-25)$$

为了使 ε_1 最小化，就要通过适当地选择某个正数 μ 使 ε 最小。

$$\varepsilon_1 = 1 - \int_{R_1} \rho(x | w_1) dx \quad (2-26)$$

$$\varepsilon_2 = 1 - \int_{R_2} \rho(x | w_2) dx \quad (2-27)$$

把 (2-26) 式和 (2-24) 式代入 (2-25) 式，得到：

$$\varepsilon = 1 + \int_{R_1} [\mu \rho(x | w_2) - \rho(x | w_1)] dx \quad (2-28)$$

把 (2-27) 式和 (2-23) 式代入 (2-25) 式，得到：

$$\varepsilon = \mu + \int_{R_2} [\rho(x | w_1) - \mu \rho(x | w_2)] dx \quad (2-29)$$

为了使 ε 最小化，上两式中的被积函数最好为负数，从而得到聂曼—皮尔逊判决规则为：

$$\left. \begin{array}{l} \text{若 } \frac{\rho(x | w_1)}{\rho(x | w_2)} > \mu, \text{ 则 } x \in w_1 \\ \text{若 } \frac{\rho(x | w_1)}{\rho(x | w_2)} < \mu, \text{ 则 } x \in w_2 \end{array} \right\} \quad (2-30)$$

显然正数 μ 是 x 的函数，根据上式，要求 $\mu(x)$ 为：

$$\mu(x) = \frac{\rho(x | w_1)}{\rho(x | w_2)} \quad (2-31)$$

为了最后确定各特征坐标上的门限值，还需要利用给定的正数 ε_2 ，参考上图，得到：

$$\varepsilon_2 = \int_{-\infty}^{\mu^{-1}(x)} \rho(x | w_2) dx \quad (2-32)$$

式中， $\mu^{-1}(x)$ 表示函数 $\mu(x)$ 的逆函数。

例：两类的概率密度函数是正态的，两类的均值向量分别为 $\mu_1 = (-1, 0)^T$ 和

$\mu_2 = (0,1)^T$ ，协方差矩阵相等且为单位矩阵。给定 $\varepsilon_2 = 0.046$ ，试确定 N-P 判决门限 t 。

解：根据给定的条件，很容易写出两类的类概率密度函数，即：

$\rho(x | w_1) \sim N(\mu_1, \Sigma)$ 和 $\rho(x | w_2) \sim N(\mu_2, \Sigma)$ 。 $\Sigma = E[(x - \mu)(x - \mu)^T]$ 。

而 $\mu_1 = (-1, 0)^T$ ， $\mu_2 = (1, 0)^T$ ， $\Sigma = I$ ，得到：

$$\begin{aligned}\rho(x | w_1) &= \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu_1)^T \cdot \Sigma^{-1}(x - \mu_1)\right] \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}[(x_1 + 1)^2 + x_2^2]\right) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + 2x_1 + 1 + x_2^2)\right) \\ \rho(x | w_2) &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}[(x_1 - 1)^2 + x_2^2]\right) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 - 2x_1 + x_2^2)\right) \\ \frac{\rho(x | w_1)}{\rho(x | w_2)} &= \exp(-2x_1)\end{aligned}$$

故： $\mu(x) = \exp(-2x_1)$ ， μ 只是 x_1 的函数，与 x_2 无关。有 $x_1 = -\frac{1}{2} \ln \mu$

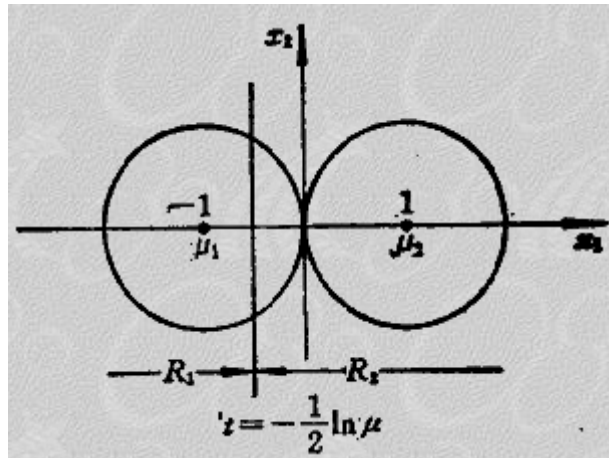


图 2-4 选择门限

又 $\rho(x | w_2)$ 的边缘密度为 $\rho(x_1 | w_2)$ ：

$$\begin{aligned}\rho(x_1 | w_2) &= \int_{-\infty}^{\infty} \rho(x | w_2) dx_2 \\ &= \int \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x_1^2 - 2x_1 + 1 + x_2^2)\right] dx_2 \\ &= \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x_1^2 - 2x_1 + 1)\right] \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x_2^2\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_1 - 1)^2\right]\end{aligned}$$

对于给定的正数 ε_2 ，可由下式计算：

$$\varepsilon_2 = \int_{-\infty}^{-\frac{1}{2}\ln\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_1-1)^2}{2}\right] dx_1 = \int_{-\infty}^{-\frac{1}{2}\ln\mu-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

显然， y 是服从标准正态分布的随机变量，令 $y_1 = -\frac{1}{2}\ln\mu-1$ 则： $\varepsilon_2 = \Phi(y_1)$

ε_2 与 y_1 具有一一对应的关系，有表可查。当 $\varepsilon_2 = 0.046$ 时， $y_1 = -1.693$ ， $\mu = 4$ ， $x_1 = -0.693$ ，因此判决门限 $t = x_1 = -0.693$ ，如上图所示。分区界线是 $x_1 = -0.693$ 的一条直线，对于样本 $x = (x_1, x_2)^T$ 的分类判决，只需考察特征 x_1 ，判决规则为，若： $x_1 < -0.693$ ，则 $x \in w_1$ 。否则， x 属于 w_2 。

2.6 最小最大判决规则

在实际应用中，有时分类器处理的各种类型样本的“先验概率是变化的”，此时再按照某个固定的 $P(w_i)$ 条件下的决策规则来进行决策，就得不到最小错误率或最小风险所需要得出的结果。这时就要用“最小最大判决规则”了。

在2.3节里，介绍了最小风险判决规则，研究了条件平均风险 $R(\alpha_i | x)$ 的概念和计算公式：

$$R(\alpha_i | x) = E[L(\alpha_i | w_j)] = \sum_{j=1}^c L(\alpha_i | w_j) \cdot P(w_j | x), i=1, 2, \dots, a \quad (2-33)$$

并且把模式样本归属于条件平均风险 $R(\alpha_i | x)$ 最小的那一种类型。由上式可以看出， $R(\alpha_i | x)$ 与类概率密度 $\rho(x | w_j)$ 、损失函数 $L(\alpha_i | w_j)$ 、先验概率 $P(w_j)$ 有关。如果上述因素是不变的，由足够的样本对分类器进行训练，就可以把特征空间划分成不同的类型区域 R_i 。如果先验概率 $P(w_j)$ 不是确切知道，在训练过程中，采用多组先验概率，就会得到多组类型区域 R_i 的划分结果。

另外，条件平均风险 $R(\alpha_i | x)$ 仅反映在样本 x 条件下，判决为 α_i 的平均风险，而不能反映把整个特征空间划分成某种类型空间的总的平均风险。

又由于 x 的观测值是随机向量，决策结果又依赖于 x ，所以决策 α 作为 x 的

函数可以记为 $\alpha(x)$ ，它也是一个随机变量。因此，可以定义“平均风险”为：

$$\bar{R} = \int_{E_d} R(\alpha(x) | x) p(x) dx \quad (2-34)$$

其中 E_d 为 x 的取值空间，实际上 E_d 就是整个特征空间。当特征空间被划分成 c 个类型区域 $R_i (i=1, 2, \dots, c)$ 之后(2-35)变为：

$$\bar{R} = \int_{R_1} R(\alpha_1 | x) p(x) dx + \int_{R_2} R(\alpha_2 | x) p(x) dx + \dots + \int_{R_c} R(\alpha_c | x) p(x) dx \quad (2-35)$$

由上式看出，如果类型区域的划分不同，则平均风险 \bar{R} 也不同。由于先验概率不同，对分类器训练结果，有不同的类型区域划分。所以，平均风险 \bar{R} 可作为先验概率的函数。（因为对于各类先验概率组合，有一系列的类型区域划分结果，从而可以计算出一系列的平均风险 \bar{R} ，可以得到 \bar{R} 与先验概率的函数关系。）

下面研究一下两类问题，用 w_1 和 w_2 表示不同的类型，它们的先验概率满足：

$$P(w_2) = 1 - P(w_1)$$

所以，上述平均风险 \bar{R} 与先验概率的关系就是 \bar{R} 与 $P(w_1)$ 的关系，一般是非线性关系。假定已经得到这个关系，如下图曲线所示。曲线上的任一点都对应着训练过程中所使用的先验概率 $P(w_1)$ 及所有决策的平均风险 \bar{R} 。

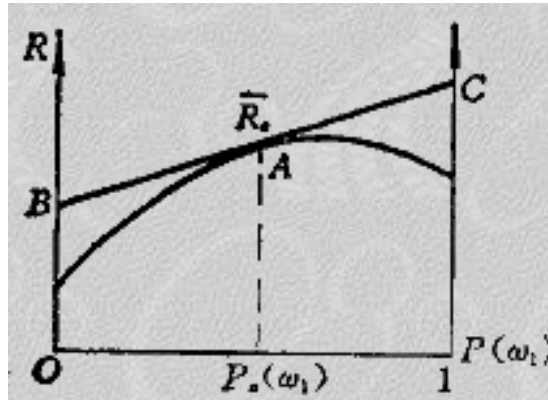


图2-5 平均风险与先验概率的关系

如果预先不确切知道先验概率，能否按照使平均风险 \bar{R} 最小来选择决策方案呢？

这是不可以的。这涉及所谓最小最大判决规则。为了说明这个问题，下面

针对两类问题进一步研究平均风险 \bar{R} 。由（2-35）可得：

$$\bar{R} = \int_{R_1} R(\alpha_1 | x) p(x) dx + \int_{R_2} R(\alpha_2 | x) p(x) dx$$

$$\text{将 } R(\alpha_i | x) = E[L(\alpha_i | w_j)] = \sum_{j=1}^c L(\alpha_i | w_j) \cdot P(w_j | x),$$

$$P(w_j | x) = \frac{\rho(x | w_j) P(w_j)}{\sum_{j=1}^c \rho(x | w_j) P(w_j)}, \text{ 令 } \alpha = c = 2 \text{ 代入上式, 得到:}$$

$$\bar{R} = \int_{R_1} [L(\alpha_1 | w_1) P(w_1) p(x | w_1) + L(\alpha_1 | w_2) P(w_2) p(x | w_2)] dx + \quad (2-36)$$

$$\int_{R_2} [L(\alpha_2 | w_1) P(w_1) p(x | w_1) + L(\alpha_2 | w_2) P(w_2) p(x | w_2)] dx$$

又因为 $P(w_2) = 1 - P(w_1)$ ，代入上式：

$$\begin{aligned} \bar{R} = & L(\alpha_1 | w_1) P(w_1) \int_{R_1} [p(x | w_1) dx + L(\alpha_1 | w_2) \int_{R_1} \rho(x | w_2) dx - L(\alpha_1 | w_2) P(w_1) \int_{R_1} \rho(x | w_2) dx \\ & + L(\alpha_2 | w_1) P(w_1) \int_{R_2} [p(x | w_1) dx + L(\alpha_2 | w_2) \int_{R_2} \rho(x | w_2) dx - L(\alpha_2 | w_2) P(w_1) \int_{R_2} \rho(x | w_2) dx \end{aligned}$$

又因为：

$$\int_{R_1} \rho(x | w_1) dx = 1 - \int_{R_2} \rho(x | w_1) dx$$

$$\int_{R_2} \rho(x | w_2) dx = 1 - \int_{R_1} \rho(x | w_2) dx$$

代入上式，得到：

$$\begin{aligned} \bar{R} = & L(\alpha_2 | w_2) + [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)] \int_{R_1} p(x | w_2) dx + \{[L(\alpha_1 | w_1) - L(\alpha_2 | w_2)] + [L(\alpha_2 | w_1) - L(\alpha_1 | w_1)] \\ & \int_{R_2} p(x | w_1) dx - [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)] \int_{R_1} p(x | w_2) dx\} P(w_1) = a + b P(w_1) \quad (2-37) \end{aligned}$$

式中：

$$a = L(\alpha_2 | w_2) + [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)] \int_{R_1} p(x | w_2) dx$$

$$b = [L(\alpha_1 | w_1) - L(\alpha_2 | w_2)] + [L(\alpha_2 | w_1) - L(\alpha_1 | w_1)] \times$$

$$\int_{R_2} p(x | w_1) dx - [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)] \int_{R_1} p(x | w_2) dx \quad (2-38)$$

损失函数 $L(\alpha_i | w_j)$ 是给定的，由（2-38）式看出，如果已经确定类型区域 R_1

和 R_2 ，则 a 、 b 为常数。根据（2-37）式，平均风险 \bar{R} 是先验概率 $P(w_1)$ 的线性函

数。由于先验概率 $P(w_1)$ 的取值范围为 $0 \sim 1$ ，所以 \bar{R} 值变化范围为 $a \sim (a+b)$ 。

例如，在上图中，在划分类型区域时， $P(w_1) = P_a(w_1)$ ， $\bar{R} = A$ 。在分类判决过程中，类型区域不再变化，而 $P(w_1)$ 可能变化，最大可能的平均风险 $\bar{R} = C = (a+b)$ ，这是所不希望的。

如何使最大可能的平均风险为最小呢？由 (2-37) 式， $\partial R / \partial p(w_1) = b$ ，如果 $b=0$ ， $\bar{R} = a$ ，且 \bar{R} 与 $P(w_1)$ 无关，即最大可能的平均风险达到最小值。但是 $b=0$ 又意味着由于类型区域的划分使平均风险 \bar{R} 达到曲线极值，如下图所示。此时 $P(w_1) = P_d(w_1)$ ， $\bar{R} = D$ 为曲线的最大值。

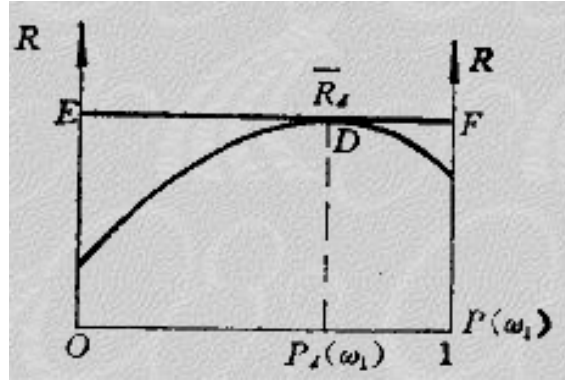


图2-6 风险关于先验概率的函数

即在训练过程中使平均风险达到最大值，恰好在分类判决中使最大可能的平均风险达到最小值，这就是最小最大判决规则的基本思想。

由上述分析，为了实施最小最大判决规则，必须令 $b=0$ 。由 (2-38) 式，有：

$$[L(\alpha_1 | w_1) - L(\alpha_2 | w_2)] + [L(\alpha_2 | w_1) - L(\alpha_1 | w_1)] \int_{R_2} p(x | w_1) dx - [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)] \int_{R_1} p(x | w_2) dx = 0 \quad (2-39)$$

此时，在分类判决中，平均风险为：

$$\bar{R} = L(\alpha_2 | w_2) + [L(\alpha_1 | w_2) - L(\alpha_2 | w_2)] \int_{R_1} p(x | w_2) dx \quad (2-40)$$

在这种情况下， \bar{R} 与先验概率的变化无关。

对于特殊情况：

$$L(\alpha_1 | w_1) = 0, \quad L(\alpha_1 | w_2) = 1$$

$$L(\alpha_2 | w_1) = 1, \quad L(\alpha_2 | w_2) = 0$$

也就是取0-1损失函数，代入(2-37到2-38)，有：

$$a = \int_{R_1} p(x | w_2) dx$$

$$b = \int_{R_2} p(x | w_1) dx - \int_{R_1} p(x | w_2) dx$$

$$\bar{R} = a + bP(w_1) = \int_{R_1} p(x | w_2) dx + [\int_{R_2} p(x | w_1) dx - \int_{R_1} p(x | w_2) dx]P(w_1)$$

$$= P(w_1) \int_{R_2} p(x | w_1) dx + P(w_2) \int_{R_1} p(x | w_2) dx$$

$$= P(w_1)P_1(e) + P(w_2)P_2(e)$$

上式就是最小错误率判决规则的错误率。

2.7 分类器设计

前面介绍了几种统计决策规则，运用这些规则对样本 x 进行分类，是分类器设计的主要问题。

1. 判别函数和决策面

定义：用于表达决策规则的函数称为判别函数。

决策面：将划分决策域的边界面称为决策面。可用数学表达式表达为决策面方程。

对两类最小错误率 Bayes 决策规则，有 4 种表达方式：

$$(1) \quad P(w_1 | x) \begin{matrix} > \\ < \end{matrix} P(w_2 | x), \quad \text{对应 } x \in \begin{cases} w_1 \\ w_2 \end{cases}$$

$$(2) \quad \rho(x | w_1) \cdot P(w_1) \begin{matrix} > \\ < \end{matrix} \rho(x | w_2) \cdot P(w_2), \quad \text{对应样本 } x \in \begin{cases} w_1 \\ w_2 \end{cases}$$

$$(3) \quad l(x) = \frac{\rho(x | w_1)}{\rho(x | w_2)} \begin{matrix} > \\ < \end{matrix} \frac{P(w_2)}{P(w_1)}, \quad \text{对应 } x \in \begin{cases} w_1 \\ w_2 \end{cases}$$

$$(4) \quad \ln \rho(x | w_1) + \ln P(w_1) \begin{matrix} > \\ < \end{matrix} \ln \rho(x | w_2) + \ln P(w_2), \quad \text{对应 } x \in \begin{cases} w_1 \\ w_2 \end{cases}$$

对多类别情况： $\Omega = \{w_1, w_2, \dots, w_c\}$ ，c 类 $x = \{x_1, x_2, \dots, x_d\}^T$

同样存在 4 个决策规则：

$$(1) P(w_i | x) > P(w_j | x), j = 1, 2, \dots, c, \text{ 且 } j \neq i, \text{ 对应样本 } x \in w_i$$

$$(2) \rho(x | w_i) \cdot P(w_i) > \rho(x | w_j) \cdot P(w_j), j = 1, 2, \dots, c, \text{ 且 } j \neq i, \text{ 对应样本 } x \in w_i$$

$$(3) l(x) = \frac{\rho(x | w_i)}{\rho(x | w_j)} > \frac{P(w_i)}{P(w_j)}, j = 1, 2, \dots, c, \text{ 且 } j \neq i, \text{ 对应 } x \in w_i$$

$$(4) \ln \rho(x | w_i) + \ln P(w_i) > \ln \rho(x | w_j) + \ln P(w_j), j = 1, 2, \dots, c, \text{ 且 } j \neq i, \text{ 对应 } x \in w_i$$

上面讨论了最小错误率 Bayes 决策，对于最小风险 Bayes 决策，同样有：

$$R(\alpha_1 | x) > R(\alpha_2 | x), \text{ 对应样本 } x \in \begin{cases} w_2 \\ w_1 \end{cases} \text{ 推广到多维情况:}$$

$$R(\alpha_i | x) > R(\alpha_j | x), i, j = 1, 2, \dots, c, \text{ 且 } i \neq j, x \in \begin{cases} w_j \\ w_i \end{cases}$$

2. 多类判别函数和分类器

(1) 判别函数

一般定义，一组函数 $g_i(x), i = 1, 2, \dots, c$ ，表示多类决策规则：

$$g_i(x) > g_j(x), j = 1, 2, \dots, c \text{ 且 } j \neq i \Rightarrow \text{样本 } x \in w_i$$

对于多类情况 $g_i(x)$ 可以定义为：

$$\textcircled{1} g_i(x) = P(w_i | x)$$

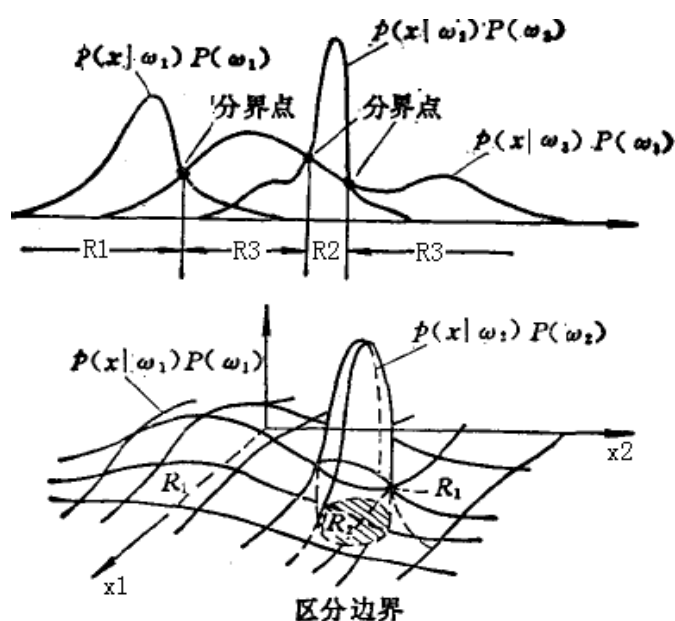
$$\textcircled{2} g_i(x) = \rho(x | w_i) \cdot P(w_i)$$

$$\textcircled{3} g_i(x) = \ln \rho(x | w_i) + \ln P(w_i)$$

(2) 决策面方程

各决策域 R_i 被决策面所分割，这些决策面是特征空间中的超曲面，对于相邻的两个决策域 R_i 和 R_j ，分割它们的决策面方程应满足： $g_i(x) = g_j(x)$ （显然

它们在决策面上相邻决策函数相等)



随机模式的判决函数及决策面
(a) 一维特征空间三个类型区域
(b) 二维特征空间两个类型区域

图 2-7 随机模式的判决函数及决策面

此时 R_2 与 R_3 的决策面: $\rho(x|w_2) \cdot P(w_2) = \rho(x|w_3) \cdot P(w_3)$

注意: 在一维空间, 对应的是点;

在二维空间, 对应的是曲线

在三维空间, 对应的是曲面

在四维空间对应的是超曲面

指 x 的维数

(3) 分类器设计

功能: 先设计出 c 个判别函数 $g_i(x)$, 再从中选出对应于判决函数为最大值的类作为决策结果。

分类器可由硬件或者软件构成 (已经模块化了)。

对于 c 类问题, $g_i(x) > g_j(x), x \in w_i$, 等效于

$$g_i(x) = \max_{j=1,2,\dots,c} \{g_j(x)\}, \text{ 则 } x \in w_i$$

多类分类器结构:

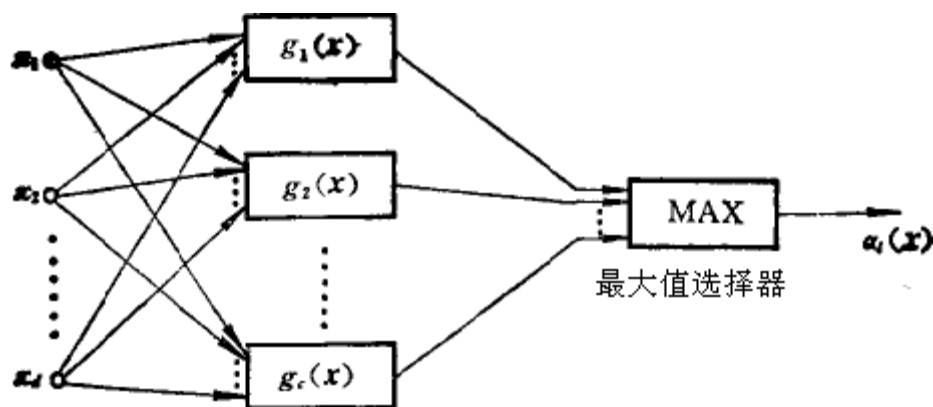


图 2-8 c 个类型分类器结构图

3. 两类情况

(1) 判决函数

$$g(x) = g_1(x) - g_2(x)$$

$$\text{决策规则: } \begin{cases} g(x) > 0, \text{ 决策 } x \in w_1 \\ g(x) < 0, \text{ 决策 } x \in w_2 \end{cases}$$

具体来说, 可定义 $g(x)$:

$$\textcircled{1} g(x) = \rho(w_1 | x) - \rho(w_2 | x)$$

$$\textcircled{2} g(x) = \rho(x | w_1) \cdot P(w_1) - \rho(x | w_2) \cdot P(w_2)$$

$$\textcircled{3} g(x) = \ln \frac{\rho(x | w_1)}{\rho(x | w_2)} - \ln \frac{P(w_1)}{P(w_2)}$$

(2) 决策面方程

$$g(x) = 0$$

也可以表示为: $\rho(x | w_1) \cdot P(w_1) - \rho(x | w_2) \cdot P(w_2) = 0$

(3) 分类器设计

通过计算 $g(x)$, 根据计算结果的符号将 x 分类。

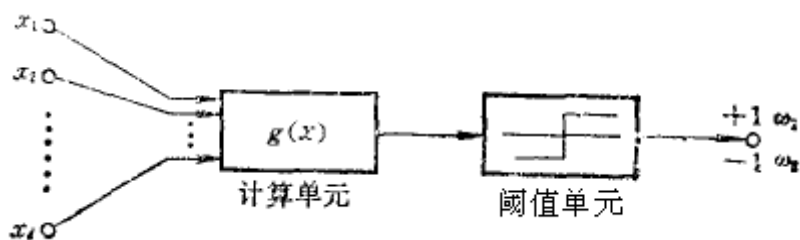


图 2-9 两类分类器结构简图

例 4：对例 1，例 2 分别写出其判决函数和决策面方程。

解：对于例 1，用判决函数： $g(x) = \rho(x | w_1) \cdot P(w_1) - \rho(x | w_2) \cdot P(w_2)$ 得到对应的判决函数为：

$$g(x) = 0.995\rho(x | w_1) - 0.005\rho(x | w_2) ;$$

决策面方程为： $0.995\rho(x | w_1) - 0.005\rho(x | w_2) = 0$ 。

对例 2，判决函数定义为：

$$g(x) = g_1(x) - g_2(x)$$

其中： $g_1(x) = 1 - R(a_1 = w_1 | x)$ ， $g_2(x) = 1 - R(a_2 = w_2 | x)$ ，带入上式：

$$g(x) = g_1(x) - g_2(x)$$

$$= R(a_2 = w_2 | x) - R(a_1 = w_1 | x)$$

$$R(a_1 = w_1 | x) = \sum_{j=1}^2 L(\alpha_1 | w_j) \cdot P(w_j | x) = 0.5 * P(w_1 | x) + 6 * P(w_2 | x)$$

$$R(a_2 = w_2 | x) = \sum_{j=1}^2 L(\alpha_2 | w_j) \cdot P(w_j | x) = 2 * P(w_1 | x) + 0.5 * P(w_2 | x)$$

$$g(x) = 1.5 * P(w_1 | x) - 5.5P(w_2 | x)$$

决策面方程为： $1.5 * P(w_1 | x) - 5.5P(w_2 | x) = 0$ 。

2.8 正态分布中的 Bayes 分类方法

在统计决策理论中，涉及到类条件概率密度函数 $P(x | w_i)$ 。对许多实际的数据集，正态分布通常是合理的近似。如果在特征空间中的某一类样本，较多地分布在这一类均值附近，远离均值点的样本比较少，此时用正态分布作为这一类的概率模型是合理的。另外，正态分布概率模型有许多好的性质，有利于作数学分析。概括起来就是：

- (1) 物理上的合理性
- (2) 数学上的简单性

下面重点讨论正态分布下的 Bayes 决策理论。

在上一章，我们已经把基于 Bayes 公式的几种分类判决规则抽象为相应的判决函数和决策面方程。这几种方法中 Bayes 最小错误率判决规则是一种最基本的方法。如果取 0-1 损失函数，最小风险判决规则和最大似然比判决规则均与最小错误判决规则等价。为了方便，我们以最小错误判决规则为例来研究 Bayes 分类方法在正态分布中的应用。

由最小错误率判决规则抽象出来的判决函数如下：

$$g_i(x) = \rho(x|w_i) \cdot P(w_i) \quad i = 1, 2, \dots, c \quad (2-41)$$

如果类概率密度是正态分布的，则 $\rho(x|w_i) \sim N(\mu_i, \Sigma_i)$ 。

$$g_i(x) = \frac{P(w_i)}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right] \quad (2-42)$$

由于对数函数是一个单调变化的函数，上式右边取对数后作为判决函数使用不会改变类型区域的划分。因此：

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i) \quad (2-43)$$

其中， $\frac{d}{2} \ln 2\pi$ 与类型无关，所有函数皆加上此项后，并不影响区域的划分，可以去掉。

下面对几种特殊情况进行讨论。

1. 情况一： $\Sigma_i = \sigma^2 I$ ， $i = 1, 2, \dots, c$

该情况下，每类的协方差矩阵相等，而且类的各特征间相互独立（由上节的性质③得知），具有相等的方差 σ^2 。

因此： $|\Sigma_i| = \sigma^{2d}$

$$\Sigma_i^{-1} = \frac{1}{\sigma^2} I$$

将上两式代入 $g_i(x)$ ：

$$g_i(x) = -\frac{(x - \mu_i)^T (x - \mu_i)}{2\sigma^2} - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^{2d} + \ln P(w_i)$$

上式中的第 2、3 项与类别无关，可以忽略，因此 $g_i(x)$ 可以简化为：

$$g_i(x) = -\frac{1}{2\sigma^2}(x - \mu_i)^T(x - \mu_i) + \ln P(w_i)$$

其中： $(x - \mu_i)^T(x - \mu_i) = \|x - \mu_i\|^2 = \sum (x_i - \mu_{ij})^2$ ， $i = 1, 2, \dots, c$ ，为 x 到类 w_i 的均值向量 μ_i 的“欧氏距离”的平方。

讨论一个特殊情况， $P(w_i) = P$ ，所有各类概率相等。则：

$$g_i(x) = -\frac{1}{2\sigma^2}(x - \mu_i)^T(x - \mu_i) = -\frac{1}{2\sigma^2}\|x - \mu_i\|^2$$

此时，对 x 的归类表示为：计算 x 到各类均值 μ_i 的欧氏距离的平方 $\|x - \mu_i\|^2$ ，然后把 x 归于具有 $\min_{i=1, \dots, c} \|x - \mu_i\|^2$ 的类。这种分类器叫最小距离分类器。

接着对 $g_i(x)$ 进一步化简：

$$\begin{aligned} g_i(x) &= -\frac{1}{2\sigma^2}(x - \mu_i)^T(x - \mu_i) + \ln P(w_i) \\ &= -\frac{1}{2\sigma^2}(x^T \cdot x - 2\mu_i^T \cdot x + \mu_i^T \cdot \mu_i) + \ln P(w_i) \end{aligned}$$

式中： $x^T \cdot x$ 与 i 无关，可以忽略：

$$g_i(x) = -\frac{1}{2\sigma^2}(-2\mu_i^T \cdot x + \mu_i^T \cdot \mu_i) + \ln P(w_i) = w_i^T x + w_{i0}$$

$$\text{式中： } w_i = \frac{1}{\sigma^2} \mu_i$$

$$w_{i0} = -\frac{1}{2\sigma^2} \mu_i^T \cdot \mu_i + \ln P(w_i)$$

$g_i(x) = w_i^T x + w_{i0}$ 是一个线性函数。

决策规则：对某个 x 计算 $g_i(x)$ ， $i = 1, 2, \dots, c$ ，若 $g_k(x) = \max_i g_i(x)$ ，则决策 $x \in w_k$ 。

由于 $g_i(x) = w_i^T x + w_{i0}$ 为线性函数，其决策面由线性方程 $g_i(x) - g_j(x) = 0$ 构成，决策面是一个超平面。下面由 $g_i(x) = w_i^T x + w_{i0}$ 推导出 $w^T(x - x_0) = 0$ ：由 $(\mu_i^T - \mu_j^T) = (\mu_i - \mu_j)^T$ ， $(\mu_i^T \mu_i - \mu_j^T \mu_j) = (\mu_i - \mu_j)^T(\mu_i + \mu_j)$ 可得：

$$\begin{aligned}
0 &= g_i(x) - g_j(x) \\
&= \frac{1}{\sigma^2}(\mu_i^T - \mu_j^T)x - \frac{1}{\sigma^2} \left[\frac{1}{2}(\mu_i^T \mu_i - \mu_j^T \mu_j) - \sigma^2 \log \frac{P(w_i)}{P(w_j)} \right] \\
&= \frac{1}{\sigma^2}(\mu_i^T - \mu_j^T)x - \frac{1}{\sigma^2}(\mu_i - \mu_j)^T \left[\frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{(\mu_i - \mu_j)^T (\mu_i - \mu_j)} (\mu_i - \mu_j) \log \frac{P(w_i)}{P(w_j)} \right] \\
&= \frac{1}{\sigma^2}(w^T x - w^T x_0)
\end{aligned}$$

继而得到, $w^T x - w^T x_0 = 0$, 其中, $w = \mu_i - \mu_j$,

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \ln \frac{P(w_i)}{P(w_j)} (\mu_i - \mu_j)$$

上述结果表示在二维特征空间里, 如下图所示:

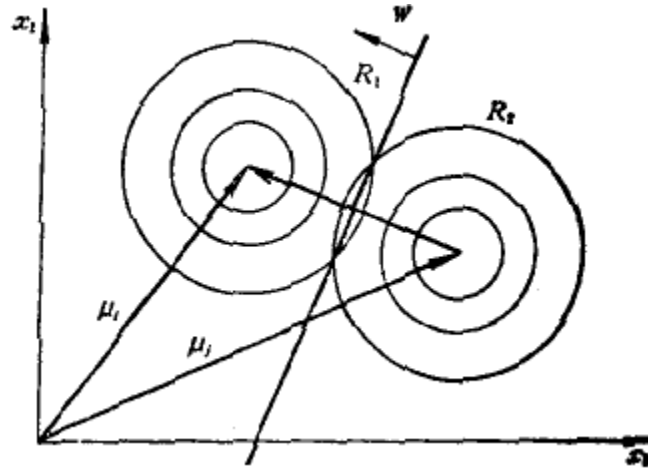


图 2-10 二类问题

两个同心圆是两类概率分布等密度点轨迹, 两个圆心就是两类的均值点。

两类的区分线 l 与 $\mu_1 - \mu_2$ 垂直, 其交点为 x_0 。 x_0 一般不是 $\mu_1 - \mu_2$ 的中点, 但当 $P(w_1) = P(w_2)$ 时, x_0 为 $\mu_1 - \mu_2$ 的中点。若 $P(w_1) \neq P(w_2)$ 时, x_0 向先验概率较小的那个类型的均值点偏移。可以推广到多类的情况, 注意这种分类方法没有不确定的区域。

2. 情况二: $\Sigma_i = \Sigma$

各类的协方差矩阵相等, 在几何上, 相当于各类样本集中在以该类均值 μ_i 为中心的同样大小和形状的超椭球内。

$$\Sigma_1 = \Sigma_2 = \dots = \Sigma$$

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i) \quad (2-44)$$

Σ 不变，与 i 无关：

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) + \ln P(w_i) \quad (2-45)$$

一个特例，当 $P(w_i) = P$ 时，各样本先验概率相等。

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) \quad (2-46)$$

其中： $\gamma^2 = (x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)$

γ^2 为 x 到均值点 μ_i 的“马氏距离”的平方（Mahalanobis）。

进一步简化： $g_i(x) = \gamma^2 = (x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)$

对于样本 x 只要计算出 γ^2 ，把 x 归于 γ^2 最小的类别。

接着对 $g_i(x)$ 化简：

$$g_i(x) = -\frac{1}{2}x^T \Sigma^{-1}x + \mu_i^T \Sigma^{-1}x - \frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \ln P(w_i) \quad (2-47)$$

去掉与 i 无关的项：

$$\begin{aligned} g_i(x) &= \mu_i^T \Sigma^{-1}x - \frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \ln P(w_i) \\ &= w_i^T x + w_{i0} \end{aligned} \quad (2-48)$$

其中： $w_i^T = \Sigma^{-1}\mu_i$ ， $w_{i0} = -\frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \ln P(w_i)$

$g_i(x) = w_i^T x + w_{i0}$ 也是一个线性函数，对应的决策面也是一个超平面。

对于 R_i 和 R_j 相邻，决策面方程：

$$g_i(x) = g_j(x) \Rightarrow w^T(x - x_0) = 0 \quad (2-49)$$

其中： $w = \Sigma^{-1}(\mu_i - \mu_j)$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln\left(\frac{P(w_i)}{P(w_j)}\right)}{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j) \quad (2-50)$$

与第一种情况不同，此时决策面通过 x_0 ，但不与 $\mu_i - \mu_j$ 正交（垂直）。二维情况：

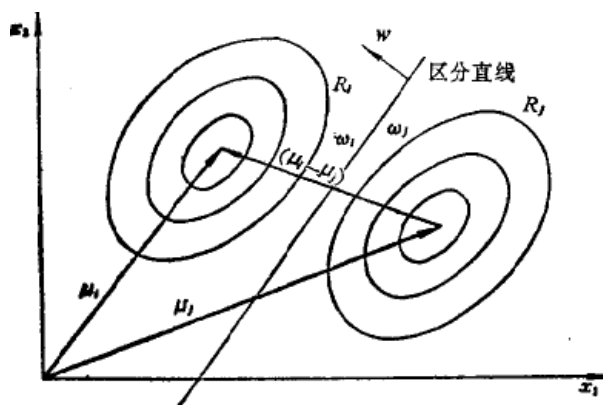


图 2-11 二维情况

当各类先验概率相等时 $P(w_i) = P(w_j)$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) \quad (2-51)$$

x_0 位于 $\mu_i - \mu_j$ 的中点上。当各类先验概率不相等时， x_0 不在 $\mu_i - \mu_j$ 的中点上，而是偏向先验概率较小的均值点。

3. 第三种情况

各类协方差矩阵不等： $\Sigma_i \neq \Sigma_j$ ， $i, j = 1, 2, \dots, c$

由于：

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i) \quad (2-52)$$

去掉与 i 无关的项 $-\frac{d}{2} \ln 2\pi$ ，得：

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i) - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i) \quad (2-53)$$

$g_i(x)$ 表示为： $g_i(x) = x^T W_i x + w_i^T x + w_{i0}$

其中： $W_i = -\frac{1}{2} \Sigma_i^{-1}$ $d \times d$ 矩阵

$w_i = \Sigma_i^{-1} \mu_i$ d 维向量

$$w_{i0} = -\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i)$$

此时 $g_i(x)$ 表示为 x 的二次型。

对于 R_i 和 R_j 相邻，决策面应为：

$$g_i(x) - g_j(x) = 0 \Rightarrow x^T(W_i - W_j)x + (w_i - w_j)^T x + w_{i0} - w_{j0} = 0 \quad (2-54)$$

该曲线为超二次曲面。随 Σ_i 、 μ_i 、 $P(w_i)$ 的不同，超二次曲面为：超球面、超椭球面、超抛物面、超双曲面，或超平面等。

假设特征空间是二维的，模式样本的两个分量之间是相互独立的，所以协方差矩阵是 2×2 维的对角矩阵。令各类的先验概率相等，那么不同类型区域的划分取决于各类的均值向量和两个方差项的差异，而决策面的形状主要取决于两个方差项的差异。

$$\Sigma_i = \begin{bmatrix} \sigma_{i1}^2 & 0 \\ 0 & \sigma_{i2}^2 \end{bmatrix}, \quad \Sigma_j = \begin{bmatrix} \sigma_{j1}^2 & 0 \\ 0 & \sigma_{j2}^2 \end{bmatrix}$$

(1) 若 $\sigma_{i1} = \sigma_{i2} = \sigma_i$ ， $\sigma_{j1} = \sigma_{j2} = \sigma_j$ ，且 $\sigma_i > \sigma_j$ ，则两类的概率分布等密度线分别是以各自均值点为圆心的同心圆，圆的大小与相应的方差相一致。由于 $\sigma_i > \sigma_j$ ，所以来自类型 w_j 的样本更密集于它的均值点附近；同时，由于圆的对称性，决策面为包围均值点 μ_j 的一个圆。

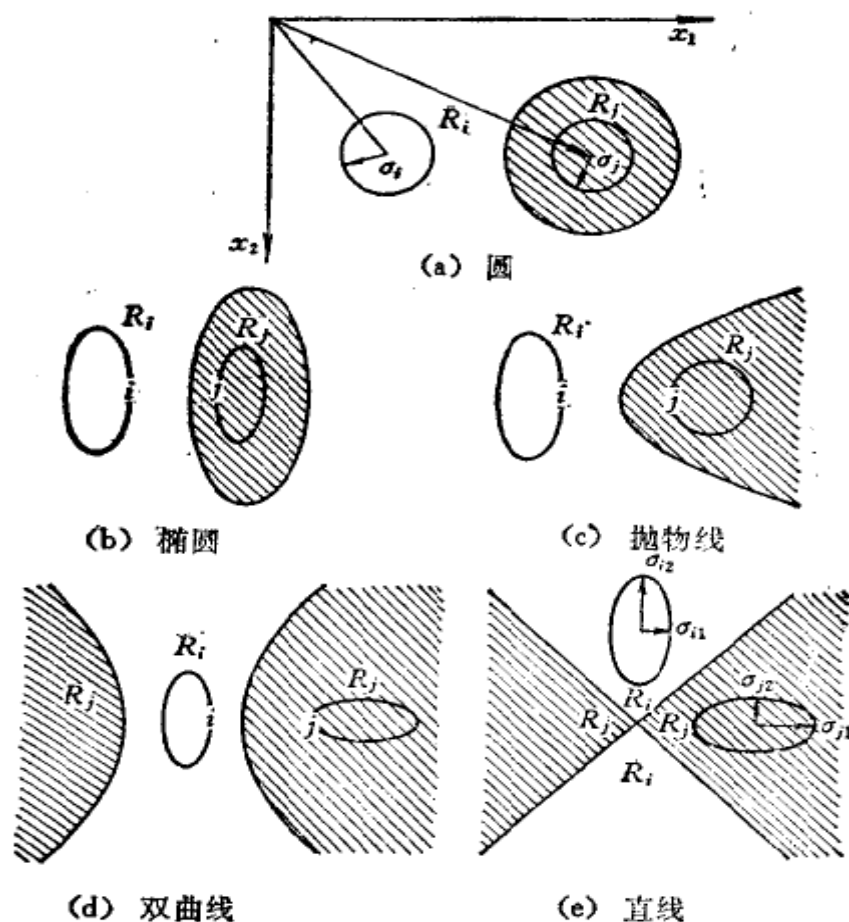


图 2-12 各种情况

(2) 若在上图的(a)的基础上增大分量 x_2 的方差 σ_{i2}^2 和 σ_{j2}^2 , 使 $\sigma_{i1} < \sigma_{i2}$ 和 $\sigma_{j1} < \sigma_{j2}$, 这样图(a)中的圆在 x_2 方向上伸展, 而变成椭圆, 如图(b)所示, 决策面也变成了椭圆。

(3) 若 $\sigma_{i1} = \sigma_{j1} = \sigma_{j2}$, $\sigma_{i1} < \sigma_{i2}$, 在这种情况下, 分量 x_2 大的样本 x 很可能来自类型 w_i , 使决策面变成一条抛物线, 如图(c)所示。

(4) 若在(c)的基础上增大 σ_{j1} , 使 $\sigma_{i1} = \sigma_{j2}$, $\sigma_{i1} < \sigma_{i2}$, $\sigma_{j1} > \sigma_{j2}$, 在这种情况下, 决策面变成双曲线, 如图(d)所示。

(5) 在一非常特殊的对称条件下, 使(d)中的双曲线向一对互相垂直的直线退化, 如图(e)所示。在这种情况下, 两种类型是线性可分的。

例 1: 设在三维特征空间里, 两类的类概率密度是正态分布的, 分别在两个

类型中获得 4 个样本，位于一个单位立方体的顶点上，如下图。两类的先验概率相等，试确定两类之间的决策面及相应的类型区域 R_1 和 R_2 。

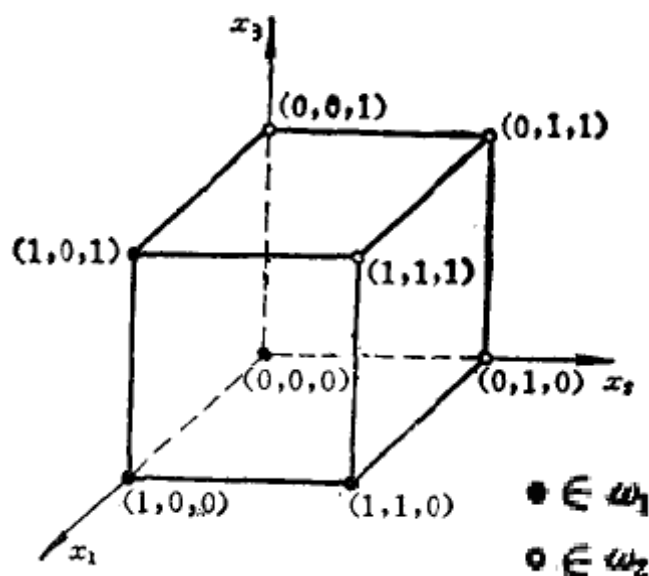


图 2-13 两类点的分布

解： w_1 和 w_2 表示两个类型，由图可知，两个类型的样本：

$$w_1: (0,0,0)^T, (1,0,0)^T, (1,1,0)^T, (0,1,0)^T$$

$$w_2: (0,1,1)^T, (0,0,1)^T, (0,1,1)^T, (1,1,1)^T$$

用各类样本的算术平均值近似代替各类均值向量，也就是：

$$\mu_i \approx \frac{1}{N_i} \sum_{k=1}^{N_i} x_{ik}$$

N_i 为 w_i 中的样本数， x_{ik} 表示 w_i 的第 k 个样本。

协方差矩阵由其定义求得：

$$\Sigma_i = R_i - \mu_i \mu_i^T = \frac{1}{N_i} \sum_{k=1}^{N_i} x_{ik} \cdot x_{ik}^T - \mu_i \mu_i^T$$

式中 R_i 为类 w_i 的自相关函数。

由题中所给条件： $i = 1, 2$ ， $N_1 = N_2 = 4$

$$\text{有： } \mu_1 = \frac{1}{4}(3,1,1)^T, \quad \mu_2 = \frac{1}{4}(1,3,3)^T$$

$$\mu_1 \mu_1^T = \left(\frac{3}{4}, \frac{1}{4}, \frac{1}{4}\right)^T \cdot \left(\frac{3}{4}, \frac{1}{4}, \frac{1}{4}\right) = \begin{pmatrix} \frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix} \cdot \left(\frac{3}{4}, \frac{1}{4}, \frac{1}{4}\right) = \frac{1}{16} \begin{bmatrix} 9 & 3 & 3 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

$$\mu_2 \mu_2^T = \begin{pmatrix} \frac{1}{4} \\ \frac{3}{4} \\ \frac{3}{4} \end{pmatrix} \cdot \left(\frac{1}{4}, \frac{3}{4}, \frac{3}{4}\right) = \frac{1}{16} \begin{bmatrix} 1 & 3 & 3 \\ 3 & 9 & 9 \\ 3 & 9 & 9 \end{bmatrix}$$

$$R_1 = \frac{1}{4} \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \cdot (0,0,0) + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot (1,0,0) + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \cdot (1,1,0) + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \cdot (1,0,1) \right] = \frac{1}{4} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\text{同理: } R_2 = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

$$\Sigma_1 = R_1 - \mu_1 \mu_1^T = \frac{1}{4} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} - \frac{1}{16} \begin{bmatrix} 9 & 3 & 3 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \end{bmatrix} = \frac{1}{16} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

$$\Sigma_2 = \frac{1}{16} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

因此, $\Sigma_1 = \Sigma_2 = \Sigma$ 符合情况二。用情况二的公式确定决策面。

$$\Sigma^{-1} = 4 \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

决策面为 $g_1(x) - g_2(x) = 0 \Rightarrow w^T(x - x_0) = 0$, $w = \Sigma^{-1}(\mu_1 - \mu_2)$,

$x_0 = \frac{1}{2}(\mu_1 + \mu_2)$, 先验概率相等 $P(w_1) = P(w_2)$

$$w = \Sigma^{-1}(\mu_1 - \mu_2) = 4 \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix} \cdot \frac{1}{4} \begin{bmatrix} 2 \\ -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 8 \\ -8 \\ -8 \end{bmatrix}$$

$$x_0 = \frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{2}(1,1,1)^T$$

决策方程: $w^T(x - x_0) = 0$

$$(8, -8, -8) \begin{bmatrix} x_1 - \frac{1}{2} \\ x_2 - \frac{1}{2} \\ x_3 - \frac{1}{2} \end{bmatrix} = 0$$

也就是: $8(x_1 - \frac{1}{2}) - 8(x_2 - \frac{1}{2}) - 8(x_3 - \frac{1}{2}) = 0$

$$8x_1 - 8x_2 - 8x_3 + 4 = 0$$

$2x_1 - 2x_2 - 2x_3 + 1 = 0$ 如下图所示。

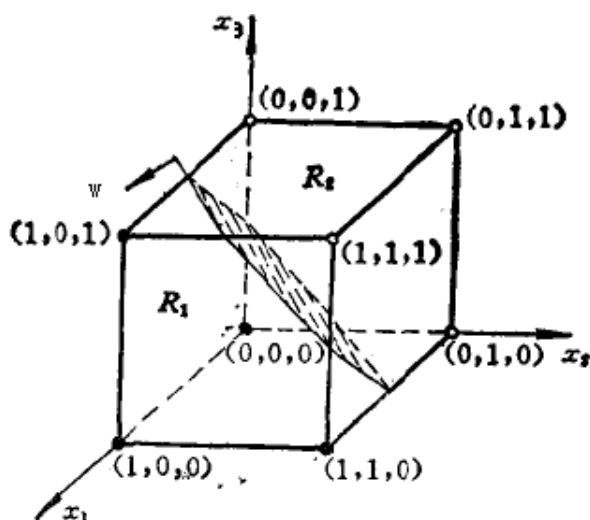


图 2-14 决策方程

w 指向的一侧为正, 是 w_1 的区域 R_1 , 负向的一侧为 w_2 。

2.9 小结

贝叶斯决策论的基本思想非常简单。为最小化总风险, 总是选择那些能够最小化条件风险 $R(\alpha | x)$ 的行为, 尤其是, 为了最小化分类问题中的误差概率, 总是选择那些使后验概率 $\rho(w_i | x)$ 最大的类别。贝叶斯公式允许我们通过先验概率和条件密度来计算后验概率。如果对在模式 w_i 中所做的误分的惩罚与模式 w_j 的不同, 那么在做出判决行为之前, 必须先根据该惩罚函数对后验概率加权。如果内在的分布为多元的高斯分布, 判决边界将是超二次型, 其形状和位置取决于先

验概率、该分布的均值和协方差。

习题

1. 利用概率论中的乘法定理和全概率公式，证明：

(1) 贝叶斯公式

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

(2) 在两类情况下： $P(w_1 | x) + P(w_2 | x) = 1$ 。

2. 分别写出在以下 2 种情况下，贝叶斯最小错误率判决规则：

(1) 两类情况，且 $p(x | w_1) = p(x | w_2)$ 。

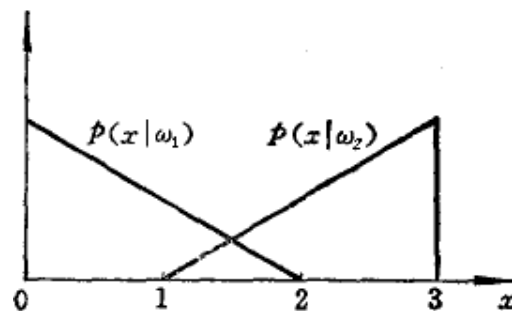
(2) 两类情况，且 $p(w_1) = p(w_2)$ 。

3. 两个一维模式类型，类概率密度函数如下图所示，假定先验概率相等，试用 0-1 损失函数：

(1) 导出贝叶斯判决函数；

(2) 求出分界点的位置；

(3) 判断下列样本各属于哪一个类型：0, 2.5, 0.5, 2, 1.5。



4. 试写出两类情况下的贝叶斯最小风险判决规则及其判决函数和决策面方程，并且证明该判决规则可以表示为：

若 $\frac{p(x | w_1)}{p(x | w_2)} > \frac{(\lambda_{12} - \lambda_{22})P(w_2)}{(\lambda_{21} - \lambda_{11})P(w_1)}$ ，则 $x \in \begin{cases} w_1 \\ w_2 \end{cases}$ 式中， λ_{12} 、 λ_{22} 、 λ_{21} 、 λ_{11} ，为损失函数

$L(\alpha_i | w_j)$, $i, j = 1, 2$ 。若 $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = \lambda_{21}$, 证明此时最小最大决策面使来自两类的错误率相等。

5. 似然比 $l(x)$ 是随机变量, 对两类问题 $l(x) = \frac{p(x | w_1)}{p(x | w_2)}$, 试证明:

$$(1) E[l^n(x) | w_1] = E[l^{n+1}(x) | w_2]$$

$$(2) E[l(x) | w_2] = 1$$

$$(3) E[l(x) | w_1] - E[l(x) | w_2] = \text{Var}[l(x) | w_2]$$

注意: 方差 $D(x) = \text{Var}(x) = E\{[x - E(X)]^2\}$ 。

6. 属于两类的一维模式, 每类都是正态分布的, 并且已知两类的均值分别为 $\mu_1 = 0$ 和 $\mu_2 = 2$, 均方差分别为 $\sigma_1 = 2$ 和 $\sigma_2 = 2$ 先验概率相等, 可用 0—1 损失函数, 试绘出类概率密度函数及判决边界; 若已获得样本: $-3, -2, 1, 3, 5$, 试判断它们各属于哪一种类型。

7. 假设已经获得两类二维的模式样本:

$$w_1 : \{(0,0)^T, (2,0)^T, (2,2)^T, (0,2)^T\}$$

$$w_2 : \{(4,4)^T, (6,4)^T, (6,6)^T, (4,6)^T\}$$

两类均服从正态分布, 且先验概率相等。

(1) 求两类之间的决策面方程;

(2) 绘出决策面。