模式识别

第5讲 统计决策分类法（I）
（统计决策准则）

2018~2019学年

内容安排

一、绪论、数学基础（第1讲）
二、聚类分析（第2讲）
三、判别函数分类法（几何分类法）（第3、4讲）
四、统计决策分类法（概率分类法）（第5、6讲）
五、特征提取与选择（第7讲）
六、模糊模式识别（第8讲）
七、神经网络模式识别（第9讲）

期末考试（平时作业：40%，期末考试：60%）

四、统计决策分类法

4.1 研究对象及相关概率

4.2 贝叶斯决策

4.3 贝叶斯分类器的错误率

4.4 耐曼-皮尔逊决策

4.5 概率密度函数的参数估计

4.6 概率密度函数的非参数估计

4.7 后验概率密度分类的势函数方法

4.1 研究对象及相关概率

1. 两类研究对象

获取模式的观察值时，有二种情况：

- 确定性事件：事物间有确定的因果关系。第三章内容。
- 随机事件：事物间没有确定的因果关系，观察到的特征具有统计特性，是一个随机向量。只能利用模式集的统计特性进行分类，使分类器发生分类错误的概率最小。

2. 相关概率

1）概率的定义

设Ω是随机试验的**基本事件空间**（所有可能实验结果或基本事件的全体构成的集合，也称**样本空间**），A为随机事件，$P(A)$为定义在所有随机事件组成的集合上的实函数，若$P(A)$满足：

（1）对任一事件A有：$0 \le P(A) \le 1$。

（2）$P(\Omega)=1$， Ω——事件的全体

（3）对于两两互斥的事件A_1，A_2，…有

$$P(A_1+A_2+\cdots)=P(A_1)+P(A_2)+\cdots$$

则称函数$P(A)$为事件A的概率。

2）概率的性质

（1）不可能事件V的概率为零，即$P(V)=0$。

（2）$P(\overline{A})=1-P(A)$

(3) $P(A \cup B)=P(A)+P(B)-P(AB)$

联合概率$P(AB)$：

A，B同时发生的概率

3）条件概率定义

设A、B是两个随机事件，且$P(B)>0$，则称

$$P(A|B)=\frac{P(AB)}{P(B)} \qquad (4\text{-}1)$$

为事件B发生的条件下事件A发生的条件概率。

4）条件概率的三个重要公式：

（1）**概率乘法公式**：如果$P(B)>0$，则联合概率

$$P(AB)=P(B)P(A|B)=P(A)P(B|A)=P(BA) \qquad (4\text{-}2)$$

（2）**全概率公式**：设事件A_1，A_2，…，A_n，两两互斥，且

$$\bigcup_{i=1}^{n} A_i=\Omega, \quad P(A_i)>0, \quad i=1,2,\cdots,n$$

则对任一事件B有：

$$P(B)=\sum_{i=1}^{n} P(A_i)P(B|A_i) \qquad (4\text{-}3)$$

（3）**贝叶斯公式（定理）**：在全概率公式的条件下，如果$P(B)>0$，则将 (4-2)，(4-3)式代入(4-1)式中，有：

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} \qquad (4\text{-}4)$$

$$P(A|B)=\frac{P(AB)}{P(B)} \qquad (4-1)$$

补充说明：

（1）对固定的B，只要$P(B)>0$，$P(\cdot|B)$ 也是一个概率测度，满足概率公理。

$$P(A_1A_2|B)= P(A_2|B)\, P(A_1|A_2B)$$

（2）对贝叶斯定理的解释

$$P(A_i \mid B)=\frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

如果把B视为某事件的观察**结果**（或**事件证据**），各个A_i视为事件可能的互斥**原因**之一，则$P(A_i)$称为**原因A_i的先验概率——未获得观察结果前，对原因A_i在未来产生结果的概率的认定（根据经验、专业知识等）**，而 $P(A_i|B)$ 则称为**原因A_i的后验概率——获得观察结果B之后，结合先验知识，计算出的对结果B是由原因A_i引起的概率。**

5）模式识别中这三个概率的解释

设随机样本向量**X**，相关的三个概率：

（1）先验概率$P(\omega_i)$：根据以前的知识和经验得出的、未来样本**X**可能属于ω_i类的概率——属**事先猜测**，故称**先验概率**。

（2）后验概率$P(\omega_i|\boldsymbol{X})$：指获得样本**X**后，根据样本提供的信息计算出的、当前**X** 可能属于ω_i类的概率。虽然仍然是概率意义上的**事后猜测**，但这是在得到样本**X**后的概率，故称为**后验概率**。

（3）条件概率$P(\boldsymbol{X}|\omega_i)$：对属于$\omega_i$类的样本，其取值为**X**的概率。例：对一批得病患者进行化验，结果为阳性的概率为95%，ω_1代表人群得病， 则“患病者$(\boldsymbol{X} \in \omega_1)$的化验结果**X**为阳性这一事件的概率”可表示为

$$P(\boldsymbol{X}=\text{阳}|\omega_1)=0.95$$

今后的分类中常用到类概率密度 $p(\boldsymbol{X}|\omega_i)$：$\omega_i$类的条件概率密度函数、 ω_i类模式的总体分布，通常也称为ω_i的似然函数。

例如：一个两类问题，ω_1表示患有某病，ω_2表示未患病，

则：$P(\omega_1)$ 表示某地区的人患有此病的概率，
$P(\omega_2)$ 表示该地区人无此病的概率。

通过统计资料得到

若用某种方法检测是否患有某病，假设 $\boldsymbol{X}$ 表示“试验反应呈阳性”。则：

$P(\boldsymbol{X}\,|\omega_2)$ 表示无病的人群做该试验时反应呈阳性(显示有病)的概率。

值低 / 高？

$P(\omega_2|\,\boldsymbol{X})$ 表示试验呈阳性的人中，根据计算得出的无病的概率。

值低 / 高？

$P(\boldsymbol{X}\,|\omega_1)$ 表示患病人群做该试验时反应呈阳性的概率。

$P(\omega_1|\,\boldsymbol{X})$ 表示试验呈阳性的人中，根据计算得出的有病的概率。

（4）三者关系：根据(4-4)贝叶斯公式有

$$P(A_i \mid B)=\frac{P(A_i)P(B \mid A_i)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

$$P(\omega_i \mid \boldsymbol{X})=\frac{p(\boldsymbol{X} \mid \omega_i)P(\omega_i)}{p(\boldsymbol{X})}=\frac{p(\boldsymbol{X} \mid \omega_i)P(\omega_i)}{\sum_{i=1}^{M} p(\boldsymbol{X} \mid \omega_i)P(\omega_i)} \qquad （4-5）$$

M：类别数

P(X)：边缘密度，与 $P(\omega_i)$ 的选择有关，因此不能视为独立的，虽然记号没有表示这点！

贝叶斯统计 (Bayesian Statistics)

- **贝叶斯方法是基于贝叶斯定理而发展起来用于系统地阐述和解决统计问题的方法。**
  ——Samuel Kotz 和吴喜之，2000年。

- **贝叶斯推断的基本方法是将关于未知参数的先验信息与样本信息综合，再根据贝叶斯定理，得出后验信息，然后根据后验信息去推断未知参数。**
  ——茆诗松和王静龙等，1998年。

- **贝叶斯提出了一种归纳推理的理论(贝叶斯定理)，以后被一些统计学者发展为一种系统的统计推断方法，称为贝叶斯方法。**
  ——摘自《中国大百科全书》（数学卷）

托马斯·贝叶斯

（Thomas Bayes）(1702–1761)

贝叶斯是英国数学家。**1702**年生于伦敦；**1761**年**4**月**17**日卒于坦布里奇韦尔斯。

贝叶斯是一位自学成才的数学家。曾助理宗教事务，后来长期担任坦布里奇韦尔斯地方教堂的牧师。**1742**年，贝叶斯被选为英国皇家学会会员。

如今在概率、数理统计学中以贝叶斯姓氏命名的有贝叶斯公式、贝叶斯风险、贝叶斯决策函数、贝叶斯决策规则、贝叶斯估计量、贝叶斯方法、贝叶斯统计等等。

贝叶斯**1763**年在《论有关机遇问题的求解》中提出一种归纳推理的理论，后被一些统计学者发展为一种系统的统计推断方法，称为贝叶斯方法。采用这种方法作统计推断所得的全部结果，构成贝叶斯统计的内容。认为贝叶斯方法是唯一合理的统计推断方法的统计学者，组成数理统计学中的贝叶斯学派，其形成可追溯到 **20**世纪 **30**年代。到**50～60**年代，已发展为一个有影响的学派。时至今日，其影响日益扩大。

统计学中有两个主要学派：频率学派与贝叶斯学派。下面从统计推断的三种信息来说明他们之间的区别与联系。

- 频率学派的观点：统计推断是根据样本信息对总体分布或总体的特征数进行推断，这里用到两种信息：总体信息和样本信息。
- 贝叶斯学派的观点：除了上述两种信息以外，统计推断还应该使用第三种信息：先验信息。

（1）**总体信息**:总体分布形式本身提供的信息（模型信息）。

（2）**样本信息**:抽取的样本观测值提供的信息（数据信息）。

（3）**先验信息**:人们在试验之前对要做的问题在经验上和资料上总是有所了解的，这些信息对统计推断是有益的。先验信息即是抽样之前有关统计问题的一些信息。一般说来，先验信息来源于经验和历史资料。先验信息在日常生活和工作中是很重要的。

（经验信息、历史信息）

三种统计信息

- 一、总体信息，即总体分布或总体所属分布给我们的信息。
- 例如：“总体是正态分布”“总体是均匀分布”
- 说明：总体信息是很重要的信息，为了获取此种信息往往耗资巨大。
- 二、样本信息，即从总体抽取的样本给我们的信息。（愈多愈好）
- 人们希望通过对样本的加工和处理对总体的某些特征做出较为精确的统计推断。
- 例：有了样本观察值，我们可根据它大概知道总体的一些特征数（均值、方差等）在一个什么范围内。

三、先验信息，即是抽样（试验）之前有关统计问题（尤其分布参数）的一些信息。

一般说来，先验信息来源于经验和历史资料。先验 信息在日常生活和工作中是很重要的。

例(1) 英国统计学家Savage曾考察如下2个统计实验：

A。一位常饮牛奶加茶的妇女声称，她能辨别先倒进杯子里的是茶还是牛奶。对此做了10次试验，她都正确地说出了。

B。一位音乐家声称，他能从一页乐谱辨别出是海顿还是莫扎特的作品。在10次这样的试验中，他都能正确辨别。

在这两个统计试验中，假如认为被试验者是在猜测，每次成功的概率为0.5，那么10次都猜中的概率为2^{-10}=0.0009766，这是一个很小的概率，是几乎不可能发生的，所以“每次成功概率为0.5”的假设应该被拒绝。

被试验者每次成功的概率要比0.5大得多。说明：这不是50-50猜测，而是他们的经验帮了他们的忙。

例(2) “免检产品”是怎样决定的？某厂的产品每天都有抽验几件，获得不合格品率 θ 的估计。在经过一段时间后就积累大量的资料，根据这些历史资料（先验信息的一种）对过去产品的不合格品率可构造一个分布：

$$P(\theta=\frac{i}{n})=\pi_i, i=0,1,\ldots,n$$

这个对先验信息进行加工获得的分布今后称为先验分布。

这个先验分布是综合了该厂过去产品的质量情况。如果这个分布的概率大部分集中在θ=0附近，那么该产品可认为是“信得过产品”。假如以后的多次抽检结果与历史资料提供的先验分布是一致的。使用单位就可以对它做出“免检产品”的决定，或者每月抽检一、二次就足够了，这就省去了大量的人力和物力。

可见历史资料在统计推断中应加以利用。

Bayes统计学派与经典统计学派在很多问题上都有分歧但是它们最根本的分歧是:

第一，是否利用先验信息。由于产品的设计、生产都有一定的继承性，这样就存在许多相关产品的信息以及先验信息可以利用，Bayes统计学派认为利用这些先验信息不仅可以减少样本容量，而且在很多情况还可以提高统计精度；而经典统计学派忽略了这些信息。

第二，是否将参数θ看成随机变量。Bayes统计学派的最基本的观点是:任一未知量θ都可以看成随机变量，关于这个未知量的任何现有知识一般用一个概率分布去描述，这个分布就是先验信息（因为任一未知量都具有不确定性，而在表述不确定性时，概率与概率分布是最好的语言）。相反，经典统计学派却把未知量θ 就简单看成一个确定但未知参数，来对它进行统计推断。

贝叶斯公式

一、贝叶斯公式的两种形式（密度函数、质量函数）

1. 总体分布依赖于参数θ的概率函数在贝叶斯统计中记为$P(x\mid\theta)$，它表示在随机变量 θ取某个给定值时总体的条件概率函数；

2. 根据参数θ的先验信息可确定先验分布$\pi(\theta)$；

3. 从贝叶斯观点看，样本 $x=$（$x_1, x_2, \ldots, x_n$ ）的产生分两步进行:首先从先验分布$\pi(\theta)$产生一个样本θ_0，然后从$P(x\mid\theta_0)$中产生一组样本。这时样本的联合条件概率函数正比于

$$p(x\mid\theta_0)=\prod_{i=1}^{n}p(x_i\mid\theta_0)$$

这个分布综合了总体信息和样本信息，合称为抽样信息，也被称为：参数θ的（关于样本x 的）似然函数。

4. θ_0 是未知的，它是按先验分布$\pi(\theta)$产生的。为把先验信息综合进去，不能只考虑θ_0，对θ的其它值发生的可能性也要加以考虑，故要用$\pi(\theta)$进行综合。这样一来，样本x=（$x_1, \ldots, x_n$）和参数θ的联合分布为:

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

这个联合分布把总体信息、样本信息和先验信息三种可用信息都综合进去了。

5. 我们的任务是对未知数θ作出推断。在没有样本信息时，人们只能依据先验分布对θ作出推断。在有了样本观察值$x=(x_1, x_2, \ldots, x_n)$之后，则应依据$h(x, \theta)$对$\theta$作出推断。由于

$$h(x, \theta) = \pi(\theta \mid x) m(x),$$

其中

$$m(x_1,\cdots,x_n) = \int_\Theta h(x_1,\cdots,x_n,\theta)d\theta = \int_\Theta p(x_1,\cdots,x_n \mid \theta)\pi(\theta)d\theta$$

是$x=(x_1, x_2, \ldots, x_n)$的边际概率分布，它与$\theta$无关，不含$\theta$的任何信息。因此能用来对$\theta$作出推断的仅是后验分布$\pi(\theta \mid x)$，它的计算公式是

$$\boxed{\pi(\theta \mid x) = \frac{h(x,\theta)}{m(x)} = \frac{p(x \mid \theta)\pi(\theta)}{\int_\Theta p(x \mid \theta)\pi(\theta)d\theta} \qquad (1.1)}$$

这就是贝叶斯公式的概率密度函数形式。

这个“在观察到x的条件下θ的条件分布”称为θ的后验分布，它集中了总体、样本和先验中有关θ的一切信息，而又是排除一切与θ无关的信息之后得到的结果。

后验分布$\pi(\theta|x)$的计算公式就是用密度函数表示的贝叶斯公式。它是用总体和样本对先验分布$\pi(\theta)$作调整的结果，贝叶斯统计的一切推断都基于后验分布进行。

6. 在θ是离散型随机变量时，先验分布可用先验分布列$\pi(\theta_i)$，$i=1,2,\ldots$，表示。这时后验分布也是离散形式（贝叶斯公式的概率质量函数形式）

$$\boxed{\pi(\theta_i \mid x)=\frac{p(x\mid\theta_i)\pi(\theta_i)}{\sum_j p(x\mid\theta_j)\pi(\theta_j)} \qquad i=1,2,\ldots.(1.2)}$$

假如总体X也是离散的，把（1.1）或（1.2）中的密度函数$p(x|\theta)$作为概率质量函数$p(X=x|\theta)$即可。

二、后验分布是三种统计信息的综合

一般说来，先验分布$\pi(\theta)$反映了人们抽样前对θ的认识，后验分布$\pi(\theta|x)$反映了人们在抽样后对θ的认识。它们之间的差异是由于样本x出现后人们对θ认识的一种调整。所以后验分布$\pi(\theta|x)$可以看做是人们用总体信息$p(\cdot|\theta)$和样本信息 x（统称为抽样信息）对先验信息$\pi(\theta)$作调整后的结果。

例(3) 设某事件A在一次试验中发生的概率为θ，为估计θ，对试验进行了n次独立观测，其中事件A发生了X次，显然 $X|\theta \sim b(n,\theta)$，即

$$P(X=x|\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\quad x=0,1,\cdots,n$$

这是似然函数。

假若我们在试验前对事件A没有什么了解，从而对其发生概率θ也没有任何信息。此时，贝叶斯本人建议采用“同等无知”的原则使用区间（0,1）上的均匀分布$U(0,1)$作为θ的先验分布，因为它取（0,1）上的每一点的机会均等。贝叶斯的这个建议被后人称为贝叶斯假设。

θ 的先验分布为$\pi(\theta)=\begin{cases}1, 0<\theta<1\\ 0, \text{其它}\end{cases}$

由此即可利用贝叶斯公式求出θ的后验分布。具体如下：先写出X和θ的联合分布

$$h(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\quad x=0,1,\cdots,n,\quad \boxed{0<\theta<1}$$

然后求X的边际分布

$$m(x)=\int_0^1\theta^{x}(1-\theta)^{n-x}d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

最后求出θ的后验分布

$$\pi(\theta\,|\,x)=\frac{h(x,\theta)}{m(x)}=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},\quad 0<\theta<1$$

最后的结果说明θ的后验分布为$Be(x+1,n-x+1)$，

4.2 贝叶斯决策

4.2.1 最小错误率贝叶斯决策

1. 问题分析

讨论模式集的分类，目的是确定$\boldsymbol{X}$属于那一类，所以要看$\boldsymbol{X}$来自哪类的概率大。在下列三种概率中：

① 先验概率$P(\omega_i)$

② 类(条件)概率密度$p(\boldsymbol{X}|\omega_i)$

③ 后验概率$P(\omega_i|\boldsymbol{X})$

采用哪种概率进行分类最合理？ 后验概率$P(\omega_i|\boldsymbol{X})$

2. 决策规则

设有M类模式，

若 $P(\omega_i|\boldsymbol{X}) = \max\{P(\omega_j|\boldsymbol{X})\},\ j=1,2,\cdots,M$ 则 $\boldsymbol{X}\in\omega_i$类 (4-6)

—— 最小错误率贝叶斯决策规则

若 $P(\omega_i \mid \boldsymbol{X}) = \max\{P(\omega_j \mid \boldsymbol{X})\}, j = 1, 2, \cdots, M$

则 $\boldsymbol{X} \in \omega_i$类

几种等价形式：

虽然后验概率$P(\omega_i | \boldsymbol{X})$可以提供有效的分类信息，但直接获得后验概率是困难的。根据Bayes定理，后验概率可以由类概率密度函数和先验概率计算得出。实际上，先验概率$P(\omega_i)$和类概率密度函数$p(\boldsymbol{X} | \omega_i)$更容易从统计资料中容易获得。

$$P(\omega_i | \boldsymbol{X}) = \frac{p(\boldsymbol{X} \mid \omega_i) P(\omega_i)}{\sum_{i=1}^{M} p(\boldsymbol{X} \mid \omega_i) P(\omega_i)}$$

可知，分母与i无关，即与分类无关，故分类规则又可表示为：

若 $p(\boldsymbol{X} \mid \omega_i) P(\omega_i) = \max\{p(\boldsymbol{X} \mid \omega_j) P(\omega_j)\}\ j = 1, 2, \cdots, M$， 则 $\boldsymbol{X} \in \omega_i$类

(4-7)

对两类问题，(4-7)式相当于

$$\text{若}\ p(\boldsymbol{X}\mid\omega_1)P(\omega_1) > p(\boldsymbol{X}\mid\omega_2)P(\omega_2)\quad，\ \text{则}\ \boldsymbol{X}\in\omega_1$$

$$\text{若}\ p(\boldsymbol{X}\mid\omega_1)P(\omega_1) < p(\boldsymbol{X}\mid\omega_2)P(\omega_2)\quad，\ \text{则}\ \boldsymbol{X}\in\omega_2$$

可改写为：

$$\text{若}\ l_{12}(\boldsymbol{X}) = \frac{p(\boldsymbol{X}\mid\omega_1)}{p(\boldsymbol{X}\mid\omega_2)} \gtrless \frac{P(\omega_2)}{P(\omega_1)}，\ \text{则}\ \boldsymbol{X}\in\begin{cases}\omega_1\\ \omega_2\end{cases} \tag{4-8}$$

统计学中称$l_{12}(\boldsymbol{X})$为X处的似然比，$P(\omega_2)/P(\omega_1)$为似然比阈值。

对(4-8)式取自然对数，有：

$$\text{若}h(\boldsymbol{X}) = \ln l_{12}(\boldsymbol{X})$$

$$= \ln p(\boldsymbol{X}\mid\omega_1) - \ln p(\boldsymbol{X}\mid\omega_2) \gtrless \ln\frac{P(\omega_2)}{P(\omega_1)}，\ \text{则}\ \boldsymbol{X}\in\begin{cases}\omega_1\\ \omega_2\end{cases} \tag{4-9}$$

(4-7)，(4-8)，(4-9)都是最小错误率贝叶斯决策规则的等价形式。

例4.1 假定在细胞识别中，病变细胞的先验概率和正常细胞的先验概率分别为$P(\omega_1)=0.05,\ P(\omega_2)=0.95$。现有一待识别细胞，其观察值为$\boldsymbol{X}$，从类条件概率密度分布曲线上查得：

$$p(\boldsymbol{X}\mid\omega_1)=0.5 \qquad p(\boldsymbol{X}\mid\omega_2)=0.2$$

试对细胞$\boldsymbol{X}$进行分类。

解：[方法1] 通过后验概率计算。

$$P(\omega_1\mid\boldsymbol{X})=\frac{p(\boldsymbol{X}\mid\omega_1)P(\omega_1)}{\sum_{i=1}^{2}p(\boldsymbol{X}\mid\omega_i)P(\omega_i)}$$

$$=\frac{0.5\times0.05}{0.05\times0.5+0.95\times0.2}\approx0.16$$

$$P(\omega_2\mid\boldsymbol{X})=\frac{0.2\times0.95}{0.05\times0.5+0.95\times0.2}\approx0.884$$

$$\because P(\omega_2\mid\boldsymbol{X})>P(\omega_1\mid\boldsymbol{X}) \qquad \therefore \boldsymbol{X}\in\omega_2$$

[方法2]：利用先验概率和类概率密度计算。

$$p(\boldsymbol{X} \mid \omega_1)P\left(\omega_1\right) = 0.5 \times 0.05 = 0.025$$

$$p(\boldsymbol{X} \mid \omega_2)P\left(\omega_2\right) = 0.2 \times 0.95 = 0.19$$

$$\because\ p(\boldsymbol{X} \mid \omega_2)P\left(\omega_2\right) > p(\boldsymbol{X} \mid \omega_1)P(\omega_1)$$

$\therefore\ \boldsymbol{X} \in \omega_2$，是正常细胞。

4.2.2 最小风险贝叶斯决策

1. 风险的概念

* 自动灭火系统
* 疾病诊断

不同的错判造成的损失不同，因此风险不同，两者紧密相连 。

考虑到对某一类的错判要比对另一类的错判更为关键，把最小错误率的贝叶斯判决做一些修改，提出了“条件风险”的概念。

最小风险贝叶斯决策基本思想：

以各种错误分类所造成的条件风险（后验风险）最小为规则，进行分类决策。

2. 决策规则

对M类问题，如果把观察样本$\boldsymbol{X}$判定为ω_i类记为决策a_i，则**决策a_i的条件风险$r_i(\boldsymbol{X})$**是指$\boldsymbol{X}$判定为ω_i类造成的**平均损失**（**条件**：指观察到$\boldsymbol{X}$）。**风险：**指**平均损失**（把X实属其他类别却被硬判为ω_i类时引起的**损失，**通过后验概率进行加权**平均**）

L_{ij}对P作加权平均

$$r_i(\boldsymbol{X}) = \sum_{j=1}^{M} L_{ij}(\boldsymbol{X}) P(\omega_j \mid \boldsymbol{X})$$

对观测X，根据总体信息、先验信息“计算出的”属于j类概率。因此，如果X被划分到i类，损失就是该概率与损失函数L_{ij}之积

式中，

i ——分类判决后指定的判决号；

j ——样本实际属于的类别号；

L_{ij}——将实属ω_j类的样本决策为ω_i类时的误判代价，即损失函数。注意下标顺序及其含义：$L_{i\leftarrow j}$

$$L_{ij}(\boldsymbol{X}) = \begin{cases} 0\text{或负值} & i = j\text{时} \\ \text{正值} & i \neq j\text{时} \end{cases}$$

注：这里的L_{ij}与孙即祥书中的λ_{ij}的含义相反！

对每个$\boldsymbol{X}$有M种可能的类别划分，$\boldsymbol{X}$被判决为各个类的条件风险分别为$r_1(\boldsymbol{X})$，$r_2(\boldsymbol{X})$，…，$r_M(\boldsymbol{X})$。决策规则：

若 $r_k(\boldsymbol{X}) = \min\{r_i(\boldsymbol{X}), \quad i = 1, \cdots, M\}$ 则 $\boldsymbol{X} \in \omega_k$

每个$\boldsymbol{X}$ 都按条件风险最小决策，则总的条件风险也最小。总的条件风险称为总的风险。

条件风险 与
总的风险 的 区别

条件风险：对某个样本而言。
总的风险：对总体样本而言。

1）多类情况

设有M 类，对于任一X 对应 M个条件风险（后验风险）：

$$r_i(\boldsymbol{X}) = \sum_{j=1}^{M} L_{ij}(\boldsymbol{X}) P(\omega_j \mid \boldsymbol{X}) \quad , \quad i=1,2, \ldots, M$$

用先验概率和条件概率的形式：

$$r_i(\boldsymbol{X}) = \sum_{j=1}^{M} L_{ij} P(\omega_j \mid \boldsymbol{X}) = \sum_{j=1}^{M} L_{ij} \frac{p(\boldsymbol{X} \mid \omega_j) P(\omega_j)}{p(\boldsymbol{X})}$$

$$= \frac{1}{p(\boldsymbol{X})} \sum_{j=1}^{M} L_{ij} p(\boldsymbol{X} \mid \omega_j) P(\omega_j)$$

$\because$ $p(\boldsymbol{X})$对所有类别一样，不提供分类信息。

$$\therefore\ r_i(\boldsymbol{X}) = \sum_{j=1}^{M} L_{ij} p(\boldsymbol{X} \mid \omega_j) P(\omega_j) \quad , \quad i=1,2,\ldots,M$$

决策规则为：

若 $r_k(\boldsymbol{X}) < r_i(\boldsymbol{X}),\quad i = 1,2,\cdots,M;\ i \neq k$，则 $\boldsymbol{X} \in \omega_k$

$$\therefore\ r_i(\boldsymbol{X})=\sum_{j=1}^{M}L_{ij}\,p(\boldsymbol{X}\mid\omega_j)P(\omega_j)$$

2）两类情况：对样本 $\boldsymbol{X}$

当 $\boldsymbol{X}$ 被判为 ω_1 类时：

$$r_1(\boldsymbol{X})=L_{11}p(\boldsymbol{X}\mid\omega_1)P(\omega_1)+L_{12}p(\boldsymbol{X}\mid\omega_2)P(\omega_2)$$

当 X 被判为 ω_2 类时：

$$r_2(\boldsymbol{X})=L_{21}p(\boldsymbol{X}\mid\omega_1)P(\omega_1)+L_{22}p(\boldsymbol{X}\mid\omega_2)P(\omega_2)$$

决策规则：

若 $r_1(\boldsymbol{X})<r_2(\boldsymbol{X})$ 　则 $\boldsymbol{X}\in\omega_1$ 　（4-15）

若 $r_1(\boldsymbol{X})>r_2(\boldsymbol{X})$ 　则 $\boldsymbol{X}\in\omega_2$ 　（4-16）

由（4-15）式：

$$\underline{L_{11}p(\boldsymbol{X}\mid\omega_1)P(\omega_1)}+L_{12}p(\boldsymbol{X}\mid\omega_2)P(\omega_2)<\underline{L_{21}p(\boldsymbol{X}\mid\omega_1)P(\omega_1)}+L_{22}p(\boldsymbol{X}\mid\omega_2)P(\omega_2)$$

$$(L_{12}-L_{22})p(\boldsymbol{X}\mid\omega_2)P(\omega_2)<(L_{21}-L_{11})p(\boldsymbol{X}\mid\omega_1)P(\omega_1)$$

$$\frac{p(\boldsymbol{X}\mid\omega_1)}{p(\boldsymbol{X}\mid\omega_2)}>\frac{(L_{12}-L_{22})P(\omega_2)}{(L_{21}-L_{11})P(\omega_1)}$$

令L_{ii}=0。观察L_{12}与L_{21}*的大小引起的阈值的变化*

$$\frac{p(\boldsymbol{X}\mid\omega_1)}{p(\boldsymbol{X}\mid\omega_2)} > \frac{(L_{12}-L_{22})P(\omega_2)}{(L_{21}-L_{11})P(\omega_1)}$$

令： $l_{12}(\boldsymbol{X}) = \dfrac{p(\boldsymbol{X}\mid\omega_1)}{p(\boldsymbol{X}\mid\omega_2)}$ ，称X处的似然比；

$\theta_{12} = \dfrac{(L_{12}-L_{22})P(\omega_2)}{(L_{21}-L_{11})P(\omega_1)}$ ，为似然比阈值

类概率密度函数 $p(\cdot\mid\omega_i)$ 称为 ω_i的似然函数

判别步骤：

① 定义损失函数L_{ij}。

② 计算似然比阈值 θ_{12}。

③ 计算似然比 $l_{12}(\boldsymbol{X})$ 。

④ 若 $l_{12}(\boldsymbol{X}) > \theta_{12}$， 则 $\boldsymbol{X} \in \omega_1$

若 $l_{12}(\boldsymbol{X}) < \theta_{12}$， 则 $\boldsymbol{X} \in \omega_2$

若 $l_{12}(\boldsymbol{X}) = \theta_{12}$， 任意判决

例4.2　在细胞识别中，病变细胞和正常细胞的先验概率 分别为

$$P(\omega_1)=0.05,\quad P(\omega_2)=0.95$$

现有一待识别细胞，观察值为$\boldsymbol{X}$， 从类概率密度分布曲线上查得

$$p(\boldsymbol{X}\mid\omega_1)=0.5,\quad p(\boldsymbol{X}\mid\omega_2)=0.2$$

损失函数分别为L_{11}=0，L_{21}=10， L_{22}=0，L_{12}=1。按最小风险贝叶斯决策分类。

解：计算 θ_{12} 和 $l_{12}(\boldsymbol{X})$ 得：

$$\theta_{12}=\frac{(L_{12}-L_{22})P(\omega_2)}{(L_{21}-L_{11})P(\omega_1)}=\frac{(1-0)\times 0.95}{(10-0)\times 0.05}=1.9$$

$$l_{12}(\boldsymbol{X})=\frac{p(\boldsymbol{X}\mid\omega_1)}{p(\boldsymbol{X}\mid\omega_2)}=\frac{0.5}{0.2}=2.5$$

$$\because l_{12}(\boldsymbol{X})>\theta_{12}\qquad \therefore \boldsymbol{X}\in\omega_1$$

为病变细胞。

思考：为什么同样检查结果，这里被判为病变细胞？

3. (0-1)损失最小风险贝叶斯决策

损失函数为特殊情况：$L_{ii}=0 \quad L_{ij}=1,\ i\neq j$

1) 多类情况

一般形式：$r_i(\boldsymbol{X})=\sum_{j=1}^{M}L_{ij}\,p(\boldsymbol{X}\mid\omega_j)P(\omega_j)$ ，$i=1,2,\ldots,M$

若 $r_k(\boldsymbol{X})<r_i(\boldsymbol{X}),\ i\neq k$，则 $\boldsymbol{X}\in\omega_k$

(0-1)情况下，$r_i(\boldsymbol{X})$ 可改写成：

$$r_i(\boldsymbol{X})=\sum_{j=1}^{M}p(\boldsymbol{X}\mid\omega_j)P(\omega_j)-p(\boldsymbol{X}\mid\omega_i)P(\omega_i)$$

$$=p(\boldsymbol{X})-p(\boldsymbol{X}\mid\omega_i)P(\omega_i)$$

当 $\boldsymbol{X}\in\omega_k$ 时，应满足

$$p(\boldsymbol{X})-p(\boldsymbol{X}\mid\omega_k)P(\omega_k)<p(\boldsymbol{X})-p(\boldsymbol{X}\mid\omega_i)P(\omega_i)$$

$$p(\boldsymbol{X}\mid\omega_k)P(\omega_k)>p(\boldsymbol{X}\mid\omega_i)P(\omega_i)$$

$$p(\boldsymbol{X} \mid \omega_k)P(\omega_k) > p(\boldsymbol{X} \mid \omega_i)P(\omega_i)$$

判别函数的等价形式：$d_i(\boldsymbol{X}) = p(\boldsymbol{X} \mid \omega_i)P(\omega_i), \qquad i = 1,2,\cdots,M$

决策规则的等价形式：

若 $d_k(\boldsymbol{X}) > d_i(\boldsymbol{X}), \qquad i = 1,2,\cdots,M,\ i \neq k$

则 $\boldsymbol{X} \in \omega_k$

——最小错误率贝叶斯决策

2) 两类情况

$$L_{11} = L_{22} = 0, \qquad L_{12} = L_{21} = 1$$

$$d_1(\boldsymbol{X}) = p(\boldsymbol{X} \mid \omega_1)P(\omega_1) \qquad\qquad d_2(\boldsymbol{X}) = p(\boldsymbol{X} \mid \omega_2)P(\omega_2)$$

决策规则为

$$\begin{cases} 若\ d_1(\boldsymbol{X}) > d_2(\boldsymbol{X}), & 则\ \boldsymbol{X} \in \omega_1 \\ 若\ d_2(\boldsymbol{X}) > d_1(\boldsymbol{X}), & 则\ \boldsymbol{X} \in \omega_2 \end{cases} \qquad (4\text{-}20)$$

$$d_i(\boldsymbol{X}) = p(\boldsymbol{X} \mid \omega_i) P(\omega_i)$$

$$\begin{cases} 若\ d_1(\boldsymbol{X}) > d_2(\boldsymbol{X}),\quad 则\ \boldsymbol{X} \in \omega_1 \\ 若\ d_2(\boldsymbol{X}) > d_1(\boldsymbol{X}),\quad 则\ \boldsymbol{X} \in \omega_2 \end{cases}$$

或从式(4-20) 导出似然比形式：

对$\boldsymbol{X} \in \omega_1$的情况：$p(\boldsymbol{X} \mid \omega_1) P(\omega_1) > p(\boldsymbol{X} \mid \omega_2) P(\omega_2)$

$$\frac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$$

式中：$l_{12}(\boldsymbol{X}) = \dfrac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)}, \theta_{12} = \dfrac{P(\omega_2)}{P(\omega_1)}$

类似地，对$\boldsymbol{X} \in \omega_2$的情况有：$\dfrac{p(\boldsymbol{X} \mid \omega_1)}{p(\boldsymbol{X} \mid \omega_2)} < \dfrac{P(\omega_2)}{P(\omega_1)}$

决策规则： 若 $l_{12}(\boldsymbol{X}) > \theta_{12}$， 则 $\boldsymbol{X} \in \omega_1$

若 $l_{12}(\boldsymbol{X}) < \theta_{12}$， 则 $\boldsymbol{X} \in \omega_2$

$L_{ij}(\boldsymbol{X})$的确定：根据错误造成损失的严重程度，及专家经验确定。

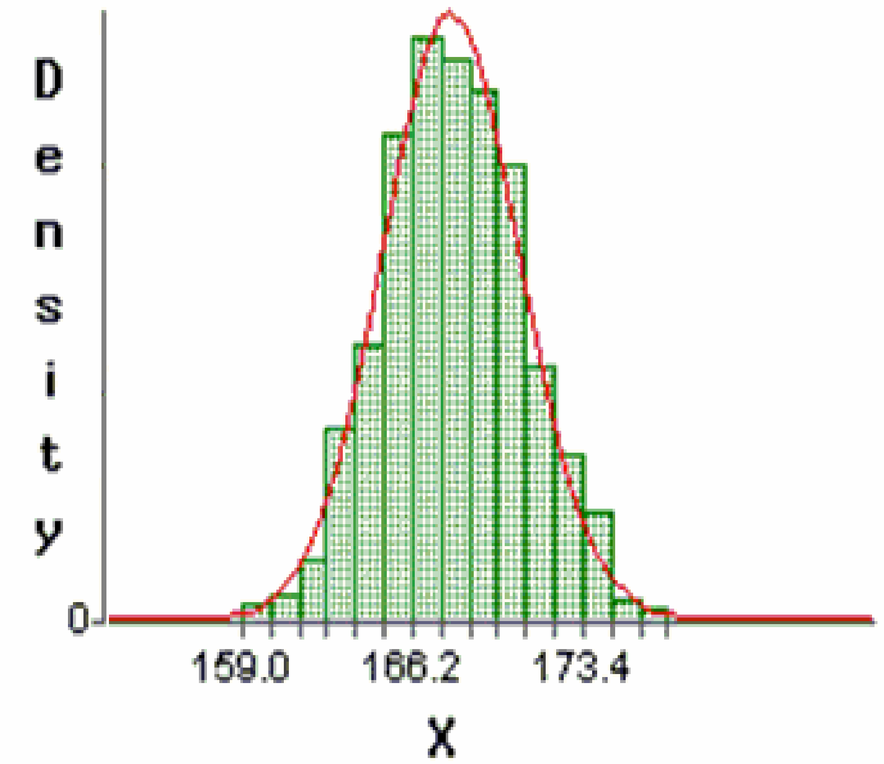
4.2.3 正态分布模式的贝叶斯决策

许多实际的数据集：

均值附近分布较多的样本；

距均值点越远，样本分布越少。

此时正态分布（高斯分布）是一种合理的近似。



正态分布概率模型的优点：

* 物理上的合理性。

* 数学上的简单性。

图中为某大学男大学生的身高数据，红线是拟合的密度曲线。可见，其身高应服从正态分布。

1. 相关知识概述

1）二次型

设一向量$\boldsymbol{X}=\left[x_1,\cdots x_n\right]^{\mathrm{T}}$，矩阵 $\boldsymbol{A}=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$

则 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}$ 称为二次型。

含义：是一个二次齐次多项式，$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}=\sum_{i,j=1}^{n}a_{ij}x_i x_j$

二次型中的矩阵$\boldsymbol{A}$是一个对称矩阵，即$a_{ij}=a_{ji}$ 。
若不是对称阵，可以简单化为对称阵——如何做到？

2）正定二次型

对于$\forall \boldsymbol{X}\neq \boldsymbol{0}$（即$\boldsymbol{X}$分量不全为零），总有$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}>0$，则称此二次型是正定的，而其对应的矩阵称为正定矩阵。

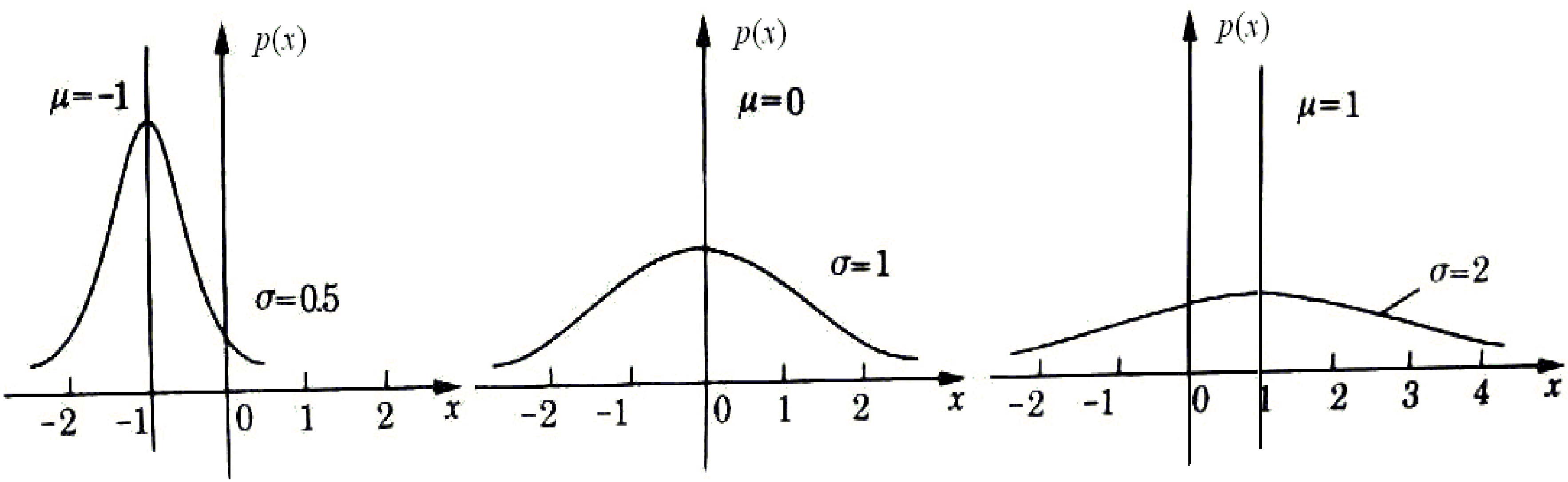
3）单变量（一维）的正态分布

密度函数定义为：

$$p(x)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right\}=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\qquad -\infty<x<\infty$$

曲线如图示：

①μ= -1，σ=0.5 ； ②μ= 0，σ=1 ； ③μ= 1，σ=2 .



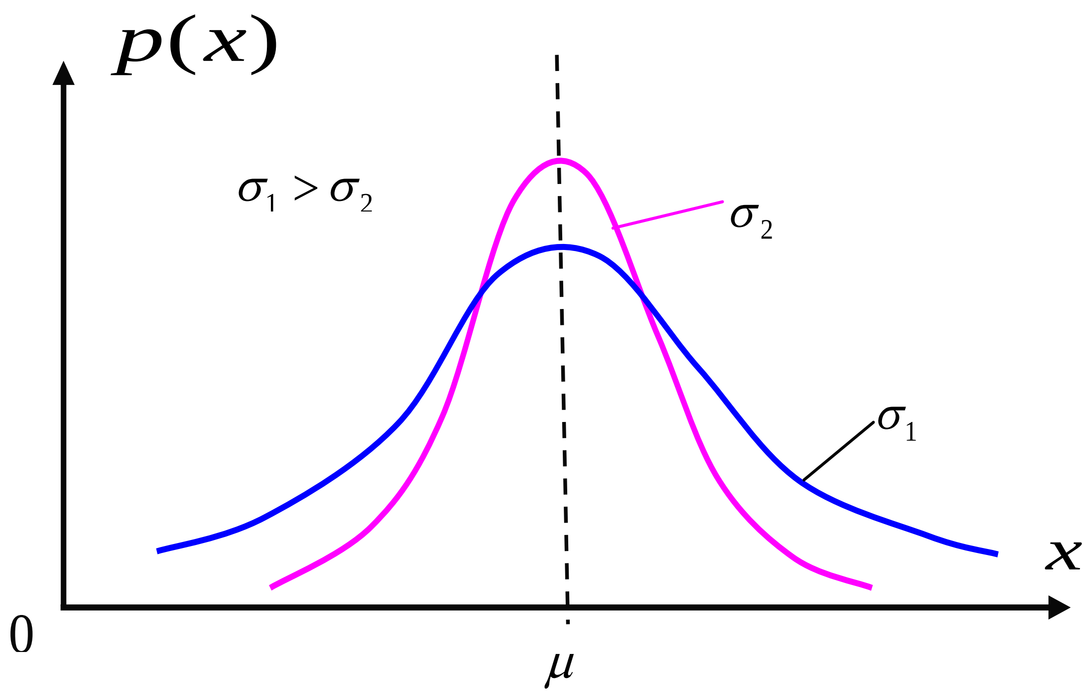
一维正态曲线的性质：

（1）曲线在 x 轴的上方，与x轴不相交。

（2）曲线关于直线 $x=\mu$对称。

（3）当 $x=\mu$时，曲线位于最高点。

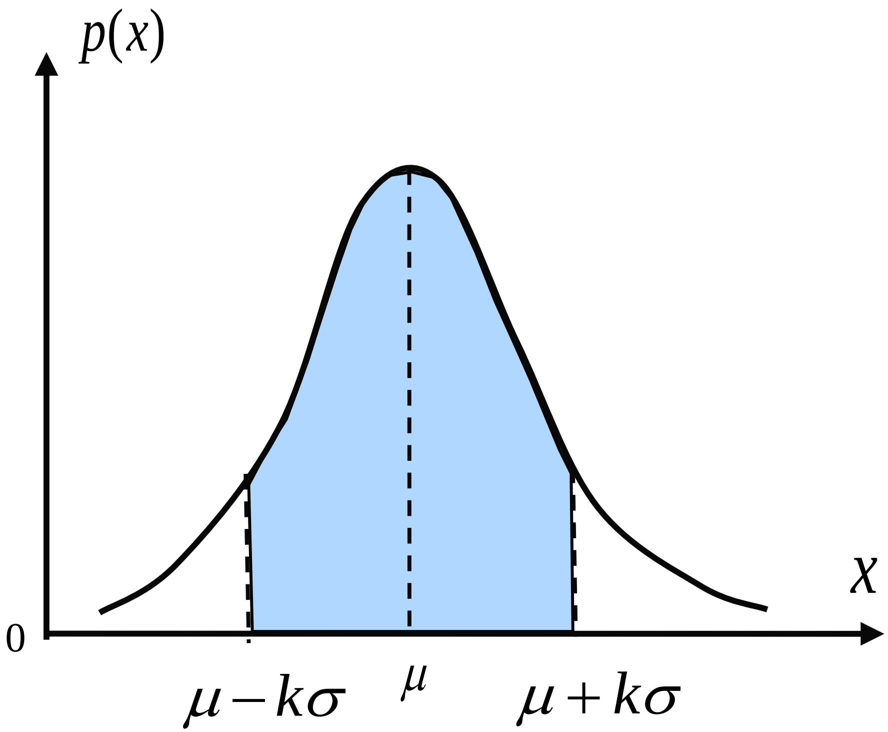
（4）当$x<\mu$时，曲线上升；当$x>\mu$时，曲线下降.并且当曲线向左、右两边无限延伸时，以x轴为渐近线，向它无限靠近。



（5）μ一定时，曲线的形状由σ确定。σ越大，曲线越“矮胖”，表示总体的分布越分散；σ越小，曲线越“瘦高”，表示总体的分布越集中。

4）3σ规则

$$P\{\mu-k\sigma\le x\le\mu+k\sigma\}=\begin{cases}0.683, & 当k=1时\\ 0.954, & 当k=2时\\ 0.997, & 当k=3时\end{cases}$$



即：绝大部分样本都落在了均值μ附近$\pm 3\sigma$的范围内，因此正态密度曲线完全可由均值和方差来确定，常简记为：

$$p(x)\sim N\left(\mu,\sigma^2\right)$$

5）多变量（n维）正态随机向量

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

密度函数定义为：

$$p(\boldsymbol{X})=\frac{1}{(2\pi)^{n/2}|\boldsymbol{C}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{M})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right\}$$

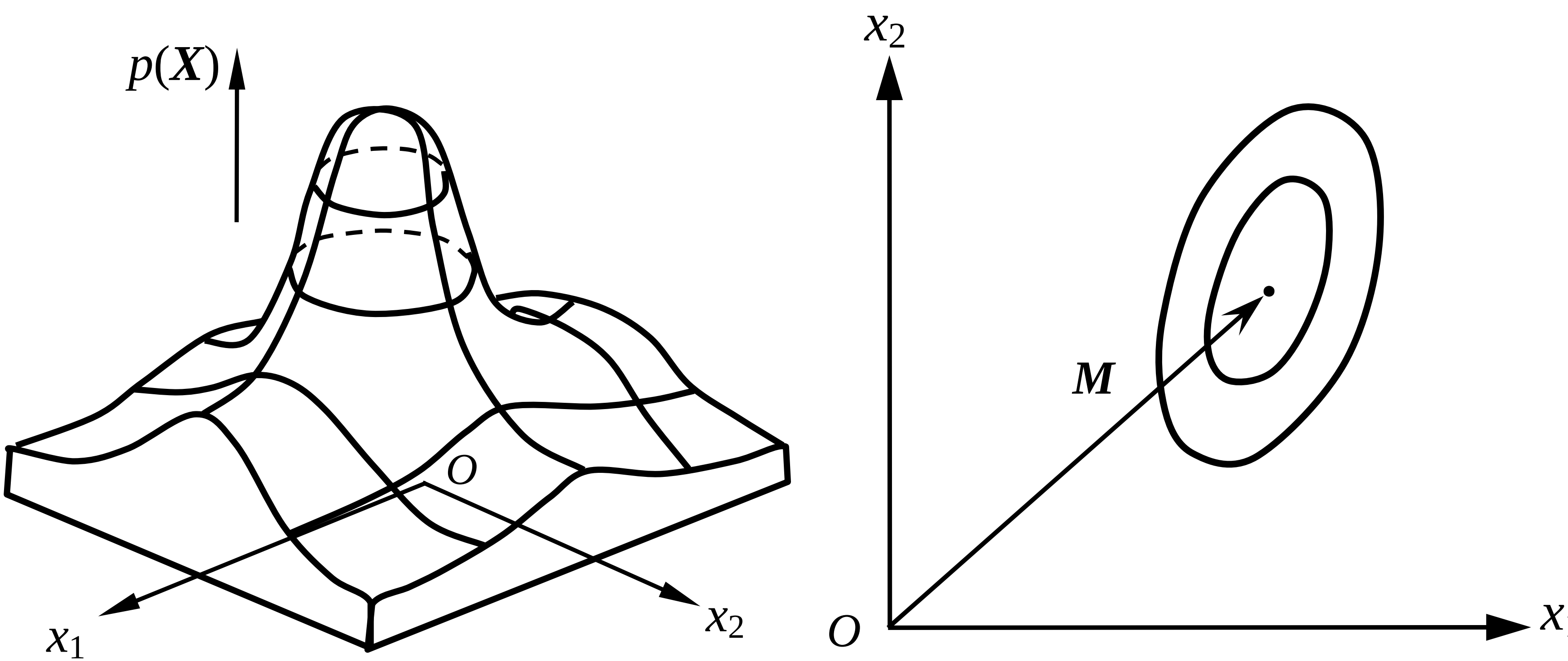
式中：$\boldsymbol{X}=[x_1,\cdots,x_n]^{\mathrm{T}}$； $\boldsymbol{M}=[m_1,\cdots,m_n]^{\mathrm{T}}$；

$$\boldsymbol{C}=\begin{bmatrix}\sigma_{11}^2 & \cdots & \sigma_{1n}^2\\ \vdots & & \vdots\\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2\end{bmatrix}$$

为协方差矩阵，是对称正定矩阵，独立元素有$n(n+1)/2$个；

$|\boldsymbol{C}|$：协方差矩阵$\boldsymbol{C}$的行列式。

多维正态密度函数完全由它的均值$\boldsymbol{M}$和协方差矩阵$\boldsymbol{C}$所确定，简记为：$p(\boldsymbol{X})\sim N(\boldsymbol{M},\boldsymbol{C})$



（a）　　　　（b）

以二维正态密度函数为例：

　　等高线（等密度线）投影到x_1ox_2面上为椭圆，从原点O到点M 的向量为均值$\boldsymbol{M}$。

　　椭圆的位置：由均值向量$\boldsymbol{M}$决定；

　　椭圆的形状：由协方差矩阵$\boldsymbol{C}$决定。

2. 正态分布的最小错误率贝叶斯决策规则

前面介绍的Bayes方法事先必须求出$p(\boldsymbol{X}|\omega_i)$, $P(\omega_i)$ 。而当$p(\boldsymbol{X}|\omega_i)$呈正态分布时，只需要知道 $\boldsymbol{M}$ 和 $\boldsymbol{C}$ 即可。

1）多类情况

具有M 种模式类别的多变量正态密度函数为：

$$p(\boldsymbol{X}\mid\omega_i)=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_i\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\}$$

$$i=1,2,\cdots,M$$

每一类模式的分布密度都完全被其均值向量$\boldsymbol{M}_i$和协方差矩阵$\boldsymbol{C}_i$所规定，其定义为：

$$\boldsymbol{M}_i=E_i[\boldsymbol{X}]$$

$$\boldsymbol{C}_i=E_i[(\boldsymbol{X}-\boldsymbol{M}_i)(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}]$$

协方差矩阵$\boldsymbol{C}_i$：反映样本分布区域的形状；

均值向量$\boldsymbol{M}_i$：表明了区域中心的位置。

$$p(\boldsymbol{X}\mid\omega_i)=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_i\right|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\}$$

最小错误率Bayes决策中，ω_i类的判别函数为 $p(\boldsymbol{X}\mid\omega_i)P(\omega_i)$，对正态密度函数，为了方便计算，取对数：

$$\ln[p(\boldsymbol{X}\mid\omega_i)P(\omega_i)]=\ln\left[p(\boldsymbol{X}\mid\omega_i)\right]+\ln\left[P(\omega_i)\right]$$

$$=\ln P(\omega_i)-\frac{n}{2}\ln 2\pi-\frac{1}{2}\ln\left|\boldsymbol{C}_i\right|-\frac{1}{2}\left\{(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\}$$

对数是单调递增函数，取对数后仍有相对应的分类性能。

去掉与i无关的项，得判别函数：

$$d_i(\boldsymbol{X})=\ln P(\omega_i)-\frac{1}{2}\ln\left|\boldsymbol{C}_i\right|-\frac{1}{2}\left\{(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\} \quad (4\text{-}25)$$

$$i=1,2,\cdots,M$$

—— 正态分布的最小错误率Bayes决策的判别函数。

$$d_i(\boldsymbol{X}) = \ln P(\omega_i) - \frac{1}{2}\ln\left|\boldsymbol{C}_i\right| - \frac{1}{2}\left\{(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\}$$

判决规则同前：

若 $d_j(\boldsymbol{X}) > d_i(\boldsymbol{X}), \quad i = 1,2,\cdots,M,\ i \neq j$ 则 $\boldsymbol{X} \in \omega_j$

$d_i(\boldsymbol{X})$为超二次曲面。可见对正态分布模式的Bayes分类器，两类模式之间用一个二次判别界面分开，就可以求得最优的分类效果。

2）两类问题

(1) 当$\boldsymbol{C}_1 \neq \boldsymbol{C}_2$时： $p(\boldsymbol{X} \mid \omega_1) \sim N(M_1, C_1)$ $p(\boldsymbol{X} \mid \omega_2) \sim N(M_2, C_2)$

对应判别函数

$$\begin{cases} d_1(\boldsymbol{X}) = \ln P(\omega_1) - \dfrac{1}{2}\ln|\boldsymbol{C}_1| - \dfrac{1}{2}\left\{(\boldsymbol{X}-\boldsymbol{M}_1)^{\mathrm{T}}\boldsymbol{C}_1^{-1}(\boldsymbol{X}-\boldsymbol{M}_1)\right\} \\ d_2(\boldsymbol{X}) = \ln P(\omega_2) - \dfrac{1}{2}\ln|\boldsymbol{C}_2| - \dfrac{1}{2}\left\{(\boldsymbol{X}-\boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}_2^{-1}(\boldsymbol{X}-\boldsymbol{M}_2)\right\} \end{cases}$$

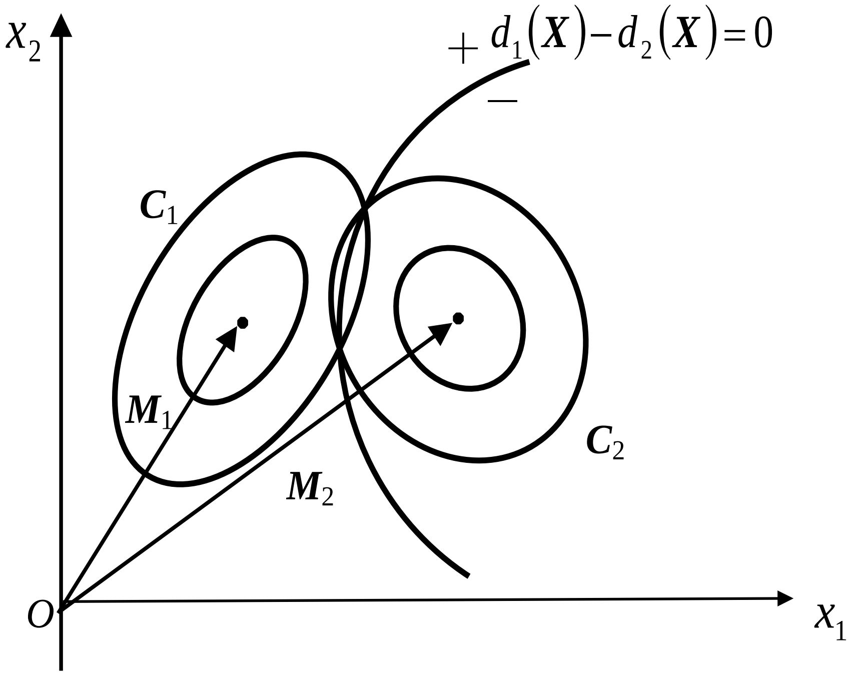


图4.3 $\boldsymbol{C}_1 \neq \boldsymbol{C}_2$时

决策规则：

若$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) \begin{cases} > 0, & \text{则 } \boldsymbol{X} \in \omega_1 \\ < 0, & \text{则 } \boldsymbol{X} \in \omega_2 \end{cases}$

判别界面$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) = 0$是$\boldsymbol{X}$的二次型方程决定的超曲面。二维判别界面如图4.3所示。

$$d_i(\boldsymbol{X}) = \ln P(\omega_i) - \frac{1}{2}\ln\left|\boldsymbol{C}_i\right| - \frac{1}{2}\left\{(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\} \quad (4-25)$$

(2) 当$\boldsymbol{C}_1$=$\boldsymbol{C}_2$=$\boldsymbol{C}$时：由式(4-25) 有

$$d_i(\boldsymbol{X}) = \ln P(\omega_i) - \frac{1}{2}\ln\left|\boldsymbol{C}\right| - \frac{1}{2}\left\{(\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{M}_i^{\mathrm{T}})\left(\boldsymbol{C}^{-1}\boldsymbol{X}-\boldsymbol{C}^{-1}\boldsymbol{M}_i\right)\right\}$$

$$= \ln P(\omega_i) - \frac{1}{2}\ln\left|\boldsymbol{C}\right| - \frac{1}{2}\left\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X} - \underline{\underline{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_i}} - \underline{\underline{\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X}}} + \boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_i\right\}$$

展开相同，合并

$$= \ln P(\omega_i) - \frac{1}{2}\ln\left|\boldsymbol{C}\right| - \frac{1}{2}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X} + \underline{\underline{\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X}}} - \frac{1}{2}\boldsymbol{M}_i^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_i \qquad i=1,2$$

由此导出判别界面为：

两类相同，抵消

$$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) \tag{4-28}$$

$$= \ln P(\omega_1) - \ln P(\omega_2) + \left(\boldsymbol{M}_1-\boldsymbol{M}_2\right)^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X} - \frac{1}{2}\boldsymbol{M}_1^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_1 + \frac{1}{2}\boldsymbol{M}_2^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_2 = 0$$

为$\boldsymbol{X}$的线性函数，是一超平面。当为二维时，判别界面为一直线，如图4.4所示。

$$d_1(\boldsymbol{X})-d_2(\boldsymbol{X})$$
$$=\ln P(\omega_1)-\ln P(\omega_2)+\left(\boldsymbol{M}_1-\boldsymbol{M}_2\right)^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X}-\frac{1}{2}\boldsymbol{M}_1^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_1+\frac{1}{2}\boldsymbol{M}_2^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_2=0$$

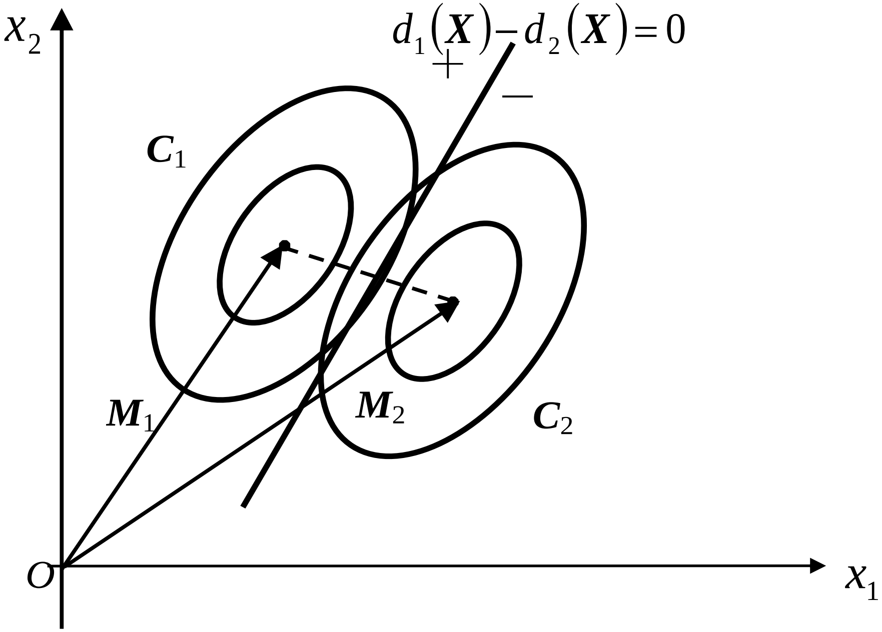


图4.4 $\boldsymbol{C}_1=\boldsymbol{C}_2=\boldsymbol{C}$，$P(\omega_1)>P(\omega_2)$

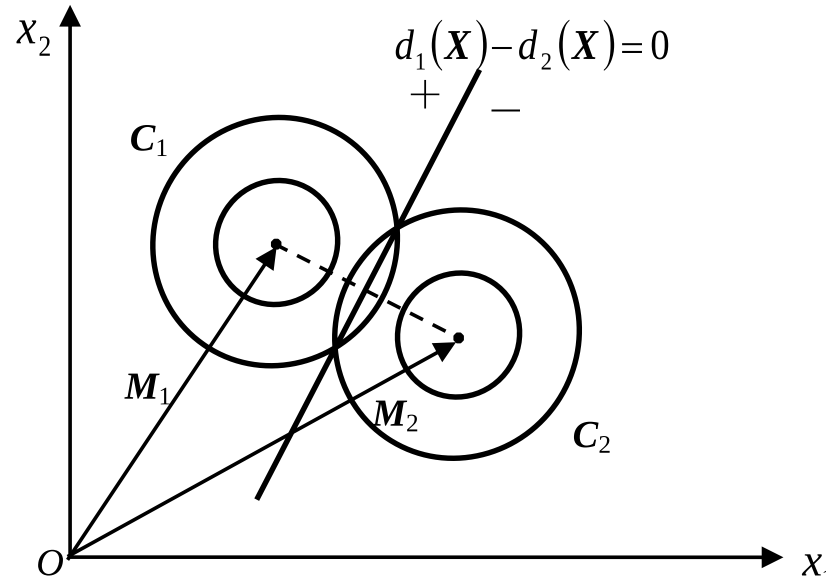


图4.5 $\boldsymbol{C}_1=\boldsymbol{C}_2=\boldsymbol{I}$且先验概率相等

(3) 当$\boldsymbol{C}_1=\boldsymbol{C}_2=\boldsymbol{I}$且$P(\omega_1)=P(\omega_2)=\dfrac{1}{2}$时：

$$d_1(\boldsymbol{X})-d_2(\boldsymbol{X})=\left(\boldsymbol{M}_1-\boldsymbol{M}_2\right)^{\mathrm{T}}\boldsymbol{X}-\frac{1}{2}\left(\boldsymbol{M}_1^{\mathrm{T}}\boldsymbol{M}_1-\boldsymbol{M}_2^{\mathrm{T}}\boldsymbol{M}_2\right)$$

判别界面如图4.5所示。

例4.3　设在三维特征空间里，有两类正态分布模式，每类各有4个样本，分别为

$$\omega_1:\quad [1,0,1]^{\mathrm{T}}\quad [1,0,0]^{\mathrm{T}}\quad [0,0,0]^{\mathrm{T}}\quad [1,1,0]^{\mathrm{T}}$$

$$\omega_2:\quad [0,0,1]^{\mathrm{T}}\quad [0,1,1]^{\mathrm{T}}\quad [1,1,1]^{\mathrm{T}}\quad [0,1,0]^{\mathrm{T}}$$

其均值向量和协方差矩阵可用下式估计：

$$\boldsymbol{M}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}\boldsymbol{X}_{ij} \tag{4-30}$$

$$\boldsymbol{C}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^{\mathrm{T}}-\boldsymbol{M}_i\boldsymbol{M}_i^{\mathrm{T}} \tag{4-31}$$

式中，N_i为类别ω_i中模式的数目，$\boldsymbol{X}_{ij}$代表在第i类中的第j个模式。两类的先验概率$P(\omega_1)=P(\omega_2)=1/2$。试确定两类之间的判别界面。

解：$\boldsymbol{M}_1 = \frac{1}{4}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}+\begin{bmatrix}1\\0\\0\end{bmatrix}+\begin{bmatrix}0\\0\\0\end{bmatrix}+\begin{bmatrix}1\\1\\0\end{bmatrix}\right\} = \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix} = \frac{1}{4}\begin{bmatrix}3 & 1 & 1\end{bmatrix}^{\mathrm{T}}$

$$\boldsymbol{M}_2 = \frac{1}{4}\begin{bmatrix}1 & 3 & 3\end{bmatrix}^{\mathrm{T}}$$

$$\boldsymbol{C}_1 = \boldsymbol{C}_2 = \frac{1}{16}\begin{bmatrix}3 & 1 & 1\\1 & 3 & -1\\1 & -1 & 3\end{bmatrix}$$ 经计算有 $\boldsymbol{C}^{-1} = \begin{bmatrix}8 & -4 & -4\\-4 & 8 & 4\\-4 & 4 & 8\end{bmatrix}$

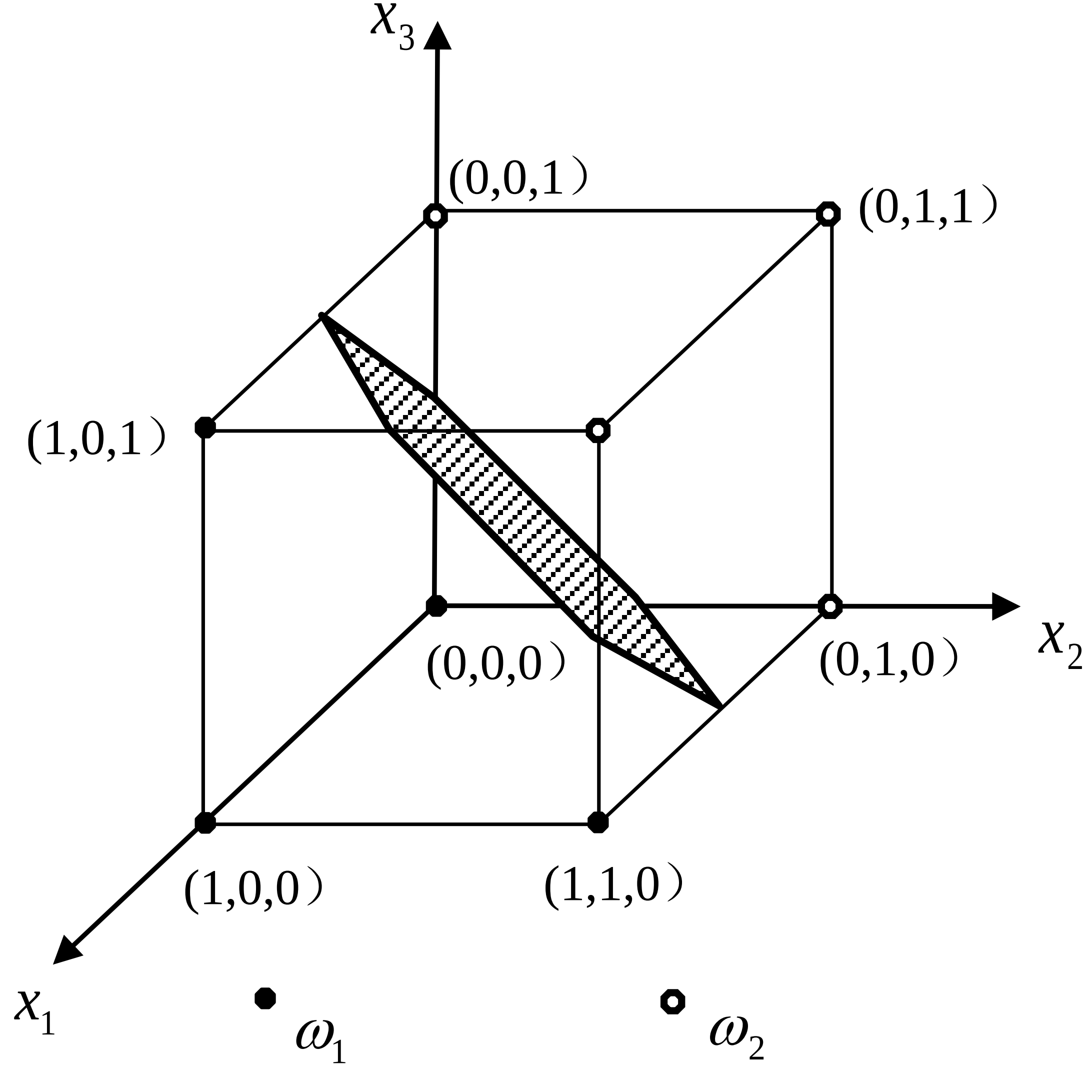
因协方差矩阵相等，故(4-28)为其判别式。由于$P(\omega_1) = P(\omega_2) = \frac{1}{2}$

$$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) = (\boldsymbol{M}_1 - \boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X} - \frac{1}{2}\boldsymbol{M}_1^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_1 + \frac{1}{2}\boldsymbol{M}_2^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_2$$

将 $\boldsymbol{X} = [x_1, x_2, x_3]^{\mathrm{T}}$ 代入：$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) = 8x_1 - 8x_2 - 8x_3 + 4 = 0$

$$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) \qquad (4\text{-}28)$$
$$= \ln P(\omega_1) - \ln P(\omega_2) + (\boldsymbol{M}_1 - \boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{X} - \frac{1}{2}\boldsymbol{M}_1^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_1 + \frac{1}{2}\boldsymbol{M}_2^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{M}_2$$

图中画出判别平面的一部分。



4.3 贝叶斯分类器的错误率

4.3.1 错误率的概念

错误率：将应属于某一类的模式错分到其他类中的概率。

是衡量分类器性能优劣的重要参数。

定义为 $P(e)=\int_{-\infty}^{\infty} P(e|\boldsymbol{X})p(\boldsymbol{X})d\boldsymbol{X}$

平均错误率

式中：$\boldsymbol{X}=[x_1,\cdots,x_n]^{\mathrm{T}}$；$P(e|\boldsymbol{X})$ 是$\boldsymbol{X}$的条件错误概率，与具体的判决规则有关；

$\int_{-\infty}^{\infty}(\ )d\boldsymbol{X}$表示$n$重积分，即整个$n$维模式空间上的积分。

错误率的计算或估计方法：

①按理论公式计算；②计算错误率上界；③实验估计。

4.3.2 错误率分析

1．两类问题的错误率

设R_1为ω_1类的判决区，R_2为ω_2类的判决区，分类中可能会发生两种错误：

① 来自ω_1类的模式落入R_2中：$\Pr(X \in R_2 | X \in \omega_1)$

② 来自ω_2类的模式落入R_1中：$\Pr(X \in R_1 | X \in \omega_2)$

错误率为两种错误之和：

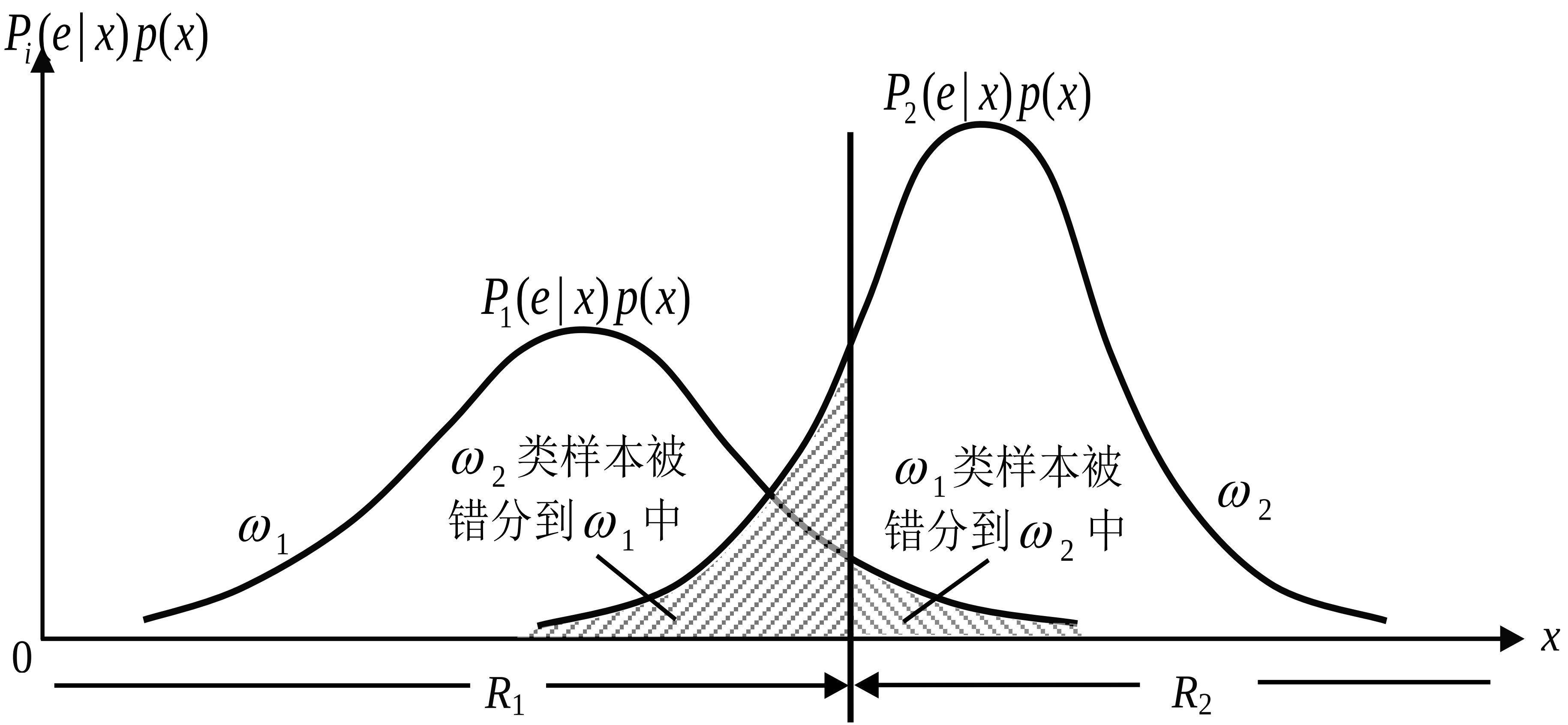
$$P(e) = \Pr(\boldsymbol{X} \in \omega_2 | \boldsymbol{X} \in \omega_1) + \Pr(\boldsymbol{X} \in \omega_1 | \boldsymbol{X} \in \omega_2)$$

$$= \int_{R_2} P(\boldsymbol{X} \in \omega_1 | \boldsymbol{X}) p(\boldsymbol{X}) d\boldsymbol{X} + \int_{R_1} P(\boldsymbol{X} \in \omega_2 | \boldsymbol{X}) p(\boldsymbol{X}) d\boldsymbol{X}$$

$$= \int_{R_2} P_1(e | \boldsymbol{X}) p(\boldsymbol{X}) d\boldsymbol{X} + \int_{R_1} P_2(e | \boldsymbol{X}) p(\boldsymbol{X}) d\boldsymbol{X} \qquad (4\text{-}33)$$

$P_1(e | \boldsymbol{X}) = P(\boldsymbol{X} | \boldsymbol{X} \in \omega_1)$ 理解为：在观察到的$\boldsymbol{X}$实际属于ω_1时，$\boldsymbol{X}$落在R_2区域（因此被判为ω_2）的概率--第1种误判概率。

$$P(e)=\int_{R_2} P_1(e\,|\,\boldsymbol{X})\,p(\boldsymbol{X})\,d\boldsymbol{X}+\int_{R_1} P_2(e\,|\,\boldsymbol{X})\,p(\boldsymbol{X})\,d\boldsymbol{X} \qquad (4\text{-}33)$$



一维情况图示

两类问题的最小错误率贝叶斯决策规则：

用后验概率密度表示为

$$\begin{cases} 若 \quad P(\omega_1 \mid \boldsymbol{X}) > P(\omega_2 \mid \boldsymbol{X}), & 则 \quad \boldsymbol{X} \in \omega_1 \\ 若 \quad P(\omega_1 \mid \boldsymbol{X}) < P(\omega_2 \mid \boldsymbol{X}), & 则 \quad \boldsymbol{X} \in \omega_2 \end{cases}$$

用先验概率和类概率密度函数表示为

$$\begin{cases} 若 \quad p(\boldsymbol{X} \mid \omega_1) P(\omega_1) > p(\boldsymbol{X} \mid \omega_2) P(\omega_2), & 则 \quad \boldsymbol{X} \in \omega_1 \\ 若 \quad p(\boldsymbol{X} \mid \omega_1) P(\omega_1) < p(\boldsymbol{X} \mid \omega_2) P(\omega_2), & 则 \quad \boldsymbol{X} \in \omega_2 \end{cases}$$

判别界面为:

$$P(\omega_1 \mid \boldsymbol{X}) = P(\omega_2 \mid \boldsymbol{X})$$

或 $$p(\boldsymbol{X} \mid \omega_1) P(\omega_1) = p(\boldsymbol{X} \mid \omega_2) P(\omega_2)$$

$$P(e)=\int_{R_2}P_1(e|\boldsymbol{X})p(\boldsymbol{X})d\boldsymbol{X}+\int_{R_1}P_2(e|\boldsymbol{X})p(\boldsymbol{X})d\boldsymbol{X} \qquad (4\text{-}33)$$

样本被划入第2类

两类问题最小错误率贝叶斯决策中错误率$P(e|\boldsymbol{X})$为：

$$P(e|\boldsymbol{X})=\begin{cases}P_1(e|\boldsymbol{X})=P(\omega_1|\boldsymbol{X}), & 若\ P(\omega_2|\boldsymbol{X})>P(\omega_1|\boldsymbol{X})\\ P_2(e|\boldsymbol{X})=P(\omega_2|\boldsymbol{X}), & 若\ P(\omega_1|\boldsymbol{X})>P(\omega_2|\boldsymbol{X})\end{cases}$$

$$\therefore\quad P(e)=\int_{R_2}P(\omega_1|\boldsymbol{X})p(\boldsymbol{X})d\boldsymbol{X}+\int_{R_1}P(\omega_2|\boldsymbol{X})p(\boldsymbol{X})d\boldsymbol{X}$$

$$P(e)=\int_{R_2}p(\boldsymbol{X}|\omega_1)P(\omega_1)d\boldsymbol{X}+\int_{R_1}p(\boldsymbol{X}|\omega_2)P(\omega_2)d\boldsymbol{X}$$

$$=P(\omega_1)\int_{R_2}p(\boldsymbol{X}|\omega_1)d\boldsymbol{X}+P(\omega_2)\int_{R_1}p(\boldsymbol{X}|\omega_2)d\boldsymbol{X}$$

令 $P_1(e)=\int_{R_2}p(\boldsymbol{X}|\omega_1)d\boldsymbol{X}$ ，$P_2(e)=\int_{R_1}p(\boldsymbol{X}|\omega_2)d\boldsymbol{X}$ ，则

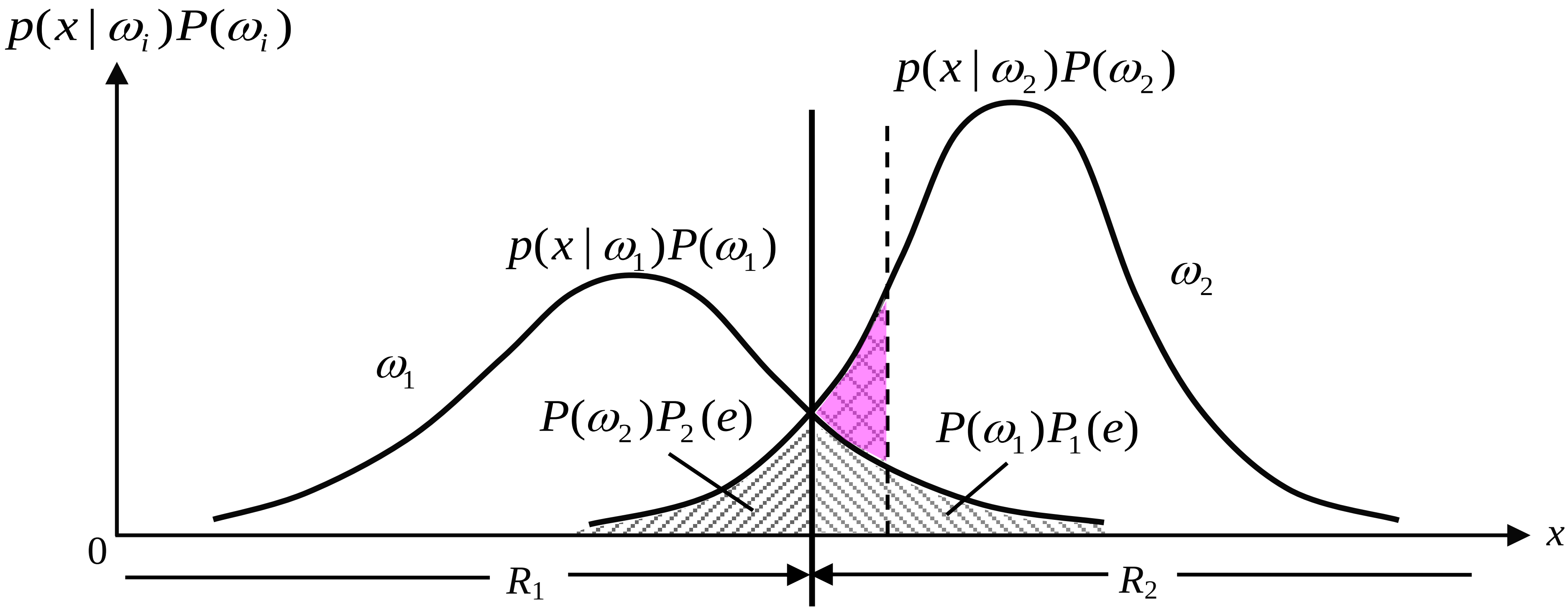
$$P(e)=P(\omega_1)P_1(e)+P(\omega_2)P_2(e)$$

$$P(\omega_i|\boldsymbol{X})=\frac{p(\boldsymbol{X}|\omega_i)P(\omega_i)}{p(\boldsymbol{X})}$$

$$P(e)=\int_{R_2} p(\boldsymbol{X}\mid\omega_1)P(\omega_1)d\boldsymbol{X}+\int_{R_1} p(\boldsymbol{X}\mid\omega_2)P(\omega_2)d\boldsymbol{X}$$

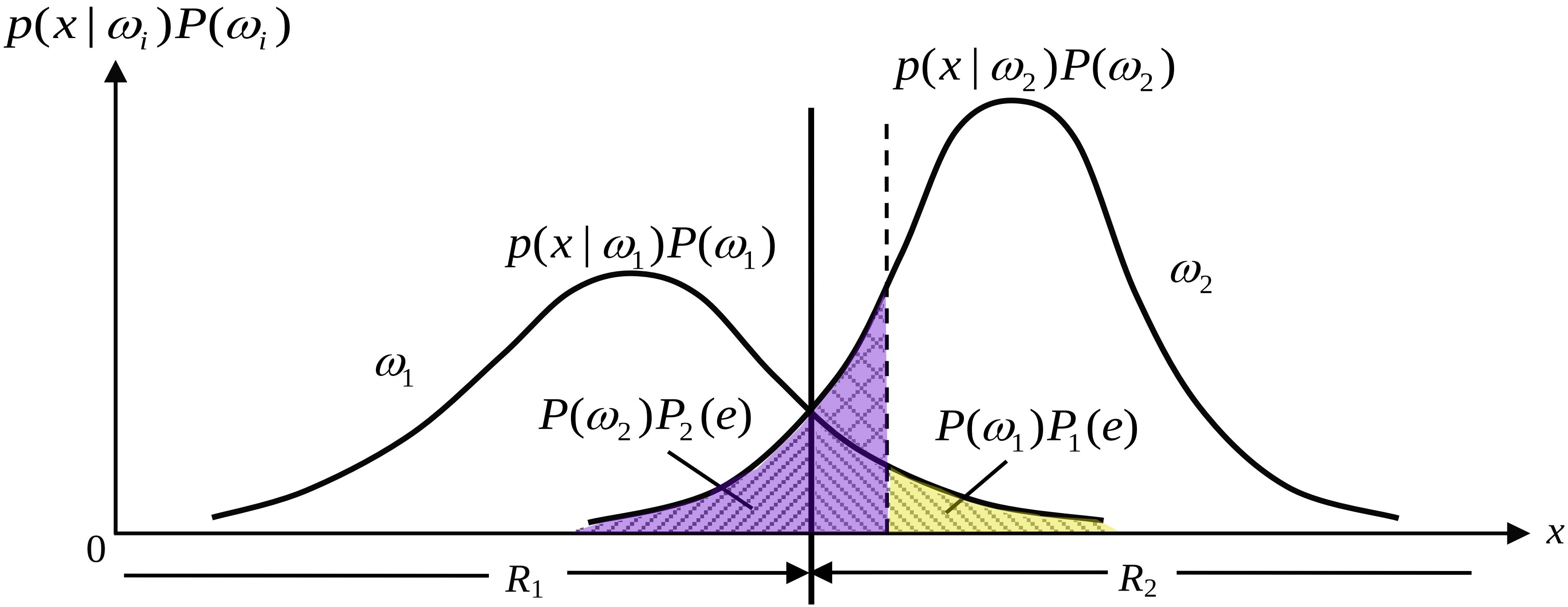
$$P(e)=P(\omega_1)P_1(e)+P(\omega_2)P_2(e)$$

在**最小错误率**贝叶斯决策中，判别界面位于两曲线的交点处，即：$p(\boldsymbol{X}\mid\omega_1)P(\omega_1)=p(\boldsymbol{X}\mid\omega_2)P(\omega_2)$



可以看出这个错误率是所有**错误率**中**最小的**（图中三角形的面积减小到0），但总错误概率不可能为零。

其他情况下的错误率：



通常需要考虑总错误概率，仅使一类样本的错误概率最小是没有意义的，因为这时另一类的错误概率可能很大。

2. 多类情况错误率

设共有M类，当判决$\boldsymbol{X} \in \omega_i$时：

$$\text{错误率}=\sum_{\substack{j=1\\j\neq i}}^{M}\int_{R_i} P(\omega_j \mid \boldsymbol{X})p(\boldsymbol{X})d\boldsymbol{X}=\sum_{\substack{j=1\\j\neq i}}^{M}\int_{R_i} p(\boldsymbol{X} \mid \omega_j)P(\omega_j)d\boldsymbol{X}$$

当$\boldsymbol{X}$判为任何一类时，都存在这样一个可能的错误，故总错误率为

$$P(e)=\sum_{i=1}^{M}\sum_{\substack{j=1\\j\neq i}}^{M}\int_{R_i} p(\boldsymbol{X} \mid \omega_j)P(\omega_j)d\boldsymbol{X}$$

正确分类概率

$$P(c)=\sum_{i=1}^{M}\int_{R_i} p(\boldsymbol{X} \mid \omega_i)P(\omega_i)d\boldsymbol{X}$$

则：$P(e)=1-P(c)$

4.3.3 正态分布贝叶斯决策的错误率计算

1．正态分布的对数似然比

设 $p(\boldsymbol{X}\mid\omega_1) \sim N(M_1, C_1)$ $\quad p(\boldsymbol{X}\mid\omega_2) \sim N(M_2, C_2)$

简化计算，假定 $C_1 = C_2 = C$。

对数似然比决策规则：

$$\text{若 } h(\boldsymbol{X}) = \ln l_{12}(\boldsymbol{X}) = \ln p(\boldsymbol{X}\mid\omega_1) - \ln p(\boldsymbol{X}\mid\omega_2) \gtrless \ln\frac{P(\omega_2)}{P(\omega_1)}$$

$$\text{则 } \boldsymbol{X} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

令 $t = \ln\dfrac{P(\omega_2)}{P(\omega_1)}$ ，有

$$\text{若 } h(\boldsymbol{X}) \gtrless t\text{，则 } \boldsymbol{X} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

由正态分布概率密度函数

$$p(\boldsymbol{X}|\omega_i)=\frac{1}{(2\pi)^{n/2}|\boldsymbol{C}_i|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{M}_i)^{\mathrm{T}}\boldsymbol{C}_i^{-1}(\boldsymbol{X}-\boldsymbol{M}_i)\right\}\quad i=1,2$$

有 $h(\boldsymbol{X})=\ln p(\boldsymbol{X}|\omega_1)-\ln p(\boldsymbol{X}|\omega_2)$

$$=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1-\boldsymbol{M}_2)-\frac{1}{2}(\boldsymbol{M}_1+\boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1-\boldsymbol{M}_2)$$

$$=-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{M}_1)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{M}_1)+\frac{1}{2}(\boldsymbol{X}-\boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{X}-\boldsymbol{M}_2)$$

—— $h(\boldsymbol{X})$是$\boldsymbol{X}$的线性函数，故$h(\boldsymbol{X})$是正态分布的一维随机变量。计算错误率较为方便。

2．对数似然比的概率分布

对 $\boldsymbol{X}\in\omega_1$， 设 $h(\boldsymbol{X})$ 的概率密度函数为 $p(h|\omega_1)$

均值：$\mu_1 = E\{h(\boldsymbol{X})\}$

$$= \boldsymbol{M}_1^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1 - \boldsymbol{M}_2) - \frac{1}{2}(\boldsymbol{M}_1 + \boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1 - \boldsymbol{M}_2)$$

$$= \frac{1}{2}(\boldsymbol{M}_1 - \boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1 - \boldsymbol{M}_2)$$

令 $r_{12}^2 = (\boldsymbol{M}_1 - \boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1 - \boldsymbol{M}_2)$ ，有

ω_1和ω_2类心间
马氏距离平方

$$\mu_1 = \frac{1}{2}r_{12}^2$$

方差：$\sigma_1^2 = E\{[h(\boldsymbol{X}) - \mu_1]^2\} = (\boldsymbol{M}_1 - \boldsymbol{M}_2)^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{M}_1 - \boldsymbol{M}_2) = r_{12}^2$

对 $\boldsymbol{X} \in \omega_2$，设 $h(\boldsymbol{X})$ 的概率密度函数为 $p(h|\omega_2)$

$$\mu_2 = E\{h(\boldsymbol{X})\} = -\frac{1}{2}r_{12}^2$$

$$\sigma_2^2 = E\{[h(\boldsymbol{X}) - \mu_2]^2\} = r_{12}^2$$

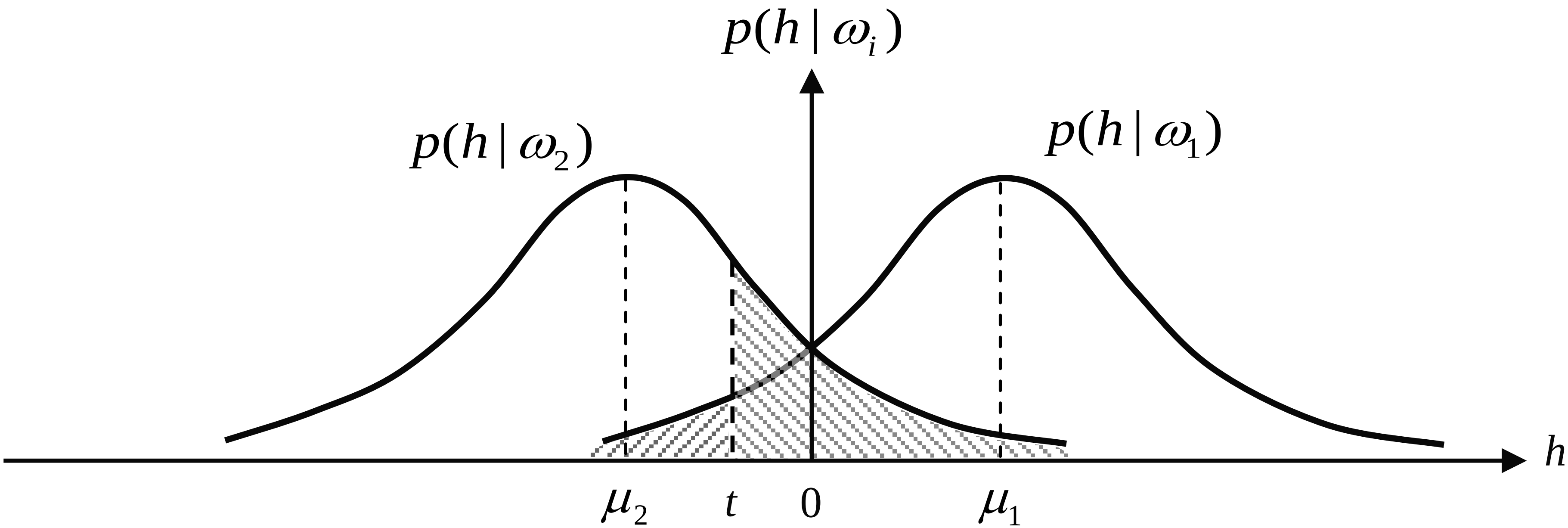


图4.9 对数似然比$h(X)$的概率分布

$$p(h|\omega_1) \sim N(\frac{1}{2}r_{12}^2, r_{12}^2)$$

$$p(h|\omega_2) \sim N(-\frac{1}{2}r_{12}^2, r_{12}^2)$$

3．正态分布最小错误率贝叶斯决策的错误率

两类问题最小错误率贝叶斯决策的错误率：

$$P(e)=P(\omega_1)P_1(e)+P(\omega_2)P_2(e)$$

其中，$P_1(e)=\int_{-\infty}^{t}p(h\,|\,\omega_1)dh$ ，$P_2(e)=\int_{t}^{\infty}p(h\,|\,\omega_2)dh$

$$P(e)=P(\omega_1)\int_{-\infty}^{t}p(h\,|\,\omega_1)dh+P(\omega_2)\int_{t}^{\infty}p(h\,|\,\omega_2)dh$$

$$=P(\omega_1)\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}r_{12}}\exp\left[-\frac{\left(h-\frac{1}{2}r_{12}^2\right)^2}{2r_{12}^2}\right]dh$$

$$+P(\omega_2)\int_{t}^{\infty}\frac{1}{\sqrt{2\pi}r_{12}}\exp\left[-\frac{\left(h+\frac{1}{2}r_{12}^2\right)^2}{2r_{12}^2}\right]dh$$

令 $\Phi(\zeta)=\int_{-\infty}^{\zeta}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy$

$$P(e)=P(\omega_1)\Phi\left(\frac{t-\frac{1}{2}r_{12}^2}{r_{12}}\right)+P(\omega_2)\left[1-\Phi\left(\frac{t+\frac{1}{2}r_{12}^2}{r_{12}}\right)\right]$$

若 $P(\omega_1)=P(\omega_2)=\frac{1}{2}$ ，则 $t=0$

$$P(e)=\frac{1}{2}\Phi\left(-\frac{1}{2}r_{12}\right)+\frac{1}{2}\left[1-\Phi\left(\frac{1}{2}r_{12}\right)\right]$$

$$=\int_{\frac{r_{12}}{2}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy$$

利用：$\Phi(\zeta)=1-\Phi(-\zeta)$

计算结果通过查标准正态分布表求得。

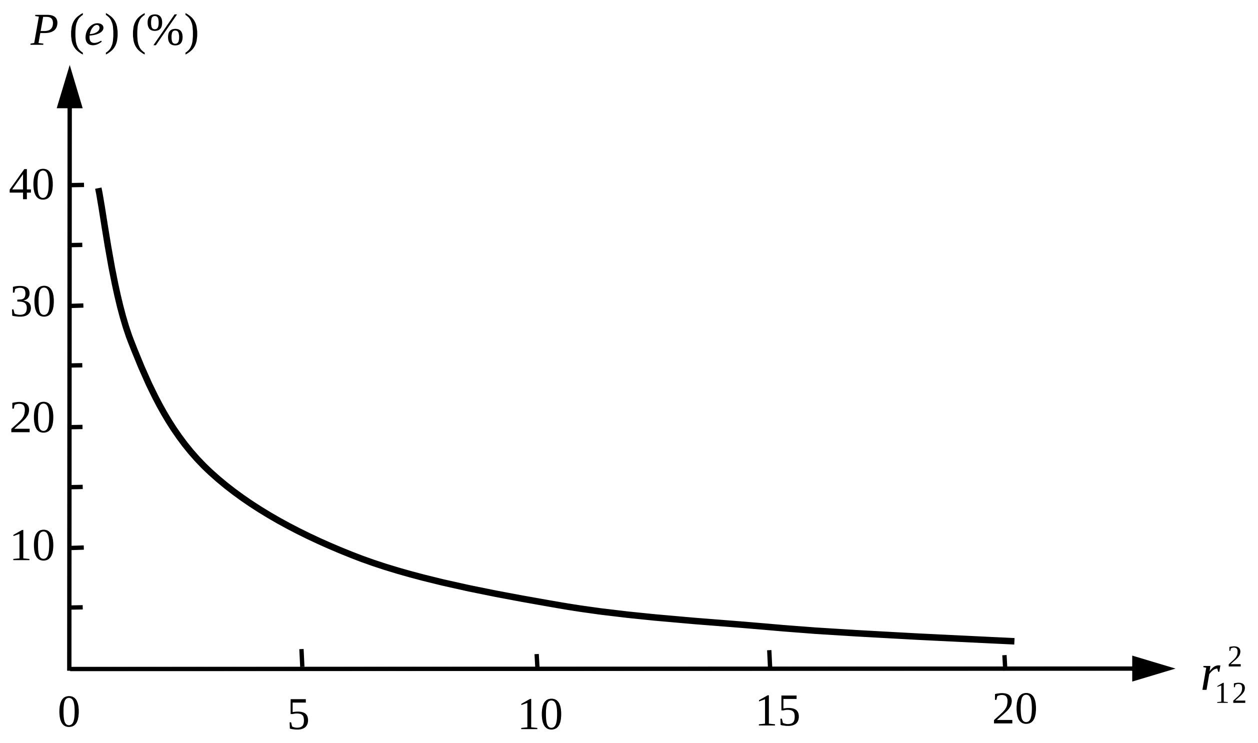


图4.10 错误率与马氏距离的关系

$P(e)$随着 r_{12}^2 的增大而单调递减，只要两类模式的马氏距离足够大，错误率就可以减到足够小。**这个结论与几何直观是一致的。**

4.3.4 错误率的估计

1. 已设计好分类器时错误率的估计

1）先验概率未知——随机抽样

$$\hat{\varepsilon} = \frac{k}{N}$$

N：随机抽取的样本数；

k：错分样本数。

2）先验概率已知——选择性抽样

分别从ω_1类和ω_2类中抽取出N_1和N_2个样本，

$$N_1 = P(\omega_1)N \qquad\qquad N_2 = P(\omega_2)N$$

用$N_1+N_2 = N$个样本对设计好的分类器作分类检验。

设ω_1类被错分的个数为k_1，ω_2类错分的个数为k_2。

k_1、k_2统计独立，联合概率为

$$P(k_1, k_2) = P(k_1)P(k_2) = \prod_{i=1}^{2} C_{N_i}^{k_i} \varepsilon_i^{k_i} (1-\varepsilon_i)^{N_i-k_i}$$

式中，ε_i是ω_i类的真实错误率。

总错误率的最大似然估计为

$$\hat{\varepsilon}' = \sum_{i=1}^{2} P(\omega_i) k_i / N_i$$

2. 未设计好分类器时错误率的估计

要求：用收集到的有限的N个样本设计分类器并估计其性能。

错误率的函数形式：$\varepsilon(\theta 1, \theta 2)$。

θ_1：用于设计分类器的样本的分布参数；

θ_2：用于检验分类器性能的样本的分布参数。

设θ是全部训练样本分布的真实参数集；

$\hat{\theta}_N$ 为全部样本中N个样本分布的参数估计量。

有
$$E\{\varepsilon(\hat{\theta}_N, \hat{\theta}_N)\} \le \varepsilon(\theta, \theta)$$

$$E\{\varepsilon(\hat{\theta}_N, \theta)\} \ge \varepsilon(\theta, \theta)$$

将有限样本划分为设计样本集和检验样本集的两种基本方法：

1）样本划分法

将样本分成两组，其中一组用来设计分类器，另一组用来检验分类器，求其错误率。取不同划分方法的平均值作为错误率的估计。

缺点：需要的样本数N很大。

2）留一法

将N个样本每次留下其中的一个，用其余的(N-1)个设计分类器，用留下的那个样本进行检验，检验完后重新放回样本集。重复进行N次。注意，每次留下的一个样本应当是不同的样本。

适用于样本数较小的情况。

缺点：计算量大。

课后作业

- 见另文。
- 下次上课前提交。
- 最好使用电子档。

End of This Part