模式识别

(Pattern Recognition)

武汉大学计算机学院

Email: 18986211797@189.cn

第3章 Bayes决策理论

3.1 Bayes决策的引入
3.2 分类器的描述方法
3.3 最小错误率Bayes决策
3.4 最小风险Bayes决策
3.5 Neyman-Pearson决策
3.6 最小最大风险决策
3.7 正态分布Bayes决策
3.8 Bayes分类器算法和例题
3.9 序惯分类（*:了解）

3.1 Bayes决策的引入

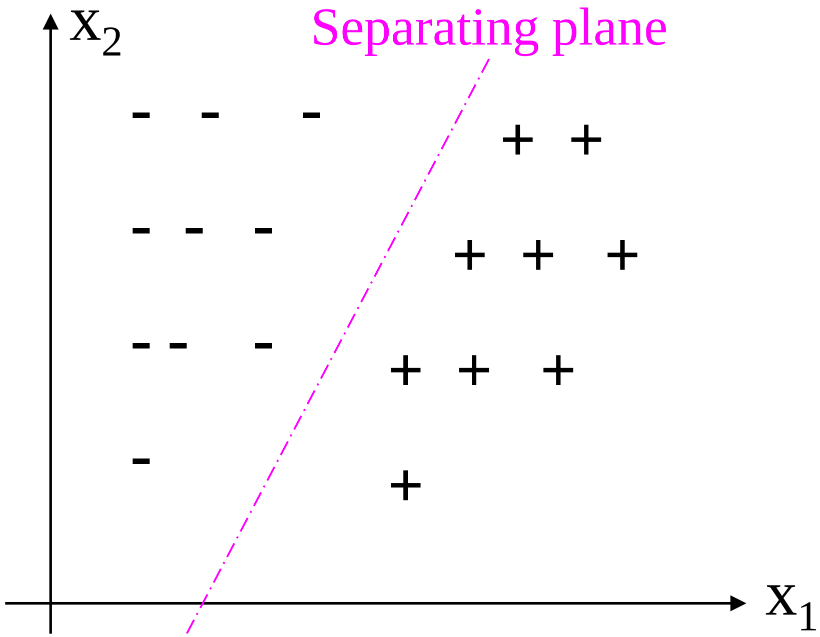
当分类器设计完成后，对待测样本进行分类，一定能分类正确吗？若有错分情况发生，是在哪种情况下出现的？错分类的可能性有多大？这些是模式识别所涉及的重要问题。

3.1.1 Bayes决策要解决的问题

这里以某制药厂生产的药品检验识别为例，说明Bayes决策要解决的问题。

+:正常药品

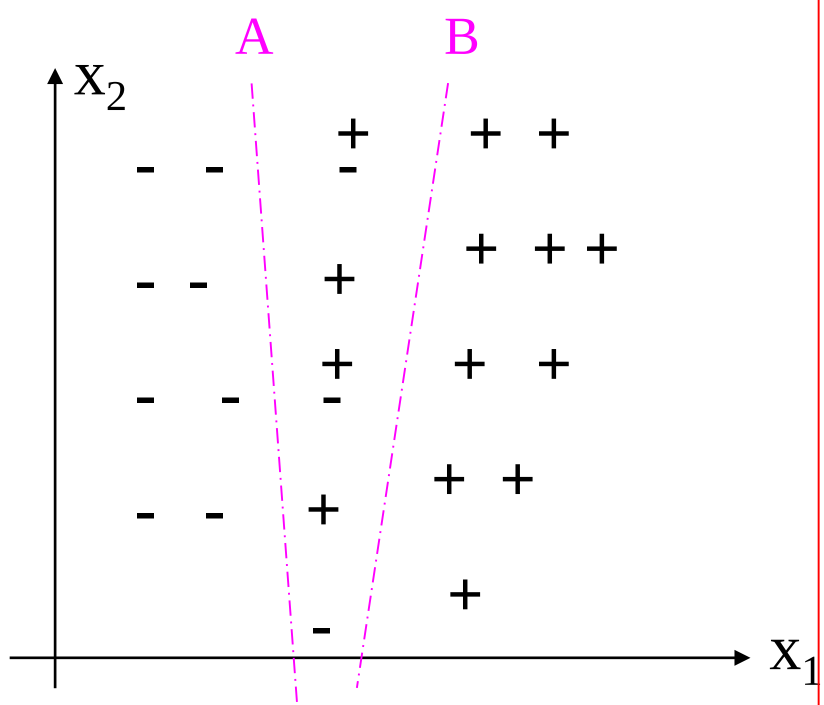
-: 异常药品



线性可分示意图

对于左图的线性分类器，可用一直线作为决策面。若待识别的药品模式向量X被划分到直线的右侧，则其为正常药品，若被划分到直线的左侧，则其为异常药品，可见对其作出决策是很容易的，也不会出现什么差错。

问题：可能出现模棱两可的情况，见下图。此时，任何决策都存在判错的可能性。



线性不可分示意图

在直线**A**、**B**之间，属于不同类的样本在特征空间中相互穿插，很难用简单的分界线将它们分开。

如果以错分类最小为原则分类，则图中**A**可能是最佳分界线，它使错分类的样本数量最小。但是，若将一个“-”样本错分成“+”类，所造成的损失要比将“+”分成“-”类严重，这是由于将异常药品误判正常药品，则会让病人失去治疗时机而遭受大的损失；把正常药品误判为异常药品会给企业带来一点损失，则偏向使对“-”类样本的错分类进一步减少，可能使总的损失为最小，那么**B**直线就可能比**A**直线更适合作为分界线。

可见，分类器参数的选择或者学习过程得到的结果取决于设计者选择什么么样的准则函数。不同的准则函数的最优解对应于不同的学习结果，得到性能不同的分类器。

错误分类难以避免，这种可能性可用$P(\omega_i|X)$，如何做出合理的判决就是Bayes决策所要讨论的问题。其中，最具代表性的是基于最小错误的bayes决策与基于最小风险的Bayes决策。

(1)基于最小错误的Bayes决策

它指出机器自动识别出现错分类的条件，错分类的可能性如何计算，如何实现使错分类出现的可能性最小。

(2)基于最小风险的Bayes决策

错误分类有不同情况，从第二个图可看出，两种错误造成的损失不一样，不同的错误分类造成的损失会不相同，前一种错误更可怕，因此就要考虑减少因错误分类造成的危害损失。为此，引入一种“风险”与“损失”的概念，希望做到使风险最小，减少危害大的错分类情况。

3.1.2 Bayes公式

Thomas Bayes 贝叶斯 (1702-1763)，英国数学家，做过神父，1742年成为英国皇家学会会员。 Bayes在数学方面主要研究概率论。他首先将归纳推理法用于概率论基础理论，并创立了Bayes统计理论，对于统计决策函数、统计推断、统计的估算等做出了贡献。

3.1.2 Bayes公式

若已知总体共有M类物体，以及各类在这n维特征空间的统计分布，具体说就是已知各类别ω_i，i=1,2,……,M的先验概率$P(\omega_i)$及类条件概率密度函数$P(X|\omega_i)$。对于待测样本X，Bayes公式可以计算该样本分属各类别的概率$P(\omega_i|X)$，叫做后验概率。看X属于哪个类的可能性最大，就把X归于可能性最大的那个类，后验概率作为识别对象归属的依据。Bayes公式为：

$$\boxed{P(\omega_i \mid X) = P(X \mid \omega_i)P(\omega_i) \Big/ \sum_{j=1}^{M} P(X \mid \omega_j)P(\omega_j) = P(X \mid \omega_i)P(\omega_i) / P(X)}$$

类别的状态是一个随机变量，而某种状态出现的概率是可以估计的。Bayes公式体现了先验概率、类条件概率密度函数和后验概率三者关系的式子。

3.1.2 Bayes公式

Bayes公式可用较直观的非正式的英文表示为：

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

也称$P(X|\omega_i)$为ω_i关于X的似然函数，或简称为“似然”(likelihood)，表明在其他条件都相等的条件下，使得$P(X|\omega_i)$较大ω_i的更有可能是真实的类别。注意到后验(posterior)概率主要是由似然函数和先验概率的乘积所决定的，证据(evidence)因子P(X)可以看成是一个标量因子(scalar factor)，以保证各类别的后验概率总和为1从而满足概率条件。

1.先验概率$P(\omega_i)$-Piror probability

先验概率$P(\omega_i)$针对M个事件出现的可能性而言，不考虑其他任何条件。如，由统计资料表明总药品数为N，其中正常药品数为N1，异常药品数为N2，则：

$$P(\omega_1)=N_1/N$$

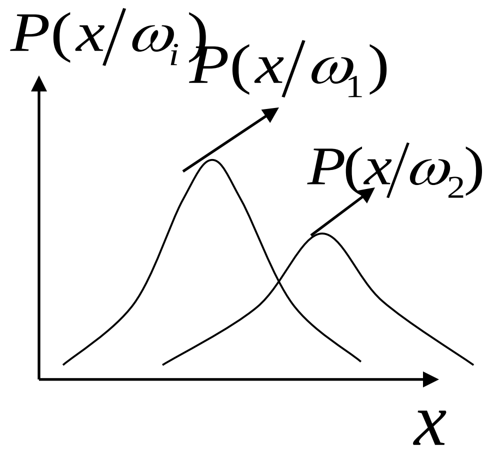
$$P(\omega_2)=N_2/N$$

称$P(\omega 1)$及$P(\omega 2)$为先验概率。显然，在一般情况下，正常药品占药品比例大，即$P(\omega 1)>P(\omega 2)$。仅按先验概率来决策，就会把所有药品都划归为正常药品，并没有达到将正常药品和异常药品区别分开的目的。这表明由先验概率所提供的信息太少。

2. 类条件概率密度函数$P(X|\omega i)$

类条件概率密度函数(Class-conditional probability density function)$P(X|\omega_i)$是指在ωi类条件下X的概率密度，即ωi类模式X的概率分布密度，简称为类概密/似然。

设只用一个特征进行分类，即n=1(特征数目)，并已知这两类的类条件概率密度函数分布，见右图。类概密$P(X|\omega_1)$是正常药品的属性（此处n=1，故为特征数值）分布，类概密$P(X|\omega_2)$是异常药品的属性分布。



类条件概率密度分布

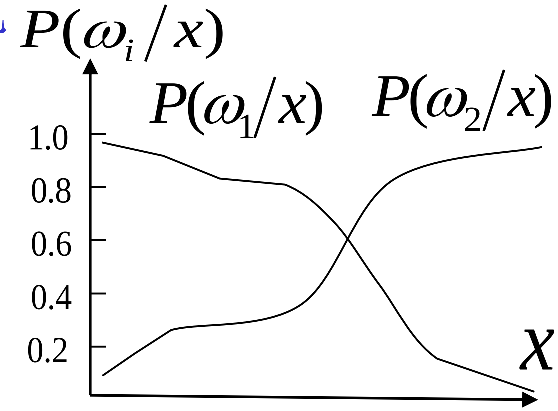
在工程问题中，统计数据往往满足正态分布规律。若采用正态密度函数作为类概密的函数形式，则函数内的参数，如期望、方差是未知的。那么问题就变成如何利用大量样本对这些参数进行估计，只要估计出这些参数，类概密$P(X|\omega i)$就确定了。

3. 后验概率$P(\omega_i|X)$-Posterior probability

条件概率P(ωi|X)表示X出现的条件下ωi类出现的概率，称其为类别ωi的后验概率，对模式识别而言可理解为X来自ωi 类的概率。即若所观察到的某一样本的特征向量为X，而在类中又有不止一类可能呈现这一X特征数值，它属于各类的概率就是P(ωi|X)。

4. $P(\omega_1|X)$、 $P(\omega_2|X)$与$P(X|\omega 1)$、 $P(X|\omega 2)$的区别

（1） $P(\omega 1|X)$和$P(\omega 2|X)$是指在同一条件X下，比较ω1与ω1出现的概率，若$P(\omega 1|X) > P(\omega 2|X)$，则有结论：在X条件下，事件ω1出现的可能性大，见右图。对于两类的情况，则有$P(\omega 1|X) + P(\omega 2|X)=1$。



后验概率分布

(2) $P(X|\omega 1)$和$P(X|\omega 2)$都是指各自条件下出现X的可能性，两者间没有联系，比较两者没有意义。$P(X|\omega 1)$和$P(X|\omega 2)$是在不同条件下讨论问题，即使只有两类ω1与ω2， $P(X|\omega 1)+P(X|\omega 2)\neq 1$。不能仅因为$P(X|\omega 1)>P(X|\omega 2)$，就认为X是ω1类的可能性就大。只有考虑先验概率这一因素，才能决定X条件下，判别为ω1类或ω2类的可能性哪个大。

3.2 分类器的描述方法

3.2.1 基本假设

给定模式空间S，由m个互不相交的模式类集合$\omega 1$，$\omega 2$，…，ωm组成：

（1）假定类ωi的先验概率为$P(\omega i)$；

（2）样本（或模式）$\boldsymbol{x}$由特征向量来表示，同样记为$\boldsymbol{x}$，假设为d维，即$\boldsymbol{x}=(x1, x2, \cdots, xd)$；

（3）特征向量$\boldsymbol{x}$的取值范围构成特征空间，记为R^d；

（4）特征向量$\boldsymbol{x}$的类条件概率密度函数为$p(\boldsymbol{x}|\omega i)$，表示当样本$\boldsymbol{x}\in\omega i$时，特征向量$\boldsymbol{x}$的概率密度函数；

（5）特征向量$\boldsymbol{x}$的后验概率为$P(\omega i|\boldsymbol{x})$，表示在特征向量$\boldsymbol{x}$出现的条件下，样本$\boldsymbol{x}$来自类$\omega i$的概率，即类$\omega i$出现的概率。

模式识别就是根据特征向量$\boldsymbol{x}$的取值，依据某个判决准则把样本$\boldsymbol{x}$划分到$\omega 1$，$\omega 2$，…，ωm中的一个。

3.2.2　模式分类器的描述

　模式分类器的描述方法有多种，这里仅介绍以下三种描述方法，它们之间是统一的。

1. 映射描述法

由于我们获取的有关观察对象的数据总量是有限的，因此，可用一个d+1维向量表示，即：

$$(x1,\ x2,\ \cdots,\ xd;\ \alpha)$$

其中：$(x_1, x_2, \ldots, x_d)$为特征向量，是特征空间$\mathbf{R}^d$中的一个点；α取值于集合$\{1, 2, \ldots, m\}$，表示模式的真实类别号，是未知的量，m为类别数。模式分类的实质在于实现特征空间$\mathbf{R}^d$到类别号空间$\{1, 2, \ldots, m\}$的一个映射，即

$$\mathbf{R}^d \to \{1, 2, \ldots, m\}$$

给定一个映射f，就给出了一种模式识别方法，不同的映射对应不同的分类方法，这就是模式识别问题的映射描述法。

2. 划分描述法

由于每个特征向量是R^d空间的一个点，且$\mathbf{R}^d \to \{1, 2, \ldots, m\}$是一个多对一的映射，通过映射，本质上实现了对空间$R^d$的一种划分，即把$R^d$划分成个不相重叠的区域，每一个区域对应一个类别。区域Ri对应第i类ωi。

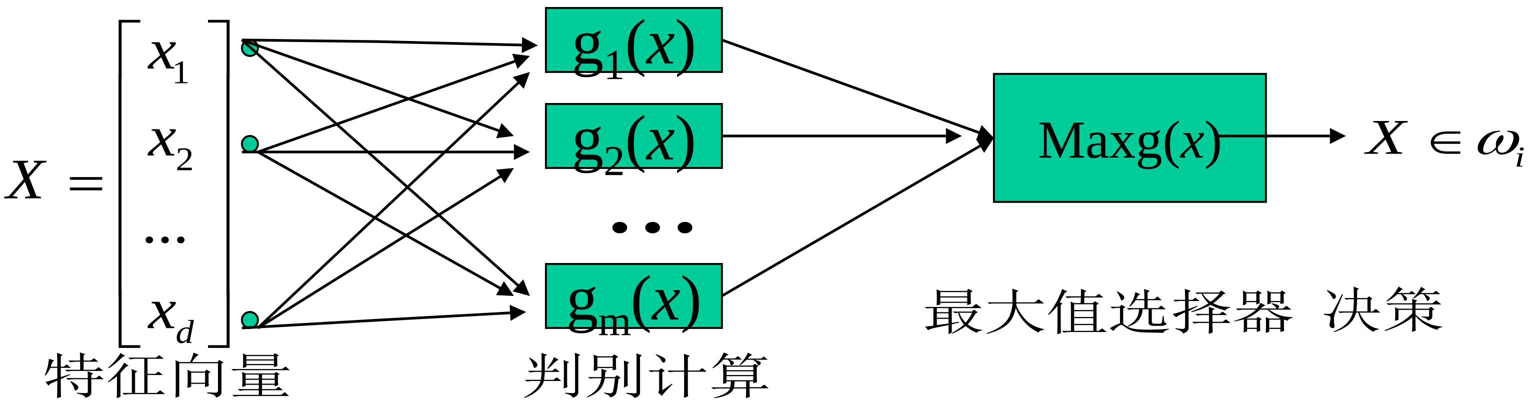
3. 判别函数法

把分类问题对应为R^d空间上的多元函数，通常称为判别函数(或称判决函数)$g_i(\boldsymbol{x})$，　　i=1，2，…，m。

将样本x判属有极大(或极小)函数值的那一类。

模式分类实际上是将特征空间划分为不同的决策区域，相邻决策区域被决策面所分割，这些决策面是特征空间中的超曲面，其决策面方程满足相邻两个决策域的判别函数相等，即 $g_i(\boldsymbol{x})=g_j(\boldsymbol{x})$

分类器可被看作是一个计算m类个判别函数并选取最大(或最小)判决值对应的类别的网络或者机器，见下图。



3.3 最小错误率Bayes决策
(Minimum-error-rate classification)

假设待识别的特征为X,样本分为m类，各类的先验概率和各类的类概密均已知，就有m个判别函数，由Bayes公式可知：

$$P(\omega_i \mid x) = P(x \mid \omega_i)P(\omega_i) \Big/ \sum_{j=1}^{m} P(x \mid \omega_j)P(\omega_j) = P(x \mid \omega_i)P(\omega_i) / P(x)$$

在取得一个观察特征X后，在特征X的条件下，看哪个类的概率最大，应该把X归于概率最大的那个类。由此，可得到最大后验概率判决准则的几种等价形式：

(1) 若 $p(\mathbf{x} \mid \omega_j)P(\omega_j) = \max\limits_{i \in \{1,2,\cdots,m\}} p(\mathbf{x} \mid \omega_i)P(\omega_i)$, 则$x \in \omega_j$

(2) 若 $L(\mathbf{x}) = \dfrac{p(\mathbf{x} \mid \omega_j)}{p(\mathbf{x} \mid \omega_i)} > \dfrac{P(\omega_i)}{P(\omega_j)}, i = 1, 2, \cdots, m, i \neq j$, 则$x \in \omega_j$

(3) 若 $\ln L(\mathbf{x}) = \ln p(\mathbf{x} \mid \omega_j) - \ln p(\mathbf{x} \mid \omega_i) > \ln \dfrac{P(\omega_i)}{P(\omega_j)}, i = 1, 2, \cdots, m, i \neq j$

则$x \in \omega j$;

其中， $L(\boldsymbol{x})$ 称为似然比， $\ln L(\boldsymbol{x})$ 称为对数似然比

$P(\omega 1)/P(\omega 2)$ 称为似然比阈值

【例 3.1】 假设在某个局部地区的细胞识别中，第一类表示正常，第二类表示异常，两类的先验概率分别为：正常 $P(\omega 1)=0.9$， $P(\omega 2)=0.1$。现有一个待识别样本细胞，其观察值为 $\boldsymbol{x}$，从类条件概率密度函数曲线 $p(\boldsymbol{x}|\omega i)$ 上可查得： $p(\boldsymbol{x}|\omega 1)=0.2$， $p(\boldsymbol{x}|\omega 2)=0.4$，试判断该细胞是否正常。

解：该细胞属于正常细胞还是异常细胞，先计算后验概率：

计算

$$p(\boldsymbol{x}|\omega 1)P(\omega 1)=0.2\times 0.9=0.18$$

$$p(\boldsymbol{x}|\omega 2)P(\omega 2)=0.4\times 0.1=0.04<0.18$$

根据 Bayes 判决准则将该细胞判为第一类ω1，即为正常细胞。

最大后验概率判决准则使决策的错误率最小

最大后验概率判决准则的一个优良性质就是使平均错误概率达到最小。因此，最大后验概率判决准则又称为最小错误概率判决准则。

这里以二分类情况为例进行分析。此时，$m=2$，任意一个判决准则对应于特征空间R^d的一个划分：$R=R1\cup R2$，$R1\cap R2=\Phi$。为了直观，假设$\boldsymbol{x}$只有一个特征，n=1。错误分类有两种情况：①若$\boldsymbol{x}$原属于$\omega 1$类，却落入$R2$，称为第1类错误；②若$\boldsymbol{x}$原属于$\omega 2$类，却落入$R1$，称为第2类错误。

第1类错误概率$P1(\mathrm{e})$为：

$$P1(e) = P(\mathbf{x} \in R_2 \mid \omega_1) = \int_{R_2} p(\mathbf{x} \mid \omega_1) d\mathbf{x}$$

第2类错误概率$P2(\mathrm{e})$为：

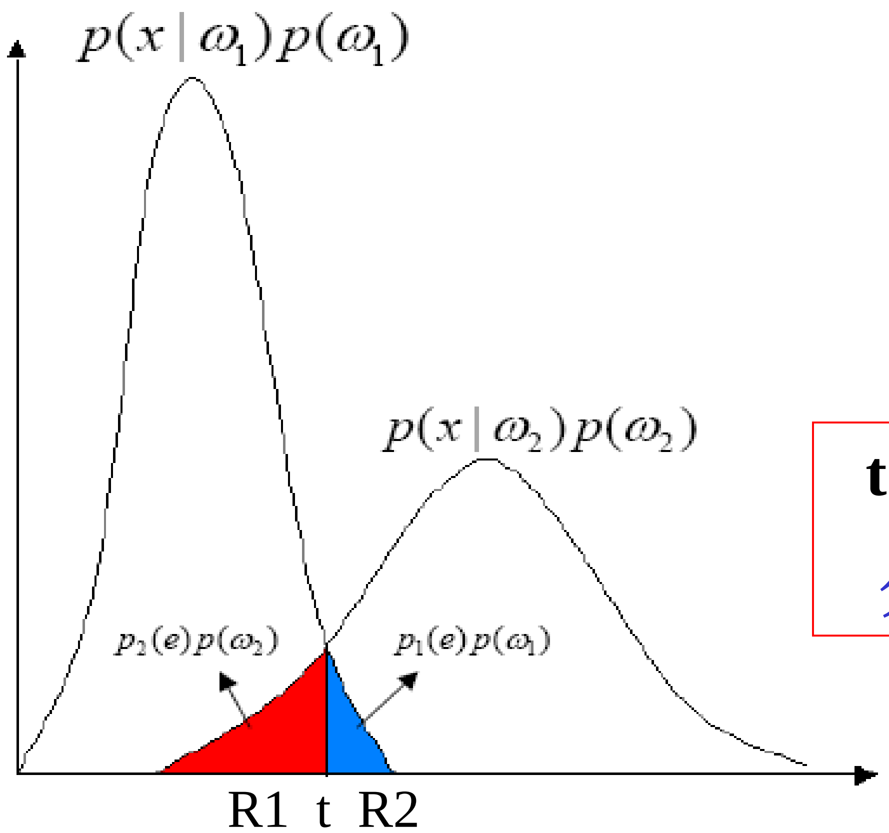
$$P2(e) = P(\mathbf{x} \in R_1 \mid \omega_2) = \int_{R_1} p(\mathbf{x} \mid \omega_2) d\mathbf{x}$$

因此，平均错误概率P (e)为：

$$P(e) = P(\mathbf{x} \in R_2 \mid \omega_1)P(\omega_1) + P(\mathbf{x} \in R_1 \mid \omega_2)P(\omega_2)$$

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

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



平均错误概率计算示意图

t为两类的分界点，将**x**轴分成两个区域**R1**和**R2**。

若要使Pe达到最小，则$\boldsymbol{x}\in\omega 1$的决策区域$R$1必须满足：

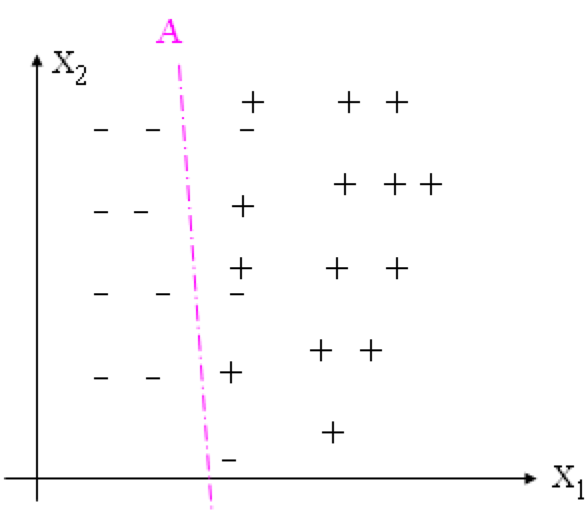
$$R_1=\{\mathbf{x}\,|\,P(\omega_2)p(\mathbf{x}\,|\,\omega_2)-P(\omega_1)p(\mathbf{x}\,|\,\omega_1)<0\}$$ 证明略

即：$$R_1=\left\{\mathbf{x}\left|\frac{p(\mathbf{x}\,|\,\omega_1)}{p(\mathbf{x}\,|\,\omega_2)}>\frac{P(\omega_2)}{P(\omega_1)}\right.\right\}$$

上式与最大后验概率判决准则中$\boldsymbol{x}\in\omega 1$的决策区域是一致的，也就是说，最大后验概率判决使平均错误概率达到最小。

3.4 最小风险Bayes决策

上节的探讨了使错误概率最小的Bayes决策规则。然而，当接触到实际问题时，可能发现使错误率最小并非是普遍适用的最佳选择。见下图。



(a)基于最小错误　　(b)基于最小风险(扩大错误率，减小损失)

图 基于最小错误分类和基于最小风险分类的比较

图中，直线B的划分把正常药品误判为异常药品，这样扩大了总错误率，会给企业带来一定的损失；直线A的划分将异常药品误判为正常药品，虽然使错误分类最小，但会使病人因失去正确的治疗而遭受极大的损失。可见使错误率最小并不一定是最佳选择。

实际应用时，从根据不同性质的错误会引起不同程度的损失考虑出发，宁可扩大一些总的错误率，但也要使总的损失减少。这时图中的直线B的划分最为实用。这会引进一个与损失有关联的概念-风险。在做决策时，要考虑所承担的风险。基于最小风险的Bayes决策规则正是为了体现这一点而产生的。

- 若要判断某颗药品是正常(ω_1)还是异常(ω_2)，在判断中可能出现如下情况：
  - 第1种，判对(正常药品→正常药品)λ_{11}；
  - 第2种，判错(正常药品→异常药品)λ_{21}；
  - 第3种，判对(异常药品→异常药品)λ_{22}；
  - 第4种，判错(异常药品→正常药品)λ_{12}。
- 在判断时，除了能做出“是” ω_i类或“不是” ω_i类的动作以外，还可以做出“拒识”的动作。

对于**两类分类问题**，最小风险Bayes决策的基本思想是：对于模式x，如果将其决策为模式类ω1的风险大于决策为模式类ω2的风险，则决策模式x属于类ω2；反之，决策模式x属于模式类ω1。

为了更好地研究最小风险Bayes分类器，下面说明几个概念。

几个概念

- 设样本**x**来自类ωi, 可能被判为$\omega 1, \omega 2, \ldots, \omega m$ 中的任何一种, 若允许拒绝判决, 可将拒绝类看成是独立的一类, 记为第m+1类, 即ωm+1。
- 行动(决策)αi：表示把模式x判决为ωi类的一次动作（决策）。

  不同的动作对应于特征空间的不同决策区域Rj, $j \in \{1, 2, \ldots, m\}$。 若$\mathbf{x} \in Rj$, 则判决$\mathbf{x} \in \omega j (j=1, 2, \ldots, m)$。这里未考虑拒识情况。
- 损失函数$\lambda_{ii}=\lambda(\alpha_i, \omega_i)$表示模式$\boldsymbol{x}$本来属于$\omega_i$类而错判为$\omega_i$所受损失。因为这是正确判决，故损失最小。
- 损失函数$\lambda_{ij}=\lambda(\alpha_i, \omega_j)$表示模式$\boldsymbol{x}$本来属于$\omega_j$类错判为$\omega_i$所受损失。因为这是错误判决，故损失最大。
- 风险R(期望损失)：对未知模式x采取一个判决行动$\alpha(x)$所付出的代价(损失)。

几个概念(Cont.)

- 条件风险(也叫条件期望损失):

$$R(\alpha_i \mid x) = E\left[\lambda(\alpha_i, \omega_j)\right] = \sum_{j=1}^{M} \lambda(\alpha_i, \omega_j) P(\omega_j \mid x), i = 1, 2, \ldots, c.(c \le M)$$

- 在整个特征空间中定义期望风险R:

$$R = \int_{R^d} R(\alpha(x) \mid x) P(x) dx$$

$$= \sum_{j}^{M} \int_{R_j} R(\alpha_j \mid x) P(x) dx, \text{(平均风险)}$$

$P(x)$为样本向量x在R^d空间中的概率密度

- 条件风险只反映对某x取值的决策行动α_i所带来的风险。期望风险R则反映在整个特征空间不同的x取值$a(x)$ {决策可看成是随机向量x的函数，记为$a(x)$}的决策行动所带来的平均风险。

几个概念(Cont.)

- 在实际应用时，可以将损失函数λij写成如下矩阵形式：

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1M} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2M} \\ & \vdots & & \\ \lambda_{c1} & \lambda_{c2} & \cdots & \lambda_{cM} \end{bmatrix} (c \le M)$$

称之为损失矩阵。

基于最小风险的Bayes决策：

决策带来的损失的平均值—风险最小。

最小风险Bayes决策规则：

若 $R(\alpha_k \mid \mathbf{x}) = \min\limits_{i=1,2,\cdots,M} R(\alpha_i \mid \mathbf{x})$, 则判决$\mathbf{x} \in \omega_k$。

损失函数根据实际问题和经验确定．若损失函数取：

$$\lambda(\alpha_{\mathrm{i}}, \omega_{\mathrm{j}}) = \begin{cases} 0 & (\mathrm{i}=\mathrm{j}) \\ 1 & (\mathrm{i} \neq \mathrm{j}) \end{cases} \quad (\mathrm{i}, \mathrm{j}=1,2,\ldots, \mathrm{M})$$

则称这种损失函数为0−1损失函数．此时决策α_{j}的条件风险为

(zero-one loss function)

$$R(\alpha_j \mid \mathbf{x}) = \sum_{i=1}^{M} \lambda(\alpha_j, \omega_i) P(\omega_i \mid x) = \sum_{i \neq j} P(\omega_i \mid x) = 1 - P(\omega_j \mid x) \quad \textbf{(2-17)}$$

从上式可以看出, $R(\alpha_j \mid \mathbf{x})$最小实际上对应于$P(\omega_j \mid x)$最大.

因此, 当取0-1损失函数时, 最小风险贝叶斯判决准则等价于最大后验概率判决准则(教材P13公式2-17)。即最大后验概率判决准则是最小风险Bayes判决准则的特例.

对于两分类问题, 条件风险为:

$$R(\alpha_1|\mathbf{x}) = \lambda(\alpha_1,\omega_1)P(\omega_1|\mathbf{x}) + \lambda(\alpha_1,\omega_2)P(\omega_2|\mathbf{x}) \quad \text{(2-18)}$$

$$R(\alpha_2|\mathbf{x}) = \lambda(\alpha_2,\omega_1)P(\omega_1|\mathbf{x}) + \lambda(\alpha_2,\omega_2)P(\omega_2|\mathbf{x}) \quad \text{(2-19)}$$

按最小风险的*Bayes' classifier*,有:

$$[\lambda(\alpha_2,\omega_1) - \lambda(\alpha_1,\omega_1)]P(\omega_1|\mathbf{x})$$

$$> [\lambda(\alpha_1,\omega_2) - \lambda(\alpha_2,\omega_2)]P(\omega_2|\mathbf{x}) \Rightarrow \mathbf{x} \in \omega_1 \quad \text{(2-20a)}$$

$$[\lambda(\alpha_2,\omega_1) - \lambda(\alpha_1,\omega_1)]P(\omega_1|\mathbf{x})$$

$$< [\lambda(\alpha_1,\omega_2) - \lambda(\alpha_2,\omega_2)]P(\omega_2|\mathbf{x}) \Rightarrow \mathbf{x} \in \omega_2 \quad \text{(2-20b)}$$

根据Bayes公式, 有:

$$[\lambda(\alpha_2,\omega_1)-\lambda(\alpha_1,\omega_1)]P(\mathbf{x}|\omega_1)P(\omega_1)$$
$$>[\lambda(\alpha_1,\omega_2)-\lambda(\alpha_2,\omega_2)]P(\mathbf{x}|\omega_2)P(\omega_2)\Rightarrow \mathbf{x}\in\omega_1 \qquad \textbf{(2-21a)}$$

$$[\lambda(\alpha_2,\omega_1)-\lambda(\alpha_1,\omega_1)]P(\mathbf{x}|\omega_1)P(\omega_1)$$
$$>[\lambda(\alpha_1,\omega_2)-\lambda(\alpha_2,\omega_2)]P(\mathbf{x}|\omega_2)P(\omega_2)\Rightarrow \mathbf{x}\in\omega_2 \qquad \textbf{(2-21b)}$$

$$L(\mathbf{x})=\frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)}\gtrless\frac{P(\omega_2)}{P(\omega_1)}\frac{\lambda(\alpha_1,\omega_2)-\lambda(\alpha_2,\omega_2)}{\lambda(\alpha_2,\omega_1)-\lambda(\alpha_1,\omega_1)}\Rightarrow \mathbf{x}\in\begin{cases}\omega_1\\ \omega_2\end{cases} \qquad \textbf{(2-22)}$$

【例 3.2】 在例3.1的基础上, 增加条件λ11=0, λ12=6, λ21=1, λ22=0, 请判断该细胞是否正常

解： 若按最小风险的Bayes判决进行判断, 先计算后验概率：

$$P(\omega_1/x)=\frac{P(x/\omega_1)P(\omega_1)}{\sum_{j=1}^{2}P(x/\omega_j)P(\omega_j)}=\frac{0.2\times0.9}{0.2\times0.9+0.4\times0.1}=0.818$$

$$P(\omega_2/x)=1-P(\omega_1/x)=0.182$$

条件风险： $R(\alpha_1\,|\,\mathbf{x})=\sum_{i=1}^{2}\lambda_{1i}P(\omega_i\,|\,\mathbf{x})=1.092$

$$R(\alpha_2\,|\,\mathbf{x})=\sum_{i=1}^{2}\lambda_{2i}P(\omega_i\,|\,\mathbf{x})=0.818$$

$\because R(\alpha_1\,|\,x)>R(\alpha_2\,|\,x)\ \therefore x\in$ 异常细胞（第2类），因此决策ω_1类风险大。

因λ_{12}=6较大，决策损失起决定作用。

3.5 Neyman-Pearson决策

-------在一类错误率固定使另一类错误率最小的判别准则

在实际遇到的模式识别问题中有可能出现这样的问题：对于两类情形，不考虑总体的情况，而只关注某一类的错误概率，要求在其中一类错误概率小于给定阈值的条件下，使另一类错误概率尽可能小。

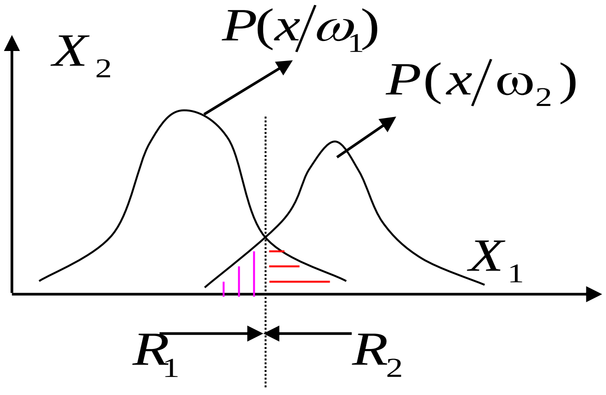
例如，在雷达目标检测中，人们可能不仅对目标出现的先验概率未知，而且对错误判断的代价也是难以估计的，甚至是难以定义的。雷达目标检测中存在两种错误：一种是虚警，即没有目标判为有目标；另一种是漏警，即有目标判为没有目标。适当的方法就是观察者确定一个允许的虚警概率，使漏警概率尽可能得小。

Neyman-Pearson判决准则解决的就是上述问题，它只适用于两类情形。 见图，在两类情况下，有两种错误概率： 第1类错误概率是，样本真实类别为 $\omega 1$，但落到了 $\omega 2$的判决区域$R2$内，从而被判为 $\omega 2$的概率，记为$E1$； 第2类错误概率是，样本真实类别为 $\omega 2$，但落到了 $\omega 1$的判决区域$R1$内，从而被判为 $\omega 1$的概率，记为$E2$。 平均错误概率为

$$Pe = P(\omega 1)E1 + P(\omega 2)E2$$

设限定$E2$不超过某个阈值ε，即：$E2 \leq \varepsilon$

在此前提下，求判决区域使$E1$达到最小值。



在不等式条件下，难以求解$E1$的最小值，

因此，可以选择$\varepsilon 0<\varepsilon$。

$E2\leq\varepsilon$条件下的求解问题转化为

$E2=\varepsilon 0$　条件下的$E1$最小值求解问题。

这是一个典型的条件极值问题。

采用Lagrange乘数法来求解，其中，约束条件为

$E2-\varepsilon 0=0$。

构造目标函数

$$r=E1+\mu(E2-\varepsilon 0)$$

其中，μ为Lagrange乘数。由$\omega 1$与$\omega 2$的决策区域分别为$R1$与$R2$，可得：

$$E_1 = \int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x}$$

$$E_2 = \int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x}$$

从而，目标函数可改写为：

$$\begin{aligned} r &= \int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x} + \mu(\int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x} - \varepsilon_0) \\ &= (1\text{-}\int_{R_1} p(\mathbf{x}|\omega_1)d\mathbf{x}) + \mu(\int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x} - \varepsilon_0) \\ &= (1-\mu\varepsilon_0) + \int_{R_1} [\mu p(\mathbf{x}|\omega_2) - p(\mathbf{x}|\omega_1)]d\mathbf{x} \end{aligned}$$

为了使r达到最小，则要求使被积函数$\mu p(\mathbf{x}|\omega 2)-p(\mathbf{x}|\omega 1)<0$，所以$R1=\{\mathbf{x}|\mu p(\mathbf{x}|\omega 2)-p(\mathbf{x}|\omega 1)<0\}$的点全部落入$R1$中，且$R1$中的点使被积函数$\mu p(\mathbf{x}|\omega 2)-p(\mathbf{x}|\omega 1)<0$。

因此，可得到如下Neyman-Pearson判决准则：

若$\mu p(\mathbf{x}|\omega 2)<p(\mathbf{x}|\omega 1)$，则$\mathbf{x}\in\omega 1$

若$\mu p(\mathbf{x}|\omega 2)>p(\mathbf{x}|\omega 1)$，则$\mathbf{x}\in\omega 2$

写成似然比形式：

$$\frac{p(\mathbf{x}\,|\,\omega_1)}{p(\mathbf{x}\,|\,\omega_2)} \gtrless \mu \Rightarrow \mathbf{x}\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

阈值μ是Lagrange乘数，是一个不确定的量，需要根据约束条件求解，即：

$$E_2=\int_{R_1} p(\mathbf{x}\,|\,\omega_2)d\mathbf{x}=\varepsilon_0$$

$$R_1=\{\mathbf{x}\,|\,L(\mathbf{x})=\frac{p(\mathbf{x}\,|\,\omega_1)}{p(\mathbf{x}\,|\,\omega_2)}>\mu\}$$

因此，Neyman-Pearson准则归结为求阈值μ。

$\therefore$ 当 $\dfrac{P(x/\omega_1)}{P(x/\omega_2)} = \mu$时，$\mu$作$R_1R_2$的分界线。

$\therefore E_2 = \int_{-\infty}^{\mu} P(x/\omega_2)dx$

$\therefore \mu$为E_2的函数在取E_2为常数时，μ可确定，这时E_2确定E_1最小。

- **例**：两类的模式分布为二维正态 $\mu_1=(-1,0)^T, \mu_2=(1,0)^T$
  协方差矩阵为单位矩阵$\Sigma_1=\Sigma_2=I$，设$E_2=0.09$求聂曼-皮尔逊准则的阈值μ。（教材P15）
- **解**：

因为是两类二元正态

$$\text{所以}P(x/\omega_1)=\frac{1}{2\pi}\exp\left[-(x-\mu_1)^T(x-\mu_1)\Big/2\right]=\frac{1}{2\pi}\exp\left[-\left[(x_1+1)^2+x_2^2\right]\Big/2\right]$$

$$\text{同理：}P(x/\omega_2)=\frac{1}{2\pi}\exp\left[-(x-\mu_2)^T(x-\mu_2)\Big/2\right]=\frac{1}{2\pi}\exp\left[-\left[(x_1-1)^2+x_2^2\right]\Big/2\right]$$

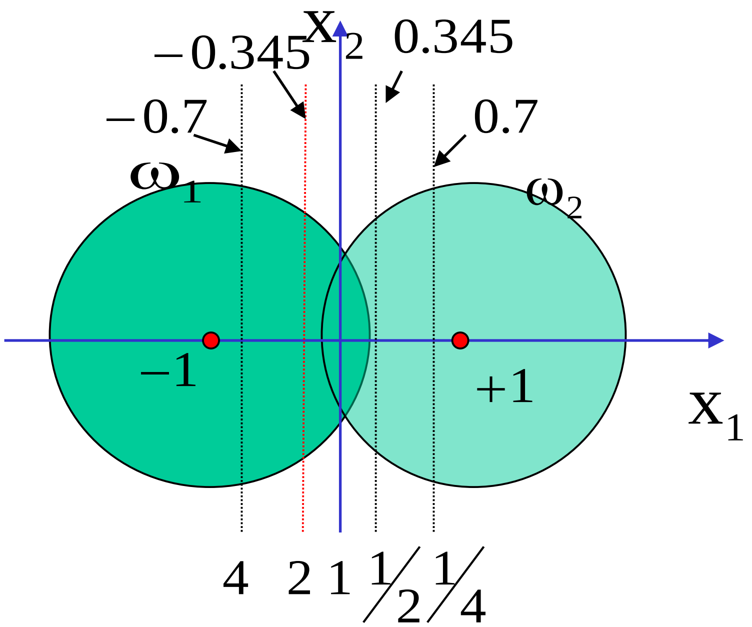
如右图所示：

$$\therefore \frac{P(x/\omega_1)}{P(x/\omega_2)} = \exp(-2x_1) = \mu$$

$\therefore$判别边界为：$\mu = \exp(-2x_1)$

$\therefore$判别式为：$\exp(-2x_1) \gtrless \mu = x \in \begin{matrix}\omega_1 \\ \omega_2\end{matrix}$

即$x_1 \lessgtr -\frac{1}{2}\ln\mu \Rightarrow x \in \begin{matrix}\omega_1 \\ \omega_2\end{matrix}$，有了判别边界和判别形式



$\therefore$对于不同μ判别边界是平行于x_2的不同直线。

如图所示：

μ大E_2小，但E_1大；μ小E_2大，但E_1小。

求似然比阈值μ：

由E_2与μ的关系有

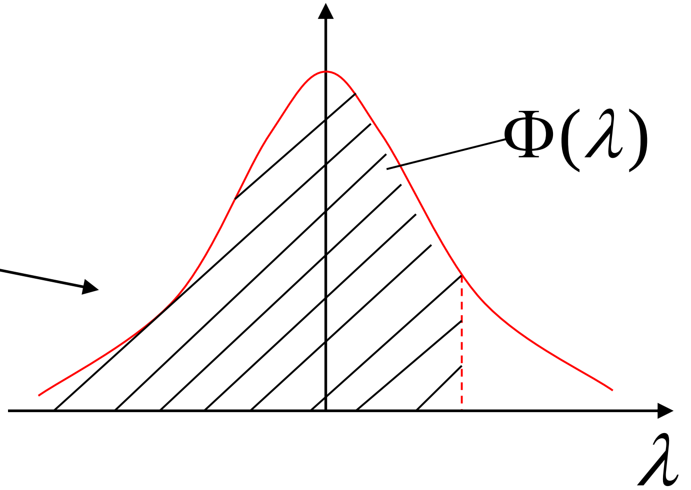
$$E2=\int_{R1}p(\mathbf{x}\mid\omega_2)d\mathbf{x}$$

$$=\int_{-\infty}^{x1}\int_{-\infty}^{x2}\frac{1}{2\pi}\exp\left[-\frac{(x_1-1)^2+x_2^2}{2}\right]dx_1dx_2$$

向正态分布表的标准形式

$$\boxed{\Phi(\lambda)=\int_{-\infty}^{\lambda}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \qquad \lambda\geq 0}$$

转换.



即：

$$E_2 = \int_{-\infty}^{-\frac{1}{2}\ln\mu} \int_{-\infty}^{+\infty} p(\mathbf{x} \mid \omega_2) d\mathbf{x}$$

$$= \int_{-\infty}^{-\frac{1}{2}\ln\mu} \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left[-\frac{(x_1-1)^2 + x_2^2}{2} \right] dx_1 dx_2$$

$$= \int_{-\infty}^{-\frac{1}{2}\ln\mu} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_1-1)^2}{2} \right] dx_1 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x_2^2}{2} \right] dx_2$$

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

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

显然，y服从标准正态分布，通过查标准正态分布表便可得到μ和$E2$间的对应关系。

查正态分布表，要求E2=0.09。在标准正态分布表(可参考概率统计相关书籍的附录)中查Φ(λ)=1-0.09=0.91， 可取两个值0.9099(相应的λ=1.34)或0.9115(相应的λ=1.35)，这里λ取两者的均值，即：

$$-1/2\ln\boldsymbol{\mu}-1=-(1.34+1.35)/2$$

计算得：$\boldsymbol{\mu}=e^{0.69}=1.9937\approx 2$

❖ μ与E_2的关系表如右：

μ	4	2	1	½	¼
E_2	0.04	0.09	0.16	0.25	0.38

$\therefore$ 给定$E_2 = 0.09$由表查得$\mu = 2$

$\therefore$ 判别式为：$\dfrac{P(x/\omega_1)}{P(x/\omega_2)} \underset{<}{>} \mu = 2 \Rightarrow x \in \underset{\omega_2}{\omega_1}$，此时$E_2 = 0.09$上式使$E_1$最小

这就是在给定E_2时使E_1最小的判别规则。

❖此时聂曼——皮尔逊分类器的分界线为：

$$x_1 = -\frac{1}{2}\ln \mu,$$

因为$\ln \mu = \ln 2 = 0.69$，所以$x_1 \underset{>}{<} -0.345 \Rightarrow x \in \underset{\omega_2}{\omega_1}$

❖由图可知为保证E_2足够小，边界应向E_1一侧靠，则E_1↑。