人工智能与自动化学院

模式识别与机器学习

13 基于统计决策的概率分类方法

统计决策的核心技术是贝叶斯方法。

所有需要作出概率预测的地方都可以见到贝叶斯方法的影子
贝叶斯是机器学习的核心方法之一。

对于不确定的事物，通过“猜测”或者“假设”得出结论

不是瞎蒙！

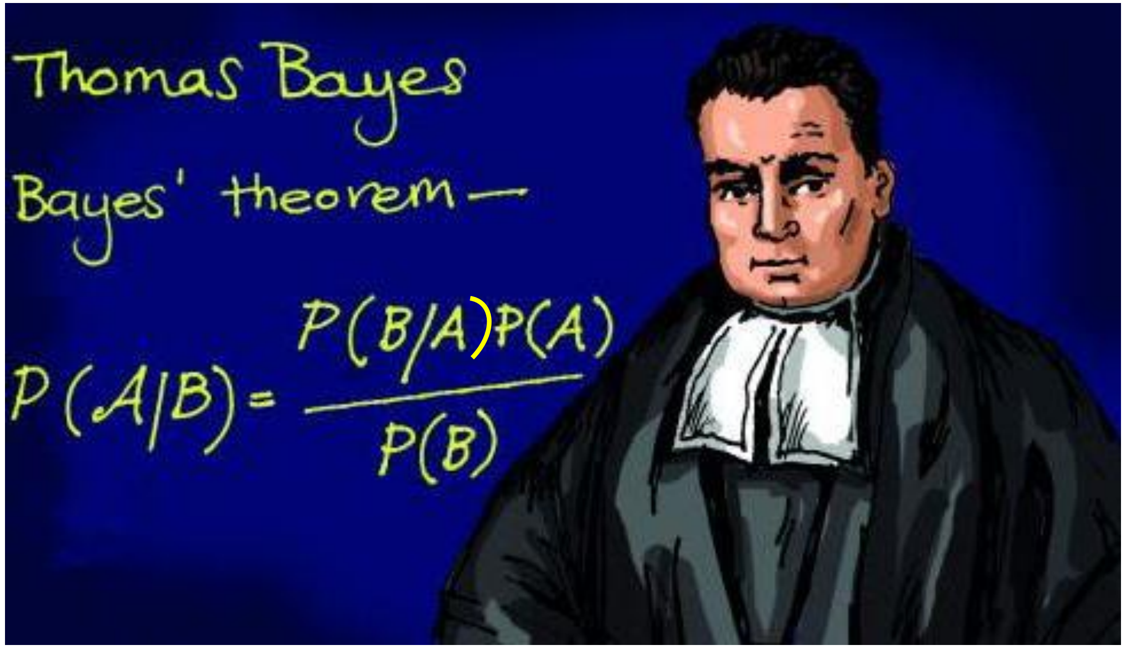
统计决策需要做两件事情：

1. 算出各种不同猜测的可能性大小。

   计算特定猜测的后验概率，对于连续的猜测空间则是计算猜测的概率密度函数。

2. 算出最靠谱的猜测是什么。

   模型比较，模型比较如果不考虑先验概率的话就是最大似然方法。



生活在18世纪的贝叶斯生前是位受人尊敬英格兰长老会牧师，为了证明上帝的存在，他发明了概率统计学原理，遗憾的是，他的这一美好愿望至死也未能实现。

引言

举例说明贝叶斯公式的来历

一所学校里面有 60% 的男生，40% 的女生。男生总是穿长裤，女生则一半穿长裤一半穿裙子。

“随机选取一个学生，他（她）穿长裤的概率和穿裙子的概率是多大”，这个就是“正向概率”的计算问题。

然而，假设你走在校园中，迎面走来一个穿长裤的学生（很不幸由于视力受限，只看得见他（她）穿的是否长裤，而无法确定他（她）的性别），你能够推断出他（她）是男生的概率是多大吗？

用频率形式重新叙述成：你在校园里面随机行走，遇到了 N 个穿长裤的人（仍然假设你无法直接观察到他们的性别），问这 N 个人里面有多少个女生多少个男生？

$$P(B|A)=\frac{U*P(B)*P(A|B)}{U*P(\tilde{B})*P(A|\tilde{B})+U*P(B)*P(A|B)}$$

$$=\frac{P(B)*P(A|B)}{P(\tilde{B})*P(A|\tilde{B})+P(B)*P(A|B)}$$

$$=\frac{P(AB)}{P(A)}$$

$$P(B|A)=\frac{P(AB)}{P(A)}$$

分析：假设学校里面人的总数是 U 个，

穿长裤的男生个数：U * P(Boy) * P(Pants|Boy)

（其中：P(Boy) = 60%；P(Pants|Boy) 是条件概率，即在 Boy 这个条件下穿长裤的概率是多大，P(Pants|Boy)= 100%）

穿长裤的女生个数： U * P(Girl) * P(Pants|Girl)

（其中，P(Girl) =40%， P(Pants|Girl)=50%）

穿长裤的总人数是：

U * P(Boy) * P(Pants|Boy) + U * P(Girl) * P(Pants|Girl)

我们要求的是 P(Girl|Pants)：

$$P(Girl|Pants)=\frac{U*P(Girl)*P(Pants|Girl)}{U*P(Boy)*P(Pants|Boy)+U*P(Girl)*P(Pants|Girl)}$$

$$=\frac{P(Girl)*P(Pants|Girl)}{P(Boy)*P(Pants|Boy)+P(Girl)*P(Pants|Girl)}$$

$$=\frac{P(Girl,Pants)}{P(Pants)}$$

“概率论”有关概念复习

划分： 设S为实验E的样本空间，$B_1, B_2, \ldots, B_n$为E的一组事件，若 $B_i \cap B_j = \Phi$，$i \neq j$，$i, j=1, 2, \ldots, n$； $B_1 \cup B_2 \cup \ldots \cup B_n = S$ 则称，$B_1, B_2, \ldots, B_n$是样本空间S的一个划分。对每次实验，$B_1, B_2, \ldots, B_n$中必有一个且仅有一个发生。

全概率公式： 设实验E的样本空间为S，A为E的事件， $B_1, B_2, \ldots, B_n$为S的一个划分，且$P(B_i)>0$（$i=1, 2, \ldots, n$），则

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\ldots+P(A|B_n)P(B_n)$$

Bayes公式： 设实验E的样本空间为S，A为E的事件，$B_1, B_2, \ldots, B_n$为S的一个划分，且$P(A)>0$，$P(B_i)>0$，（$i=1, 2, \ldots, n$），则：

$$P(B_i|A)=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)}=\frac{P(A|B_i)P(B_i)}{P(A)}$$

$$\boxed{P(A)P(B_i|A)=P(B_i)P(A|B_i)}$$

后验概率 posterior

似然函数 likelihood function

先验概率 prior

$$P(\omega_i|\vec{x})=\frac{p(\vec{x}|\omega_i)P(\omega_i)}{p(\vec{x})}=\frac{p(\vec{x}|\omega_i)P(\omega_i)}{\sum_{i=1}^{n}p(\vec{x}|\omega_i)P(\omega_i)}$$

证据因子 Evidence

全概率

模式识别领域贝叶斯公式形式

13.1 最小错误率贝叶斯决策

“概率论”有关概念复习

$$P(A)P(B_i|A) = P(B_i)P(A|B_i)$$

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

先验概率：指没有对样本获取任何观测值情况下的概率。$P(\omega_i)$表示类ω_i出现的先验概率，简称类ω_i的概率。

后验概率：$P(\omega_i|x)$表示在x出现条件下类ω_i出现的概率，称其为类别的后验概率，对于模式识别来讲可理解为x来自类ω_i的概率。

类概密：$p(x|\omega_i)$表示在类ω_i条件下的概率密度，即类ω_i模式x的概率分布密度，简称为类概密。

条件概率：在事件B出现的条件下，事件A发生的概率记做为P(A|B)，称为在B出现的条件下A出现的条件概率，而称 P(A)为无条件概率。

统计概率：若在大量重复试验中，事件A发生的频率稳定地接近于一个固定的常数p，它表明事件A出现的可能性大小，则称此常数p为事件A发生的概率，记为P(A)，即 P(A)=p

注：实际应用中实验次数足够多时可以用频率近似概率。

13.1 最小错误率贝叶斯决策

Bayes公式：设实验E的样本空间为S，A为E的事件，$B_1, B_2, \ldots, B_n$为S的一个划分，

且P(A)>0，$P(B_i)$>0，(i=1, 2, ..., n)，则：

$$P(B_i \mid A)=\frac{P(B_i A)}{P(A)}=\frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}=\frac{P(A \mid B_i)P(B_i)}{P(A)}$$

$$\boxed{P(A)P(B_i|A)=P(B_i)P(A|B_i)}$$

后验概率
posterior

似然函数
likelihood
function

先验概率
prior

$$P(\omega_i \mid \vec{x})=\frac{p(\vec{x} \mid \omega_i)P(\omega_i)}{p(\vec{x})}=\frac{p(\vec{x} \mid \omega_i)P(\omega_i)}{\sum_{i=1}^{n} p(\vec{x} \mid \omega_i)P(\omega_i)}$$

模式识别领域贝叶斯公式

证据因子
Evidence

全概率

Q:贝叶斯公式有什么用？是用来干什么的?

AW：贝叶斯公式实质上是通过观察A,把状态的先验概率$P(B_i)$转化为状态的后验概率$P(B|A_i)$

贝叶斯决策

贝叶斯决策理论方法是统计模型决策的一个基本方法，基本思想：

1. 已知类条件概率密度函数和先验概率
2. 利用贝叶斯公式转换成后验概率
3. 根据后验概率大小进行决策分类

类条件概率密度函数　先验概率

$$P(\omega_i \mid \mathrm{x}) = \frac{p(\mathrm{x} \mid \omega_i) P(\omega_i)}{p(\mathrm{x})}$$

证据因子（全概率）

- 多个特征共同构成了待识别量X的描述，$X=(x_1,x_2,\dots,x_n)^T$

最小错误率贝叶斯决策

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

若 $P(\omega_i|X)=\max\limits_{j=1,\cdots,c} P(\omega_j|X)$，$(j=1\cdots n)$，则X属于$\omega_i$类。

(1) 两分类情况下的最小错误率贝叶斯决策

若 $P(\omega_1|X)>P(\omega_2|X)$，则X属于$\omega_1$类；若 $P(\omega_1|X)<P(\omega_2|X)$，则X属于$\omega_2$类。

两分类情况下的最小错误率决策，下面四种等价规则的决策 $X\in\omega_1$，否则$X\in\omega_2$。

① 后验概率判决 $P(\omega_1 \mid X)>P(\omega_2 \mid X)$

② $p(X \mid \omega_1)P(\omega_1)>p(X \mid \omega_2)P(\omega_2)$

③ 似然比 $l(X)=\dfrac{p(X|\omega_1)}{p(X|\omega_2)}>\dfrac{P(\omega_2)}{P(\omega_1)}$，$\dfrac{P(\omega_2)}{P(\omega_1)}$ 称为似然比阈值

④ $h(X)=\ln[l(X)]=\ln p(X \mid \omega_1)-\ln p(X \mid \omega_2)>\ln\left(\dfrac{P(\omega_2)}{P(\omega_1)}\right)$

13.1 最小错误率贝叶斯决策

(1) 两分类情况下的最小错误率贝叶斯决策（续）

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

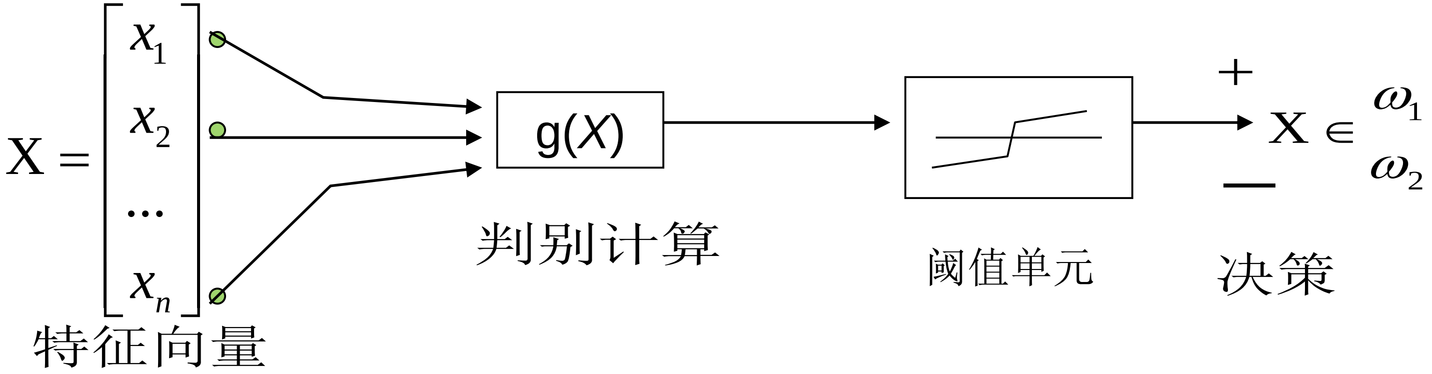
四种等价规则相应的判别函数：

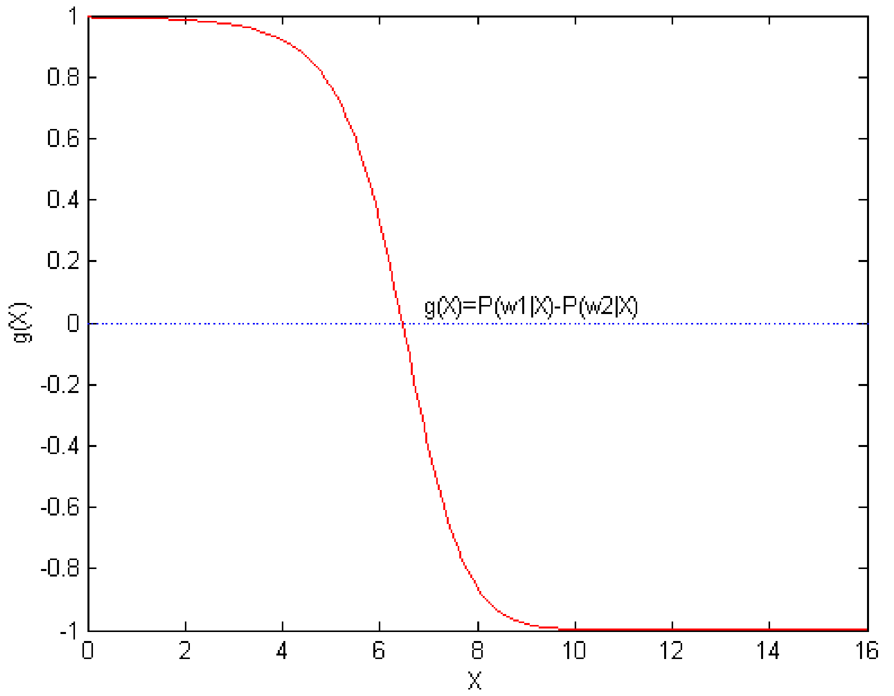
① $g(X) = P(\omega_1 | X) - P(\omega_2 | X)$, (后验概率)

② $g(X) = p(X|\omega_1)P(\omega_1) - p(X|\omega_2)P(\omega_2)$, (类条件概率密度)

③ $g(X) = \dfrac{p(X|\omega_1)}{p(X|\omega_2)} - \dfrac{P(\omega_2)}{P(\omega_1)}$, (似然比形式)

④ $g(X) = \ln\dfrac{p(X|\omega_1)}{p(X|\omega_2)} - \ln\dfrac{P(\omega_2)}{P(\omega_1)}$, (取对数方法)





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

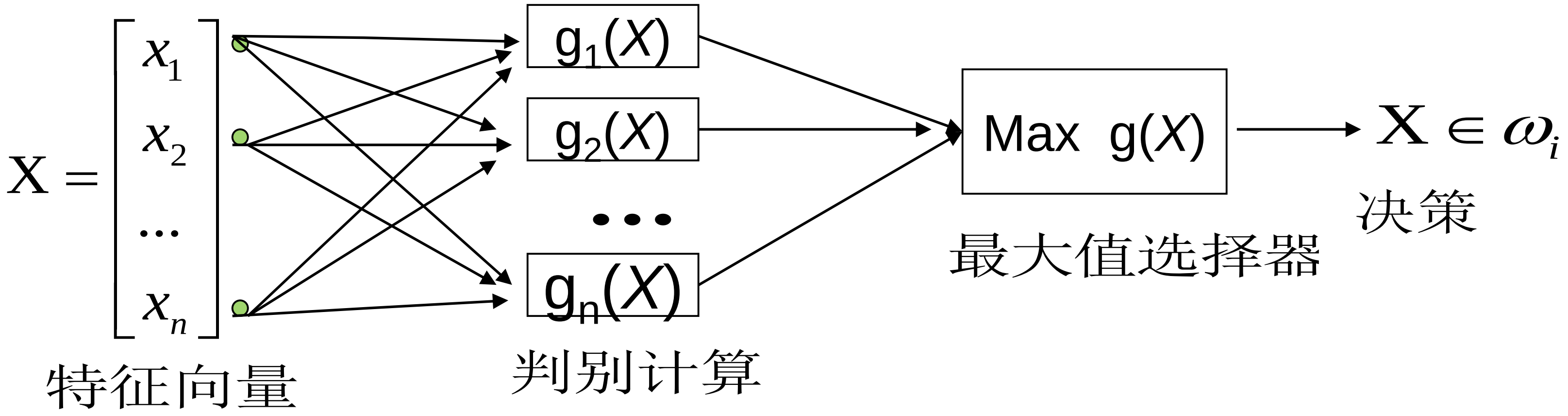
(2) 多分类情况下的最小错误率贝叶斯决策

若 $P(\omega_i|X) = \max_{j=1,\cdots,c} P(\omega_j|X)$，$(j=1\cdots n)$，则X属于$\omega_i$类。

判别函数：

$$g_i(X) = p(X \mid \omega_i)P(\omega_i) = \max_{1\le j\le M} p(X \mid \omega_j)P(\omega_j) \Rightarrow x \in \omega_i, (i=1,2,\ldots,M)$$

$$g_i(X) = \ln p(X \mid \omega_i) + \ln P(\omega_i) = \max_{1\le j\le M}\{\ln p(X \mid \omega_j) + \ln P(\omega_j)\} \Rightarrow x \in \omega_i$$



13.1 最小错误率贝叶斯决策

例1 先验概率 P(导弹)=0.2, P(飞机)=0.8，有一待识别的目标,其亮度为x,从类条件概率密度分布曲线上查到p(x|导弹)=0.2,p(x|飞机)=0.4,试用最小错误率贝叶斯决策对该目标进行分类。

解:利用贝叶斯公式，分别计算出目标x判定为飞机和导弹的后验概率。

$$
\begin{aligned}
P(\text{导弹} \mid \mathrm{X}) &= \frac{p(X \mid \text{导弹})P(\text{导弹})}{p(X \mid \text{导弹})P(\text{导弹}) + p(X \mid \text{飞机})P(\text{飞机})} \\
&= \frac{0.2 \times 0.2}{0.2 \times 0.2 + 0.4 \times 0.8} \\
&= 0.11
\end{aligned}
$$

$$
P(\text{飞机} \mid \mathrm{X}) = 1 - P(\text{导弹} \mid \mathrm{X}) = 0.89
$$

根据贝叶斯决策规则，有P(飞机|X)>P(导弹|X)，故判定这个亮度为x的待识别目标为飞机。

13.1 最小错误率贝叶斯决策

采用上述决策规则，为什么能保证判决具有最小错误率？ 简证

分类平均错误率为 $P(error)=\int_{-\infty}^{\infty}P(error,x)dx=\int_{-\infty}^{\infty}P(error\,|\,x)\,p(x)dx$

而因为决策规则为 $\boxed{\text{若}\quad P(\omega_i|X)=\max_{j=1,\cdots,c}P(\omega_j|X),\quad (j=1\cdots n)，\quad \text{则}X\text{属于}\omega_i\text{类。}}$

则 $P(error|X)=1-\max_{j=1,\cdots,c}P(\omega_j|X),\quad (j=1\cdots n)$

在该决策规则条件下，对于每一个 x， 我们都可以确保 $P(error|x)$ 最小。

此时，**$P(error)$ 最小**。

13.2 最小风险贝叶斯决策

条件风险

在实际工作中，有时仅考虑错误率最小是不够的，故引入一个与损失有关联的、比错误率更为广泛的概念—**风险**。

条件风险: $R(\alpha_i|X)=E\left[\lambda(\alpha_i,\omega_j)\right]$

α_i代表将观测样本X判定为ω_i类的决策。

$\lambda(\alpha_i,\omega_j)$ 表示观测样本$X\in\omega_j$，但被判为ω_i的**损失**。

如果i=j，则为正确判决；

否则，为错误判决。

$$\lambda_{ij}\triangleq\lambda(\alpha_i,\omega_j)$$

对于给定的X,如果采用决策 α_i(即判定X为ω_i)，从决策表可见，对于决策α_i，损失 λ 可以从 c 个$\lambda(\alpha_i,\omega_j)$ $(j=1,2,\ldots,c)$中任意选择一个，其相应的概率为 $P(\omega_j|X)$，所以有：

$$R(\alpha_i|X)=E\left[\lambda(\alpha_i,\omega_j)\right]=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j|X)\quad i=1,2,\cdots c$$

后验概率加权和

决策表

	ω_1	ω_2	…	ω_j	…	ω_c
α_1	$\lambda(\alpha_1,\omega_1)$	$\lambda(\alpha_1,\omega_2)$	…	$\lambda(\alpha_1,\omega_j)$	…	$\lambda(\alpha_1,\omega_c)$
α_2	$\lambda(\alpha_2,\omega_1)$	$\lambda(\alpha_2,\omega_2)$	…	$\lambda(\alpha_2,\omega_j)$	…	$\lambda(\alpha_2,\omega_c)$
…	…	…	…	…	…	…
α_i	$\lambda(\alpha_i,\omega_1)$	$\lambda(\alpha_i,\omega_2)$	…	$\lambda(\alpha_i,\omega_j)$	…	$\lambda(\alpha_i,\omega_c)$
…	…	…	…	…	…	…
α_a	$\lambda(\alpha_a,\omega_1)$	$\lambda(\alpha_a,\omega_2)$	…	$\lambda(\alpha_a,\omega_j)$	…	$\lambda(\alpha_a,\omega_c)$

13.2 最小风险贝叶斯决策

最小风险贝叶斯决策

$$R(\alpha_i|X)=E\left[\lambda\left(\alpha_i,\omega_j\right)\right]=\sum_{j=1}^{c}\lambda\left(\alpha_i,\omega_j\right)P(\omega_j\,|\,X)\quad i=1,2,\cdots c$$

若 $R(\alpha_k\,|\,X)=\min\limits_{i=1,2,\cdots,a}R(\alpha_i\,|\,X)$，则对应的决策 $\alpha=\alpha_k$，判定$X\in\omega_k$类。

- 在决策论中条件风险$R(\alpha_i|X)$ 为X 被判为ω_i类时损失的均值。
- 采取不同决策α_i，其条件风险$R(\alpha_i|X)$的大小是不同的。

最小风险贝叶斯决策的步骤

①已知$P(\omega_j)$, $p(X\,|\,\omega_j)$, $j=1,2,\cdots c$，根据待识别X，由Bayes公式计算后验概率$P(\omega_j\,|\,X)$

②利用决策表，计算采取α_i决策的条件风险$R(\alpha_i\,|\,X)$

$$R(\alpha_i\,|\,X)=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j\,|\,X),\ i=1,2,\cdots,a$$

③上式得到的a个条件风险值$R(\alpha_i\,|\,X)$, $i=1,2,\cdots,a$ ，找出使条件风险最小的决策a_k

$$R(\alpha_k\,|\,X)=\min_{i=1,2,\cdots,a}R(\alpha_i\,|\,X)$$

则 α_k就是最小风险Bayes决策，判决$X\in\omega_k$

两分类情况下的最小风险贝叶斯决策

$$R(\alpha_1|X) = \lambda(\alpha_1,\omega_1)P(\omega_1|X) + \lambda(\alpha_1,\omega_2)P(\omega_2|X)$$
$$R(\alpha_2|X) = \lambda(\alpha_2,\omega_1)P(\omega_1|X) + \lambda(\alpha_2,\omega_2)P(\omega_2|X)$$

此时，最小风险贝叶斯决策为：若$R(\alpha_1|X) < R(\alpha_2|X)$, 则X属于$\omega_1$, 否则属于$\omega_2$。

最小风险贝叶斯决策的另两种形式:

$$\lambda_{ij} \triangleq \lambda(\alpha_i,\omega_j)$$

若 $(\lambda_{21}-\lambda_{11})P(\omega_1|X) > (\lambda_{12}-\lambda_{22})P(\omega_2|X)$，则决策$X\in\omega_1$；否则$X\in\omega_2$。

若 $l(x) = \dfrac{p(X|\omega_1)}{p(X|\omega_2)} > \dfrac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\dfrac{P(\omega_2)}{P(\omega_1)}$，则决策$X\in\omega_1$；否则$X\in\omega_2$。

13.2 最小风险贝叶斯决策

例2 已知先验概率P(导弹)=0.2，P(飞机)=0.8，有一个待识别的目标, 某次实验测得其亮度为x, 查类概率密度函数表得概率密度为p(X|导弹)=0.2， p(X|飞机)=0.4，决策表如右所示，试按最小风险贝叶斯决策方法对本次实验进行分类。

解：后验概率P(飞机|X)=0.89，P(导弹|X)=0.11，具体计算过程见前例1。

再计算条件风险

损失λ	X∈ω_1导弹	X∈ω_2飞机
α_1 判定X为导弹	0	1
α_2 判定X为飞机	9	0

$$R(\alpha_1|\mathrm{X}) = \lambda(\alpha_1,\omega_1)P(\omega_1|\mathrm{X}) + \lambda(\alpha_1,\omega_2)P(\omega_2|\mathrm{X})$$

$$= \lambda(\alpha_1,\omega_2)P(\omega_2|\mathrm{X}) = 0.89$$

$$R(\alpha_2|\mathrm{X}) = \lambda(\alpha_2,\omega_1)P(\omega_1|\mathrm{X}) + \lambda(\alpha_2,\omega_1)P(\omega_2|\mathrm{X})$$

$$= \lambda(\alpha_2,\omega_1)P(\omega_1|\mathrm{X}) = 9\times 0.11 = 0.99$$

$$\because R(\alpha_1|\mathrm{X}) < R(\alpha_2|\mathrm{X}), \therefore \text{判定} X \in \omega_1, \text{目标是导弹}$$

13.2 最小风险贝叶斯决策

两种决策方法之间的关系：基于最小错误率的决策是基于最小风险决策的一个特例。

设损失函数为 $\lambda(\alpha_i,\omega_j)=\begin{cases}0 & i=j\\ 1 & i\neq j\end{cases},\quad i,j=1,2,\cdots,c$

式中假定对c类只有c个决策，即不考虑“拒绝”等其他情况，当作出正确决策(即$i=j$)时没有损失(损失为0)，而对于任何错误决策，其损失均为1。这样定义的损失函数称为为0-1损失函数。

根据条件风险定义，有

$$R(\alpha_i|X)=\sum_{j=1}^{c}\lambda(\alpha_i,\omega_j)P(\omega_j|X)=\sum_{j=1,j\neq i}^{c}P(\omega_j|X)\quad i=1,2,\cdots c$$

则

$$\min_{i=1,2,\cdots,c}R(\alpha_i|X)=\min_{i=1,\cdots,c}\sum_{\substack{j=1\\ j\neq i}}^{c}P(\omega_j|X)=\min_{i=1,\cdots,c}\left[1-P(\omega_i|X)\right]=1-\max_{i=1,\cdots,c}\left[P(\omega_i|X)\right]$$

最小错误率贝叶斯决策就是0--1损失函数条件下的最小风险贝叶斯决策。

两类错误率

- 不同情况的分类错误带来的损失是不同的。
- 在医学领域，人们通常用阳性（Positive）和阴性（Negative）来代表两类：阳性表示某一症状存在或检测到某一指标的异常；阴性则表示考查的症状不存在或监测的指标没有异常。
- 对疾病的诊断就是对阳性和阴性的一个两类决策问题，在不考虑拒绝的情况下，状态和决策之间可能的关系：

决策	状态	
	阳性	阴性
阳性	真阳性（TP）	假阳性（FP）
阴性	假阴性（FN）	真阴性（TN）

两类错误率

决策	状态	
	阳性	阴性
阳性	真阳性（TP）	假阳性（FP）
阴性	假阴性（FN）	真阴性（TN）

- 假阳性率（False Positive Rate） $\alpha = \frac{FP}{TN + FP}$

表示假阳性样本占总阴性样本的比例。

- 假阴性率（False Negative Rate） $\beta = \frac{FN}{TP + FN}$

表示假阴性样本占总阳性样本的比例。

- 灵敏度（Sensitivity） $S_n = \frac{TP}{TP + FN}$

表示阳性样本正确识别出来的能力。在医学应用场景就是把有病的人诊断出来的比例。

- 特异度（Specificity） $S_p = \frac{TN}{TN + FP}$

表示把阴性样本正确判断出来的能力。在医学应用场景就是把无病的人误诊为有病的比例。

$$S_n = 1 - \beta$$

$$S_p = 1 - \alpha$$

Neyman-Pearson决策

在某些问题中，某一种错误较另一种错误更为重要，即危害更为重要，这时需要在某个约束条件下最小化总风险。

比如，在目标分类器中要求将“导弹”误判为“飞机”的错误率不得超过1%，同时要求在此约束条件下最小化将“飞机”误判为“导弹”的可能性。

Neyman-Pearson准则：是严格限制较重要的一类错误概率令其等于某常数而使另一类误判概率最小。

Neyman-Pearson决策

对两类识别问题，有两种错误可能发生：1. 属于ω1的模式X被误分到类ω2；
2. 属于ω2的模式X被误分到类ω1；

设这两种错误的概率分别为$P_1(e)$和$P_2(e)$，则：

$$P_1(e)=\int_{R_2} p(X|\omega_1)dX \qquad P_2(e)=\int_{R_1} p(X|\omega_2)dX$$

其中，ω1和ω2两类的决策域分别为R_1和R_2，R是整个特征空间，$R_1+R_2=R$ 。

平均错误率：$P(e)=p(\omega_1)\int_{R_2} P(X|\omega_1)dX+p(\omega_2)\int_{R_1} P(X|\omega_2)dX=P_1(e)p(\omega_1)+P_2(e)p(\omega_2)$

显然，决策域R_1、R_2及类条件概率密度函数$p(X|\omega_1)$和$p(X|\omega_2)$会影响$P_1(e)$和$P_2(e)$的值。

Neyman-Pearson判决准则：在$P_2(e)=\varepsilon_0$（常数）的条件下，使得$P_1(e)$取得最小值。

Neyman-Pearson决策

Neyman-Pearson判决准则：在$P_2(e)=\varepsilon_0$（常数）的条件下，使得$P_1(e)$取得最小值。

运用拉格朗日乘数法求条件极值，为此作辅助函数：$L=P_1(e)+\lambda(P_2(e)-\varepsilon_0)$

利用

$$\int_{R_1} p(X|\omega_1)dX+\int_{R_2} p(X|\omega_1)dX=1$$

$$P_1(e)=\int_{R_2} p(X|\omega_1)dX$$

$$P_2(e)=\int_{R_1} p(X|\omega_2)dX$$

$$\Rightarrow$$

$$\begin{aligned}
L&=P_1(e)+\lambda(P_2(e)-\varepsilon_0)\\
&=\int_{R_2} p(X|\omega_1)dX+\lambda\left[\int_{R_1} p(X|\omega_2)dX-\varepsilon_0\right]\\
&=1-\int_{R_1} p(X|\omega_1)dX+\lambda\int_{R_1} p(X|\omega_2)dX-\lambda\varepsilon_0\\
&=(1-\lambda\varepsilon_0)-\int_{R_1}[p(X|\omega_1)-\lambda p(X|\omega_2)]dX
\end{aligned}$$

Neyman-Pearson决策

优化的目标：求解使L式最小的决策边界t

$$L=(1-\lambda\varepsilon_0)-\int_{R_1}\underline{\left[p(X|\omega_1)-\lambda p(X|\omega_2)\right]}dX$$

=0 是判决面

设决策域R_1和R_2的分界面为t，L分别对t和对λ求导，并令

$$\frac{\partial L}{\partial t}=0 \qquad \frac{\partial L}{\partial \lambda}=0$$

决策边界t上满足 $\lambda=\dfrac{p(X=t|\omega_1)}{p(X=t|\omega_2)}$

决策边界使：$\boxed{\int_{R_1}p(X|\omega_2)dX=\varepsilon_0}$

求R_1使得L取极小值

一般地讲，R_1无法直接用解析的办法求得。

R_1的求解？

但注意到λ在L式中是确定的，$p(\vec{x}|\omega_1)$、$p(\vec{x}|\omega_2)$在R空间中也是确定的，若选择满足条件$p(\vec{x}|\omega_1)-\lambda p(\vec{x}|\omega_2)>0$的全体X作为$R_1^*$，就能保证此时的L *值比其他$R_1$取法时的L值要小。

因为这种取法下，是使被积函数取正数的最大的域。 可证（见后）

即得判决准则 $$\boxed{\frac{p(X|\omega_1)}{p(X|\omega_2)}>\lambda,\ X\in\omega_1 \qquad \frac{p(X|\omega_1)}{p(X|\omega_2)}<\lambda,\ X\in\omega_2}$$

该准则限定一类错误率为常数，而使得另一类错误率最小

13.3 两类错误率、Neyman-Pearson决策和ROC曲线

简证　这是因为，对于任何其它取法，ω_1的新的判决类域

$$R_1' = (R_1^* - R_{11}) \cup R_{12}$$

R_1^*如前定义，$R_{11} \subseteq R_1^*, R_{12} \subseteq R_2^*$

于是在R_{11}上有　$p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2) > 0$

在R_{12}上有　$p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2) < 0$

这时的L值为

$$L = (1-\lambda\varepsilon_0) - \int_{R_1^*-R_{11}} (p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2))d\vec{x} - \int_{R_{12}} (p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2))d\vec{x}$$

$$= \boxed{L^* + \int_{R_{11}} \underbrace{(p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2))}_{>0} d\vec{x} - \int_{R_{12}} \underbrace{(p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2))}_{<0} d\vec{x}} \quad >L^*$$

因此，选择满足条件$p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2) > 0$的全体X作为$R_1^*$

综上，即：
$$\begin{cases} 在\ R_1\ 中 & p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2) > 0 \\ 在\ R_2 中 & p(\vec{x}|\omega_1) - \lambda p(\vec{x}|\omega_2) < 0 \end{cases}$$

Neyman-Pearson决策

Neyman-Pearson判决准则求得似然比

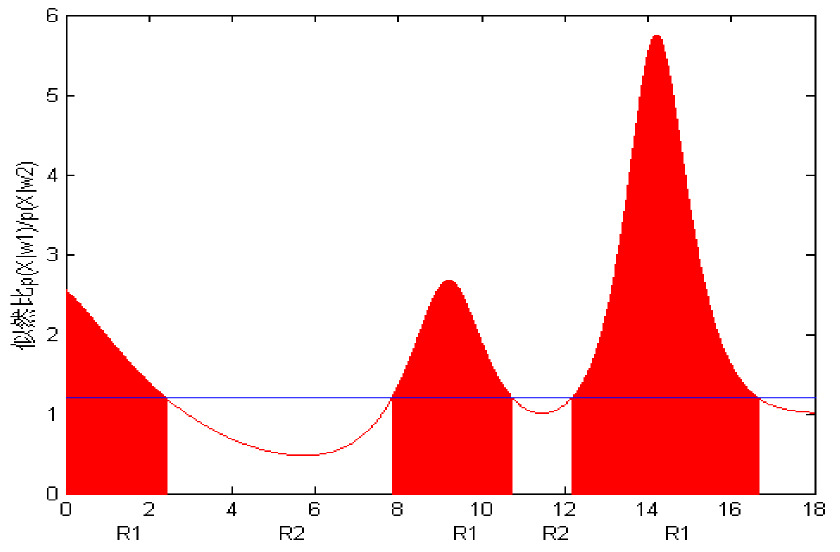
$$l(X)=\frac{p(X|\omega_1)}{p(X|\omega_2)}$$

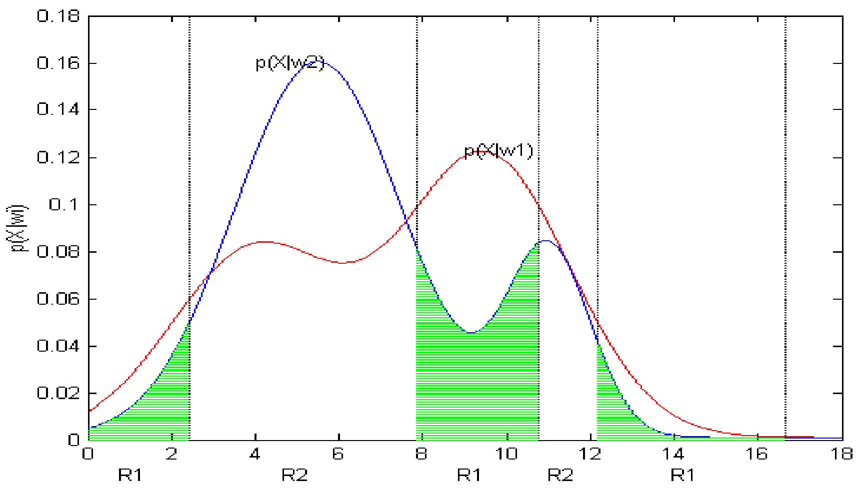
设置阈值 λ，判决边界将由阈值决定。

如果 $\dfrac{p(X|\omega_1)}{p(X|\omega_2)}>\lambda$ ， $X\in\omega_1$;

如果 $\dfrac{p(X|\omega_1)}{p(X|\omega_2)}<\lambda$ ， $X\in\omega_2$

$$\int_{R_1}p(X|\omega_2)dX=\varepsilon_0$$



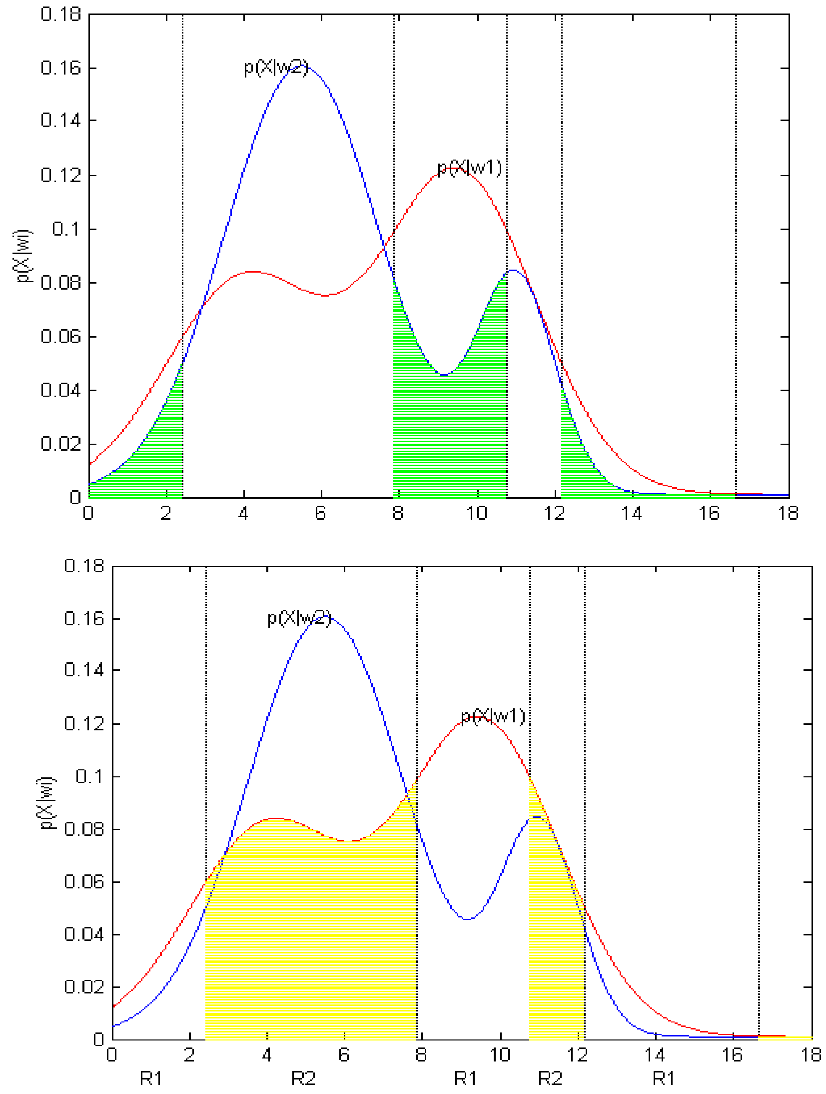


Neyman-Pearson决策

其他的判决边界划分方法即使可以保证$P_2(e)$绿色区域面积等于ε_0

$$\int_{R_1} p(X \mid \omega_2) dX = \varepsilon_0$$

但$P_1(e)$黄色区域面积却不能取到最小。



Neyman-Pearson决策与其它决策的比较

- Neyman-Pearson判决准则 $l(X)=\dfrac{p(X|\omega_1)}{p(X|\omega_2)}\begin{matrix}>\\<\end{matrix}\lambda,\quad$ 则$X\in\begin{cases}\omega_1\\\omega_2\end{cases}$

- 最小错误率贝叶斯决策规则 $l(X)=\dfrac{p(X|\omega_1)}{p(X|\omega_2)}\begin{matrix}>\\<\end{matrix}\dfrac{P(\omega_2)}{P(\omega_1)},$ 则$X\in\begin{cases}\omega_1\\\omega_2\end{cases}$

- 最小风险贝叶斯决策规则 $l(x)=\dfrac{p(X|\omega_1)}{p(X|\omega_2)}\begin{matrix}>\\<\end{matrix}\dfrac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\dfrac{P(\omega_2)}{P(\omega_1)},$ 则$X\in\begin{cases}\omega_1\\\omega_2\end{cases}$

都是以似然比为基础，目标不同。

Neyman-Pearson决策 λ 的计算

- 定义 $l(X)$： $l(\mathrm{X})=\dfrac{p(\mathrm{X}\mid\omega_1)}{p(\mathrm{X}\mid\omega_2)}$
- 定义 $p(l|\ \omega_2)$ 为似然比 $l(X)$ 在条件 $X\in\omega_2$ 下的概率密度。
- 因 $l(X)>\lambda$ 就判 $X\in\omega_1$，所以 λ 可用下式确定

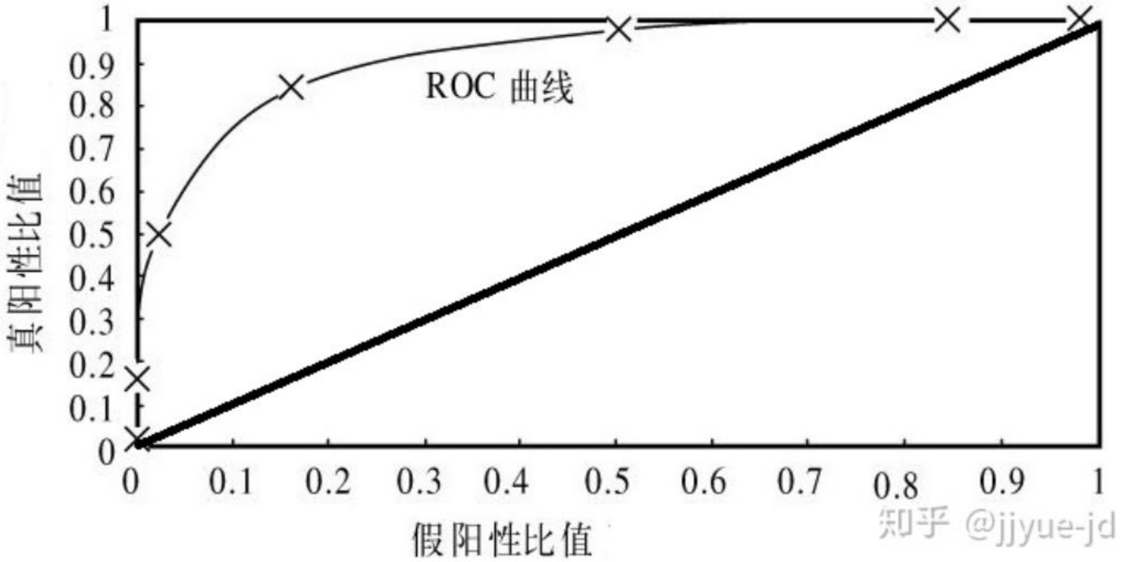
$$P_2(e)=\int_{\lambda}^{+\infty}p(l\mid\omega_2)dl=\varepsilon_0$$

λ无法显式计算得到，需要用数值方法求解

- 因为 $p(l|\ \omega_2)\geq 0$，$P_2(e)$ 是 λ 的单调函数，可采用试探法对几个不同的λ值计算出$P_2(e)$后，总可以找到一个合适的λ值，能满足 $P_2(e)=\varepsilon_0$，又使得 $P_1(e)$ 尽可能小。

ROC曲线

ROC曲线：如果把灵敏度即真阳性率（$1-P_1(e)$）作为纵坐标，假阳性率（$1-S_p=P_2(e)$）作为横坐标，可以获得随着决策面变化，两类错误率的变化情况。



- 绘制方法：在类条件概率密度已知的情况下，可以通过求不同似然比阈值条件下的两类错误率（ $P_1(e)=\int_{R_2} p(X|\omega_1)dX$ $P_2(e)=\int_{R_1} p(X|\omega_2)dX$ ），画出ROC曲线。
- 用途：① 比较不同分类判别方法的性能
  ② 评估特征与类别的相关性度量
- ROC曲线比较：可以用ROC曲线下的面积即AUC（Area under ROC curve）来定量衡量方法的性能。

Break

正态分布

从分类器设计上来看，决策面方程都涉及到类条件概率密度 $p(X|\omega_i)$ 。在连续类概率密度函数中，研究较多的是正态分布。

- ✓ 大量随机变量服从正态分布，
- ✓ 而且正态分布在数学上有很多可用的性质，数学上容易处理

因此，以正态分布为例来说明。

单变量正态分布函数

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \qquad \text{记为} p(x)\sim N(\mu,\sigma^2)$$

均值 $E\{x\}=\int_{-\infty}^{\infty}xp(x)dx=\mu$

方差 $\int_{-\infty}^{\infty}(x-\mu)^2p(x)dx=\sigma^2$

性质:

- 单变量正态分布p(x)由μ, σ^2 可以完全确定。
- 随机变量 x 集中在均值μ附近, 其分散度用标准差σ表示, 95%样本落入 $|x-\mu|<2\sigma$ 范围内。

从单变量正态分布函数到多元正态分布函数

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

Univariate Density

$$p(x)=(2\pi)^{-1/2}(\sigma^2)^{-1/2}\exp[-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)]$$

Univariate Density

Multivariate Density

$$p(x)=(2\pi)^{-d/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$

称X服从d维正态分布，记作$X\sim N_d(\mu,\Sigma)$

13.4 正态分布的统计决策

多元正态分布函数

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right] \quad 记作X\sim N_d(\mu,\Sigma)$$

式中$x=[x_1,\cdots,x_d]^T$是d维随机变量，$\mu=[\mu_1,\cdots,\mu_d]^T$是$d$维均值向量

$$\Sigma=E[(x-\mu)(x-\mu)^T]=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd}\end{pmatrix}=\begin{pmatrix}\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d}\\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2d}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_d^2\end{pmatrix}$$

Σ是$d\times d$维的协方差矩阵，为对称矩阵且正定$|\Sigma|>0$。

x_i的方差$\sigma_{ii}=\sigma_i^2$，是对角线上的元素；x_i和x_j的协方差$\sigma_{ij}=E[(x_i-\mu_i)(x_j-\mu_j)]$，是非对角线上的元素

Σ^{-1} 是Σ的逆矩阵，$|\Sigma|$ 是Σ的行列式

多元正态分布函数

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$ 记作 $X\sim N_d(\mu,\Sigma)$

$\mu=E\{x\}$, $\qquad \Sigma=E[(x-\mu)(x-\mu)^T]$

μ,Σ 分别是向量 x 和矩阵 $(x-\mu)(x-\mu)^T$ 的期望

$$x=[x_1,\cdots,x_d]^T \qquad \mu_i=E\{x_i\}=\int_{E^d}x_i p(x)\mathrm{d}x=\int_{-\infty}^{\infty}x_i p(x_i)\mathrm{d}x_i$$

其中，$p(x_i)$ 为边缘分布: $p(x_i)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{i-1}\mathrm{d}x_{i+1}\cdots\mathrm{d}x_d$

$$\sigma_{ij}=E\left[(x_i-\mu_i)(x_j-\mu_j)\right]$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_i-\mu_i)(x_j-\mu_j)p(x_i,x_j)\mathrm{d}x_i\mathrm{d}x_j$$

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd}\end{pmatrix}$$

13.4 正态分布的统计决策

多元正态分布函数的性质

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$ 记作 $X\sim N_d(\mu,\Sigma)$

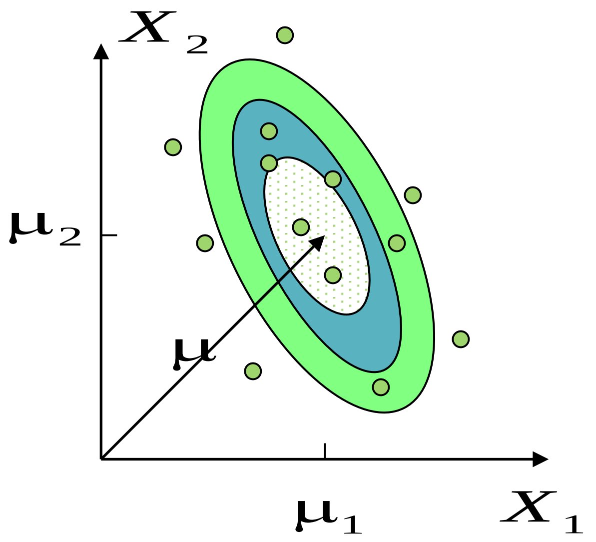
① 参数 μ 和 Σ决定分布形状

多元正态分布由μ 和 Σ完全确定，具体包括 $d+d(d+1)/2$ 个数目的参数，其中，d为均值 μ 的分量数，$d(d+1)/2$为协方差 Σ 的独立元素数。通常记为 。

$$p(x)\sim N(\mu,\Sigma)$$

② 等概率密度点的轨迹为一超椭球面

从正态分布总体中抽取的样本大部分落在以均值向量 μ 为中心，大小由协方差矩阵 Σ确定的区域。该区域中心由 μ 决定，区域形状由 Σ 决定。



多元正态分布函数的性质

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$ 记作 $X \sim N_d(\mu,\Sigma)$

② 等概率密度点的轨迹为一超椭球面

当指数项为常数时，密度 $p(x)$ 不随 x 变化，即为等概率密度点。此时:

方程 $(x-\mu)^T\Sigma^{-1}(x-\mu)=$ 常数k，对应的解是一个超椭球面。

定义 $\boldsymbol{r}^2=(\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)$，称为 x 到 μ 的 Mahalanobis（马氏）距离平方。

所以,等概率密度点的轨迹是 x 到 μ 的马氏距离为常数的超椭球面。

对应于马氏距离为r的超椭球体积为: $V=V_d|\Sigma|^{\frac{1}{2}}r^d$

∴在维数d给定的情况下，样本离散度随$|\Sigma|^{1/2}$而变

其中，V_d是d维单位超球体的体积

$$V_d=\begin{cases}\dfrac{\pi^{d/2}}{(d/2)!} & d\text{为偶数}\\[2ex] \dfrac{2^d\pi^{(d-1)/2}\left((d-1)/2\right)!}{d!} & d\text{为奇数}\end{cases}$$

13.4 正态分布的统计决策

多元正态分布函数的性质

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$ 记作 $X\sim N_d(\mu,\Sigma)$

③ 在正态分布中不相关性等价于独立性

$$\begin{cases}\text{不相关性定义：}E[x_i x_j]=E[x_i]\cdot E[x_j]\\ \text{独立性定义：}p(x_i,x_j)=p(x_i)p(x_j)\end{cases}$$

*独立性条件更强

若多元正态分布的任意两个分量 x_i 与 x_j 互不相关,则 x_i 与 x_j 一定独立。

多元正态分布下不相关性等价于独立性

推论：如果多元正态随机向量 x 的协方差阵是对角阵，则 x 的分量是互相独立的正态分布随机变量。

多元正态分布函数的性质

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]$$ 记作 $X\sim N_d(\mu,\Sigma)$

④ 多元正态分布的边缘分布和条件分布具有正态性

$x\sim N(\mu,\Sigma)$，其中，$x=\begin{bmatrix}x_1\\x_2\end{bmatrix},\mu=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix},\Sigma=\begin{bmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{12}&\sigma_{22}\end{bmatrix}$

x_1 的边缘分布为 $p(x_1)\sim N(\mu_1,\sigma_{11})$

x_2 的边缘分布为 $p(x_2)\sim N(\mu_2,\sigma_{22})$

多元正态分布的边缘分布和条件分布仍然是正态分布。

多元正态分布函数的性质

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]\quad \text{记作} X\sim N_d(\mu,\Sigma)$$

⑤ 线性变换的正态性

$\boldsymbol{x}$ 为多元正态分布的随机向量，其均值向量为 $\boldsymbol{\mu}$，协方差矩阵为 Σ。对 $\boldsymbol{x}$ 作线性变换，即

$$\boldsymbol{y} = A\,\boldsymbol{x}$$

A为线性变换矩阵且非奇异。y 服从均值向量为$A\boldsymbol{\mu}$、协方差矩阵为$A\Sigma A^T$的多元正态分布。

$$p(y)\sim N(A\boldsymbol{\mu},\ A\Sigma A^T)$$

⑥ 线性组合的正态性

$\boldsymbol{x}$ 为多元正态分布的随机向量，则线性组合 $y=a^T\boldsymbol{x}$ 是一维的正态随机变量, a是x同维向量。

$$p(y)\sim N(a^T\boldsymbol{\mu},\ a^T\Sigma a)$$

正态分布概率模型下的最小错误率贝叶斯决策

类条件概率密度为正态分布：

$$\boxed{p(x|\omega_i)=\frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}\exp[-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)]}$$

判别函数 $g_i(x)=\ln[p(x|\omega_i)P(\omega_i)]$

$$=\ln p(x|\omega_i)+\ln P(\omega_i)$$

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

决策面方程 $g_i(x)=g_j(x)$ 或 $g_i(x)-g_j(x)=0$ i, j相邻

即 $$-\frac{1}{2}[(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)-(x-\mu_j)^T\Sigma_j^{-1}(x-\mu_j)]-\frac{1}{2}\ln\frac{|\Sigma_i|}{|\Sigma_j|}+\ln\frac{P(\omega_i)}{P(\omega_j)}=0$$

正态分布概率模型下的最小错误率贝叶斯决策

下面分三种情况讨论决策面：

(1) $\Sigma_i=\sigma^2 I,\quad i=1,2,\cdots,c$　　最简单，最特殊

$$\Sigma_i=\begin{bmatrix}\sigma^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma^2\end{bmatrix}$$

(2) $\Sigma_i=\Sigma$

$$\Sigma_1=\Sigma_2=\cdots=\Sigma_c=\Sigma$$

(3) $\Sigma_i\neq\Sigma_j,\quad i,j=1,2,\cdots,c$　　最复杂，最普遍

正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

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

- 各类模式分布的协方差矩阵相等，各 x_i 统计独立且方差相同，协方差均为0。
- 几何上相当于各类样本落在以 μ_i 为中心同样大小的一些超球体中。

$$\Sigma_i=\begin{bmatrix}\sigma^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sigma^2\end{bmatrix} \Rightarrow |\Sigma_i|=\sigma^{2d},\ \Sigma_i^{-1}=I/\sigma^2=\begin{bmatrix}1/\sigma^2 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1/\sigma^2\end{bmatrix}$$

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

判别函数中第二和第三项与类别i无关

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

$$\xrightarrow{\text{欧氏距离平方}}$$

$$\|x-\mu_i\|^2=(x-\mu_i)^T(x-\mu_i)$$

$$\boxed{\boldsymbol{g_i(x)=-\frac{\|x-\mu_i\|^2}{2\sigma^2}+\ln P(\omega_i)}}$$

正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

$$g_i(x) = -\frac{\| x-\mu_i \|^2}{2\sigma^2} + \ln P(\omega_i)$$

① 如果$P(w_i)= P(w_j)$ 先验概率相等

- 测量从待分类向量x到每一类均值向量的欧氏距离，把x分到距离最近的类，即 $\min\limits_{i=1,\cdots,c} \| x-\mu_i \|^2$ ，μ_i是从训练样本集中得到的。也称最小距离分类器。
- 若把每个均值向量μ_i 看作一个典型的样本(模板)，则这种分类方法也称为模板匹配技术。

② 如果$P(w_i)\neq P(w_j)$ 先验概率不相等

- 欧氏距离的平方必须用方差σ^2规范化后减去$\ln P(w_i)$，再用于分类。
- 如果待分类的向量 x 同两类均值向量的欧氏距离相等，则最小错误概率Bayes决策把这模式归入先验概率大的那类。

正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

$$g_i(x)=-\frac{\|x-\mu_i\|^2}{2\sigma^2}+\ln P(\omega_i)$$

把 $g_i(x)$ 展开可得

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

x^T x与分类无关，可以忽略

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

其中，$W_i=\frac{1}{\sigma^2}\mu_i$， $w_{i0}=-\frac{1}{2\sigma^2}\mu_i^T\mu_i+\ln P(\omega_i)$

这是与均值有关的线性判别函数，组成线性分类器。

对待分类的样本x,分别计算 $g_i(x)$, $i=1,2,\ldots,c$，若 $g_k(x)=\max g_i(x)$，则决策 $x\in w_k$

正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

$$\boldsymbol{g}_i(\boldsymbol{x})=-\frac{1}{2\sigma^2}(-2\mu_i^T\boldsymbol{x}+\mu_i^T\mu_i)+\ln \boldsymbol{P}(\omega_i)=\boldsymbol{W}_i^T\boldsymbol{x}+\boldsymbol{w}_{i0}$$

决策面方程： $g_i(x)-g_j(x)=0$

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

$$=\frac{1}{\sigma^2}\left(\mu_i^T-\mu_j^T\right)x-\frac{1}{2\sigma^2}\left(\mu_i^T\mu_i-\mu_j^T\mu_j\right)+\ln\frac{P(\omega_i)}{P(\omega_j)}$$

欧氏距离: $\|\mu_i-\mu_j\|^2=(\mu_i-\mu_j)^T(\mu_i-\mu_j)$

$$g_i(x)-g_j(x)=\frac{1}{\sigma^2}\left(\mu_i-\mu_j\right)^T x-\frac{1}{2\sigma^2}\left(\mu_i-\mu_j\right)^T\left(\mu_i+\mu_j\right)+\frac{\left(\mu_i-\mu_j\right)^T\left(\mu_i-\mu_j\right)}{\|\mu_i-\mu_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}=0$$

$$\frac{1}{\sigma^2}\left(\mu_i-\mu_j\right)^T\left[x-\frac{1}{2}\left(\mu_i+\mu_j\right)+\frac{\sigma^2\left(\mu_i-\mu_j\right)}{\|\mu_i-\mu_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}\right]=0$$

正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

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

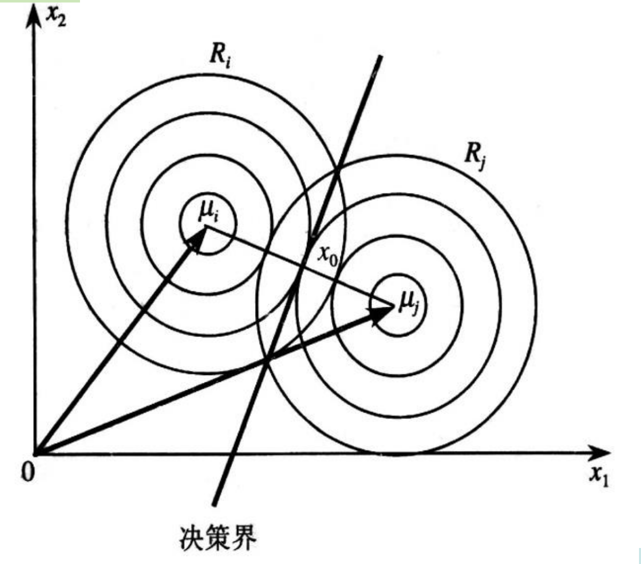
决策面方程： $g_i(x)-g_j(x)=0$

$$g_i(x)-g_j(x)=\frac{1}{\sigma^2}(\mu_i-\mu_j)^T x-\frac{1}{2\sigma^2}(\mu_i-\mu_j)^T(\mu_i+\mu_j)+\frac{(\mu_i-\mu_j)^T(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}=0$$

$$\frac{1}{\sigma^2}(\mu_i-\mu_j)^T\left[x-\frac{1}{2}(\mu_i+\mu_j)+\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}\right]=0$$

$$W^T(x-x_0)=0$$

其中，$W=\mu_i-\mu_j$， $x_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|}\ln\frac{P(\omega_i)}{P(\omega_j)}$



- 这个方程确定了决策面是通过 x_0 并正交于向量W的一个超平面。
- 由于W=$\mu_i-\mu_j$ 所以超平面正交于均值向量 μ_i 与 μ_j 之间的联线。

正态分布概率模型下的最小错误率贝叶斯决策

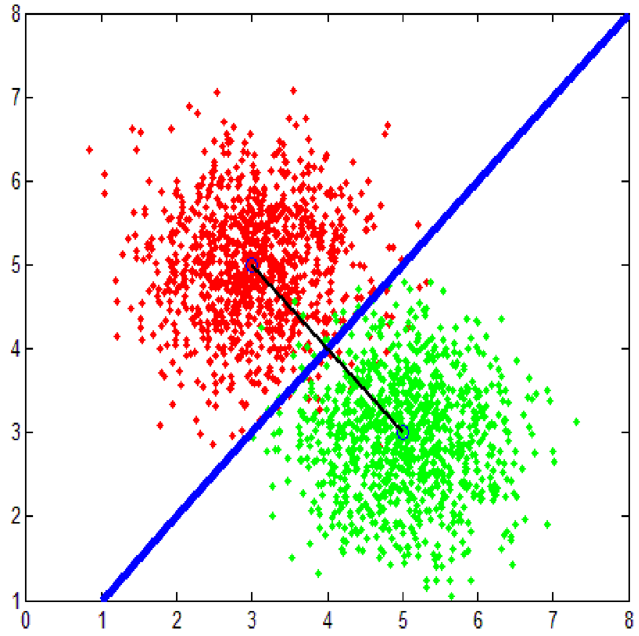
(1) $\Sigma_i=\sigma^2 I$

决策面方程： $W^T(x-x_0)=0$ 其中，$W=\mu_i-\mu_j$， $x_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|}\ln\frac{P(\omega_i)}{P(\omega_j)}$

① 若先验概率相等 $P(\omega_i)=P(\omega_j)$，则 $x_0=\frac{1}{2}(\mu_i+\mu_j)$

超平面通过 μ_i 与 μ_j 连线的中点 x_0，且与连线正交。

$\mu_1(3,5)$， $\mu_2(5,3)$， $\sigma^2=0.5$， $P(w_1)=P(w_2)=0.5$



13.4 正态分布的统计决策

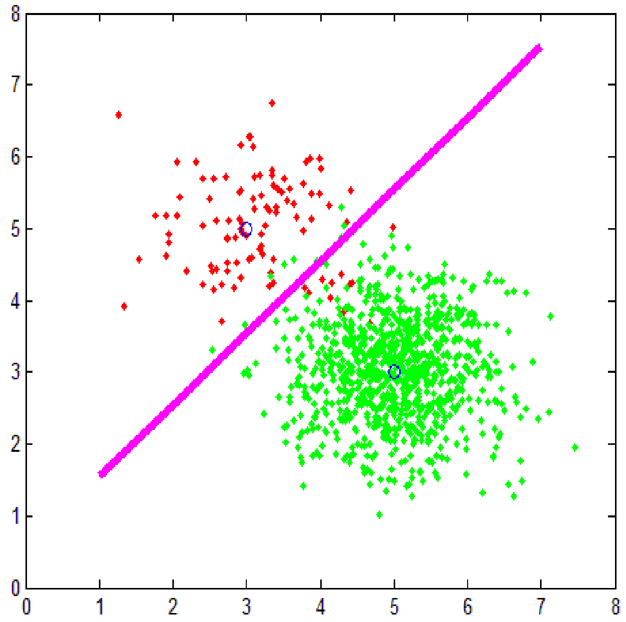
正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

决策面方程： $W^T(x-x_0)=0$ 　其中，$W=\mu_i-\mu_j$， $x_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|}\ln\frac{P(\omega_i)}{P(\omega_j)}$

① 若先验概率相等 $P(\omega_i)=P(\omega_j)$，则 $x_0=\frac{1}{2}(\mu_i+\mu_j)$

超平面通过 μ_i 与 μ_j 连线的中点 x_0，且与连线正交。



② 若先验概率不相等 $P(\omega_i)\neq P(\omega_j)$，则 x_0 不在中点，超平面向先验概率小的方向移动。

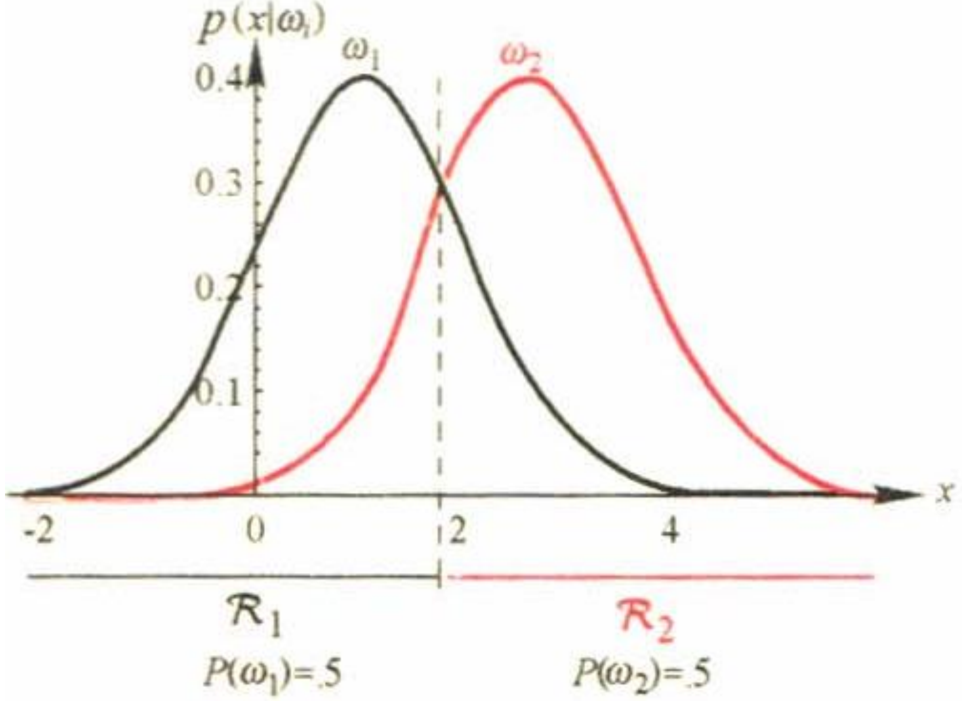
$$if \quad P(\omega_i)>P(\omega_j),\ \ln\frac{P(\omega_i)}{P(\omega_j)}>0,\ x_0<\frac{1}{2}(\mu_i+\mu_j)$$

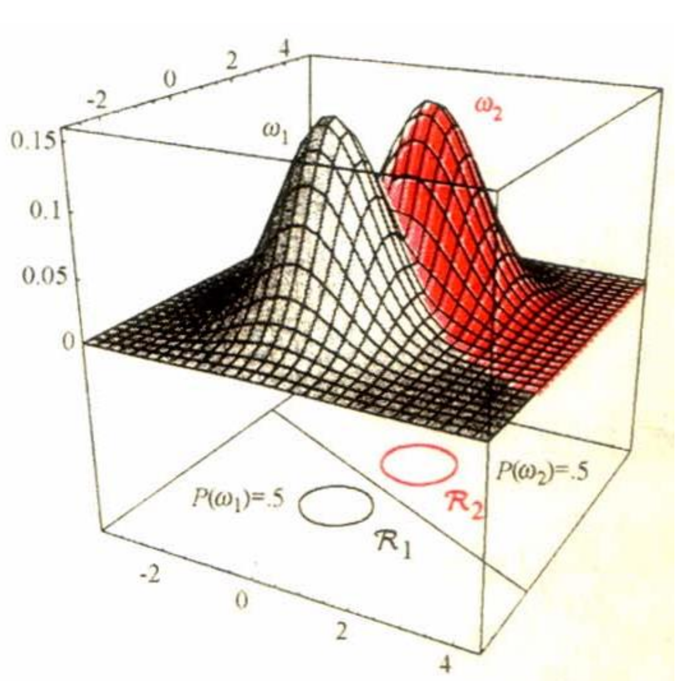
$$if \quad P(\omega_i)<P(\omega_j),\ \ln\frac{P(\omega_i)}{P(\omega_j)}<0,\ x_0>\frac{1}{2}(\mu_i+\mu_j)$$

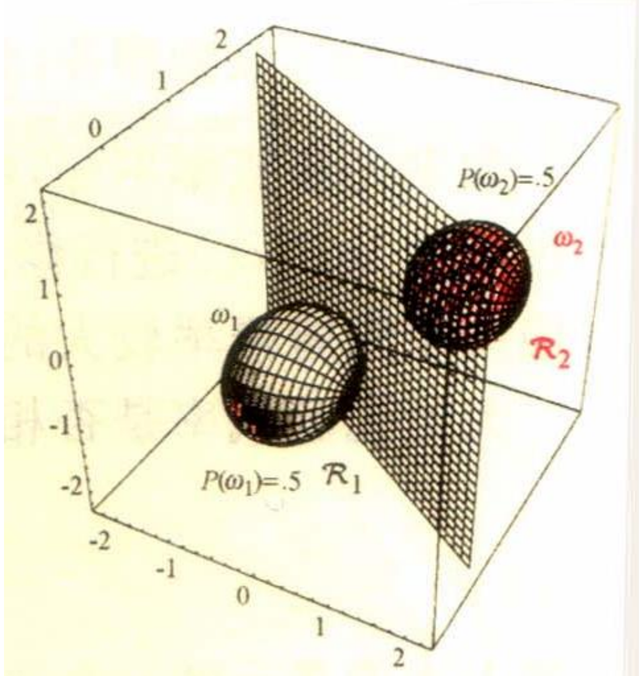
正态分布概率模型下的最小错误率贝叶斯决策

(1) $\Sigma_i=\sigma^2 I$

决策面方程：$W^T(x-x_0)=0$ 其中，$W=\mu_i-\mu_j$，$x_0=\frac{1}{2}(\mu_i+\mu_j)-\frac{\sigma^2(\mu_i-\mu_j)}{\|\mu_i-\mu_j\|}\ln\frac{P(\omega_i)}{P(\omega_j)}$







正态分布概率模型下的最小错误率贝叶斯决策

(2) $\Sigma_i=\Sigma$

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

各类的协方差矩阵相等$\Sigma_1=\Sigma_2=\cdots=\Sigma_c=\Sigma$。几何上相当于各类样本集中于以该类均值$\mu_i$点为中心的同样大小和形状的超椭球体中。

判别函数：$$\boldsymbol{g}_i(\boldsymbol{x})=-\frac{1}{2}(\boldsymbol{x}-\mu_i)^T\Sigma^{-1}(\boldsymbol{x}-\mu_i)+\ln \boldsymbol{P}(\omega_i)$$

若c类先验概率相等，则 $\ln P(\omega_1)=\ln P(\omega_2)=\cdots=\ln P(\omega_c)$ ，此时

$$g_i(x)=\underline{(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)}=r^2$$

马氏距离平方

Bayes决策

计算x到每类均值点μ_i的马氏距离平方r^2, 将x分到距离最近的类中去，或归于r^2最小的类。

正态分布概率模型下的最小错误率贝叶斯决策

(2) $\Sigma_i = \Sigma$　　$g_i(x) = (x-\mu_i)^T \Sigma^{-1}(x-\mu_i) = r^2$

马氏距离：印度统计学家马哈拉诺比斯提出表示数据的协方差距离，是两个服从同一分布并且其协方差矩阵为Σ的随机变量的差异程度。

*与欧式距离不同的是，马氏距离考虑到各种特性之间的联系，即独立于测量尺度。

- 如果协方差矩阵为单位矩阵，马氏距离简化为欧氏距离（m=2）

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \qquad r^2(X,G) = \begin{bmatrix} X_1-\mu_1 & X_2-\mu_2 \end{bmatrix} \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/\sigma^2 \end{bmatrix} \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix} = \frac{1}{\sigma^2}\left[(X_1-\mu_1)^2 + (X_2-\mu_2)^2\right]$$

- 如果协方差矩阵为对角矩阵，马氏距离则为正规化的欧式距离

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \qquad r^2(X,G) = \begin{bmatrix} X_1-\mu_1 & X_2-\mu_2 \end{bmatrix} \begin{bmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{bmatrix} \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix} = \frac{(X_1-\mu_1)^2}{\sigma_1^2} + \frac{(X_2-\mu_2)^2}{\sigma_2^2}$$

13. 4 正态分布的统计决策

正态分布概率模型下的最小错误率贝叶斯决策

(2) $\Sigma_i = \Sigma$

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

决策面方程：$g_i(x) - g_j(x) = 0$

$$g_i(x) - g_j(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + \frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j) + \ln P(\omega_i) - \ln P(\omega_j) = 0$$

$$\underline{(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) - (x-\mu_j)^T\Sigma^{-1}(x-\mu_j)} - 2\left(\ln P(\omega_i) - \ln P(\omega_j)\right) = 0 \qquad (*)$$

$(A+B)^T = A^T + B^T$

$(A+B)C = AC + BC$

$$(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) = (x^T-\mu_i^T)\Sigma^{-1}(x-\mu_i) = \left(x^T\Sigma^{-1} - \mu_i^T\Sigma^{-1}\right)(x-\mu_i)$$

$$= x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_i - \mu_i^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i$$

$$\cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_i - \mu_i^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i - \left(\cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_j - \mu_j^T\Sigma^{-1}x + \mu_j^T\Sigma^{-1}\mu_j\right)$$

$$= -x^T\Sigma^{-1}\mu_i + x^T\Sigma^{-1}\mu_j - \mu_i^T\Sigma^{-1}x + \mu_j^T\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i - \mu_j^T\Sigma^{-1}\mu_j$$

$(AB)^T = B^TA^T$

$$= \underline{x^T\Sigma^{-1}\left(\mu_j - \mu_i\right)} + \left(\mu_j^T - \mu_i^T\right)\Sigma^{-1}x + \mu_i^T\Sigma^{-1}\mu_i - \mu_j^T\Sigma^{-1}\mu_j$$

Σ对称阵　$\Sigma^{-1} = \left(\Sigma^{-1}\right)^T$

$\left(\mu_j - \mu_i\right)^T\Sigma^{-1}x$　数值

代入（*）式

$g_i(x) - g_j(x) = 0$

$$2\left(\mu_i - \mu_j\right)^T\Sigma^{-1}x + \left(\mu_j^T\Sigma^{-1}\mu_j - \mu_i^T\Sigma^{-1}\mu_i\right) + 2\left(\ln P(\omega_i) - \ln P(\omega_j)\right) = 0$$

正态分布概率模型下的最小错误率贝叶斯决策

(2) $\Sigma_i = \Sigma$

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

决策面方程：$g_i(x) - g_j(x) = 0$

$$2\left(\mu_i-\mu_j\right)^T \Sigma^{-1} x + \left(\mu_j^T \Sigma^{-1}\mu_j - \mu_i^T \Sigma^{-1}\mu_i\right) + 2\left(\ln P(\omega_i) - \ln P(\omega_j)\right) = 0$$

$$2\left(\mu_i-\mu_j\right)^T \Sigma^{-1} x + 2\left(\mu_i-\mu_j\right)^T \Sigma^{-1} \frac{1}{2\left(\mu_i-\mu_j\right)^T \Sigma^{-1}}\left(\mu_j^T \Sigma^{-1}\mu_j - \mu_i^T \Sigma^{-1}\mu_i\right)$$

$$+2\left(\mu_i-\mu_j\right)^T \Sigma^{-1} \frac{2}{2\left(\mu_i-\mu_j\right)^T \Sigma^{-1}}\left(\ln P(\omega_i) - \ln P(\omega_j)\right) = 0$$

$$2\left(\mu_i-\mu_j\right)^T \Sigma^{-1}\left[x + \frac{\left(\mu_j^T \Sigma^{-1}\mu_j - \mu_i^T \Sigma^{-1}\mu_i\right)}{2\left(\mu_i-\mu_j\right)^T \Sigma^{-1}} + \frac{2}{2\left(\mu_i-\mu_j\right)^T \Sigma^{-1}}\ln\frac{P(\omega_i)}{P(\omega_j)}\right] = 0$$

$$\boxed{W^T(x-x_0) = 0}$$

其中，$W = \Sigma^{-1}(\mu_i - \mu_j)$

$$x_0 = \frac{1}{2}(\mu_i+\mu_j) - \frac{\ln[p(\omega_i)/p(\omega_j)]}{(\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)}(\mu_i-\mu_j)$$

正态分布概率模型下的最小错误率贝叶斯决策

(2) $\Sigma_i = \Sigma$　　　决策面方程：$\boxed{W^T(x - x_0) = 0}$

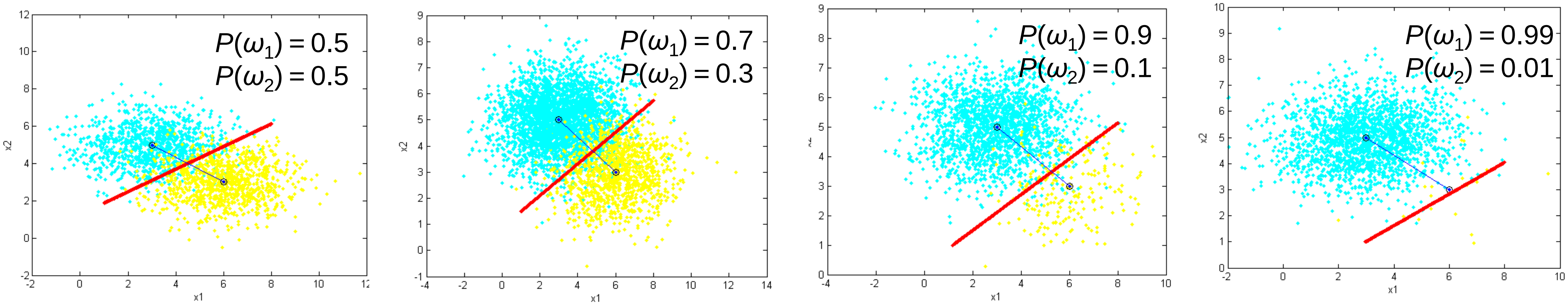
$$\text{其中，} W = \Sigma^{-1}(\mu_i - \mu_j) \qquad x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln[p(\omega_i)/p(\omega_j)]}{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$$

- W 通常不在$(\mu_i$-$\mu_j)$方向上， x-x_0表示决策面通过x_0点。
- W 与 $(x$-$x_0)$的点积为 0，表示W与$(x$-$x_0)$正交，决策面通过 x_0 点但不与均值向量连线 $(\mu_i$-$\mu_j)$ 正交。

正态分布概率模型下的最小错误率贝叶斯决策

(2) $\Sigma_i = \Sigma$　　　　**决策面方程：** $\boxed{W^T(x - x_0) = 0}$

其中，$W = \Sigma^{-1}(\mu_i - \mu_j)$　　　　$x_0 = \dfrac{1}{2}(\mu_i + \mu_j) - \dfrac{\ln[p(\omega_i)/p(\omega_j)]}{(\mu_i - \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$



- 若先验概率相等，则交点在均值向量连线的中点上；
- 若先验概率不相等则向先验概率小的均值点偏移。
- 若先验概率相差较大，判别边界不会落入球状高斯分布的中心点之间。

正态分布概率模型下的最小错误率贝叶斯决策

例 3 已知两类问题，协方差 Σ 相同，均值向量不同。先验概率相等， $\mu_1=[0\quad 0]^T$， $\mu_2=[3\quad 3]^T$，

$\Sigma_1=\Sigma_2=\begin{bmatrix}1.1 & 0.3\\ 0.3 & 1.9\end{bmatrix}$， 要求根据Bayes决策，对样本$x$=[1.0, 2.2]^{T}分类。

*

解:分别计算对两个均值向量的马氏距离，得

$$\Sigma^{-1}=\begin{bmatrix}0.95 & -0.15\\ -0.15 & 0.55\end{bmatrix}$$

$$r^2(\mu_1,x)=(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)=[1.0\quad 2.2]\begin{bmatrix}0.95 & -0.15\\ -0.15 & 0.55\end{bmatrix}\begin{bmatrix}1.0\\ 2.2\end{bmatrix}=2.952$$

$$r^2(\mu_2,x)=(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)=[-2.0\quad -0.8]\begin{bmatrix}0.95 & -0.15\\ -0.15 & 0.55\end{bmatrix}\begin{bmatrix}-2.0\\ -0.8\end{bmatrix}=3.672$$

样本x应属于距离近的类， x＝[1.0，2.2]^{T}属于第一类。

正态分布概率模型下的最小错误率贝叶斯决策

(3) $\Sigma_i \neq \Sigma_j$， $i, j = 1, 2, \cdots, c$

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

去掉与类别的无关项

判别函数：
$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$
$$= \boxed{x^T W_i x + w_i^T x + w_{i0}}$$

这是x的一个非线性二次形式

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

若决策域 Ri 和 R*j* 毗邻，则决策面方程为：

$$g_i(x) - g_j(x) = 0$$

$$x^T(W_i - W_j)x + (w_i - w_j)^T x + w_{i0} - w_{j0} = 0$$

决策面为超二次曲面：超球面、超椭球面、超抛物面、超双曲面、超平面

二元正态情况下，其决策面是二次曲线(椭圆、双曲线、抛物线、一对直线)。此时，Bayes分类器是二次曲线分类器。

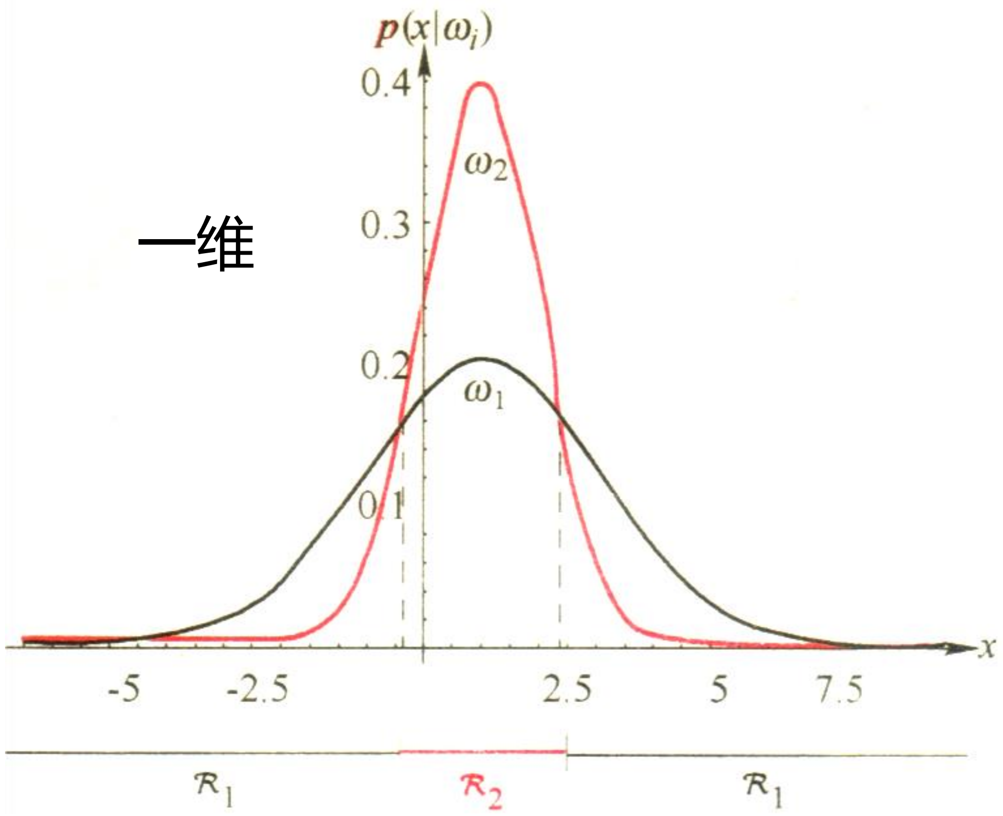
13.4 正态分布的统计决策

了解

正态分布概率模型下的最小错误率贝叶斯决策

(3) $\Sigma_i \neq \Sigma_j$， $i, j = 1, 2, \cdots, c$

一维



二维

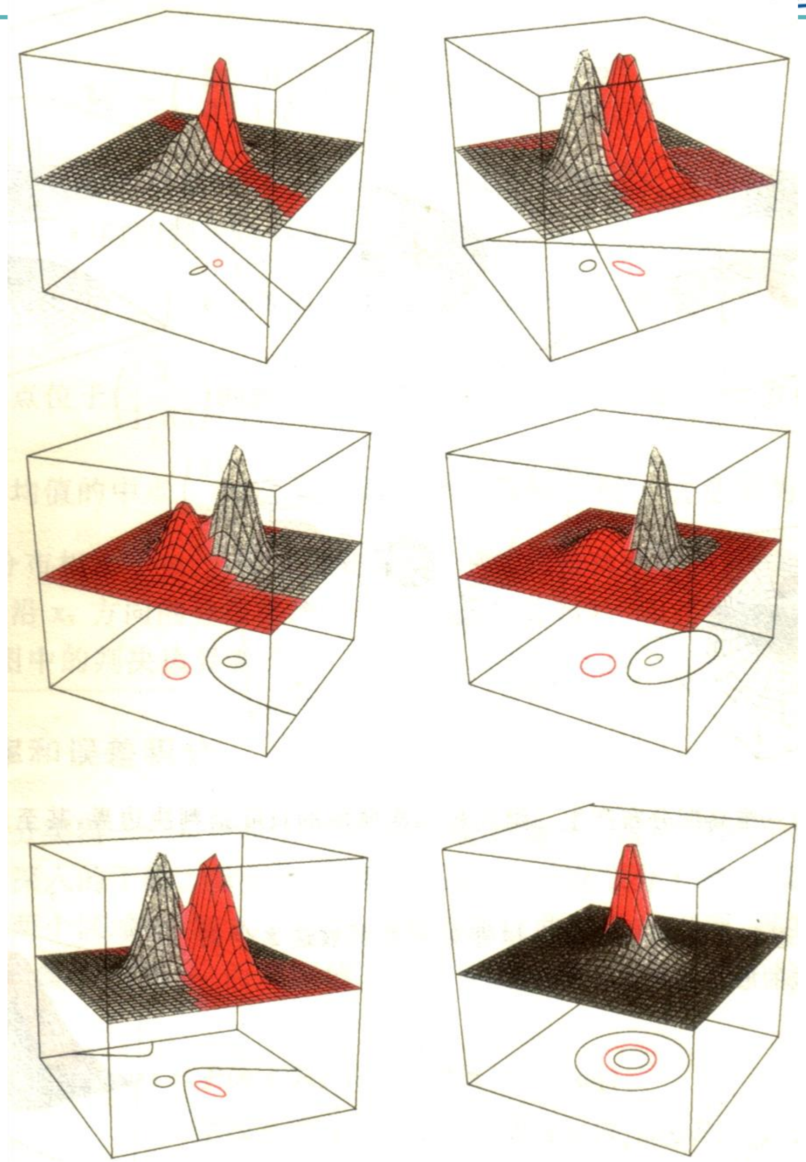


图 2-14 任意高斯分布导致一般超二次曲面的贝叶斯判决边界。反之，给定任意超二次曲面，就能求出两个高斯分布，其贝叶斯判决边界就是该超二次曲面。它们的方差由常概率密度的围线表示

了解

正态分布概率模型下的最小错误率贝叶斯决策

(3) $\Sigma_i \neq \Sigma_j$， $i, j = 1, 2, \cdots, c$

三维

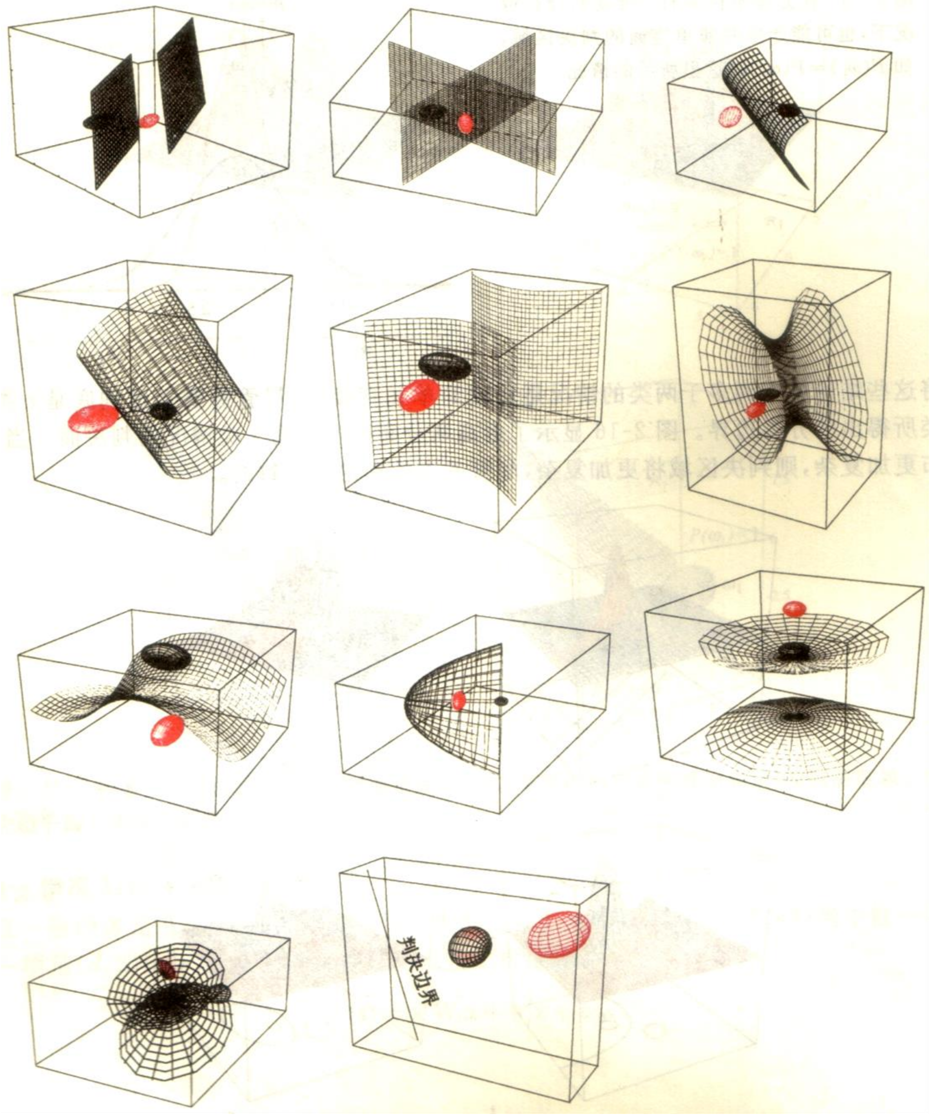


图 2-15 任意的三维高斯分布产生二维的超二次曲面的贝叶斯判决边界，甚至还有退化为单一直线的判决边界

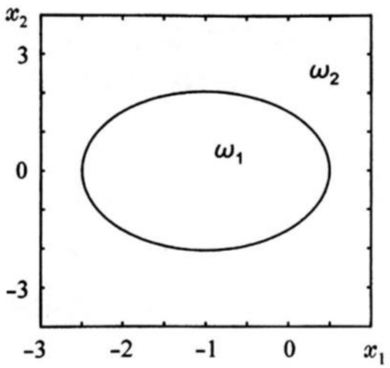
正态分布概率模型下的最小错误率贝叶斯决策

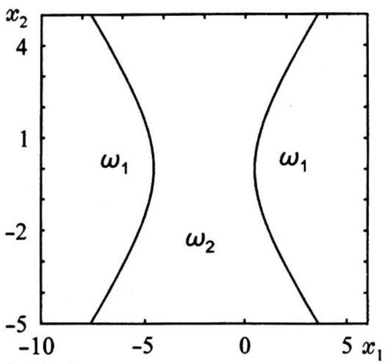
(3) $\Sigma_i \neq \Sigma_j$，$i, j = 1, 2, \cdots, c$

例 4 $P(\omega_1)=P(\omega_2)$, $\mu_1=[0,0]^T$ 和 $\mu_2=[1,0]^T$ Σ不同时的决策面。

$$\Sigma_1=\begin{bmatrix}0.1 & 0.0\\ 0.0 & 0.15\end{bmatrix},\ \Sigma_2=\begin{bmatrix}0.2 & 0.0\\ 0.0 & 0.25\end{bmatrix}$$ 左图

$$\Sigma_1=\begin{bmatrix}0.1 & 0.0\\ 0.0 & 0.15\end{bmatrix},\ \Sigma_2=\begin{bmatrix}0.15 & 0.0\\ 0.0 & 0.1\end{bmatrix}$$ 右图





- 根据决策面方程得到的二次曲线分别是椭圆和双曲线。

两类二维正态分布的决策面，由协方差Σ和均值向量μ根据判别函数和决策面方程计算决策面。

(1)判别函数

$$g_i(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x}-\mu_i)^T \Sigma_i^{-1}(\boldsymbol{x}-\mu_i) - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$
$$= \boldsymbol{x}^T \boldsymbol{W}_i \boldsymbol{x} + \boldsymbol{w}_i^T \boldsymbol{x} + w_{i0}$$

其中 $\boldsymbol{W}_i = -\frac{1}{2}\Sigma_i^{-1}$ $(\boldsymbol{d}\times\boldsymbol{d})$

$$\boldsymbol{w}_i = \Sigma_i^{-1}\mu_i$$

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

(2)决策面方程

$$g_i(\boldsymbol{x}) - g_j(\boldsymbol{x}) = 0$$

$$\boldsymbol{x}^T(\boldsymbol{W}_i - \boldsymbol{W}_j)\boldsymbol{x} + (\boldsymbol{w}_i - \boldsymbol{w}_j)^T\boldsymbol{x} + w_{i0} - w_{j0} = 0$$

13.5 关于分类器的错误率

- 在类条件概率密度及先验概率已知的条件下进行分类决策，其错误率是确定的。
- 在分类器设计完成之后，通常以错误率的大小来衡量其性能的优劣。
- 错误率是模式识别的重要参数。

最小错误率贝叶斯决策的错误率 $$P(e)=P(\omega_1)\int_{\Gamma_2} p(\vec{x}\mid\omega_1)d\vec{x}+P(\omega_2)\int_{\Gamma_1} p(\vec{x}\mid\omega_2)d\vec{x}$$

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

最小错误率贝叶斯决策可保证决策的错误率在统计意义上是最小的，但分类器性能比较需知道大小？

✓ x 为多维向量时，要进行多重积分，当概率密度表达式复杂式，难于计算，并且，积分范围中判决区域的不连续也导致直接的计算困难。

错误率计算途径：

（1）理论计算

（2）计算错误率上界

（3）实验估计

只能在某些特殊情况下才能实现错误率的理论计算：

- 正态分布且等协方差阵（$\Sigma_i = \Sigma_j = \Sigma$)
- 随机变量各分量独立，且维数足够大

正态模式分类的错误率

考虑两类问题，设两类模式为协方差阵相等的多变量正态分布，它们的密度函数分别为

$$p(x|\omega_i) \sim N(\mu_i, \Sigma) \qquad p(x|\omega_j) \sim N(\mu_j, \Sigma)$$

$$p(x|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}\exp[-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)]$$

对数似然比

$$L_{ij}(x) \triangleq \ln l_{ij}(x) = \ln p(x|\omega_i) - \ln p(x|\omega_j)$$

$$= -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + \frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)$$

$$= x^T\Sigma^{-1}(\mu_i-\mu_j) - \frac{1}{2}(\mu_i+\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)$$

$L_{ij}(x)$ 是 x 的线性函数，而 x 的各分量是正态分布的，故 $L_{ij}(x)$ 是正态分布的随机变量。

正态模式分类的错误率

$$\boxed{L_{ij}(x)=x^T\Sigma^{-1}(\mu_i-\mu_j)-\frac{1}{2}(\mu_i+\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)}$$

如果 $x\in\omega_i$，有$E\left(x\,|\,\omega_i\right)=\mu_i$

$$E_i\left[L_{ij}\right]=\mu_i^T\Sigma^{-1}(\mu_i-\mu_j)-\frac{1}{2}(\mu_i+\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)=\frac{1}{2}(\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)$$

令 $r_{ij}^{\ 2}=(\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)$ $\qquad E_i\left[L_{ij}\right]\triangleq\bar{L}_{ij}=\frac{1}{2}r_{ij}^{\ 2}$

$$\begin{aligned}Var_i\left[L_{ij}\right]&=E_i\left[(L_{ij}-E_i\left(L_{ij}\right))^2\right]=E_i\left[((x-\mu_i)^T\Sigma^{-1}(\mu_i-\mu_j))^2\right]\\&=E_i\left[(\mu_i-\mu_j)^T\Sigma^{-1}\underline{(x-\mu_i)(x-\mu_i)^T}\Sigma^{-1}(\mu_i-\mu_j)\right]\\&=(\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)\\&=r_{ij}^{\ 2}\end{aligned}$$

Σ

正态模式分类的错误率

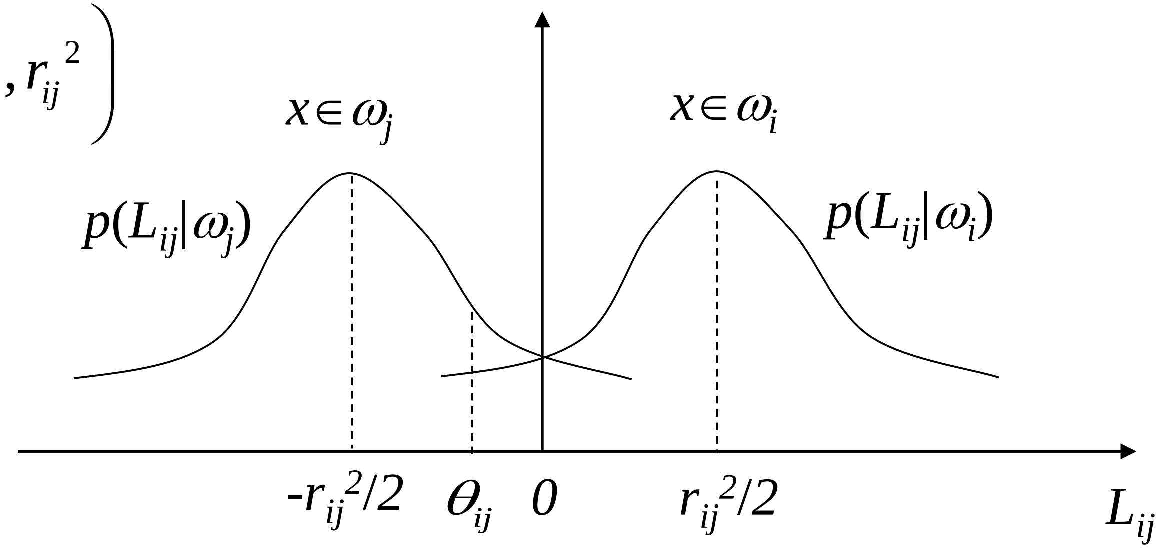
$$L_{ij}(x) = x^T \Sigma^{-1}(\mu_i - \mu_j) - \frac{1}{2}(\mu_i + \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)$$

$$E_i[L_{ij}] \triangleq \bar{L}_{ij} = \frac{1}{2} r_{ij}^2 \qquad Var_i[L_{ij}] = E_i\left[(L_{ij} - \bar{L}_{ij})^2\right] = r_{ij}^2$$

$L_{ij}(x|\omega_i)$ 是服从正态分布 $N\left(\frac{1}{2}r_{ij}^2, r_{ij}^2\right)$ 的一维随机变量。

$$L_{ij}(x|\omega_i) \sim N\left(\frac{1}{2}r_{ij}^2, r_{ij}^2\right) \qquad L_{ij}(x|\omega_j) \sim N\left(-\frac{1}{2}r_{ij}^2, r_{ij}^2\right)$$

对应判决门限 $\theta_{ij} = \ln \dfrac{\left[\lambda(\alpha_i,\omega_j) - \lambda(\alpha_j,\omega_j)\right] P(\omega_j)}{\left[\lambda(\alpha_j,\omega_i) - \lambda(\alpha_i,\omega_i)\right] P(\omega_i)}$



正态模式分类的错误率

$$L_{ij}(x|\omega_i) \sim N\left(\frac{1}{2}r_{ij}^{2}, r_{ij}^{2}\right) \qquad \Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})dy \qquad p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

将属于 ω_i 类的模式误判为属于 ω_j 类的错误概率为

$$P(L_{ij} < \theta_{ij} | \omega_i) = \int_{-\infty}^{\theta_{ij}} \frac{1}{\sqrt{2\pi}\, r_{ij}} \exp\left[-\frac{(L_{ij} - r_{ij}^{2}/2)^2}{2r_{ij}^{2}}\right] dL_{ij} = \Phi(\frac{\theta_{ij} - r_{ij}^2/2}{r_{ij}})$$

将属于 ω_j 类的模式误判为属于 ω_i 类的错误概率为

$$P(L_{ij} > \theta_{ij} | \omega_j) = \int_{\theta_{ij}}^{\infty} \frac{1}{\sqrt{2\pi}\, r_{ij}} \exp\left[-\frac{(L_{ij} + r_{ij}^{2}/2)^2}{2r_{ij}^{2}}\right] dL_{ij} = 1 - \Phi(\frac{\theta_{ij} + r_{ij}^2/2}{r_{ij}})$$

总的误判概率为

$$P(e) = P(\omega_i)P(L_{ij} < \theta_{ij} | \omega_i) + P(\omega_j)P(L_{ij} > \theta_{ij} | \omega_j) = P(\omega_i)\Phi\left[\frac{\theta_{ij} - \frac{1}{2}r_{ij}^{2}}{r_{ij}}\right] + P(\omega_j)\left\{1 - \Phi\left[\frac{\theta_{ij} + \frac{1}{2}r_{ij}^{2}}{r_{ij}}\right]\right\}$$

13.5 关于分类器的错误率

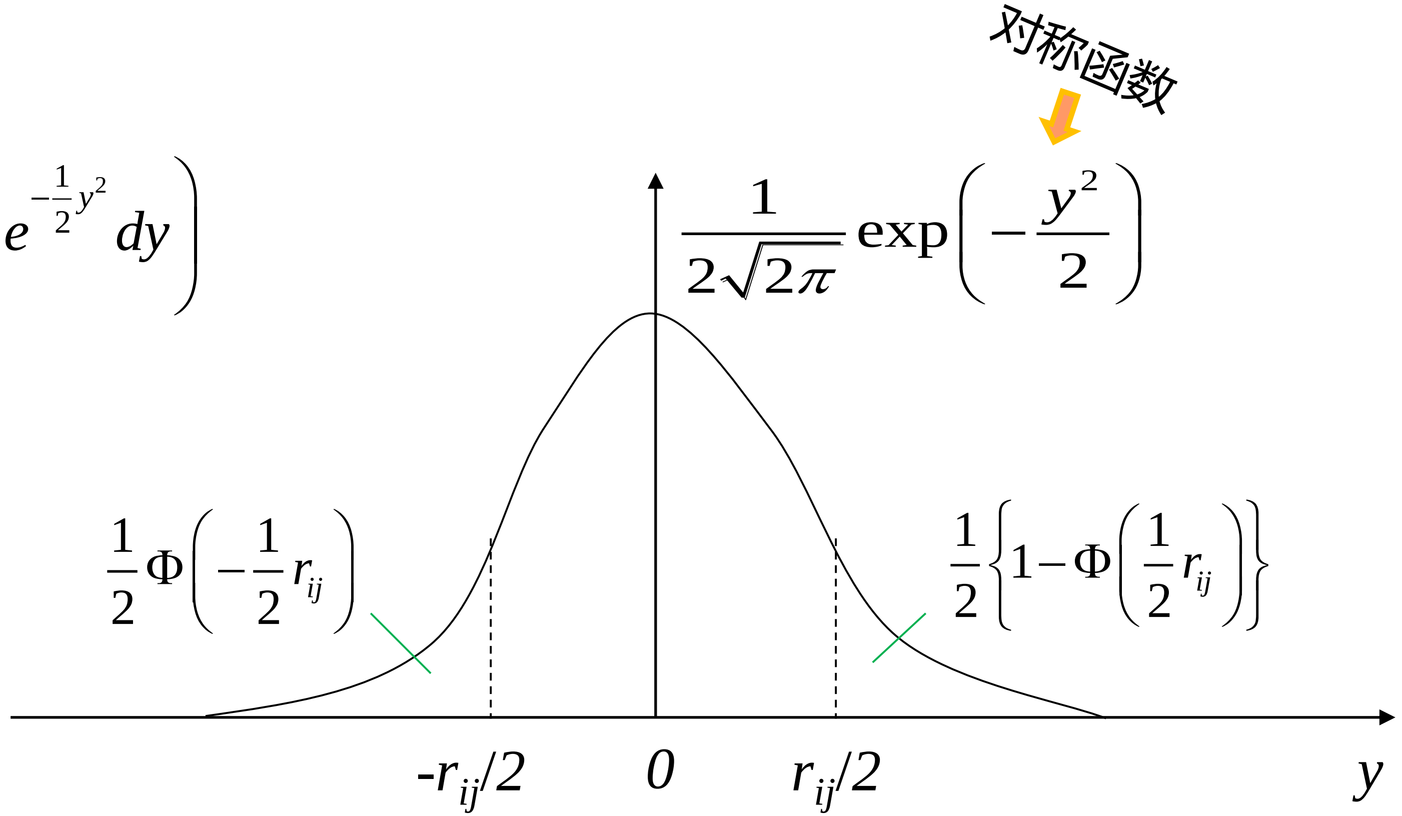
了解

正态模式分类的错误率

如果 $P(\omega_i)=P(\omega_j)=\dfrac{1}{2}$ ，0－1 损失函数，此时 $\theta_{ij}=0$ 。

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

$$=\frac{1}{2}\int_{-\infty}^{-r_{ij}/2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy+\frac{1}{2}\left(1-\int_{-\infty}^{r_{ij}/2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy\right)$$



13.5 关于分类器的错误率

了解

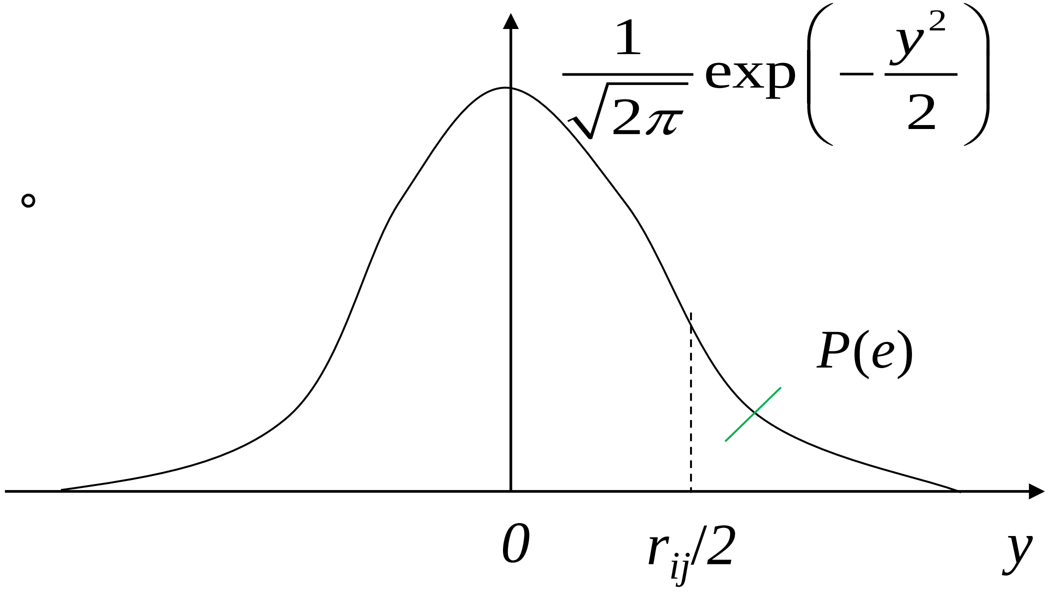
正态模式分类的错误率

如果 $P(\omega_i)=P(\omega_j)=\dfrac{1}{2}$ ，0－1 损失函数，此时 $\theta_{ij}=0$ 。

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

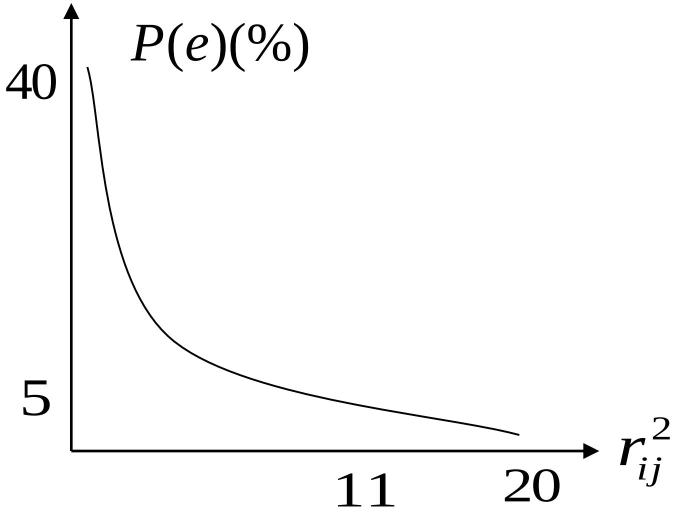
$$=\frac{1}{2}\int_{-\infty}^{-r_{ij}/2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy+\frac{1}{2}\left(1-\int_{-\infty}^{r_{ij}/2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy\right)$$

$$=\int_{r_{ij}/2}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{y^2}{2}\right]dy$$



上式表明了误判概率与两类的马氏距离的关系：

$P(e)$ 随 r_{ij}^2 的增大而单调递减，只要两类马氏距离足够大，其误判概率可足够小。



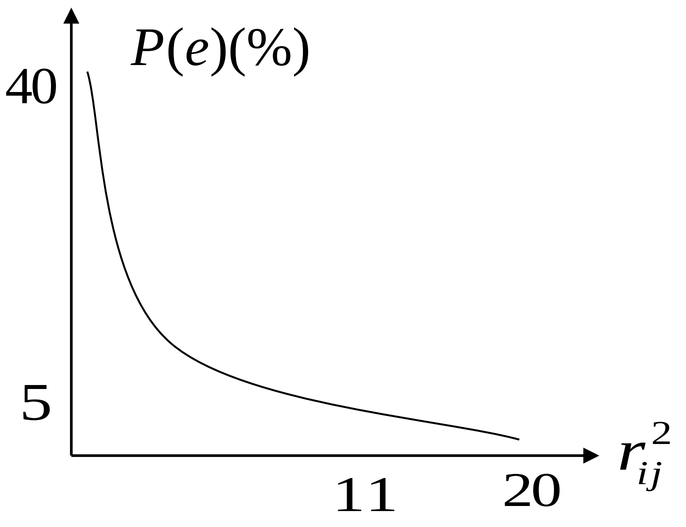
正态模式分类的错误率

如果 $P(\omega_i)=P(\omega_j)=\frac{1}{2}$，0 - 1 损失函数，此时 $\theta_{ij}=0$。

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

$$=\frac{1}{2}\int_{-\infty}^{-r_{ij}/2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy+\frac{1}{2}\left(1-\int_{-\infty}^{r_{ij}/2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy\right)$$

$$=\int_{r_{ij}/2}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{y^2}{2}\right]dy$$



上式表明了误判概率与两类的马氏距离的关系：

$P(e)$ 随 r_{ij}^2 的增大而单调递减，只要两类马氏距离足够大，其误判概率可足够小。

- 如果协方差阵Σ是对角阵

$$r_{ij}^2=(\mu_i-\mu_j)^T\Sigma^{-1}(\mu_i-\mu_j)$$

$$=\sum_{k=1}^{d}\left(\frac{\mu_{ik}-\mu_{jk}}{\sigma_k}\right)^2$$

此时，马氏距离等于欧氏距离。

独立随机变量的错误率

由于 x 各分量独立，其概率密度为各分量的边缘概率密度的乘积，通过采用负对数似然比决策，使得似然比 $\theta(X)$ 为各随机变量 $\theta(x_i)$ 之和,当维数足够大数，由中心极限定理，随机变量 $\theta(X)$ 服从正态分布，从而错误率计算转化为对一维正态分布随机变量 $\theta(X)$ 的的积分计算。

d维随机变量x的分量相互独立，则其密度函数可表示为：

$$p(X\,|\,\omega_i)=\prod_{l=1}^{d}p(x_l\,|\,\omega_i)$$

i:类别　l:维度

负对数似然比为：

$$\theta_{ij}(X)=\sum_{l=1}^{d}\theta_{ij}(x_l),\text{其中，}\quad \theta_{ij}(x_l)=-\ln\frac{p(x_l\,|\,\omega_i)}{p(x_l\,|\,\omega_j)}$$

独立随机变量的错误率

d足够大，$\theta_{ij}(\mathrm{x}) \rightarrow$ 正态分布（中心极限定理）

计算出$\theta_{ij}(\mathrm{x}|\omega_i)$的均值$\mu_{\mathrm{i}}$和方差$\sigma_i$，采用正态分布下的错误率计算公式

$$\mu_{\mathrm{i}} = E\left[\sum_{l=1}^{d}\theta_{ij}(x_l)\,|\,\omega_i\right] = \sum_{l=1}^{d}E\left[\theta_{ij}(x_l)\,|\,\omega_i\right] = \sum_{l=1}^{d}\mu_{il}$$

i:类别 l:维度

$$\sigma^2{}_i = E\left[\left(\sum_{l=1}^{d}\theta_{ij}(x_l)-\eta_i\right)^2 |\,\omega_i\right] = \sum_{l=1}^{d}E\left[\left(\theta_{ij}(x_l)-\eta_i\right)^2 |\,\omega_i\right] = \sum_{l=1}^{d}\sigma^2{}_{il}$$

同理,计算出$\theta_{ij}(\mathrm{x}|\omega_j)$的均值$\mu_j$和方差$\sigma_j$

$$\mu_j = E\left[\sum_{l=1}^{d}\theta_{ij}(x_l)\,|\,\omega_j\right] = \sum_{l=1}^{d}\mu_{jl} \quad , \qquad \sigma^2{}_j = E\left[\left(\sum_{l=1}^{d}\theta_{ij}(x_l)-\eta_i\right)^2 |\,\omega_j\right] = \sum_{l=1}^{d}\sigma^2{}_{jl}$$

独立随机变量的错误率

将属于 ω_i 类的模式误判为属于 ω_j 类的错误概率为

$$P(L_{ij}<\theta_{ij}|\omega_i)=\int_{-\infty}^{\theta_{ij}}\frac{1}{\sqrt{2\pi}\ \sigma_i}\exp\left[-\frac{(L_{ij}-\mu_i)^2}{2\sigma_i^2}\right]dL_{ij}=\Phi(\frac{\theta_{ij}-\mu_i}{\sigma_i})$$

将属于 ω_j 类的模式误判为属于 ω_i 类的错误概率为

$$P(L_{ij}>\theta_{ij}|\omega_j)=\int_{\theta_{ij}}^{\infty}\frac{1}{\sqrt{2\pi}\ \sigma_j}\exp\left[-\frac{(L_{ij}-\mu_j)^2}{2\sigma_j^{\ 2}}\right]dL_{ij}=1-\Phi(\frac{\theta_{ij}-\mu_j}{\sigma_j})$$

于是，总的误判概率为

$$P(e)=P(\omega_i)P(L_{ij}<\theta_{ij}|\omega_i)+P(\omega_j)P(L_{ij}>\theta_{ij}|\omega_j)$$

$$=P(\omega_i)\Phi\left[\frac{\theta_{ij}-\mu_i}{\sigma_i}\right]+P(\omega_j)\left\{1-\Phi\left[\frac{\theta_{ij}-\mu_j}{\sigma_j}\right]\right\}$$

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

第13章 基于统计决策的概率分类方法

13.6 离散概率模型下的统计决策

自习

到目前为止所讨论的特征向量 x 可以为 d 维欧氏空间中的任意一点。但是，在许多实际应用中，x 中的元素可能是二进制，三进制或者更高的离散整数值，以至于 x 可以被认为是 m 个离散值 $v_1,\cdots\cdots v_m$中的一个。

- 在这种情况下，$p(x|\omega_j)$ 变得奇异化，积分形式 $\int p(x|\omega_j)dx$ 转变为求和形式 $\sum_x P(x|\omega_j)$
- 概率密度函数 $p(\bullet)$ 换成 概率分布函数 $P(\bullet)$
- 其它方面与连续的情况基本相同，这里不一一赘述

自习

独立的二值特征

考虑两类问题，其中特征向量的元素为二值的，并且条件独立。

令 $x=(x_1,\ldots,x_d)'$，其中，x_i 可能为 0 或 1，且：

$$p_i=\mathrm{P}[x_i=1\mid\omega_1] \qquad q_i=\mathrm{P}[x_i=1\mid\omega_2]$$

假设条件独立，可将x元素的概率写为 $P(x\mid\omega_i)$，即

$$\begin{cases} P(x\mid\omega_1)=\prod_{i=1}^{d} p_i^{x_i}(1-p_i)^{1-x_i} \\ P(x\mid\omega_2)=\prod_{i=1}^{d} q_i^{x_i}(1-q_i)^{1-x_i} \end{cases}$$

$$\left.\begin{aligned} & p_1=\mathrm{P}[x_1=1\mid\omega_1],\ p_2=\mathrm{P}[x_2=1\mid\omega_1] \\ & P((1,1)\mid\omega_1)=p_1p_2 \\ & P((1,0)\mid\omega_1)=p_1(1-p_2) \\ & P((0,1)\mid\omega_1)=(1-p_1)p_2 \\ & P((0,0)\mid\omega_1)=(1-p_1)(1-p_2) \end{aligned}\right\}$$

那么似然比为

$$\boxed{\frac{P(x\mid\omega_1)}{P(x\mid\omega_2)}=\prod_{i=1}^{d}\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}}$$

13.6 离散概率模型下的统计决策

自习

独立的二值特征

$$\frac{P(x|\omega_1)}{P(x|\omega_2)}=\prod_{i=1}^{d}(\frac{p_i}{q_i})^{x_i}(\frac{1-p_i}{1-q_i})^{1-x_i}$$

基于最小错误率贝叶斯准则的判决函数 $g(x)=\ln\frac{p(x|\omega_1)}{p(x|\omega_2)}+\ln\frac{P(\omega_1)}{P(\omega_2)}$

$$g(x)=\sum_{i=1}^{d}[x_i\ln\frac{p_i}{q_i}+(1-x_i)\ln\frac{1-p_i}{1-q_i}]+\ln\frac{P(\omega_1)}{P(\omega_2)}$$

注意判决函数对x_i是线性的，可改写为

$$g(x)=\sum_{i=1}^{d}w_i x_i+w_0$$

其中，$w_i=\ln\frac{p_i(1-q_i)}{q_i(1-p_i)}$ $\quad i=1,\ldots,d \quad$ $w_0=\sum_{i=1}^{d}\ln\frac{1-p_i}{1-q_i}+\ln\frac{P(\omega_1)}{P(\omega_2)}$

若 $g(x)>0$判别为 ω_1，否则为 ω_2。

- $g(x)$可以看作是x的各分量的加权组合
- 特征独立的条件产生线性分类器，而如果特征不独立将产生复杂的分类器

自习

例8 三维二值特征的贝叶斯决策 $P(\omega_1)=P(\omega_2)=0.5$， $p_i=0.8, q_i=0.5$， $i=1,2,3$

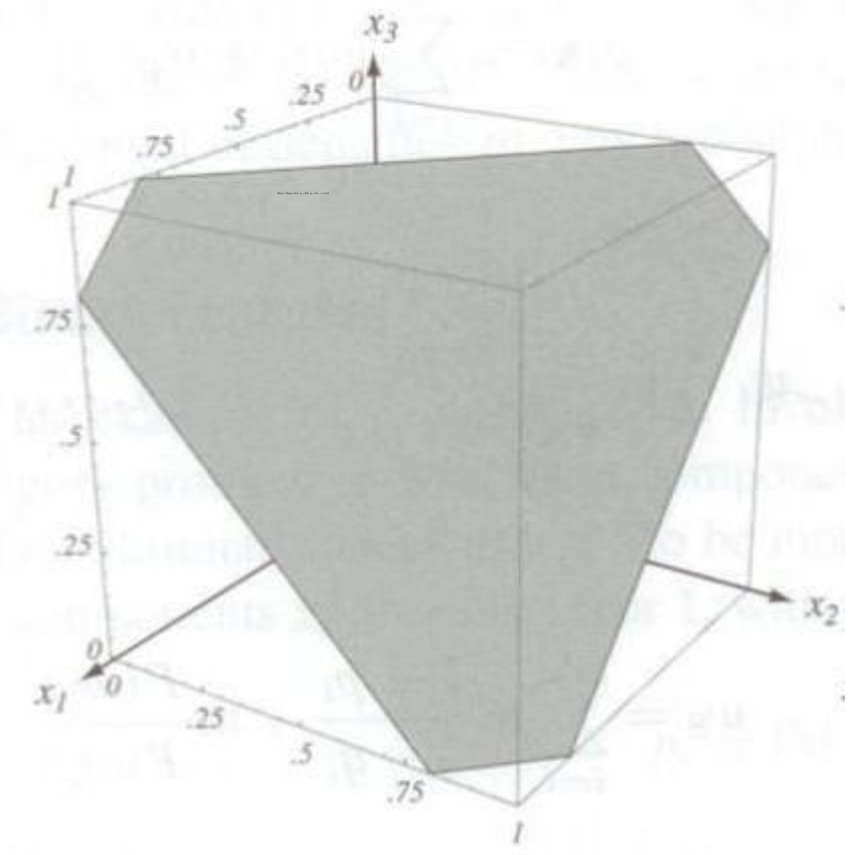
$$g(x)=\sum_{i=1}^{d} w_i x_i + w_0$$

$$w_i=\ln\frac{p_i(1-q_i)}{q_i(1-p_i)} \qquad i=1,\ldots,d$$

$$w_0=\sum_{i=1}^{d}\ln\frac{1-p_i}{1-q_i}+\ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$w_i=\ln\frac{0.8(1-0.5)}{0.5(1-0.8)}=1.3863$$

$$w_0=\sum_{i=1}^{3}\ln\frac{1-0.8}{1-0.5}+\ln\frac{0.5}{0.5}=-2.75$$



自习

例8 三维二值特征的贝叶斯决策 $P(\omega_1)=P(\omega_2)=0.5$，$p_i=0.8, q_i=0.5$，$i=1,2$，$p_3=q_3=0.5$

$$g(x)=\sum_{i=1}^{d} w_i x_i + w_0$$

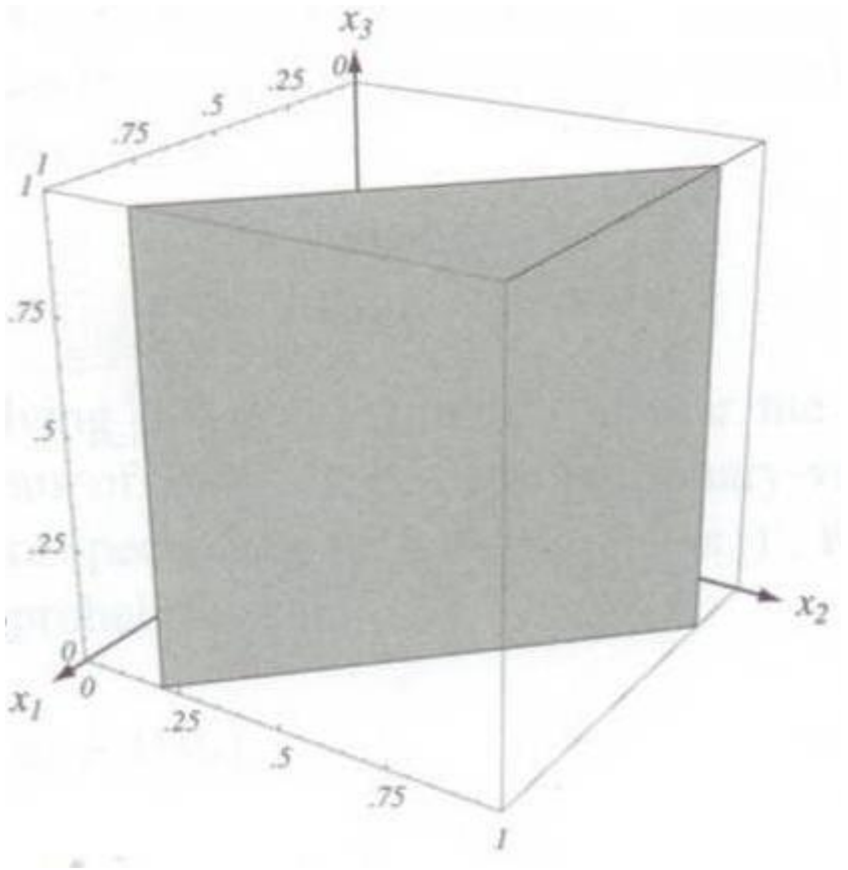
$$w_i=\ln\frac{p_i(1-q_i)}{q_i(1-p_i)} \qquad i=1,\dots,d$$

$$w_0=\sum_{i=1}^{d}\ln\frac{1-p_i}{1-q_i}+\ln\frac{P(\omega_1)}{P(\omega_2)}$$

$$w_i=\ln\frac{0.8(1-0.5)}{0.5(1-0.8)}=1.3863, \quad i=1,2$$

$$w_3=0$$

$$w_0=\sum_{i=1}^{2}\ln\frac{1-0.8}{1-0.5}+\ln\frac{0.5}{0.5}=-1.83$$



Ending

Next：14-概率密度函数的参数估计