工业和信息化部“十二五”规划教材
“十二五”国家重点图书出版规划项目

模式识别

Pattern Recognition

第5讲 统计分类器及其学习I

哈尔滨工业大学 计算机学院
模式识别与智能系统研究中心
ISBN 978-7-5603-4763-9

1 概率论与统计学习

2 贝叶斯决策理论

3 高斯分布的贝叶斯分类器

哈尔滨工业大学 计算机学院
模式识别与智能系统研究中心
ISBN 978-7-5603-4763-9

概率论与统计学习

判别式模型（Discriminative Model）

将$\mathbf{x}$看做特征空间中的点

构建判别函数 $g(\mathbf{x})$ 来决定 $\mathbf{x}$ 属于哪个类别

关键在于计算 $\mathbf{x}$ 与训练样本$\mathbf{x}_i$ 间的相对位置（内积）关系

产生式模型（Generative Model）

将$\mathbf{x}$看做随机变量

根据 $\mathbf{x}$ 属于各类别 ω_i 的概率大小 $P(\omega_i|\mathbf{x})$，来决定其类别

关键在于计算不同类别产生“待识别模式”的概率 $P(\mathbf{x}|\omega_i)$

ISBN 978-7-5603-4763-9

常用的概率表示形式

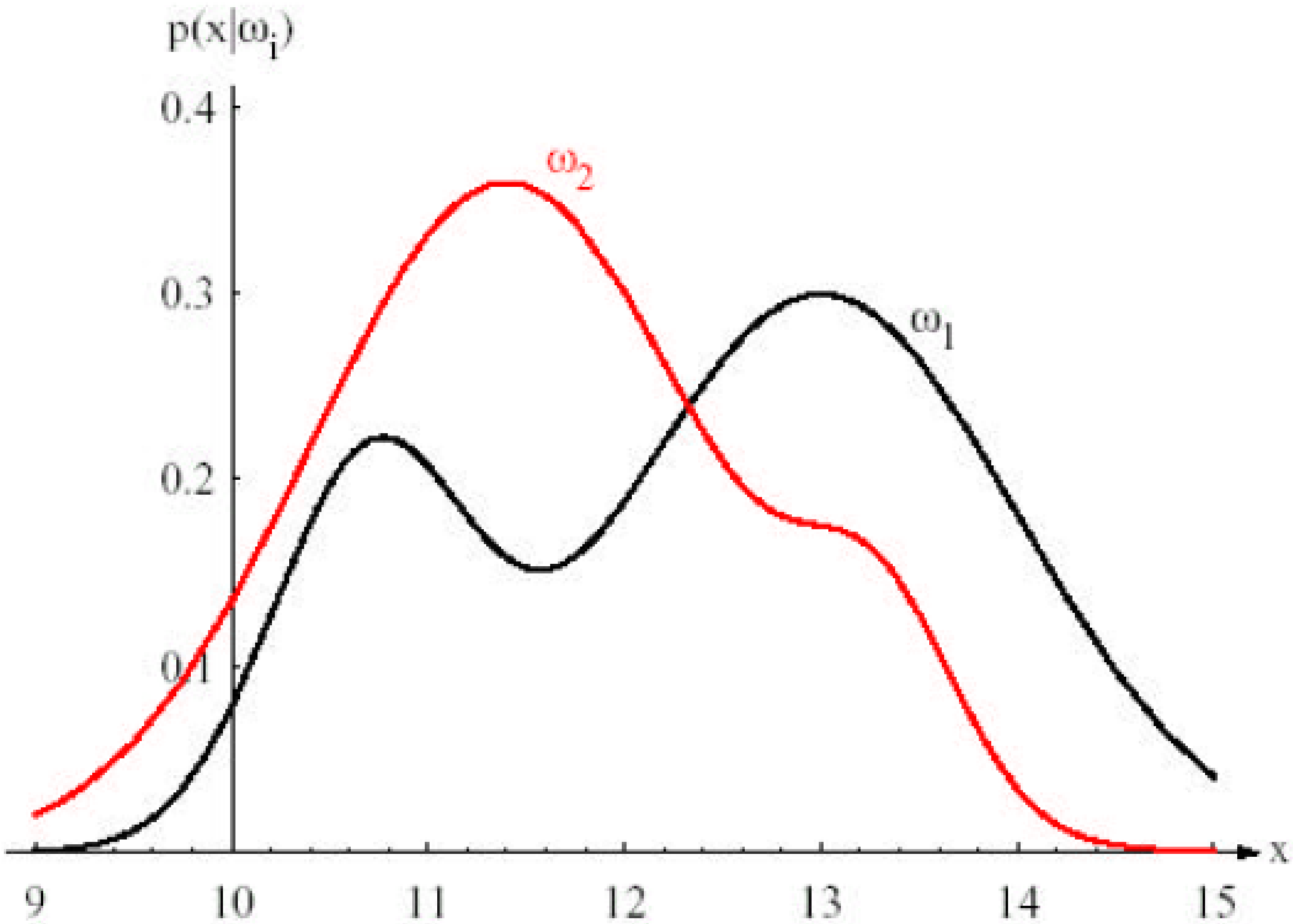
先验概率： $P(\omega_i)$

后验概率： $P(\omega_i|\mathbf{x})$

类条件概率： $P(\mathbf{x}|\omega_i)$

贝叶斯公式： $P(\omega_i|\mathbf{x})=\dfrac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})}$

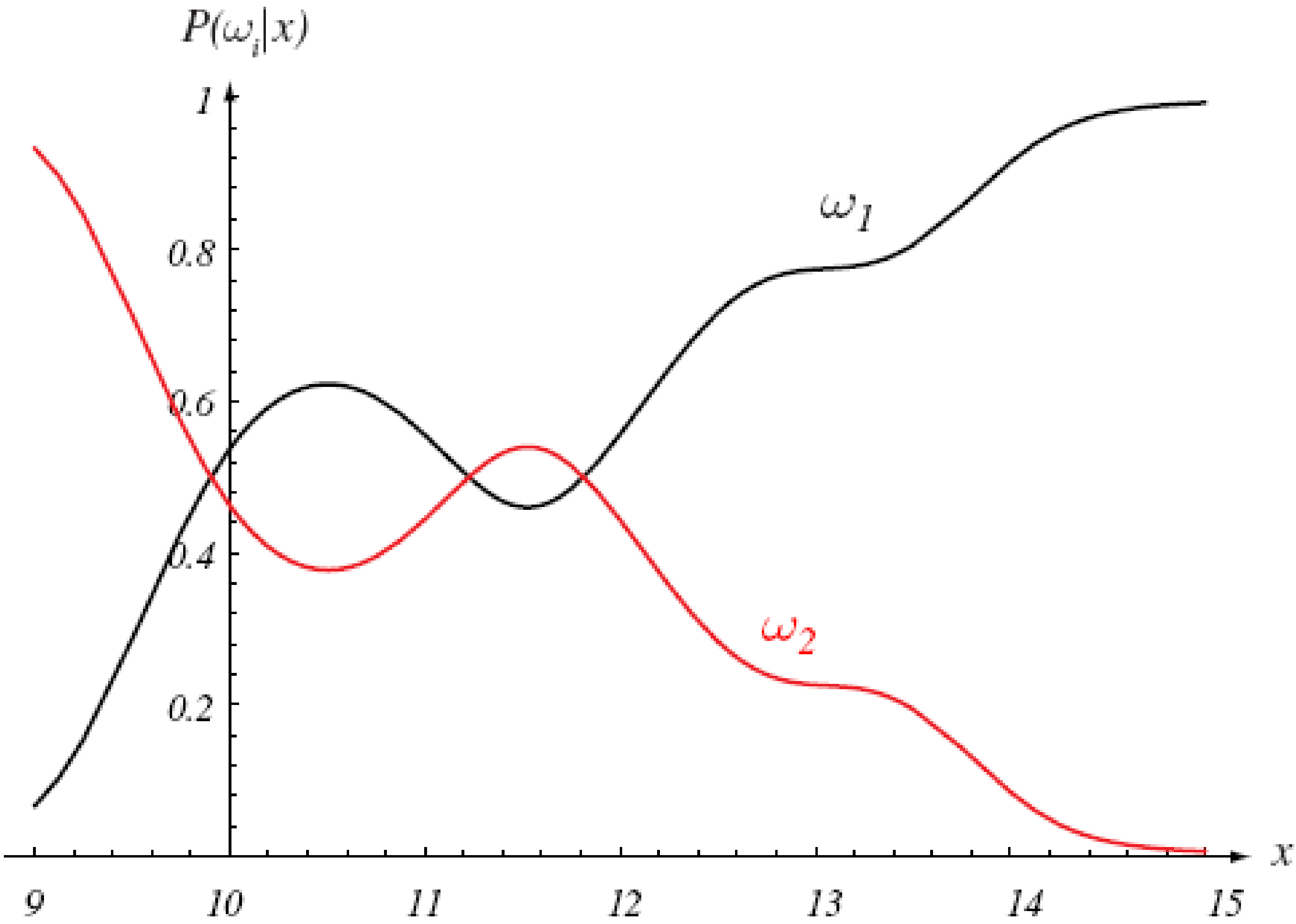
ISBN 978-7-5603-4763-9



- 类条件概率密度

Class-conditional probability density

ISBN 978-7-5603-4763-9

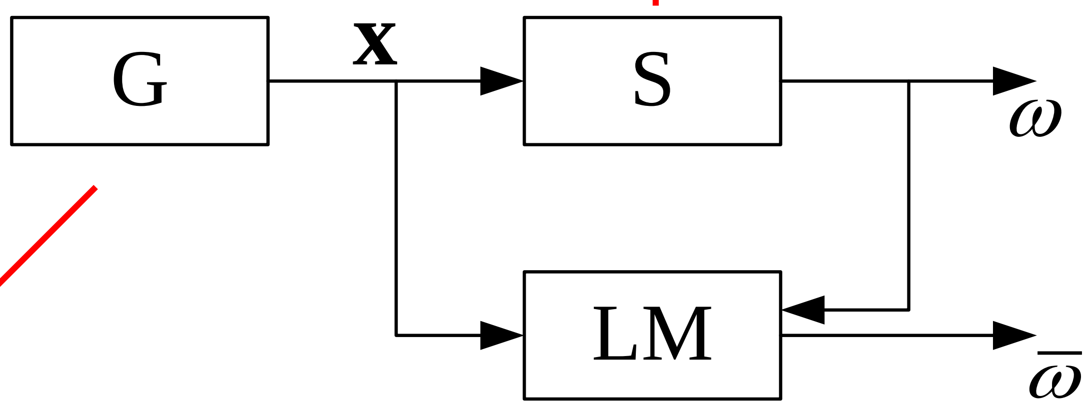


- **P(w$_1$)=2/3, P(w$_2$)=2/3 时的后验概率密度**

统计学习模型

S：目标算子（训练器）

将x变换成 ω，存在、不变、未知。



G：数据（实例）发生器

依据某一固定但未知的概率分布函数F(x)，独立同分布地产生向量x

$$F(\mathbf{x},\omega)=F(\mathbf{x})F(\omega|\mathbf{x})$$

LM：学习机器

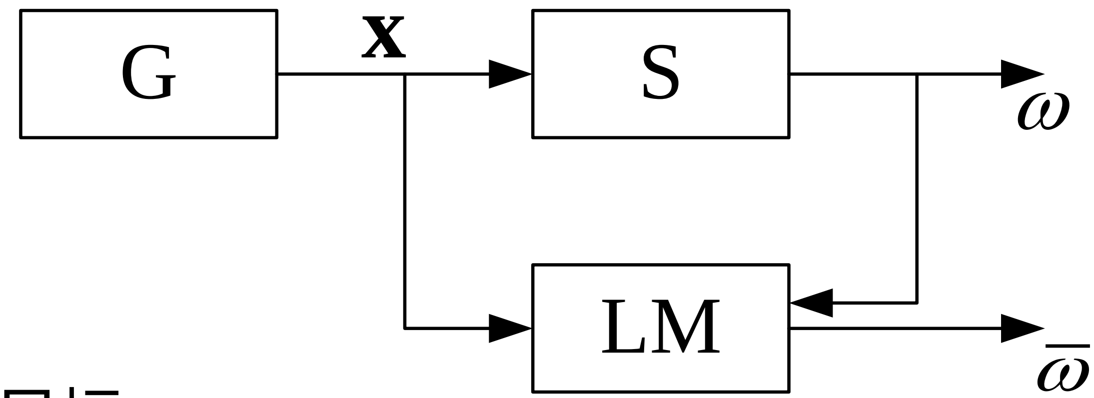
观测训练集$(\mathbf{x}_1,\omega_1)\cdots(\mathbf{x}_\ell,\omega_\ell)$

构造算子，预测特定向量$\mathbf{x}_i$的响应ω_i

学习过程：从给定的函数集中寻找一个适当的函数

ISBN 978-7-5603-4763-9

统计学习模型



学习目标：

1，模仿训练器的算子：

对给定的发生器G，预测训练器S的输出

2，辨识训练器算子：

构造一个非常接近训练器算子S的算子

ISBN 978-7-5603-4763-9

1 概率论与统计学习

2 贝叶斯决策理论

3 高斯分布的贝叶斯分类器

ISBN 978-7-5603-4763-9

贝叶斯决策理论

各类别总体的概率分布已知

要决策分类的类别数一定

已知：分类问题有c个类别 $\omega_1, \omega_2, \cdots, \omega_c$

各类别的先验概率 $P(\omega_i)$

各类条件概率密度函数 $P(\mathbf{x}|\omega_i)$

决策：对于特征空间中观测到的向量 $\mathbf{x} = [x_1, x_2, \cdots, x_d]^t$

应该将x分到哪一类?

错误率最小 $\longrightarrow$ 风险最小

贝叶斯决策理论

- 基于最小错误率的贝叶斯决策
- 基于最小风险的贝叶斯决策
- Neyman-Pearson决策规则: 限定一类错误率，最小化另一类错误率
- 极小化极大准则： 先验概率$P(\omega_i)$未知的情况下，使最大可能的风险最小化

两类问题的错误率

- 观察到特征x时作出判别的错误率：

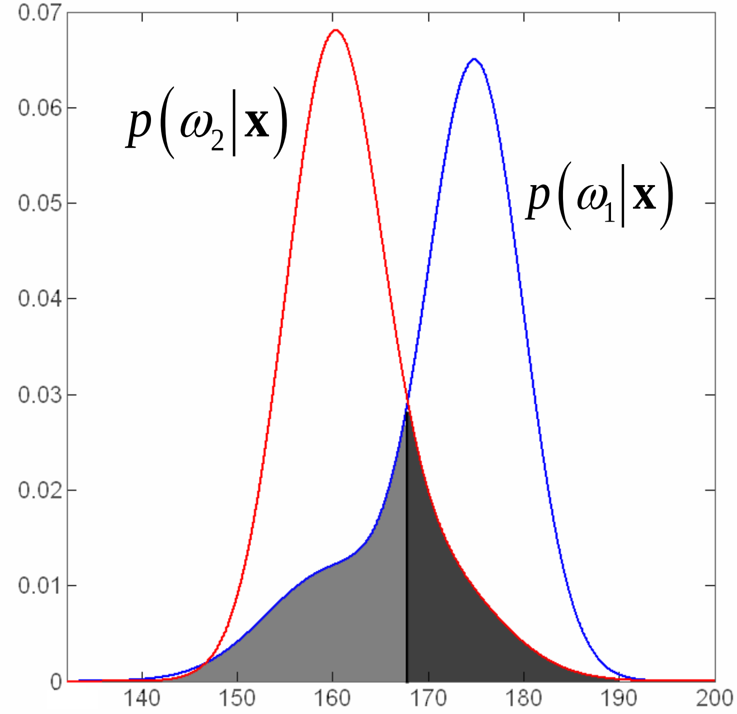
$$P(error|\mathbf{x}) = \begin{cases} P(\omega_1|\mathbf{x}), & \text{判定}\,\omega_2 \\ P(\omega_2|\mathbf{x}), & \text{判定}\,\omega_1 \end{cases}$$

- 两类问题最小错误率判别准则：

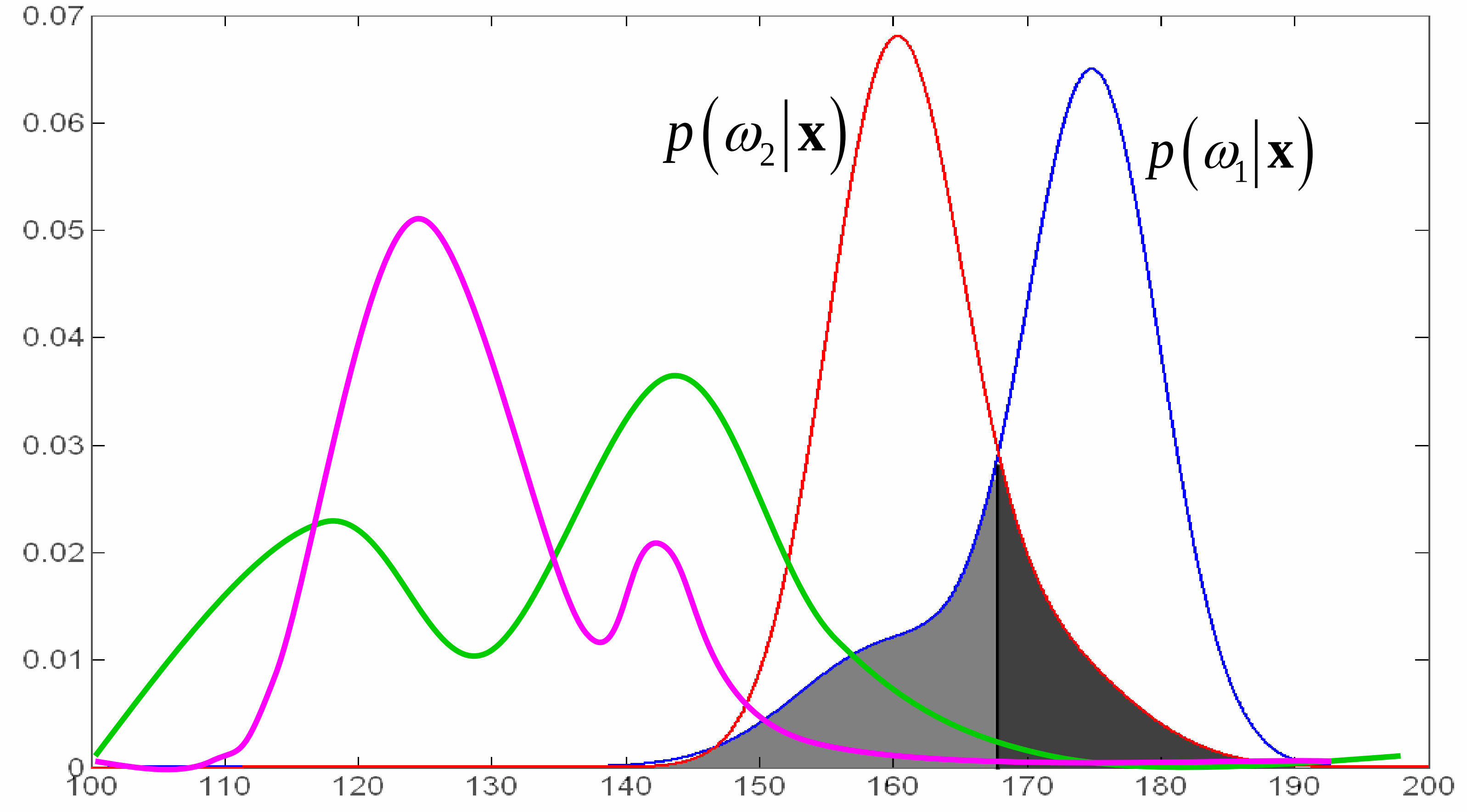
$$\begin{cases} \text{如果} P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x}), & \mathbf{x} \in \omega_1 \\ \text{如果} P(\omega_1|\mathbf{x}) < P(\omega_2|\mathbf{x}), & \mathbf{x} \in \omega_2 \end{cases}$$

$\Downarrow$

最大后验概率



贝叶斯分类器的错误率估计



多类问题最小错误率

- 判别x属于ω_i的错误率：

$$P(error|\mathbf{x})=\sum_{j\neq i}P(\omega_j|\mathbf{x})=1-P(\omega_i|\mathbf{x})$$

- 判别准则为：

$$i=\underbrace{\arg\max_{1\leq j\leq c}P(\omega_j|\mathbf{x})}_{\text{最大后验概率}},\ \text{则：}\ \mathbf{x}\in\omega_i$$

贝叶斯最小错误率准则

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

$$g_j(\mathbf{x}) = p(\mathbf{x}|\omega_j)P(\omega_j)$$

$$g_j(\mathbf{x}) = \ln p(\mathbf{x}|\omega_j) + \ln P(\omega_j)$$

Bayes判别准则：

$i = \arg\max\limits_{1 \le j \le c} g_j(\mathbf{x})$ ，则 $\mathbf{x} \in \omega_i$

ISBN 978-7-5603-4763-9

例：

对一大批人进行癌症普查，设ω1类代表患癌症，ω2类代表正常人。已知先验概率：

$$P(\omega_1)=0.005, P(\omega_2)=0.995$$

以一个化验结果作为特征x: {阳性，阴性}，患癌症的人和正常人化验结果为阳性的概率分别为：

$$P(x=\text{阳性}|\omega_1)=0.95, P(x=\text{阳性}|\omega_2)=0.01$$

现有一人化验结果为阳性，问此人是否患癌症?

最小错误率准则： $g_j(\mathbf{x})=p(\mathbf{x}|\omega_j)P(\omega_j)$

基于最小风险的贝叶斯决策

判断某人是正常($\omega_{正}$)还是患者($\omega_{患}$),将出现以下情况:

实际类别 ω_i	决策类别 ω_j	风险因子 λ_{ij}
正常人	正常人	0
正常人	患者	0.25
患者	正常人	1
患者	患者	0

何种决策风险最小?

做出正常判决的风险: $R_{正}(\mathbf{x}) = \lambda_{正\to正} P(\omega_{正}|\mathbf{x}) + \lambda_{患\to正} P(\omega_{患}|\mathbf{x})$

做出患病判决的风险: $R_{患}(\mathbf{x}) = \lambda_{正\to患} P(\omega_{正}|\mathbf{x}) + \lambda_{患\to患} P(\omega_{患}|\mathbf{x})$

做出j类判决的风险: $R_j(\mathbf{x}) = \sum_{i=1}^{c} \lambda_{ij} P(\omega_i|\mathbf{x})$

——计算各类决策风险，选则最小风险对应的决策

最小平均风险准则贝叶斯分类器

- 有c个类别$\omega_1, \omega_2, \ldots, \omega_c$, 将$\omega_i$类的样本判别为$\omega_j$类的代价为$\lambda_{ij}$。
- 将未知模式x判别为ω_j类的平均风险为：

$$\gamma_j(\mathbf{x}) = \sum_{i=1}^{c} \lambda_{ij} \underline{P(\omega_i|\mathbf{x})}$$

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

- 利用Bayes公式:

$$\gamma_j(\mathbf{x}) = \sum_{i=1}^{c} \lambda_{ij} \underline{P(\mathbf{x}|\omega_i)P(\omega_i)}$$

- 构造判别函数：

$$g_j(\mathbf{x}) = -\gamma_j(\mathbf{x})$$

ISBN 978-7-5603-4763-9

例

- 对一大批人进行癌症普查，设ω_1类代表患癌症，ω_2类代表正常人。已知先验概率：

$$P(\omega_1)=0.004, P(\omega_2)=0.996$$

以一个化验结果作为特征x：{阳性，阴性}，患癌症的人和正常人化验结果为阳性的概率分别为：

$$P(x=\text{阳性}|\omega_1)=0.96, P(x=\text{阳性}|\omega_2)=0.01$$

判别代价： $\lambda_{11}=0, \lambda_{22}=0, \lambda_{12}=95, \lambda_{21}=15$
现有一人化验结果为阳性，问此人是否患癌症？

贝叶斯最小风险判别：

$$g_j(\mathbf{x})=-\gamma_j(\mathbf{x}) \qquad \gamma_j(\mathbf{x})=\sum_{i=1}^{c}\lambda_{ij}P(\mathbf{x}|\omega_i)P(\omega_i)$$

贝叶斯决策论——举例

决策问题：

判别类别 $\{\omega_1,\cdots,\omega_c\}$ 正常人、癌症早期，癌症晚期

治疗行为 $\{\alpha_1,\cdots,\alpha_a\}$：不治疗、保守治疗、手术治疗、化疗

风险函数： 在类别状态 ω_j 时，采取行动 α_i 的风险（预期损失）

$$\lambda_{\text{手术，癌症早期}}=\lambda\left(\alpha_{\text{手术}}\middle|\omega_{\text{癌症早期}}\right)$$

条件风险： 观察到某人体检特征 x，采取行为 $\alpha_{\text{手术}}$ 的损失期望

$$\begin{aligned}\underline{R\left(\alpha_{\text{手术}}\middle|\mathbf{x}\right)}&=\lambda\left(\alpha_{\text{手术}}\middle|\omega_{\text{正常}}\right)P\left(\omega_{\text{正常}}\middle|\mathbf{x}\right)\\&+\lambda\left(\alpha_{\text{手术}}\middle|\omega_{\text{早期}}\right)P\left(\omega_{\text{早期}}\middle|\mathbf{x}\right)\\&+\lambda\left(\alpha_{\text{手术}}\middle|\omega_{\text{晚期}}\right)P\left(\omega_{\text{晚期}}\middle|\mathbf{x}\right)\\&=\sum_{j=1}^{c}\lambda\left(\alpha_i\middle|\omega_j\right)P\left(\omega_j\middle|\mathbf{x}\right)\end{aligned}$$

对某人x进行手术的预期风险

ISBN 978-7-5603-4763-9

贝叶斯决策论

引入一般的损失函数来替代误差概率

$\{\omega_1,\cdots,\omega_c\}$ 表示有限的c个类别集

$\{\alpha_1,\cdots,\alpha_a\}$ 表示有限的a种可能采取的行为集

风险函数 $\lambda_{ij}=\lambda\left(\alpha_i \mid w_j\right)$

在类别状态ω_j时，采取行动α_i的风险（预期损失）

条件风险：观察模式为x的条件下，采取行为 α_i 的损失期望

$$R\left(\alpha_i \mid \mathbf{x}\right)=E\left[\lambda\left(\alpha_i \mid w_j\right)\right]=\sum_{j=1}^{c}\lambda\left(\alpha_i \mid w_j\right)P\left(w_j \mid \mathbf{x}\right)$$

将模式x判属某类所造成的损失的条件数学期望，

反映对未知x采取一个判决行动α(x)所付出的代价

贝叶斯决策论

对于**x**的不同观察值，采取决策α_i 时，其条件风险的大小是不同的。所以，究竟采取哪一种决策将随**x**的取值而定。决策 α可以看成随机向量**x**的函数，记为 $\alpha(\mathbf{x})$

期望风险 **R**（总风险、条件期望损失）：

$$R=\int R\left(\alpha\left(\mathbf{x}\right)\middle|\mathbf{x}\right)p\left(\mathbf{x}\right)d\mathbf{x}$$

$d\mathbf{x}$是特征空间的体积元，积分在整个特征空间进行。

在整个特征空间中定义期望风险R，反映对整个特征空间所有x的取值采取相应的决策α(x)所带来的平均风险。

ISBN 978-7-5603-4763-9

贝叶斯决策论

基本思想：统计意义上使由于误判而蒙受的损失最小。

条件风险最小判决：对**x**采取决策α使其条件风险最小。

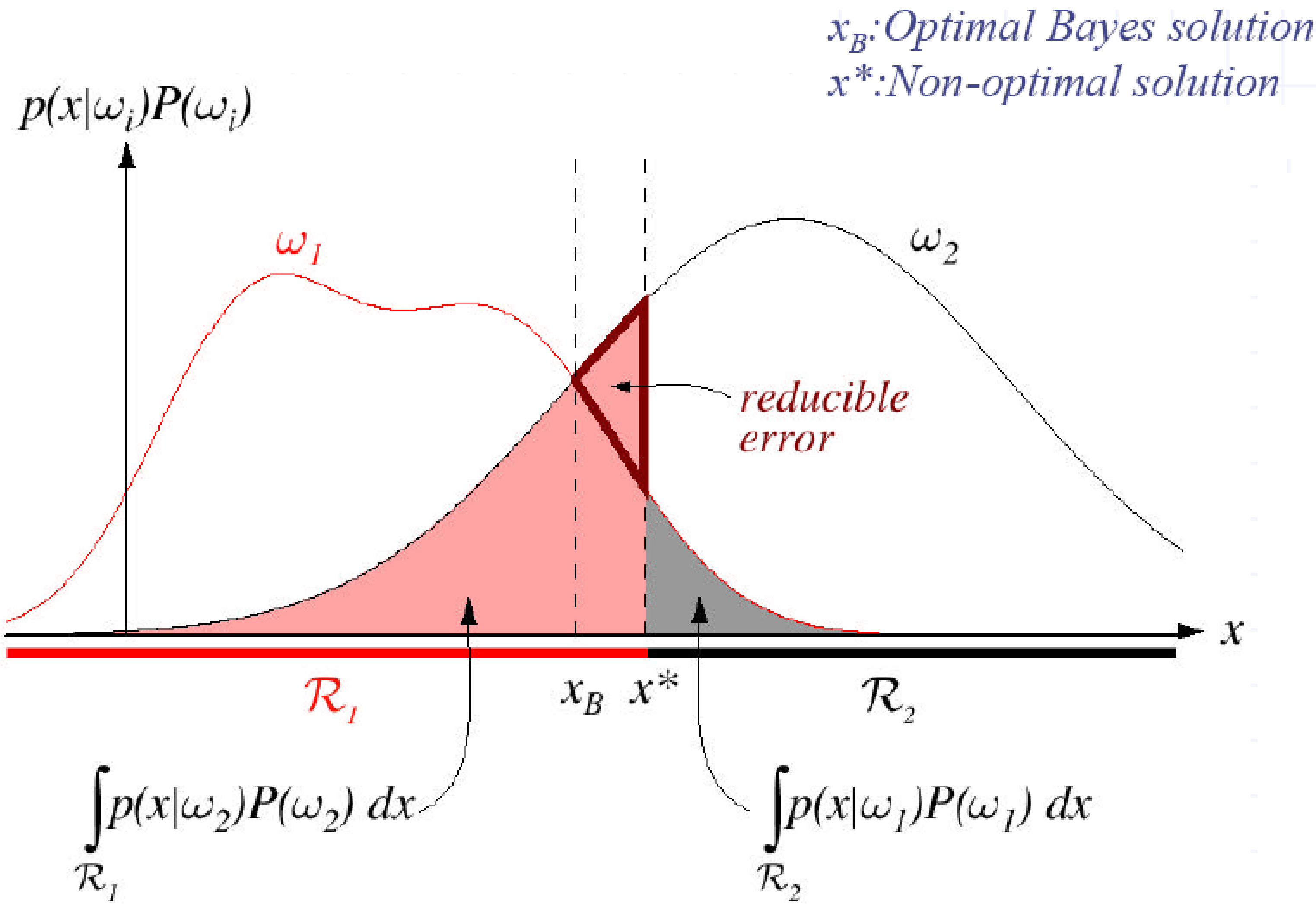
$$\alpha = \arg\min_{\alpha_i} R(\alpha_i | \mathbf{x})$$

贝叶斯风险：对所有的**x**作出决策时，均进行条件风险最小化，总的期望风险R也最小，称为贝叶斯风险，记为R* 。

R*：最小化后的总风险值，可获得的最优结果

ISBN 978-7-5603-4763-9

误差概率



$P(\mathbf{x} \in R_1, \omega_2) = P(\mathbf{x} \in R_1 | \omega_2) \cdot P(\omega_2)$

$P(\mathbf{x} \in R_2, \omega_1) = P(\mathbf{x} \in R_2 | \omega_1) \cdot P(\omega_1)$

最小风险贝叶斯决策

如果：$R\left(\alpha_k \mid x\right) = \min_{i=1,2,\cdots a} R\left(\alpha_i \mid x\right)$

则：$a = a_k$

$x \in \omega_k$

最小风险贝叶斯决策必须要有合适的损失函数 。实际工作中要列出合适的决策表很不容易，往往要根据所研究的具体问题，分析错误决策造成损失的严重程度来确定。

两类问题的最小风险贝叶斯决策

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

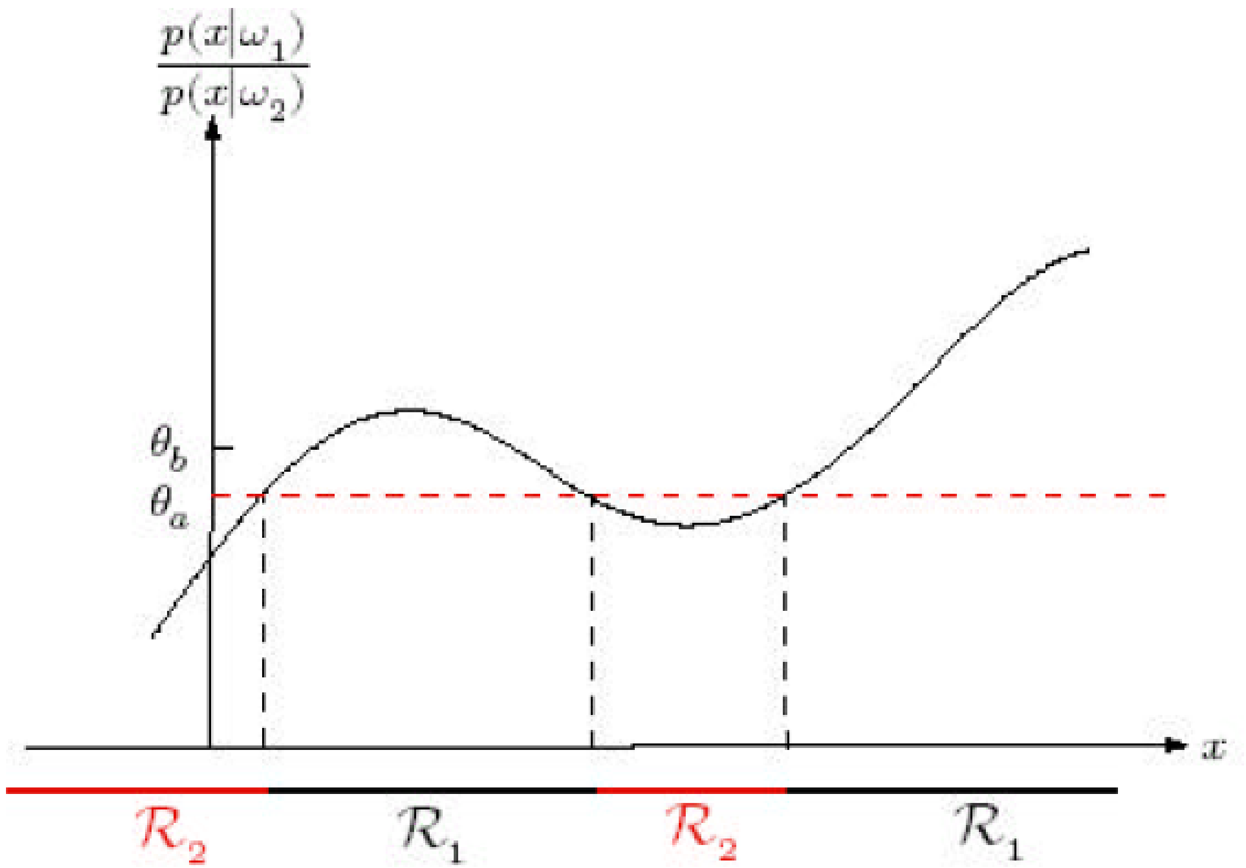
$$R(\alpha_2 \mid \mathbf{x}) = \lambda_{21} P(\omega_1 \mid \mathbf{x}) + \lambda_{22} P(\omega_2 \mid \mathbf{x})$$

$$\frac{R(\alpha_1 \mid \mathbf{x})}{R(\alpha_2 \mid \mathbf{x})} < 1 \Rightarrow \mathbf{x} \in \omega_1$$

$$\frac{P(\omega_1 \mid \mathbf{x})}{P(\omega_2 \mid \mathbf{x})} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \Rightarrow \mathbf{x} \in \omega_1$$

$$P(\omega_j \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_j) P(\omega_j)}{p(\mathbf{x})}$$

$$\underbrace{\frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)}}_{\text{似然比}} > \underbrace{\frac{(\lambda_{12} - \lambda_{22}) P(\omega_2)}{(\lambda_{21} - \lambda_{11}) p(\omega_1)}}_{\text{不依赖观测值x的阈值}} \Rightarrow \mathbf{x} \in \omega_1$$



$$\frac{p(\mathbf{x}\mid\omega_1)}{p(\mathbf{x}\mid\omega_2)} > \theta \quad \Rightarrow \quad x \in \omega_1$$

在采用0-1损失函数时，最小风险贝叶斯决策就等价于最小错误率贝叶斯决策。

0-1损失函数

$$\lambda\left(\alpha_i,\omega_j\right)=\begin{cases}0, & i=j\\1, & i\neq j\end{cases}$$

对于正确决策（即i=j），$\lambda\left(\alpha_i,\omega_j\right)=0$，就是说没有损失；而对于任何错误决策，其损失均为1

ISBN 978-7-5603-4763-9

贝叶斯分类器的其它版本

- 约束一定错误率（风险）：Neyman-Pearson准则；
- 先验概率$P(\omega_i)$未知：极小化极大准则；
- 某些特征缺失的决策：
- 连续出现的模式之间统计相关的决策：

ISBN 978-7-5603-4763-9

Neyman-Pearson准则

1.问题的提出：

(1)某些二类判决问题，某一种错误较另一种错误更为重要—危害更为严重。

(2)先验概率未知。

2.基本思想：

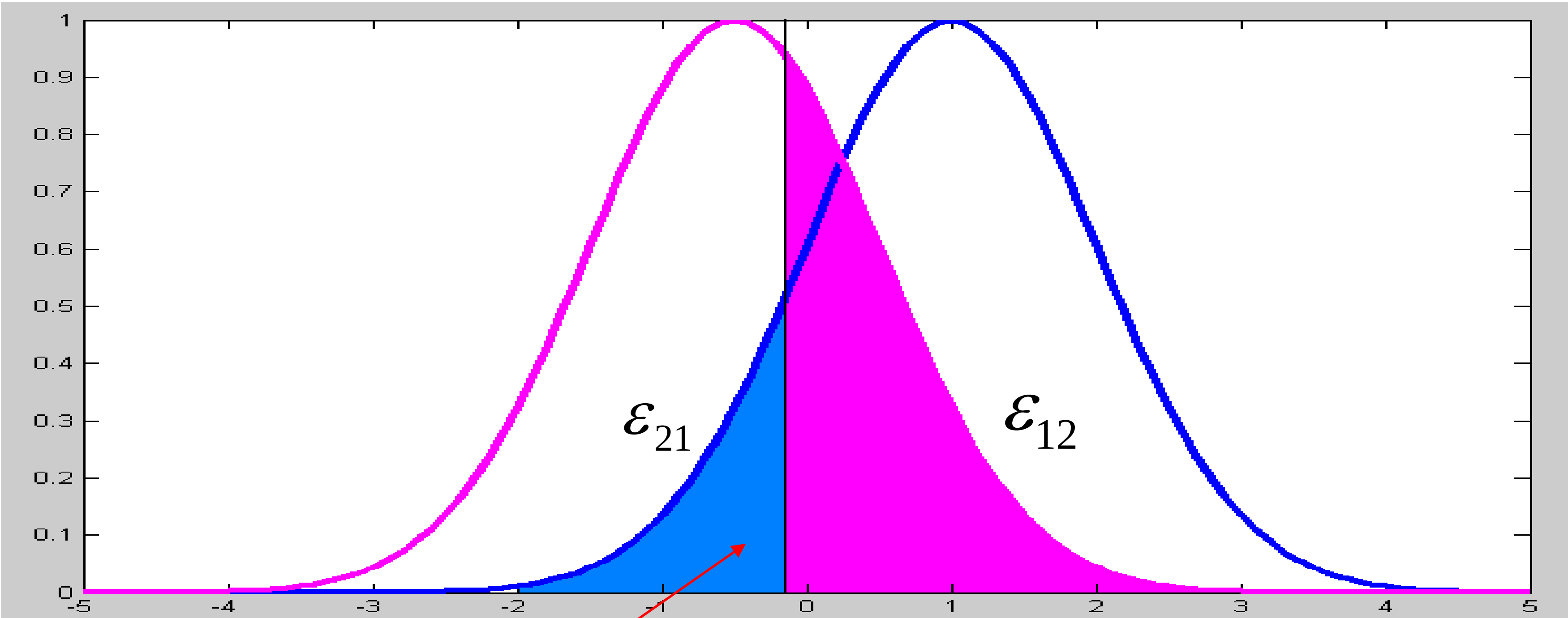
严格限制较重要的一类错误概率，在令其等于某常数的约束下使另一类误判概率最小。

在癌症诊断中，把癌症误判为正常的损失更为严重，要求这种误判错误率$P_2(e)=\varepsilon_0$，ε_0是一个很小的常数

在这样的约束下，求把正常误判为癌症的错误率$P_1(e)$极小值

ISBN 978-7-5603-4763-9

3.决策规则



$$\varepsilon_{21}=P_2(e)=\int_{R1}p(\mathbf{x}\mid\omega_2)d\mathbf{x}\qquad \varepsilon_{12}=P_1(e)=\int_{R2}p(\mathbf{x}\mid\omega_1)d\mathbf{x}$$

限定$\varepsilon_{21}=\varepsilon_0$时，最小化$\varepsilon_{12}$：构造**Lagrange**乘子

$$r=\varepsilon_{12}+\lambda(\varepsilon_{21}-\varepsilon_0)$$

$$\int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x} = 1 - \int_{R_1} p(\mathbf{x}|\omega_1)d\mathbf{x}$$

$$r = \int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x} + \lambda\left[\int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x} - \varepsilon_0\right]$$

$$= (1-\lambda\varepsilon_0) + \int_{R_1}\left[\lambda p(\mathbf{x}|\omega_2) - p(\mathbf{x}|\omega_1)\right]d\mathbf{x}$$

- 由此式分别对x和λ求导，令 $\frac{\partial r}{\partial x} = 0$ $\frac{\partial r}{\partial \lambda} = 0$ 有

$$\lambda = p(\mathbf{x}|\omega_1) \big/ p(\mathbf{x}|\omega_2) \qquad \int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x} = \varepsilon_0$$

满足 $\lambda = \frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)}$ 的最佳值 λ 和满足$\int_{R_1} P(\mathbf{x}|\omega_2)d\mathbf{x} = \varepsilon_0$ 的边界面就能使r极小。

ISBN 978-7-5603-4763-9

N-P决策规则

如果： $\dfrac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \begin{matrix} > \\ < \end{matrix} \lambda$ 则： $\mathbf{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$

N-P决策规则归结为找阈值 λ 。

当 $\dfrac{p(\mathbf{x}/\omega_1)}{p(\mathbf{x}/\omega_2)} = \lambda$ 时，λ 作 $R_1 R_2$ 的分界线.

$$\varepsilon_2 = \int_{-\infty}^{\lambda} p(\mathbf{x}|\omega_2)dx,$$

λ 为 ε_2 的函数在取 ε_2 为常数时，λ 可确定，

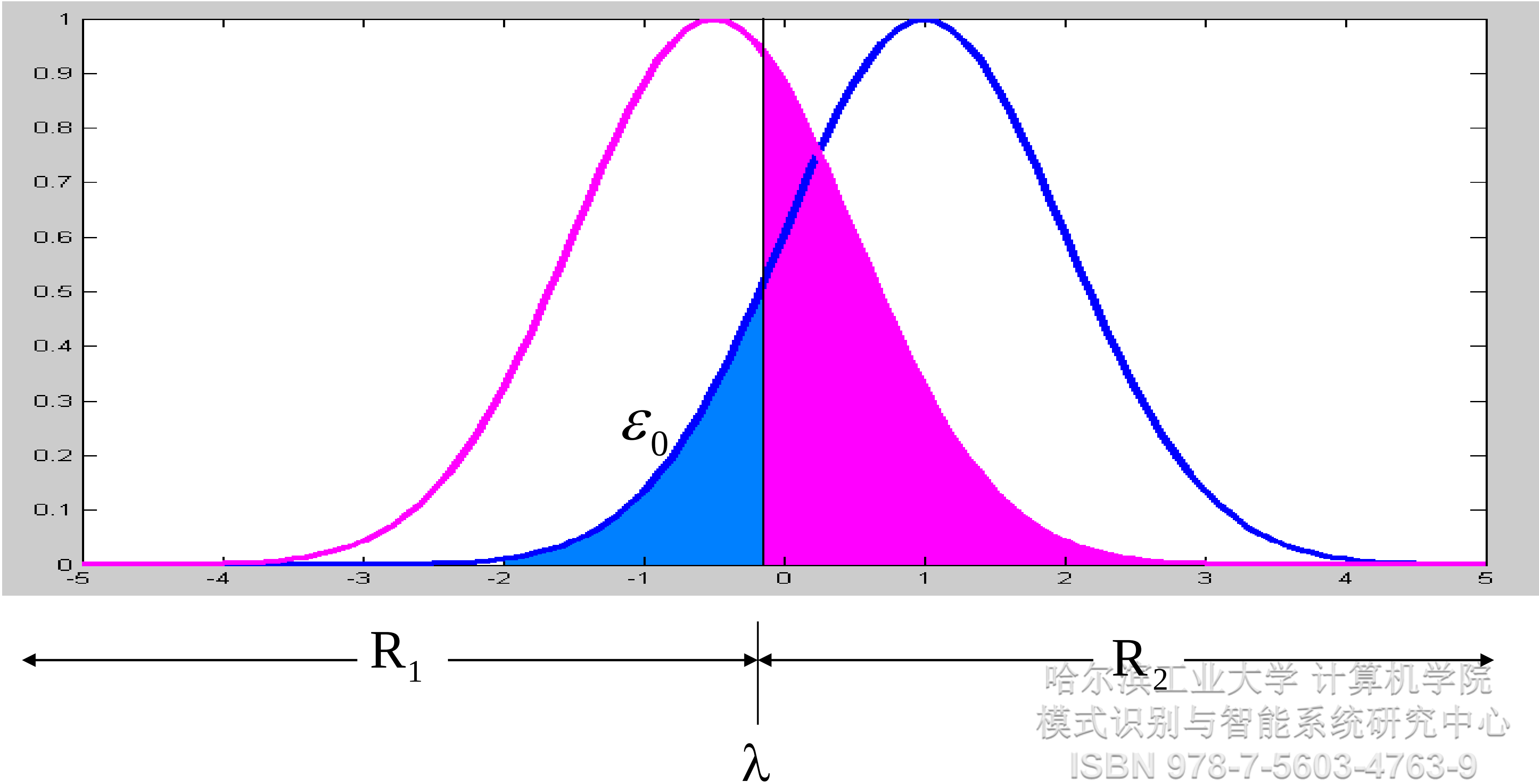
这时 ε_2 一定 ε_1 最小

ISBN 978-7-5603-4763-9

λ的确定：

$$\varepsilon_0 = \varepsilon_{21} = \int_{-\infty}^{\lambda} p(l/\omega_2)dl$$

通过建立λ－－ε对照表确定。



ISBN 978-7-5603-4763-9

4．最小错误率贝叶斯决策规则与N-P决策

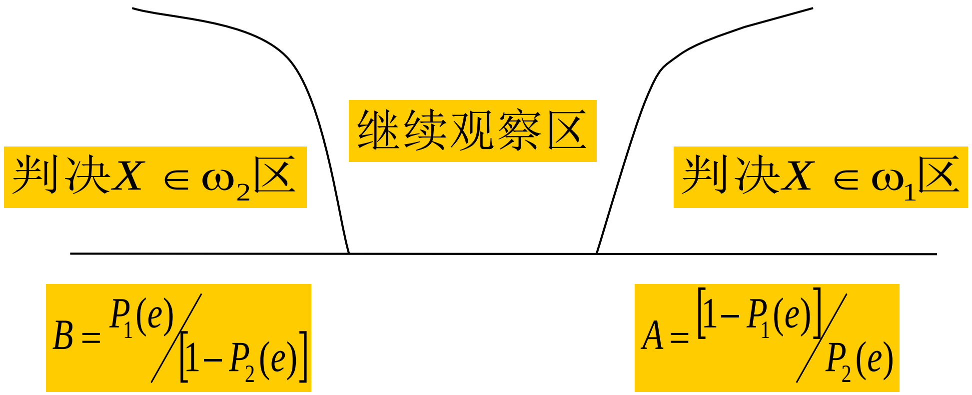
最小错误率	$\dfrac{p(\mathbf{x}\mid\omega_1)}{p(\mathbf{x}\mid\omega_2)} > \dfrac{P(\omega_2)}{p(\omega_1)} \quad \Rightarrow \mathbf{x}\in\omega_1$
最小风险	$\dfrac{p(\mathbf{x}\mid\omega_1)}{p(\mathbf{x}\mid\omega_2)} > \dfrac{(\lambda_{12}-\lambda_{22})P(\omega_2)}{(\lambda_{21}-\lambda_{11})p(\omega_1)} \quad \Rightarrow \mathbf{x}\in\omega_1$
N-P决策	$\dfrac{p(\mathbf{x}\mid\omega_1)}{p(\mathbf{x}\mid\omega_2)} > \lambda,\quad \varepsilon_2 = \int_{-\infty}^{\lambda} p(l\mid\omega_2)\,dl \quad \Rightarrow \mathbf{x}\in\omega_1$
	似然比 $\longleftrightarrow$ 不依赖观测值x的阈值

ISBN 978-7-5603-4763-9

序贯分类决策

- 在许多实际问题中，观察实际上是序贯的。
- 对样品进行第 i 次观察获取一序列特征为：$X=(x1,x2,\ldots,xi)^T$ 则对于ω1，ω2两类问题,
  - 若X ∈ ω1，则判决完毕
  - 若X∈ ω2 ，则判决完毕
  - 若X不属ω1也不属ω2 ，则不能判决，进行第i+1次观察，得 $X=(x1,x2,\ldots,xi,,x\ i+1)^T$ ，再重复上面的判决，直到所有的样品分类完毕为止。
- 使那些在二类边界附近的样本不会因某种偶然的微小变化而误判

➢序贯分类决策规则：



$$
\begin{cases}
\dfrac{P_i(X/\omega_1)}{P_i(X/\omega_2)} \geq A = \dfrac{1-P_1(e)}{P_2(e)} \text{时} X \in \omega_1 \\
\dfrac{P_i(X/\omega_1)}{P_i(X/\omega_2)} \leq B = \dfrac{P_1(e)}{1-P_2(e)} \text{时} X \in \omega_2 \\
B < \dfrac{P_i(X/\omega_1)}{P_i(X/\omega_2)} < A \text{时，继续观察}
\end{cases}
$$

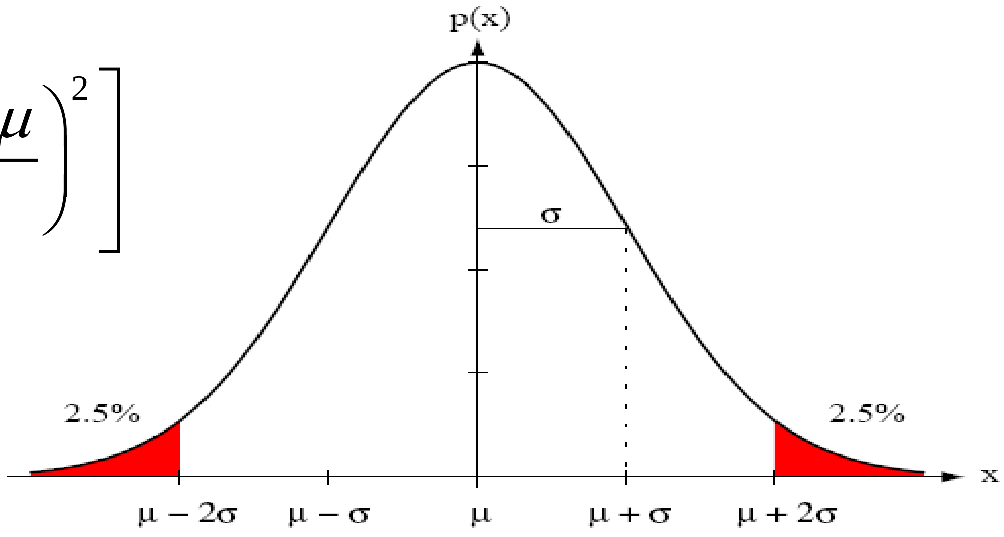
- 上下门限A、B是由设计给定的错误概率$P_1(e)$, $P_2(e)$来确定的，Wald 已证明，观察次数不会很大，它收敛的很快。

高斯分布的贝叶斯分类器

单变量正态分布密度函数（高斯分布）：

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

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



最大熵：在均值、方差确定的各种分布中，正态分布的熵最大

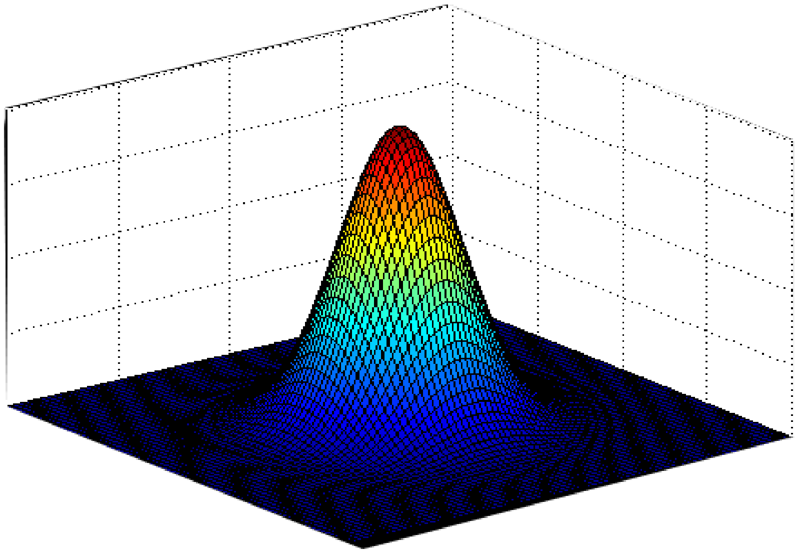
$$-\int p(x)\ln p(x)dx$$

中心极限定理：大量小的、独立随机分布的总和等效为正态分布

多元正态分布函数

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

$$p(\mathbf{x})\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$$



均值：$\boldsymbol{\mu}=E(\mathbf{x})=\int \mathbf{x}p(\mathbf{x})d\mathbf{x}$

$$\mu_i=E(x_i)$$

协方差矩阵：$\boldsymbol{\Sigma}=E\left((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t\right)=\int(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^t p(\mathbf{x})d\mathbf{x}$

$$\sigma_{ij}=E\left((x_i-\mu_i)(x_j-\mu_j)\right)$$

多元正态类条件概率密度

$$p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

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

正态分布的判别函数

贝叶斯判别函数可以写成对数形式：

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$$

$$\text{其中} p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right]$$

类条件概率密度函数为正态分布时：

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{\Sigma}_i| + \ln P(\omega_i)$$

情况一： $\mathbf{\Sigma}_i = \sigma^2\mathbf{I}, P(\omega_i) = \frac{1}{c}$ 距离分类器

- 判别函数可以写成：

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{\Sigma}_i| + \ln P(\omega_i)$$

$$\Downarrow$$

$$g_i(\mathbf{x}) = -(\mathbf{x}-\boldsymbol{\mu}_i)^t(\mathbf{x}-\boldsymbol{\mu}_i) = -\|\mathbf{x}-\boldsymbol{\mu}_i\|^2$$

- 此分类器称为距离分类器，判别函数可以用待识模式$\mathbf{x}$与类别均值$\boldsymbol{\mu}_i$之间的距离表示：

$$g_i(\mathbf{x}) = -d(\mathbf{x}, \boldsymbol{\mu}_i)$$

情况二： $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ 线性分类器

判别函数可以写成：

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

$$\Downarrow$$

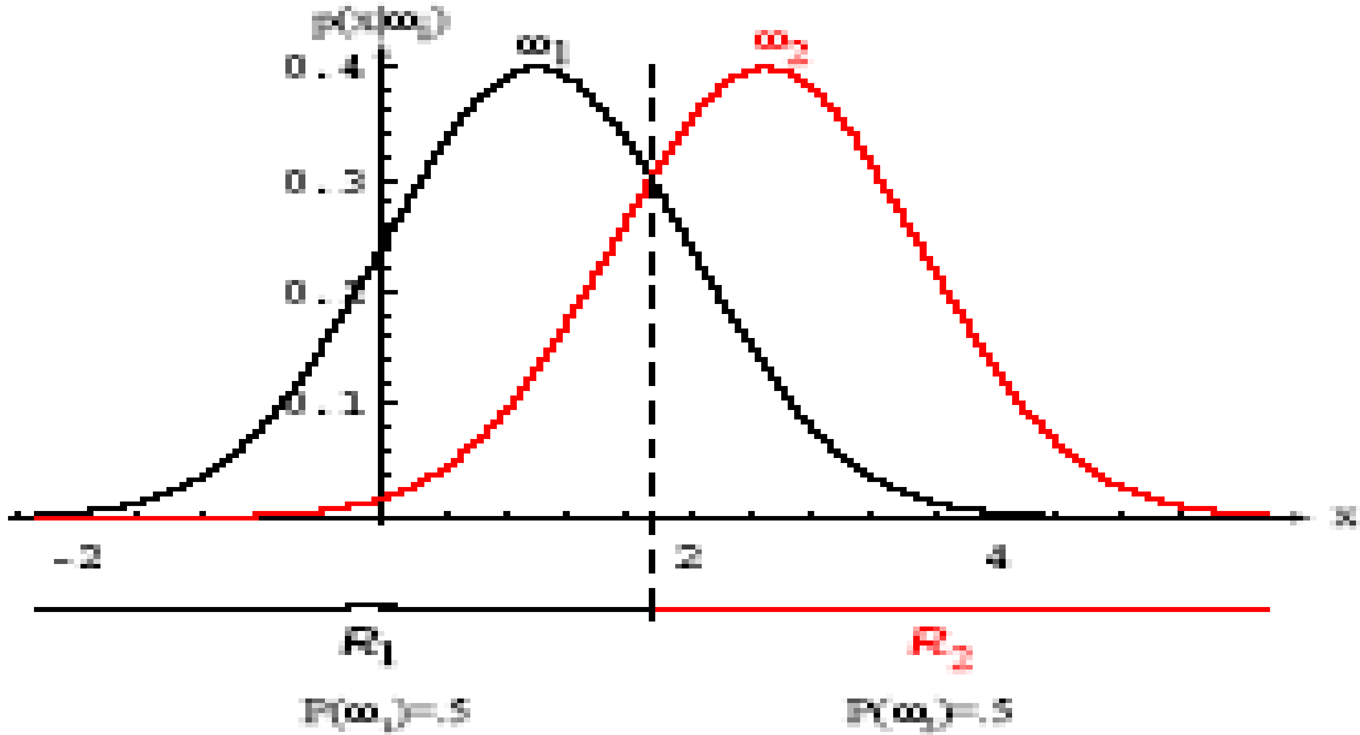
$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i)$$

$$\Downarrow$$

$$g_i(\mathbf{x}) = \boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^t \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i + \ln P(\omega_i) = \mathbf{w}_i^t \mathbf{x} + w_{i0}$$

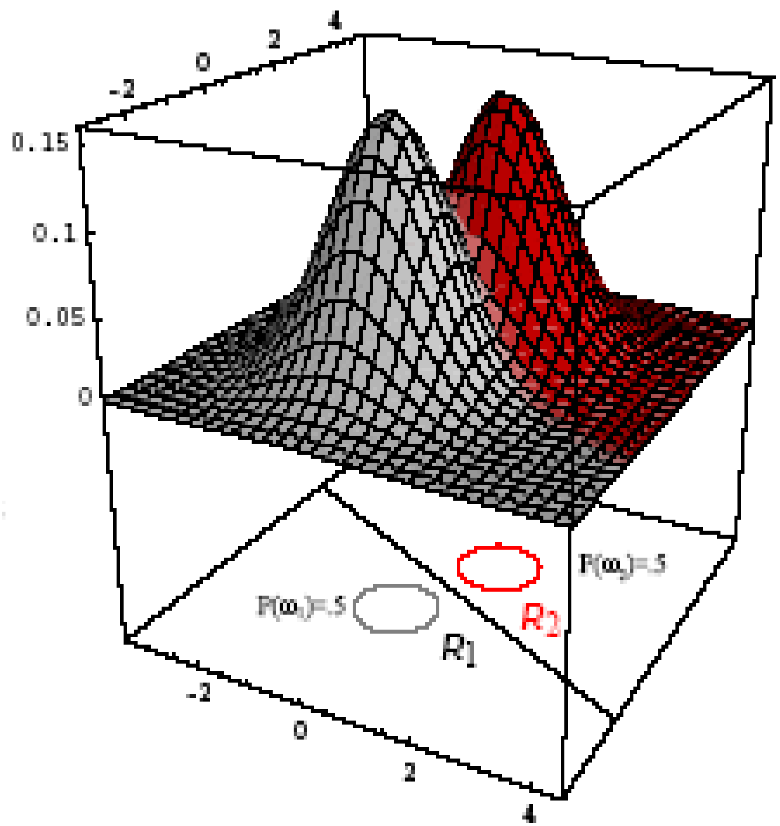
线性分类器

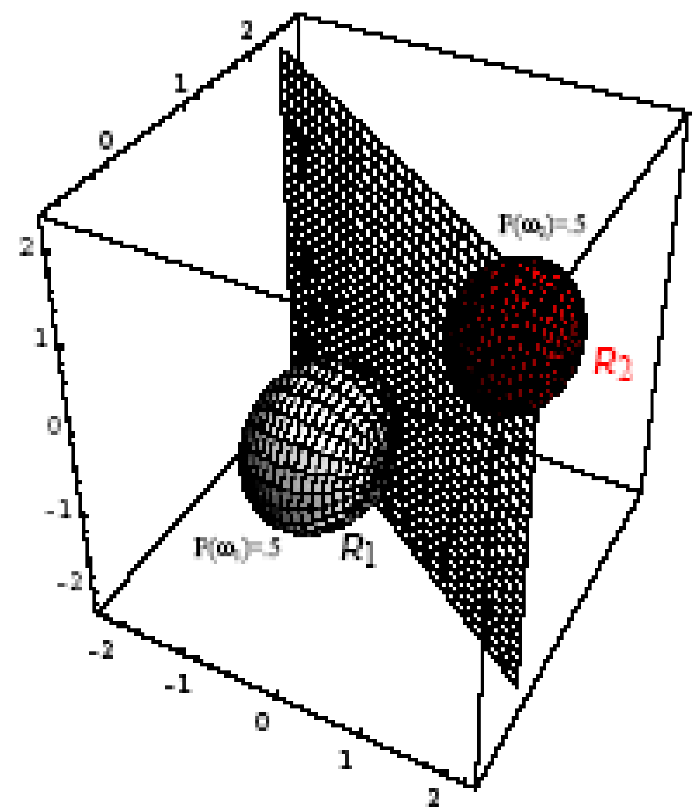
- 两类问题，1维特征，先验概率相同时：



线性分类器

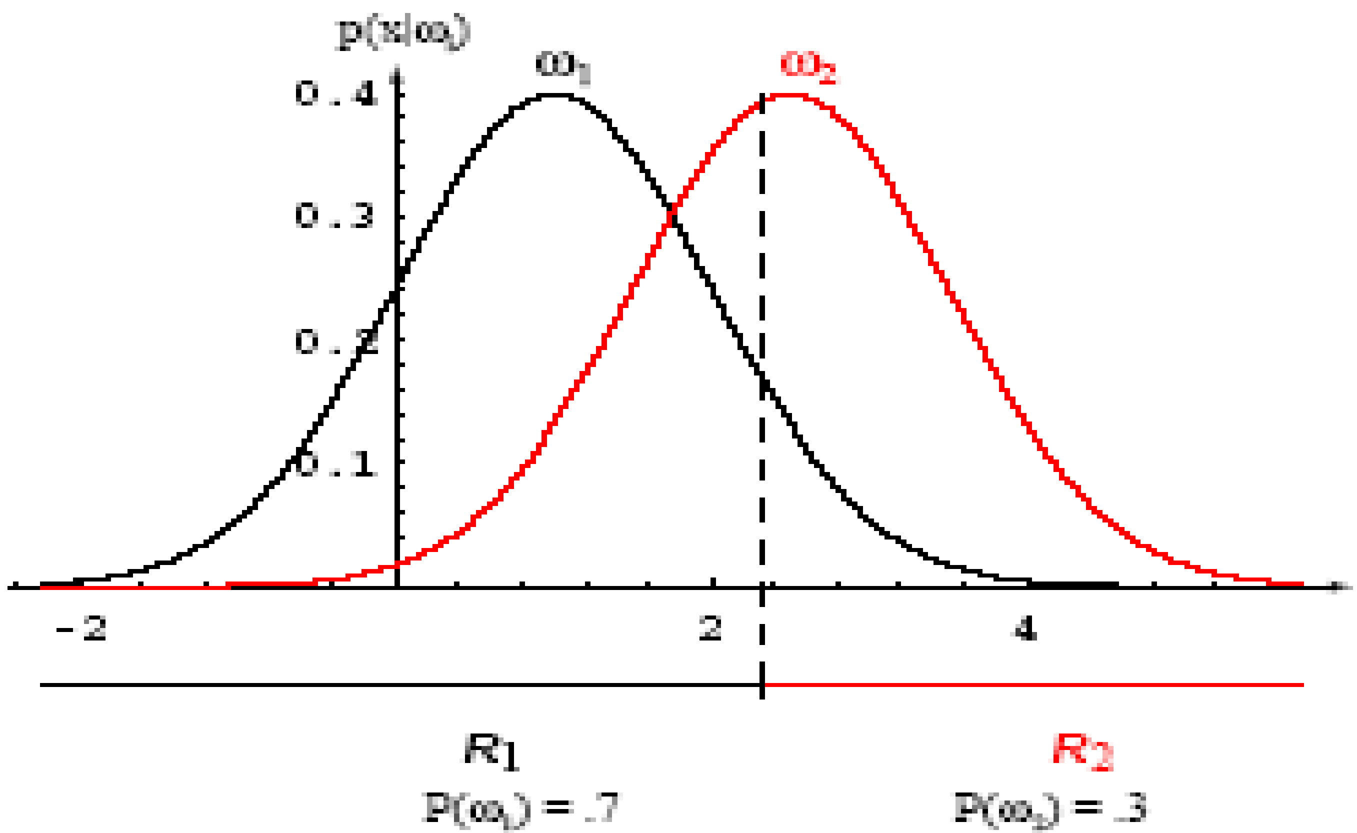
两类问题，高维特征，先验概率相同时：





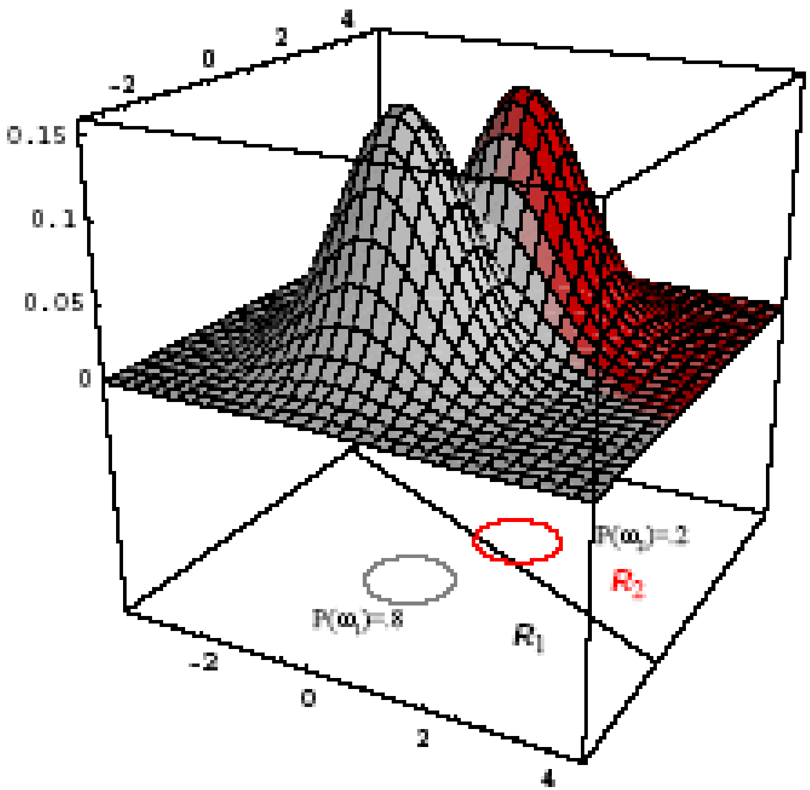
线性分类器

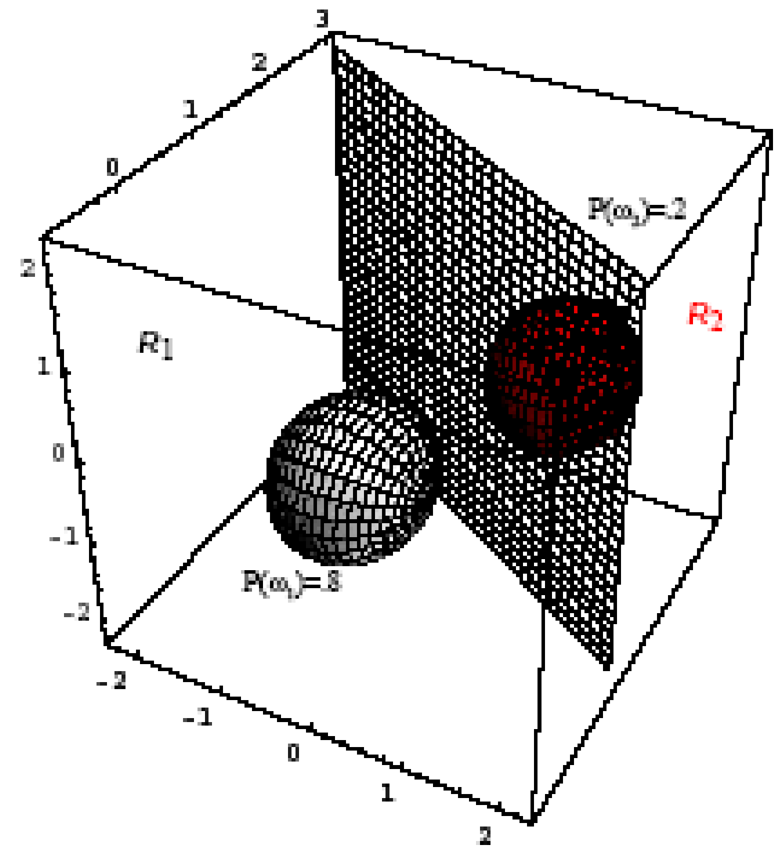
两类问题，1维特征，先验概率不同时：



线性分类器

两类问题，高维特征，先验概率不同时：





情况三：$\boldsymbol{\Sigma}_i$任意

判别函数可以写成:

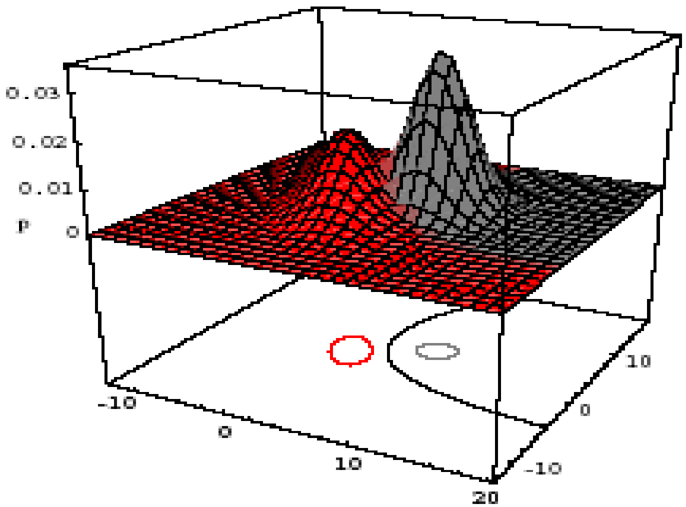
$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_i\right| + \ln P(\omega_i)$$

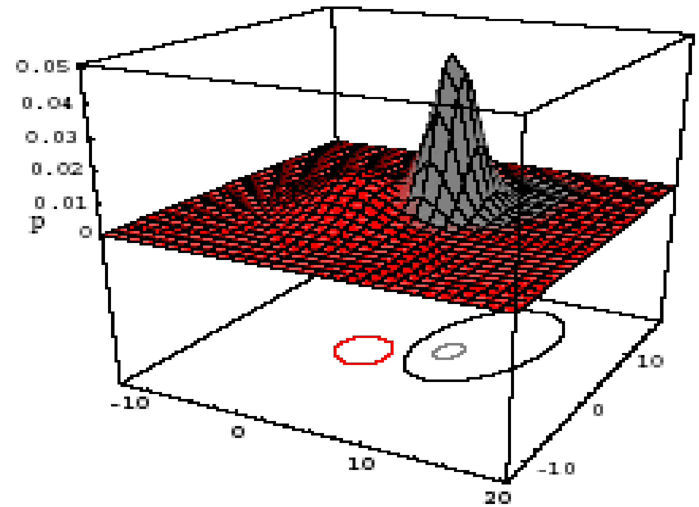
$$\Downarrow$$

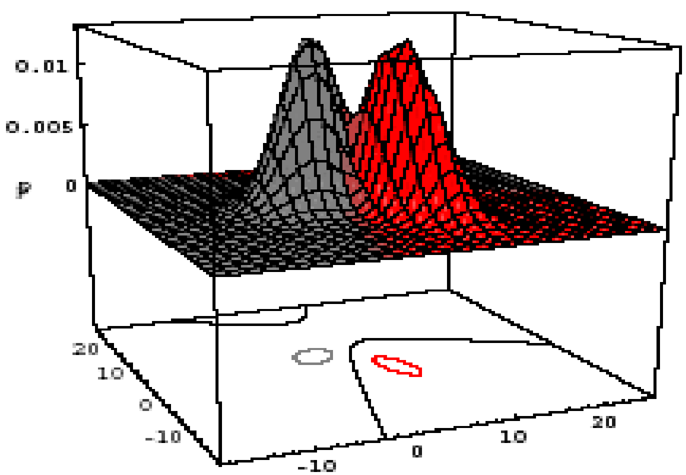
$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^t\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \left(-\frac{1}{2}\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln\left|\boldsymbol{\Sigma}_i\right| + \ln P(\omega_i)\right)$$

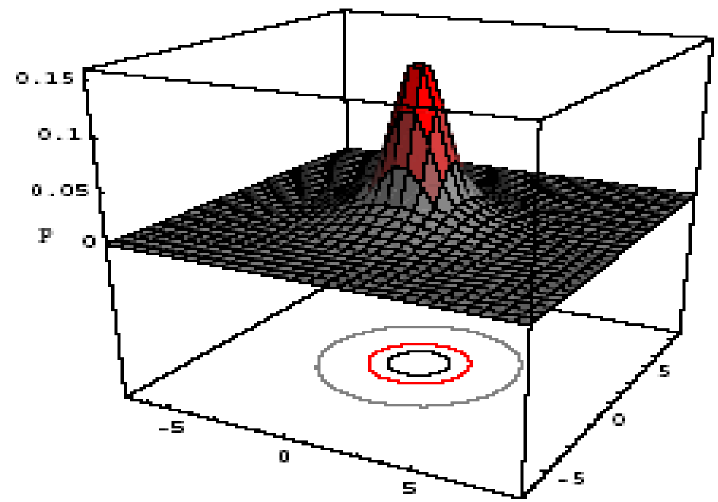
分类界面为2次曲线（面）

二次分类曲线



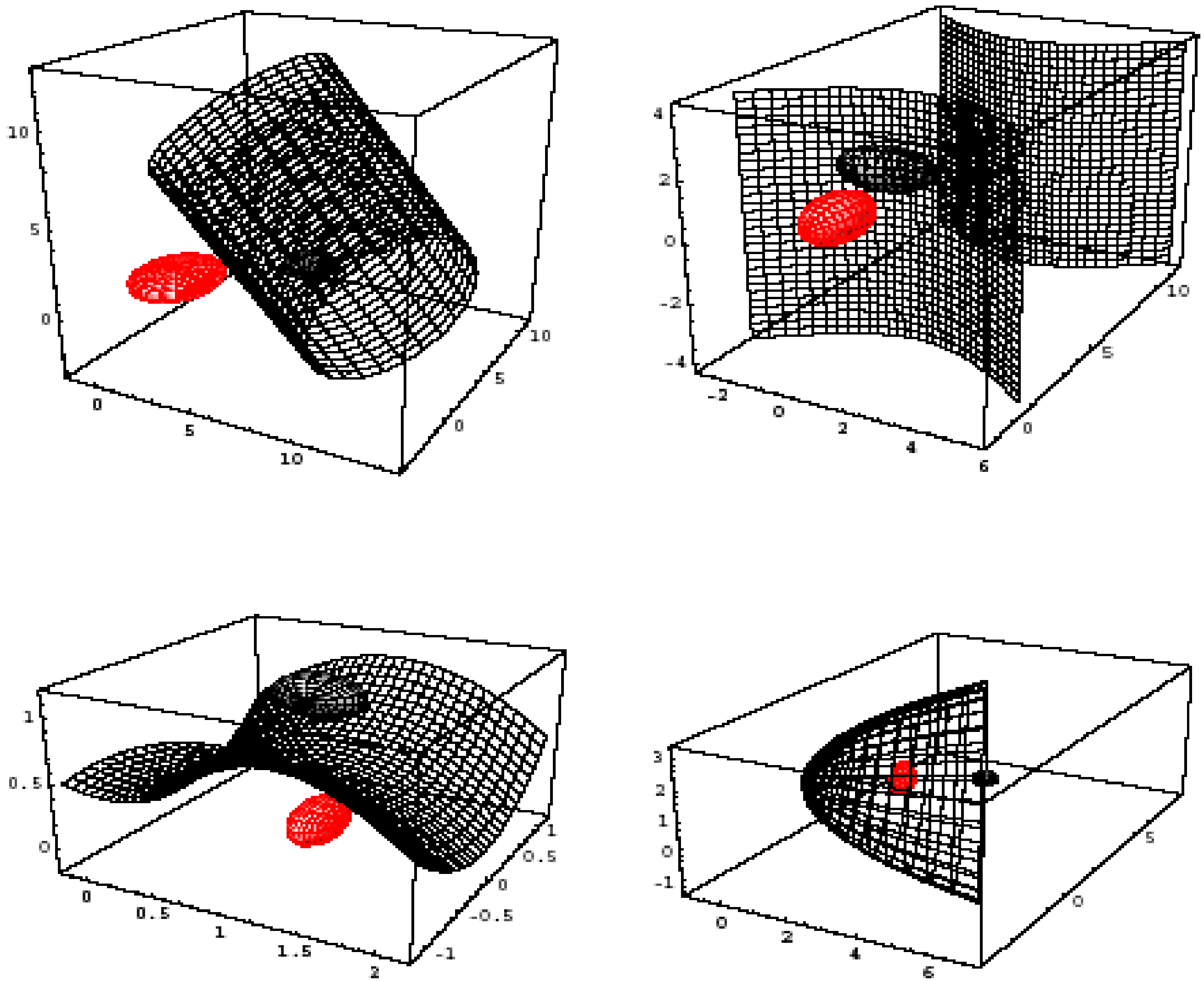






ISBN 978-7-5603-4763-9

二次分类曲面


10
5
0
0
5
10
0
5
10
4
2
0
-2
-4
-2
0
2
4
6
0
5
10
1
0.5
0
0
0.5
1
1.5
2
-1
-0.5
0
0.5
1
3
2
1
0
-1
0
2
4
6
0
5

例:正态分布二次分类曲面举例

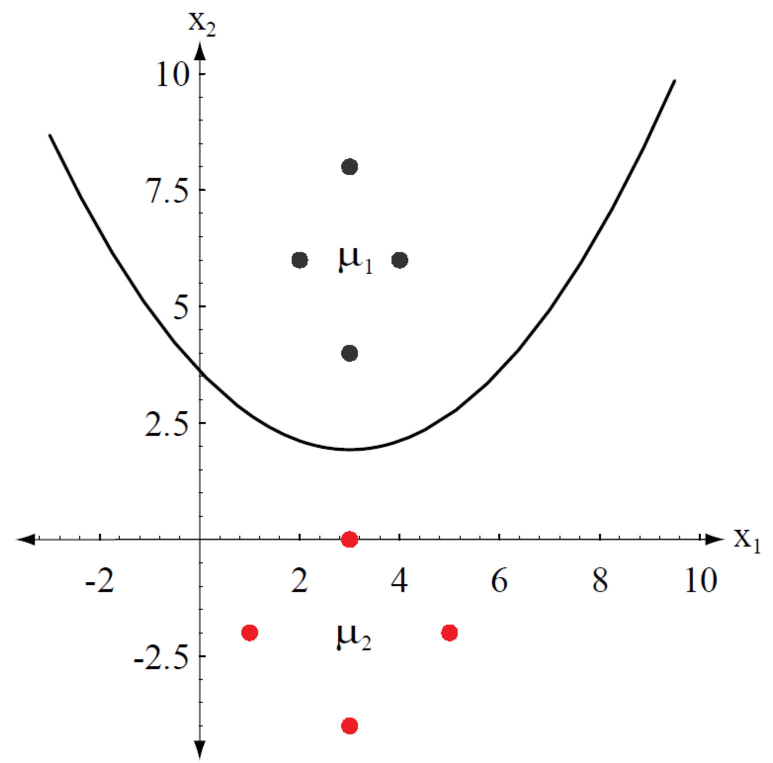
$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^t\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \left(-\frac{1}{2}\boldsymbol{\mu}_i^t\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)\right)$$

$$p(\omega_1) = p(\omega_2) = 0.5$$

$$\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \Sigma_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\Sigma_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \Sigma_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$



$$g_1(\mathbf{x}) = g_2(\mathbf{x}) \quad \Rightarrow \quad x_2 = 3.514 - 1.125x_1 + 0.1875x_1^2$$

ISBN 978-7-5603-4763-9

朴素贝叶斯分类器

难估计协方差矩阵：特征的维数较高、训练样本数量较少时，无法有效估计协方差矩阵

假设各维独立：假设各维特征服从相互独立的高斯分布

$$p(\mathbf{x}|\omega_i)=\prod_{j=1}^{d}p\left(\underline{x_j|\omega_i}\right)=\prod_{j=1}^{d}\left\{\frac{1}{\sqrt{2\pi}\sigma_{ij}}\exp\left[-\frac{\left(x_j-\mu_{ij}\right)^2}{2\sigma_{ij}^2}\right]\right\}$$

x_j：第j维特征

ω_i：第i类别

对数判别函数为

$$g_i(\mathbf{x})=\ln p(\mathbf{x}|\omega_i)+\ln P(\omega_i)$$

$$=\sum_{j=1}^{d}\left\{-\frac{1}{2}\ln 2\pi-\ln\sigma_{ij}-\frac{\left(x_j-\mu_{ij}\right)^2}{2\sigma_{ij}^2}\right\}+\ln P(\omega_i)$$

$$=\cancel{-\frac{d}{2}\ln 2\pi}-\sum_{j=1}^{d}\ln\sigma_{ij}-\sum_{j=1}^{d}\frac{\left(x_j-\mu_{ij}\right)^2}{2\sigma_{ij}^2}+\ln P(\omega_i)$$

改进二次判别函数（MQDF）

协方差矩阵不可逆、或特征值太小计算不稳定时怎么办?

对协方差矩阵进行特征值分解（主成分分析思想）

保留大特征值对应的主要成分

用常数替代小特征值，保证计算稳定

协方差矩阵特征值分解:

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{j=1}^{d} \underline{\lambda_j \mathbf{v}_j} \mathbf{v}_j^T$$

λ_j :特征值

$\mathbf{v}_j$:特征向量

协方差矩阵的逆:

$$\mathbf{\Sigma}_i^{-1} = \left(\mathbf{Q}_i\mathbf{\Lambda}_i\mathbf{Q}_i^T\right)^{-1} = \mathbf{Q}_i\mathbf{\Lambda}_i^{-1}\mathbf{Q}_i^T = \sum_{j=1}^{d} \frac{1}{\lambda_{ij}} \mathbf{v}_{ij}\mathbf{v}_{ij}^T$$

$$g_i(\mathbf{x}) = -\frac{1}{2}\underline{(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)} \cancel{-\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)$$

$$\boldsymbol{\Sigma}_i^{-1} = \sum_{j=1}^{d}\frac{1}{\lambda_{ij}}\mathbf{v}_{ij}\mathbf{v}_{ij}^t \qquad |\boldsymbol{\Sigma}_i| = \prod_{j=1}^{d}\lambda_{ij}$$

$$= -\frac{1}{2}\sum_{j=1}^{d}\frac{1}{\lambda_{ij}}(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{v}_{ij}\mathbf{v}_{ij}^T(\mathbf{x}-\boldsymbol{\mu}_i) - \underline{\frac{1}{2}\sum_{j=1}^{d}\ln\lambda_{ij}} + \ln P(\omega_i)$$

$$= \underline{-\frac{1}{2}\sum_{j=1}^{d}\frac{1}{\lambda_{ij}}\left[(\mathbf{x}-\boldsymbol{\mu}_i)^t\mathbf{v}_{ij}\right]^2} - \underline{\frac{1}{2}\sum_{j=1}^{d}\ln\lambda_{ij}} + \ln P(\omega_i)$$

令$y_j = (\mathbf{x}-\boldsymbol{\mu}_i)^T\mathbf{v}_{ij}$,代表$\mathbf{x}$在新坐标系上第$j$轴的坐标

只保留前k个最大的特征值，其他用一小正数δ_i替代

$$= -\frac{1}{2}\sum_{j=1}^{k}\frac{y_j^2}{\lambda_{ij}} - \frac{1}{2\delta_i}\underline{\sum_{j=k+1}^{d}y_j^2} - \frac{1}{2}\sum_{j=1}^{k}\ln\lambda_{ij} - \frac{d-k}{2}\ln\delta_i + \ln P(\omega_i)$$

$$= \sum_{j=1}^{d}y_j^2 - \sum_{j=1}^{k}y_j^2 = \|\mathbf{x}-\boldsymbol{\mu}_i\|^2 - \sum_{j=1}^{k}y_j^2$$

$$= -\frac{1}{2}\sum_{j=1}^{k}\frac{y_j^2}{\lambda_{ij}} - \frac{1}{2\delta_i}\left(\|\mathbf{x}-\boldsymbol{\mu}_i\|^2 - \sum_{j=1}^{k}y_j^2\right) - \frac{1}{2}\sum_{j=1}^{k}\ln\lambda_{ij} - \frac{d-k}{2}\ln\delta_i + \ln P(\omega_i)$$

$$g_i(\mathbf{x}) = -\frac{1}{2}\sum_{j=1}^{k}\frac{y_j^2}{\lambda_{ij}} - \frac{1}{2\delta_i}\left(\left\|\mathbf{x}-\boldsymbol{\mu}_i\right\|^2 - \sum_{j=1}^{k} y_j^2\right)$$

$$-\frac{1}{2}\sum_{j=1}^{k}\ln\lambda_{ij} - \frac{d-k}{2}\ln\delta_i + \ln P(\omega_i)$$

- 降低算法、存储复杂度 $k \ll d$

计算k维坐标$y_j = (\mathbf{x}-\boldsymbol{\mu}_i)^t \mathbf{v}_{ij}$，计算一次向量范数$\left\|\mathbf{x}-\boldsymbol{\mu}_i\right\|^2$

- 先验概率的估计

$$P(\omega_i) = n_i / n$$

- 保留特征值个数k的选定：

$$\sum_{k=1}^{j}\lambda_{ik} \Big/ \sum_{k=1}^{d}\lambda_{ik} \ge 95\%$$

作业：态分布的N-P判决

N-P判决举例：设两类问题中，二维样本均为正态分布，有 $\boldsymbol{\mu}_1=\left(-1,0\right)^t,\boldsymbol{\mu}_2=\left(1,0\right)^t,\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\mathbf{I}$，取$\varepsilon_{21}$=0.046，求N-P判决界面