请大家预习

第2章

贝叶斯决策理论

Bayesian Decision Theory

向 世 明

smxiang@nlpr.ia.ac.cn

http://people.ucas.ac.cn/~xiangshiming

时空数据分析与学习课题组(STDAL)

中科院自动化研究所 模式识别国家重点实验室

助教： 杨 奇 (yangqi2021@ia.ac.cn)

张 涛 (zhangtao2021@ia.ac.cn)

内容提要

- 介绍
- 最小错误率贝叶斯决策
- 最小风险贝叶斯决策
- 分类器设计

2.1.1 贝叶斯公式

问题：给定 事件B 的基础上，希望计算 事件A 发生的概率 $P(A|B)$

局限：我们可能只知道 $P(B|A)$

举例： 在某些证据下，希望知道某个事件发生的可能性有多大。比如，想知道**一个人的血液检查结果为阳性，那么得病的概率有多大**？

但是，我们只能知道**在得病的条件下，血液检查结果呈阳性的概率为99%**，即在给定事件下，知道证据发生的概率。

贝叶斯公式可以在知道$P(B|A)$的情况下，计算出$P(A|B)$

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

- ✓ $P(A|B)$： 后验概率 (posterior)，需要结合先验概率和证据才能知道
- ✓ $P(B|A)$： 似然 (likelihood)，在A发生的情况下，B(或evidence)的概率
- ✓ $P(A)$： 先验概率 (prior)，事件A发生的概率有多大
- ✓ $P(B)$： 证据 (evidence)，即无论事件如何，B(或evidence)的可能性

2.1.2 两类的例子

- **HIV血液检测**

设某市某人做了一次HIV血液检测，结果检测是阳性。

病人：“检测的灵敏度如何？”

医生：“非常灵敏99%，而且误检率也很低约1%。”

病人：“某市现在大约有百分之几的HIV患者？”

医生：“某市大约有万分之一的HIV患者。”

数学家：“这个人真的携带HIV的可能性是多大呢？”

- **采用贝叶斯分析方法**
  - **H：HIV患者，　H=Y (第一类ω_1)；　or H=N (第二类ω_2)**
  - **T：血液检测，　T=阳；or T=阴**
- **需要计算：**　　$P(\omega_1 \mid T = 阳) = ?$
- **贝叶斯公式：**

$$P(H \mid T) = \frac{P(H,T)}{P(T)} = \frac{P(T \mid H)P(H)}{P(T)}$$

全概率公式

$$= \frac{P(T \mid H)P(H)}{P(T \mid H = 阳)P(H = 阳) + P(T \mid H = 阴)P(H = 阴)}$$

(简记)

$$= \frac{P(T \mid H)P(H)}{P(T \mid \omega_1)P(\omega_1) + P(T \mid \omega_2)P(\omega_2)}$$

- **贝叶斯公式的具体应用：**

$$P(\omega_1 \mid T=\text{阳}) = \frac{P(T=\text{阳}\mid\omega_1)P(\omega_1)}{P(T=\text{阳}\mid\omega_1)P(\omega_1)+P(T=\text{阳}\mid\omega_2)P(\omega_2)}$$

$$= \frac{0.99\times 0.0001}{0.99\times 0.0001+0.01\times 0.9999}$$

$$= 0.98\%$$

误检率（假阳性）

- **我们需要知道的几个信息：**
  - 先验概率：HIV发生的先验概率 $P(\omega_1)$； $P(\omega_1)+P(\omega_2)=1$
  - 条件概率(灵敏度)：当HIV为真时，检测为阳性的概率 $P(T=\text{阳}\mid\omega)$；
  - 条件概率(误检率)：当HIV为假时，检测为阳性的概率 $P(T=\text{阳}\mid\omega_2)$；

2.1.2 两类的例子

- 更多的例子

✓ 生命科学家用贝叶斯法则来研究基因是如何被控制的；

✓ 基金经理用贝叶斯法则找到投资策略；

✓ 早期，Google用贝叶斯法则改进搜索功能，帮助用户过滤垃圾邮件；

✓ 无人驾驶汽车接收车顶传感器搜集到的路况和交通数据，运用贝叶斯法则从地图上获得信息；

✓ 人工智能、机器翻译中大量用到贝叶斯法则。

✓ 所有这些应用的原理都是一样的：

- 如果掌握事物的全部信息，则可计算一个客观概率。
- 但绝大多数决策面临的信息是不全的，仅有有限几个证据。
- 贝叶斯分析方法的精神在于，既然无法得到全面的信息，在证据有限的情况下，尽可能地做一个更好的判断。

2.1.3 识别两种鱼

- 教室门口判断进来的是男生还是女生，没有任何传感器（证据）？
- **两种鱼的分类**
  - ω 是随机变量， $\omega = \omega_1$： 鲈鱼； $\omega = \omega_2$： 三文鱼；
  - 引入先验概率$P(\omega_1)$ 和$P(\omega_2)$ 。
    - 先验概率反映了在实际的鱼没有出现之前，所拥有的对于可能出现哪种鱼的先验知识。这一点是可以做到的，因为渔场地点不同、季节不同，鲈鱼和三文鱼的数量并不相同。
    - $P(\omega_1) + P(\omega_2) =1$
  - 如果仅采用先验概率来进行决策，会得到什么样的决策规则？
    - $\omega = \omega_1$， if $P(\omega_1) > P(\omega_2)$; otherwise $\omega = \omega_2$

2.1.3 识别两种鱼

- **决策的好坏：此时只能取决于先验概率的值：**
  - 如果 $P(\omega_1) >> P(\omega_2)$，则判定为 ω_1在大多数情况下是对的。
  - **错误率：** $$P(error) = \begin{cases} P(\omega_2) & \text{if we decide } \omega_1 \\ P(\omega_1) & \text{if we decide } \omega_2 \end{cases}$$
  - **如何改进：**
    - 引入观测信息，比如鱼的长度 x。
    - 引入类条件概率密度函数：x为一个连续随机变量，其分布因类别而不同。因此表示为$p(x|\omega)$，即类别为ω时的概率密度函数。

2.1.3 识别两种鱼

- **贝叶斯分析：**
  - 某个模式属于类别ω_i并具有特征值x的联合概率密度：

$$p(\omega_i, x) = p(\omega_i \mid x)\ p(x) = p(x \mid \omega_i)\ P(\omega_i)$$

  - 于是可得贝叶斯公式：

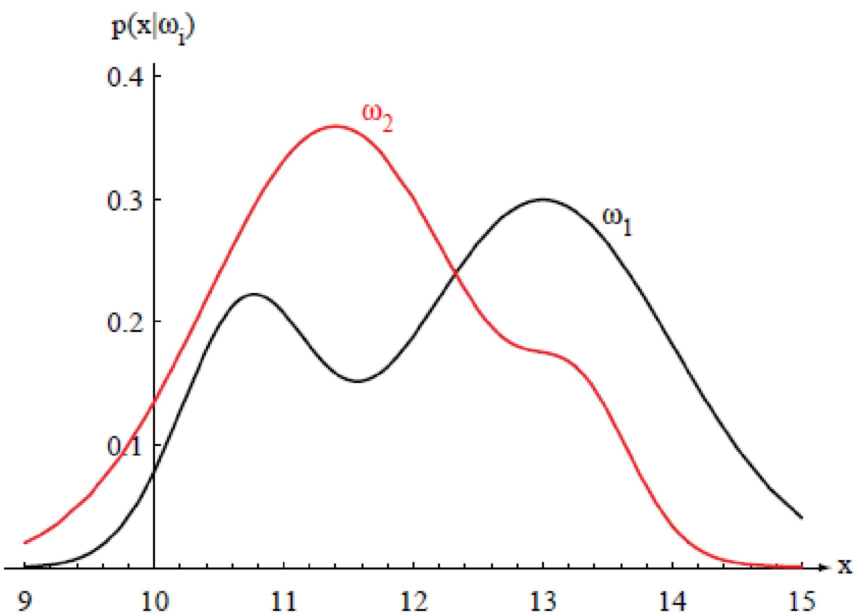
$$p(\omega_i \mid x) = \frac{p(\omega_i, x)}{p(x)} = \frac{\overbrace{p(x \mid \omega_i)}^{\text{likelihood}}\overbrace{P(\omega_i)}^{\text{prior}}}{\underbrace{p(x)}_{\text{evidenece}}}$$

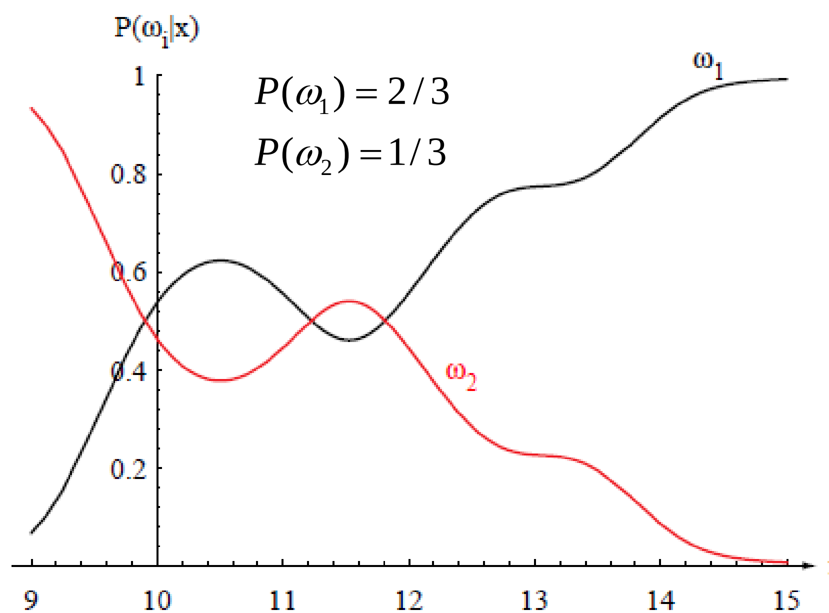
其中，$p(x) = \sum_{i=1}^{2} p(x \mid \omega_i) P(\omega_i)$

- **贝叶斯决策规则：**
  - **$\omega = \omega_1$， if $P(\omega_1|\mathrm{x}) > P(\omega_2|\mathrm{x})$; otherwise $\omega = \omega_2$**

2.1.3 识别两种鱼

- **有传感器（特征）的情况：** Decision based on posterior probabilities
- **鲈鱼和三文鱼的条件概率密度函数：**





x轴：一维特征空间

$$P(\omega_i\,|\,\mathbf{x})=\frac{p(\mathbf{x}\,|\,\omega_i)P(\omega_i)}{p(\mathbf{x})}=\frac{p(\mathbf{x}\,|\,\omega_i)P(\omega_i)}{\sum_{j=1}^{c}p(\mathbf{x}\,|\,\omega_j)P(\omega_j)}$$

2.2 最小错误率贝叶斯决策

- **问题描述**
  - 类别： $\omega_i, i=1,\ldots,c$
  - 特征矢量 $\mathbf{x}=[x_1,\ldots,x_d]\in R^d$
  - 已知：先验概率 $P(\omega_i)$, $\sum_{i=1}^{c}P(\omega_i)=1$
  - 已知：概率密度函数（条件概率） $p(\mathbf{x}\,|\,\omega_i)$
  - 任务：如果观测到一个样本**x**，应该将其分到哪一类才最合理呢？

2.2 最小错误率贝叶斯决策

- **贝叶斯决策**
  - **后验概率：**

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{p(\mathbf{x})} = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\mathbf{x} \mid \omega_j)P(\omega_j)}, \qquad \sum_{i=1}^{c} P(\omega_i \mid \mathbf{x}) = 1$$

  - **决策规则：**

如果 $p(\omega_i \mid \mathbf{x}) = \max_{j=1,2,\ldots,c} p(\omega_j \mid \mathbf{x})$，则 $\mathbf{x} \in \omega_i$

2.2 最小错误率贝叶斯决策

- **贝叶斯决策的几种等价形式：**
  - **Evidence (a.k.a. Likelihood)**

$$\text{if}\;\; p(\mathbf{x}\,|\,\omega_1)P(\omega_1) > p(\mathbf{x}\,|\,\omega_2)P(\omega_2),$$
$$\text{then}\;\mathbf{x}\in\omega_1;\;\; \text{otherwise}\;\mathbf{x}\in\omega_2$$

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

  - 其它：

$$\text{if}\;\; l(\mathbf{x}) = \frac{p(\mathbf{x}\,|\,\omega_1)}{p(\mathbf{x}\,|\,\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)},\;\; \text{then}\;\mathbf{x}\in\omega_1;\;\; \text{otherwise}\;\mathbf{x}\in\omega_2$$

$$\text{if}\;\; -\ln\big(l(\mathbf{x})\big) = -\ln\big(p(\mathbf{x}\,|\,\omega_1)\big) + \ln\big(p(\mathbf{x}\,|\,\omega_2)\big) < -\ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right),$$
$$\text{then}\;\mathbf{x}\in\omega_1;\;\; \text{otherwise}\;\mathbf{x}\in\omega_2$$

2.2 最小错误率贝叶斯决策

- **一个例子**
  - 假设在某个局部地区细胞中正常(ω_1)和异常(ω_2) 两类的先验概率为$P(\omega_1)$=0.9 和 $P(\omega_2)$=0.1。现有一待识别细胞，其观测值为x，从类条件概率密度分布曲线中查得：$P(x\,|\omega_1)$=0.2， $P(x\,|\omega_2)$=0.4。试对该细胞进行分类。
- **解：**

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

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

因为 $P(\omega_1|x)>P(\omega_2|x)$，根据贝叶斯决策规则，所以合理的决策是将$x$归为正常类。

2.3 最小风险贝叶斯决策

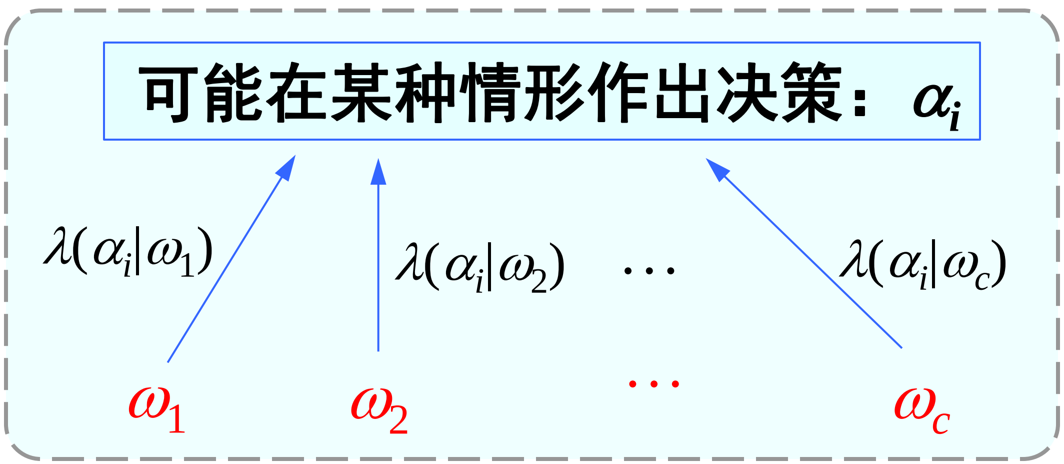
- 引言
  - 在医疗诊断、工业检测、自动商店性别判断等应用环境中，仅仅使决策的错误率最小还不足够。还应该对做出判断后才产生的风险进行考虑。
  - 风险和损失总是关联在一起的。
  - 在决策论中，称采取的决定为决策或行为。所有可能采取的各种决策组成的集合称为决策空间或行动空间。
  - 每个决策或行为均会带来一定的损失，它通常是决策和自然状态（类别）的函数。

2.3 最小风险贝叶斯决策

- **问题描述**
  - 类别： $\omega_i, i=1,\ldots,c$
  - 特征矢量 $\mathbf{x}=[x_1,\ldots,x_d]\in R^d$
  - 已知：先验概率 $P(\omega_i)$, $\sum_{i=1}^{c}P(\omega_i)=1$
  - 已知：概率密度函数(条件概率) $p(\mathbf{x}\mid\omega_i)$
  - 已知：决策空间包含a个决策α_i, i=1,2,…, a
  - 已知：损失函数$\lambda(\alpha_i|\omega_j)$，表示当类别为$\omega_j$所采取的决策$\alpha_i$所引起的损失，简记为$\lambda_{ij}$。
  - 任务：如果观测到一个样本x，应该将其分到哪一类其风险最小？

2.3 最小风险贝叶斯决策

- **最小风险决策**
  - 由于引入了“损失”，在考虑错判（错误决策）所造成的损失时，不能仅仅根据后验概率大小来作为决策，而必须考虑所采取的决策是否使损失最小。
  - 对于给定的样本**x**，考虑对其作出决策α_i，但具体是在x的哪个类别状态下作出决策目前并不知道，因此需考虑“条件风险”（加权平均）。



2.3 最小风险贝叶斯决策

- **条件风险**
  - 条件风险，即条件期望风险$R(\alpha_i|\mathrm{x})$，可计算如下：

$$R(\alpha_i\,|\,\mathbf{x}) = E[\lambda(\alpha_i\,|\,\omega_j)] = \sum_{j=1}^{c}\lambda(\alpha_i\,|\,\omega_j)P(\omega_j\,|\,\mathbf{x}),\quad i=1,2,\ldots,a$$

  - $R(\alpha_i|\mathbf{x})$是随机变量**x**的函数。
- **期望风险**
  - 将决策规则视为随机变量**x**的函数，记为$\alpha(\mathbf{x})$。对特征空间中所有可能的样本 **x** 采取的决策所造成的期望损失(平均风险)是

$$R(\alpha) = \int R(\alpha(\mathbf{x})|\mathbf{x})\cdot p(\mathbf{x})\cdot d\mathbf{x}$$

  - 最小风险贝叶斯决策就是最小化期望风险：$\min_{\alpha} R(\alpha)$
    - ✓ 期望风险$R(\alpha)$：反映对整个特征空间上所有样本所采取的相应决策所带来的平均风险；
    - ✓ 条件风险$R(\alpha_i|\mathbf{x})$：只反映对样本**x**取值采取决策α_i所带来的风险。

2.3 最小风险贝叶斯决策

- **计算步骤**
  - 利用贝叶斯公式计算后验概率：$P(\omega_j|\mathrm{x})$, j=1,2,…c
  - 利用决策计算风险

$$R(\alpha_i\,|\,\mathbf{x}) = E[\lambda(\alpha_i\,|\,\omega_j)] = \sum_{j=1}^{c}\lambda(\alpha_i\,|\,\omega_j)P(\omega_j\,|\,\mathbf{x}),\quad i=1,2,\ldots,a$$

  - 在各种决策中选择风险最小的决策:

$$a = \arg\min_{j=1,\ldots,a} R(\alpha_j\,|\,\mathbf{x})$$

2.3 最小风险贝叶斯决策

- **两类情形**
  - **考察两类情形，且没有拒识(a=c=2):**

$$R(\alpha_1|\mathbf{x}) = \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$$
$$R(\alpha_2|\mathbf{x}) = \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$$

  - **决策规则:**
    - **if $R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$, then** $\alpha = \alpha_1$; otherwise $\alpha = \alpha_2$

$$\text{if } \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x}) < \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x}) \quad \Rightarrow \quad \mathbf{x} \in \omega_1$$
$$\text{if } \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x}) > \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x}) \quad \Rightarrow \quad \mathbf{x} \in \omega_2$$

将第一类分为第一类的损失

将第二类分为第一类的损失

将第一类分为第二类的损失

将第二类分为第二类的损失

2.3 最小风险贝叶斯决策

- **其它等价形式**
  - 不失一般性，可以假设$\lambda_{11} < \lambda_{21}$, $\lambda_{22} < \lambda_{12}$，于是有：

$$(\lambda_{11} - \lambda_{21})P(\omega_1 | \mathbf{x}) < (\lambda_{22} - \lambda_{12})P(\omega_2 | \mathbf{x}) \quad \Rightarrow \quad \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

$$\frac{P(\omega_1 | \mathbf{x})}{P(\omega_2 | \mathbf{x})} = \frac{p(\mathbf{x} | \omega_1)P(\omega_1)}{p(\mathbf{x} | \omega_2)P(\omega_2)} > \frac{\lambda_{22} - \lambda_{12}}{\lambda_{11} - \lambda_{21}} \quad \Rightarrow \quad \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

$$l(\mathbf{x}) = \frac{p(\mathbf{x} | \omega_1)}{p(\mathbf{x} | \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \cdot \frac{\lambda_{22} - \lambda_{12}}{\lambda_{11} - \lambda_{21}} \quad \Rightarrow \quad \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

- **例子**
  - 假设在某个局部地区细胞中正常(ω_1)和异常(ω_2) 两类的先验概率为$P(\omega_1)$=0.9 和 $P(\omega_2)$=0.1。现有一待识别细胞，其观测值为x，从类条件概率密度分布曲线中查得： $P(x|\omega_1)$=0.2， $P(x|\omega_2)$=0.4。已知决策风险λ_{11}=0, λ_{12}=6, λ_{22}=0, λ_{21}=1。试对该细胞进行分类。
  - 解：
    - 前面已经解得$P(\omega_1|x)$=0.818; $P(\omega_2|x)$=0.182; 进一步计算条件风险：

$$R(\alpha_1 \mid \mathbf{x}) = \lambda_{11}P(\omega_1 \mid \mathbf{x}) + \lambda_{12}P(\omega_2 \mid \mathbf{x}) = 0\times 0.818 + 6\times 0.182 = 1.092$$

$$R(\alpha_2 \mid \mathbf{x}) = \lambda_{21}P(\omega_1 \mid \mathbf{x}) + \lambda_{22}P(\omega_2 \mid \mathbf{x}) = 1\times 0.818 + 0\times 0.182 = 0.818$$

    - 由于$R(\alpha_2|\mathrm{x}) < R(\alpha_1|\mathrm{x})$，即决策为$\omega_2$的条件风险小于决策为$\omega_1$的条件风险，因此采取决策行动$\alpha_2$，即判定待别的细胞为异常细胞。

2.3 最小风险贝叶斯决策

- **Zero-One Loss**

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \qquad i, j = 1, 2, \dots, c$$

$$\begin{aligned} R(\alpha_i \mid \mathbf{x}) &= \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x}) \\ &= \sum_{i \neq j} P(\omega_j \mid \mathbf{x}) \\ &= 1 - P(\omega_i \mid \mathbf{x}) \end{aligned}$$

- Minimum error decision: Maximum a posteriori (MAP):

$$\text{decide } \omega_i \text{ if } P(\omega_i \mid \mathbf{x}) > P(\omega_j \mid \mathbf{x}) \text{ for all } j \neq i$$

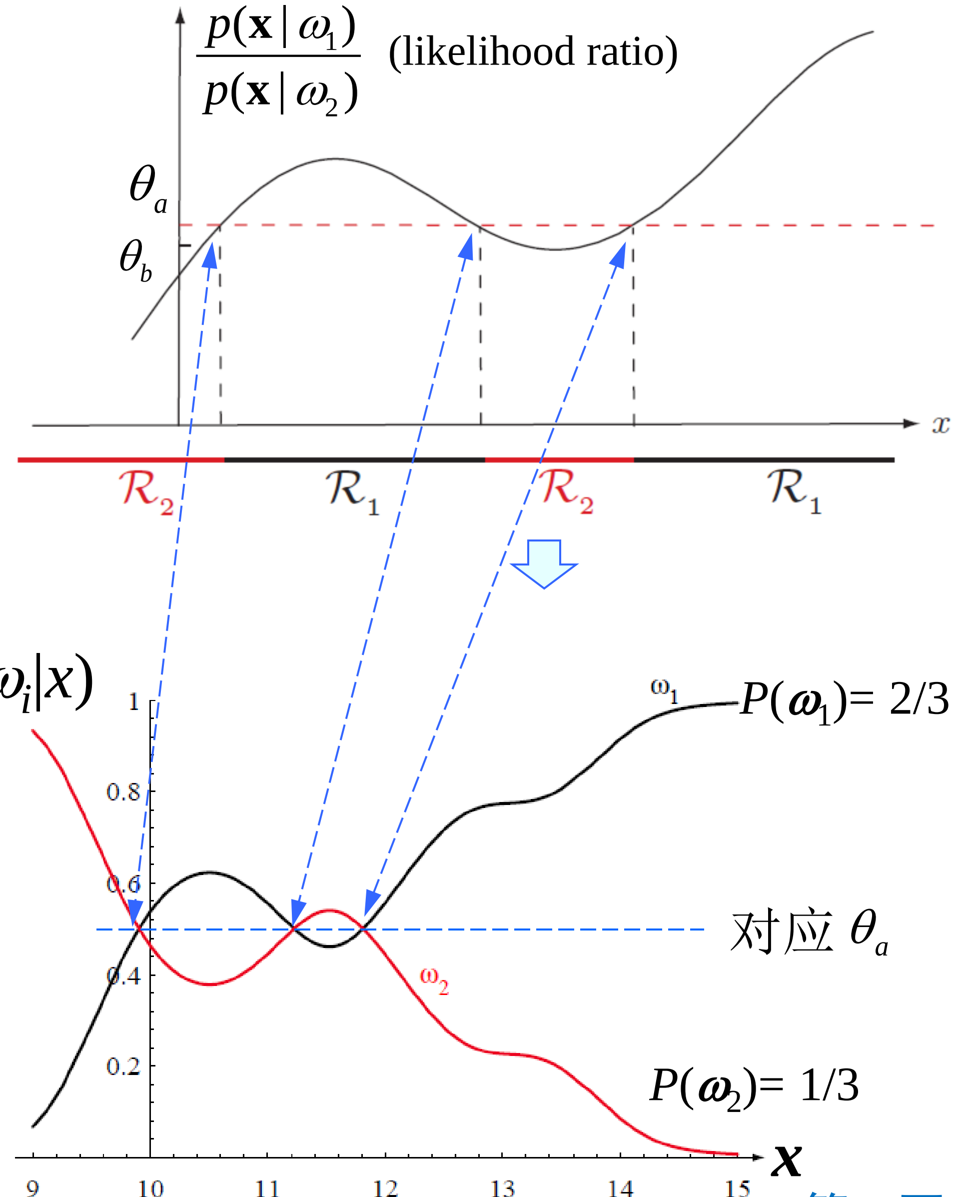
2.3 最小风险贝叶斯决策

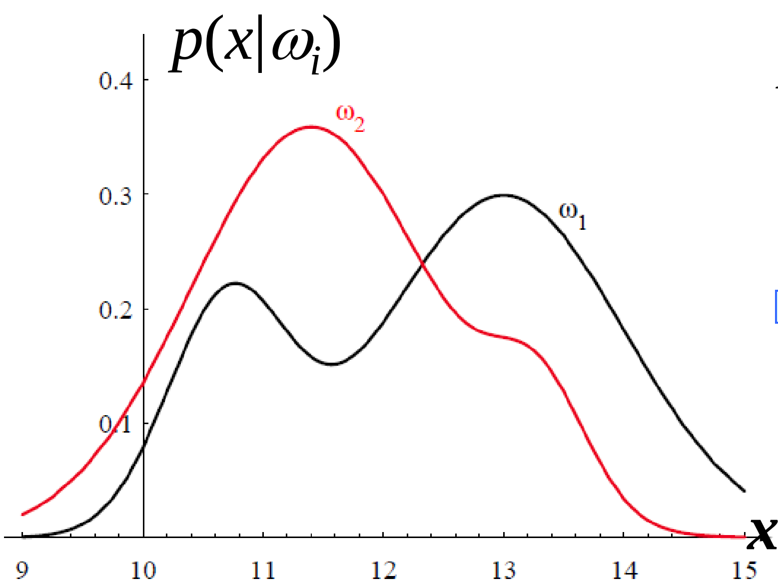
if $\dfrac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \dfrac{P(\omega_2)}{P(\omega_1)} \cdot \dfrac{\lambda_{22}-\lambda_{12}}{\lambda_{11}-\lambda_{21}} = \theta_a$

then $\mathbf{x} \in \omega_1$; otherwise $\mathbf{x} \in \omega_2$

for 0-1loss: $\dfrac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \dfrac{P(\omega_2)}{P(\omega_1)} = \theta_a$

$p(\omega_1|\mathbf{x}) > p(\omega_2|\mathbf{x})$



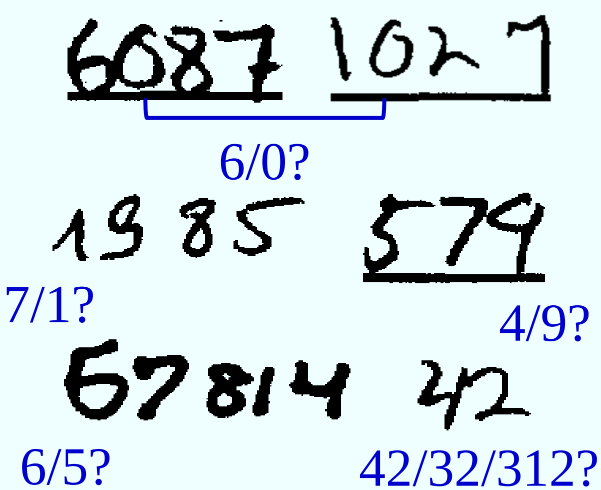


University of Chinese Academy of Sciences

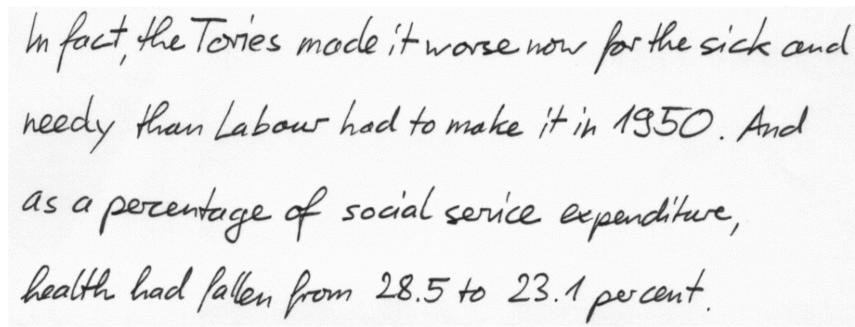
2.3 最小风险贝叶斯决策：带拒识的决策

- 为什么要拒识？错误识别可能带来严重后果
  - 是否每次一定要做出决策？在有的情况下，不做决策比做出错误率很大的决策会更好！
  - 比如：医疗诊断，金额识别

Ambiguity Rejection



Distance Rejection


In fact, the Tories made it worse now for the sick and needy than Labour had to make it in 1950. And as a percentage of social service expenditure, health had fallen from 28.5 to 23.1 percent.

English is outlier for a digit recognizer

経営不振に陥っているソニーのパソコン事業は国内 投資ファンド が買い取り、開発と製造をてがける新会社バイオをつくった。ソニーの国内向け通販サイトと同社の直営店を通じ、個人から注文をとってきた。

These are outliers for an English recognizer

带拒识的决策

分类器可以拒绝将样本判为 c 个类别中的任何一类

- Formulation: c+1 classes

假设： $$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0\,, & i = j \\ \lambda_s\,, & i \neq j \\ \lambda_r\,, & \text{reject} \end{cases}$$ （通常 $\lambda_r < \lambda_s$ ）

$$R(\alpha_i \mid \mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x})$$

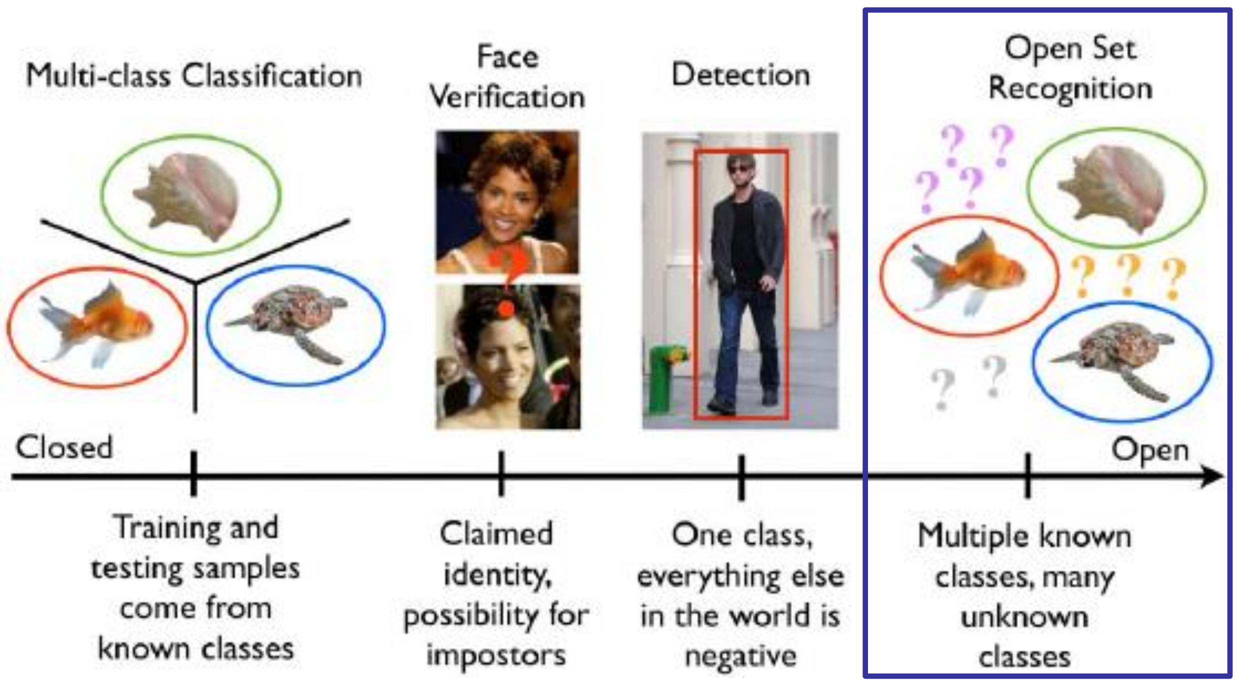
$$\Rightarrow \quad R_i(\mathbf{x}) \triangleq R(\alpha_i \mid \mathbf{x}) = \begin{cases} \lambda_s[1 - P(\omega_i \mid \mathbf{x})], & i = 1, \ldots, c \\ \lambda_r\,, & \text{reject} \end{cases}$$

$\Rightarrow$ 当 $\lambda_s[1 - P(\omega_i \mid \mathbf{x})] > \lambda_r$ 时，选择拒识。因此有以下决策规则：

$$\arg\min_i R_i(\mathbf{x}) = \begin{cases} \arg\max_i P(\omega_i \mid \mathbf{x}), & \text{if } \max_i P(\omega_i \mid \mathbf{x}) > 1 - \lambda_r / \lambda_s \\ \text{reject}, & \text{otherwise} \end{cases}$$

扩展：开放集识别的贝叶斯决策

- 传统的分类器假设训练样本和测试样本都来自预设的C个类别（闭合集，Closed set）。
- 实际环境中，测试样本可能不属于预设的C个类别（异常样本, outlier），这种情况称为开放集(Open set)。
- 开放集的难点是异常样本没有训练集，只能训练已知C类的分类器。



Closed set　　Open set

W.J. Scheirer, et al, Toward open set recognition, *IEEE Trans. PAMI*, 35(7): 1757-1772, 2013.

开放集分类贝叶斯决策

- 问题表示
  - 已知类别：$\omega_i, i=1,\dots,c$
  - 先验概率 $\sum_{i=1}^{c} P(\omega_i) \le 1$
  - 后验概率 $\sum_{i=1}^{c} P(\omega_i \mid \mathbf{x}) \le 1 \qquad \sum_{i=1}^{c+1} P(\omega_i \mid \mathbf{x}) = 1$
  - 条件概率密度 $p(\mathbf{x} \mid \omega_i), i=1,\dots,c, \quad p(\mathbf{x} \mid \omega_{c+1}) = ?$
- 分类决策
  - 假设 $p(\mathbf{x} \mid \omega_{c+1}) = \rho$ ρ为很小的常数
  - 后验概率 $P(\omega_i \mid \mathbf{x}) = \dfrac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{p(\mathbf{x})} = \dfrac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{\sum_{j=1}^{c+1} p(\mathbf{x} \mid \omega_j)P(\omega_j)}$
  - 最大后验概率决策 $\begin{cases} \text{in-class}, & \text{if } \max\limits_{i=1,\dots,c} p(\mathbf{x} \mid \omega_i)P(\omega_i) > \rho P(\omega_{c+1}) \\ \text{outlier}, & \text{otherwise} \end{cases}$

2.4 分类器设计

- **回顾：两类问题的最小错误率贝叶斯决策规则：**

$$\text{if } P(\omega_1 \mid \mathbf{x}) > P(\omega_2 \mid \mathbf{x}), \quad \text{then } \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

$$\text{if } p(\mathbf{x} \mid \omega_1)P(\omega_1) > p(\mathbf{x} \mid \omega_2)P(\omega_2), \quad \text{then } \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

$$\text{if } \frac{p(\mathbf{x} \mid \omega_1)}{p(\mathbf{x} \mid \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}, \quad \text{then } \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

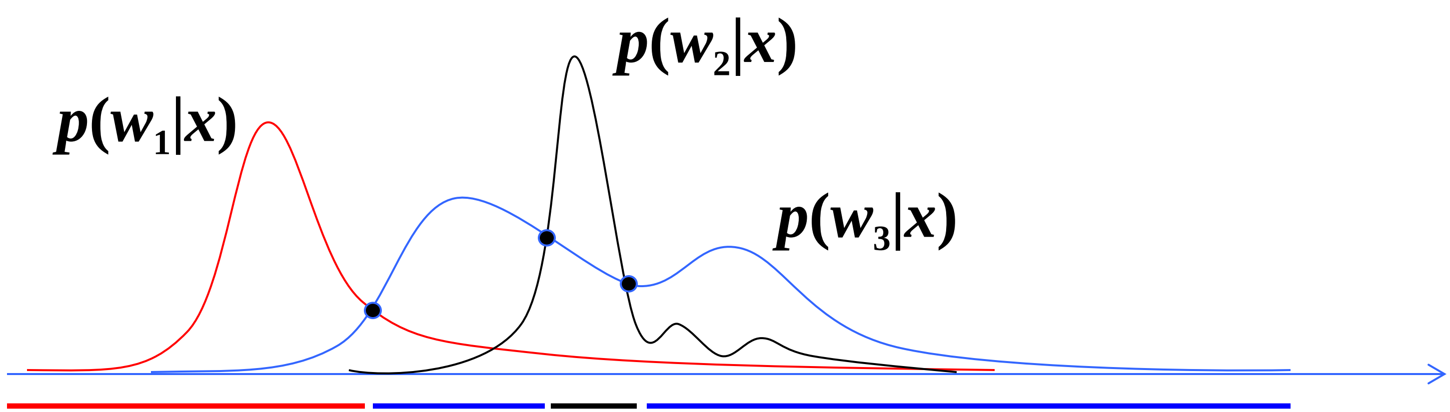
$$\text{if } \ln\big(p(\mathbf{x} \mid \omega_1)\big) - \ln\big(p(\mathbf{x} \mid \omega_2)\big) > \ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right), \quad \text{then } \mathbf{x} \in \omega_1; \ \text{otherwise } \mathbf{x} \in \omega_2$$

2.4 分类器设计

- **判别函数**
  - 用于表达决策规则的某些函数称为判别函数。
  - 判别函数：通常定义一组判别函数 $\mathbf{g}_i(\mathbf{x})$，$i=1,2,\ldots,c$ 用于表示多类决策规则。如果$\mathbf{g}_i(\mathbf{x}) > \mathbf{g}_j(\mathbf{x})$对任意 $j\neq i$ 均成立，则将$\mathbf{x}$归于ω_i类。
  - 参照贝叶斯决策规则，我们可以定义：
    - $g_i(\mathbf{x}) = P(\omega_i|\mathbf{x})$
    - $g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$
    - $g_i(\mathbf{x}) = \ln(p(\mathbf{x}|\omega_i)) + \ln(P(\omega_i))$
    - $g_i(\mathbf{x}) = f(p(\mathbf{x}|\omega_i)) + h(\mathbf{x}, \omega_i)$,　　**更一般情形**

2.4 分类器设计

- **决策面**
  - 对于 c 类分类问题，按照决策规则可以把 d 维特征空间分成 c 个决策区域 R_i，i=1,2..,c。划分决策区域的边界称为决策面。
  - 各决策域 R_i 被决策面分割而成。这些决策面是特征空间中的**超曲面**，相邻的两个决策域在决策面上其判别函数值是相等的。如果 R_i 和 R_j 是相邻的，则它们的决策面方程应满足：$g_i(\mathrm{x}) = g_j(\mathrm{x})$

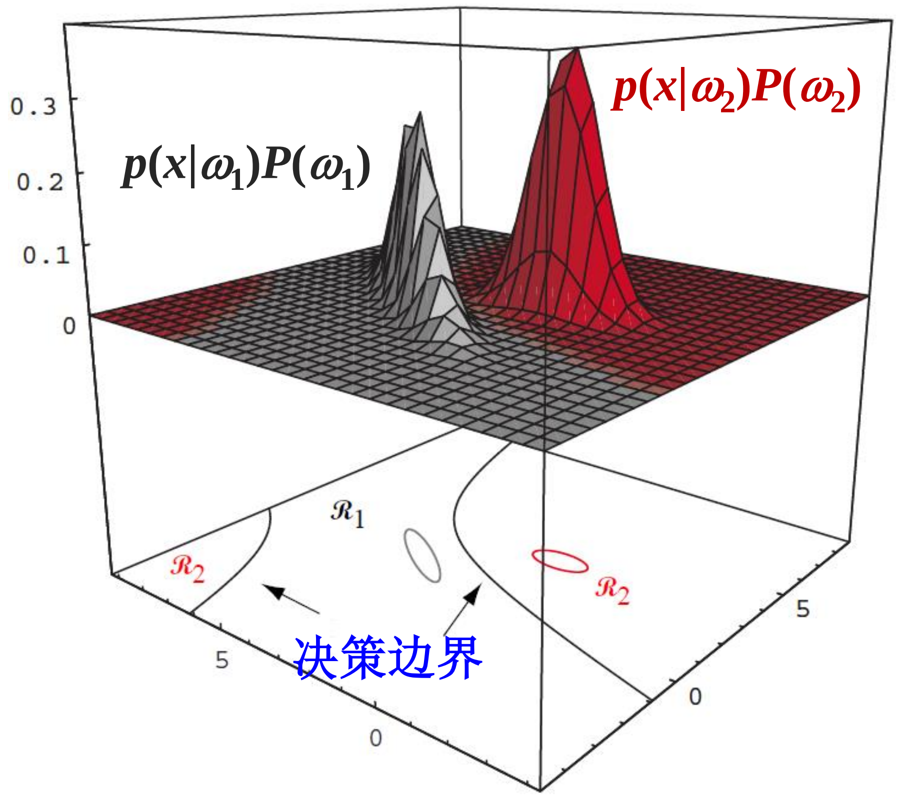


2.4 分类器设计

- 两类情形下的判别函数
  - 对于两类情形，只需要定义一个判别函数：
    - $\boldsymbol{g(\mathrm{x}) = P(\omega_1|\mathrm{x}) - P(\omega_2|\mathrm{x})}$
    - $\boldsymbol{g(\mathrm{x}) = p(\mathrm{x}|\omega_1)P(\omega_1) - p(\mathrm{x}|\omega_2)P(\omega_2)}$
    - $\boldsymbol{g(\mathrm{x})} = \ln(\boldsymbol{p(\mathrm{x}|\omega_1)}) - \ln(\boldsymbol{p(\mathrm{x}|\omega_2)}) + \ln(\boldsymbol{P(\omega_1)}) - \ln(\boldsymbol{P(\omega_2)})$
    - $\boldsymbol{g(\mathrm{x}) = R(\alpha_1|\mathrm{x}) - R(\alpha_2|\mathrm{x})}$
- 两类情形下的决策面方程
  - $g(\mathbf{x}) = 0$
  - x为一维时，决策面为一些分界点；二维时，决策面为一些曲线（曲线段）；三维时，决策面为一些曲面（曲面片）；高维则为一些超曲面（超曲面片）。
  - 若$g(\mathbf{x})$为线性判别函数，则为平面或平面片

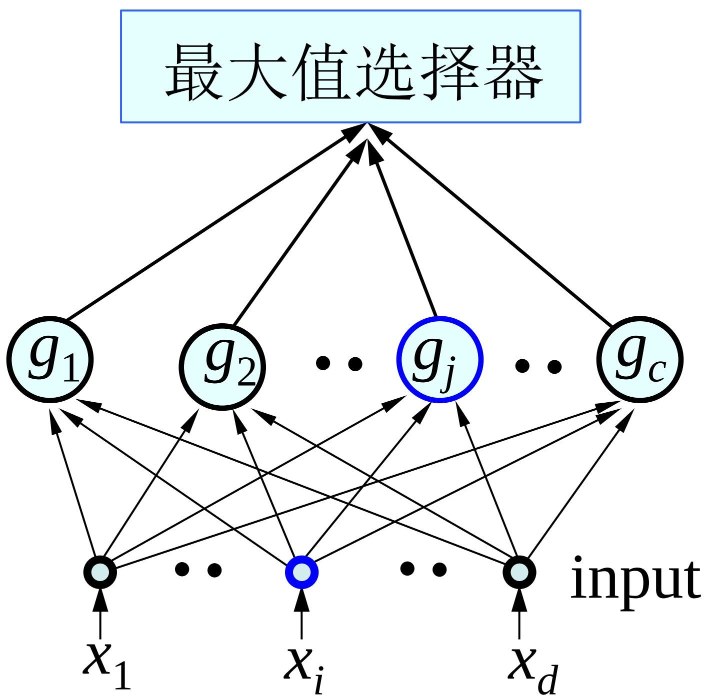
2.4 分类器设计

- **两类情形下的决策面方程**
  - 正态分布下的一个例子



2.4 分类器设计

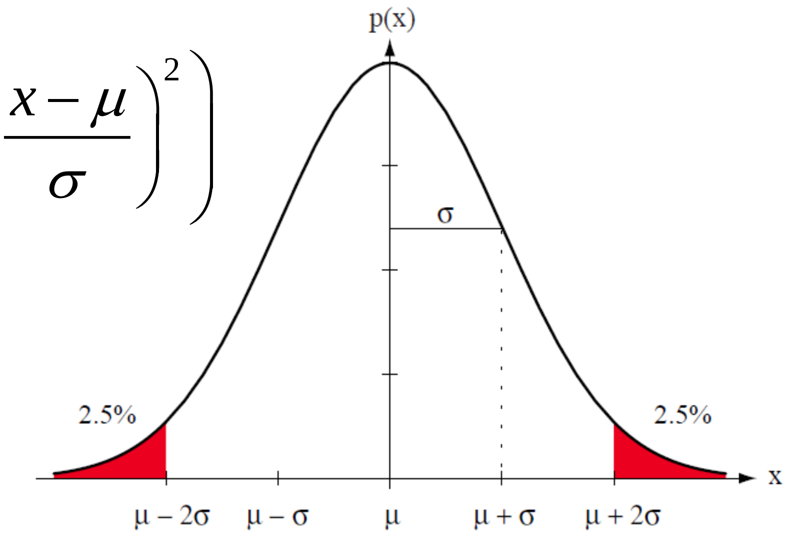
- **分类器设计**
  - 分类器可以看成一个机器，其功能是计算出 c 个判别函数，然后再从中选出对应于判别函数为最大值的类作为分类结果。



2.5 高斯密度下的判别函数

- **单变量正态分布：** $x \sim N(\mu, \sigma^2)$

公式要牢记：$p(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{1}{2} \left(\dfrac{x-\mu}{\sigma} \right)^2 \right)$



- 均值μ与标准差σ：

$$\mu = E\{x\} = \int_{-\infty}^{+\infty} x p(x) dx,$$

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

- 性质：$p(x) \geq 0 \quad (-\infty < x < +\infty)$, $\int_{-\infty}^{+\infty} p(x) dx = 1$

- 在给定均值和方差的所有分布中，正态分布的熵最大
- 根据Central Limit Theorem，大量独立随机变量之和趋近正态分布
- 实际环境中，很多类别的特征分布趋近正态分布

2.5 高斯密度下的判别函数

- **多元正态分布：** $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mathbf{x} = [x_1, x_2, ..., x_d]^T \in R^d,$$
$$\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_d]^T \in R^d,$$
$$\boldsymbol{\Sigma} \in R^{d \times d}$$

  - 密度函数：

公式要牢记：
$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

|.|--矩阵行列式

  - **均值向量$\boldsymbol{\mu}$：**

$$\boldsymbol{\mu} = E\{\mathbf{x}\} \in R^d$$

$$\mu_i = E\{x_i\} = \int_{-\infty}^{+\infty} x_i p(x_i) dx_i, \quad i = 1, 2, ..., d$$

边际分布密度函数：$p(x_i) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p(x) dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots dx_d$

2.5 高斯密度下的判别函数

- **多元正态分布**
  - 协方差矩阵Σ:

$$\boldsymbol{\Sigma} = E\left\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\right\} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1d}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 & \cdots & \sigma_{2d}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1d}^2 & \sigma_{2d}^2 & \cdots & \sigma_{dd}^2 \end{pmatrix} \in \mathrm{R}^{d\times d}$$

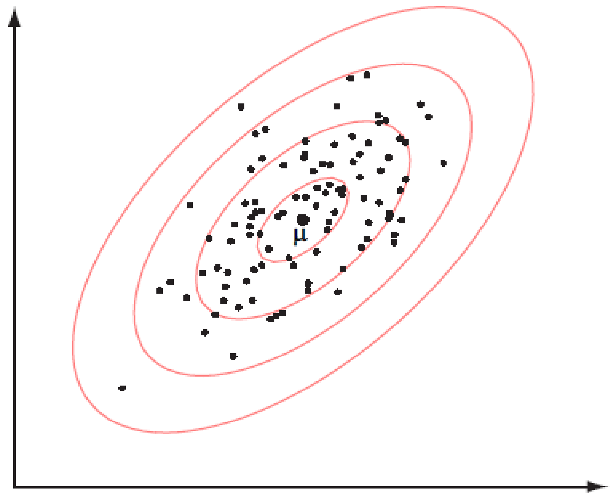
$$\sigma_{ij}^2 = E\left\{(x_i-\mu_i)(x_j-\mu_j)\right\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i-\mu_i)(x_j-\mu_j)p(x_i,x_j)dx_i dx_j$$

2.5 高斯密度下的判别函数

- **等密度轨迹**
  - 等密度轨迹为一超椭球面。从多元正态分布函数可以看出，当其指数项等于常数时，密度 $p(\mathrm{x})$ 的值不变，因此等函数点即为使如下方程为常数的点，即：

$$(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = const.$$

特殊情况下为圆形或超球面



等密度点轨迹：hyper-ellipsoid

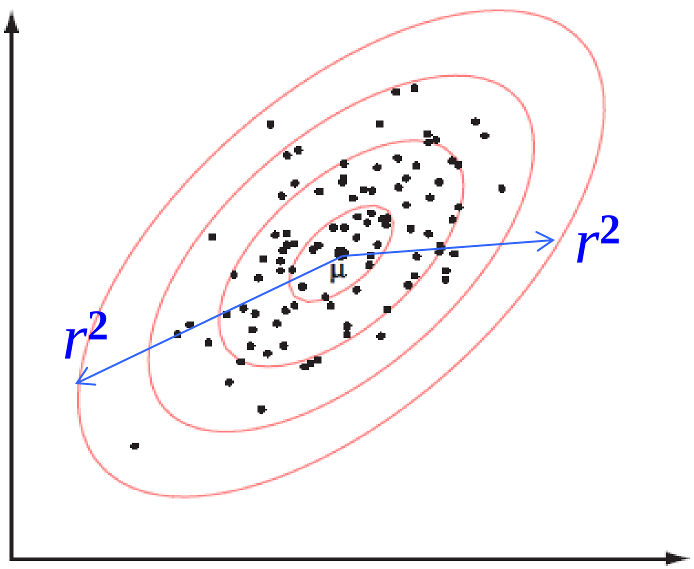
该方程为一个超椭球面，其主轴方向由矩阵Σ的特征向量所决定，主轴的长度与其本征值成正比。

2.5 高斯密度下的判别函数

- **Mahalanobis距离（马氏距离）**
  - 定义如下 样本x到中心μ的计算公式：

$$r^2 = (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})$$

  - 称上式为样本 x 到中心 μ 的 Mahalanobis距离。



2.5 高斯密度下的判别函数

- **性质**
  - 不相关性等价于独立性
  - 边缘分布与条件分布均为正态分布
  - 多元正态随机向量的线性变换（非奇异）仍为多元正态分布的随机变量
  - 线性组合的正态性：若x为多元正态随机向量，则线性组合y=a^{T}x是一个一维正态随机变量。

协方差矩阵的性质

- 协方差矩阵为实对称矩阵
  - 本征值，本征向量（Eigenvalues & eigenvecters）

$$\boldsymbol{\Sigma}\boldsymbol{\varphi}_i = \lambda_i \boldsymbol{\varphi}_i, \ \boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \cdots, \boldsymbol{\varphi}_d], \quad \boldsymbol{\Lambda} = diag[\lambda_1, \lambda_2, \cdots, \lambda_d]$$

  - Orthonormal

$$\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \mathbf{I} \qquad \boldsymbol{\Phi}^T = \boldsymbol{\Phi}^{-1} \quad \Leftrightarrow \quad \boldsymbol{\varphi}_i^T\boldsymbol{\varphi}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

  - 矩阵表示

$$\boldsymbol{\Sigma}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Lambda} \leftrightarrow \boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T \leftrightarrow \boldsymbol{\Sigma} = \sum_{i=1}^{d} \lambda_i \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^T$$

  - 矩阵对角化

$$\boldsymbol{\Phi}^T\boldsymbol{\Sigma}\boldsymbol{\Phi} = \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_d \end{bmatrix} \quad \Leftrightarrow \quad \boldsymbol{\varphi}_i^T\boldsymbol{\Sigma}\boldsymbol{\varphi}_i = \lambda_i$$

2.5 高斯密度下的判别函数

- **线性变换**
  - **Covariance matrix Σ, Eigenvalue decomposition:** $\boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T$

$$\boldsymbol{\Phi} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_d] \in R^{d\times d},$$

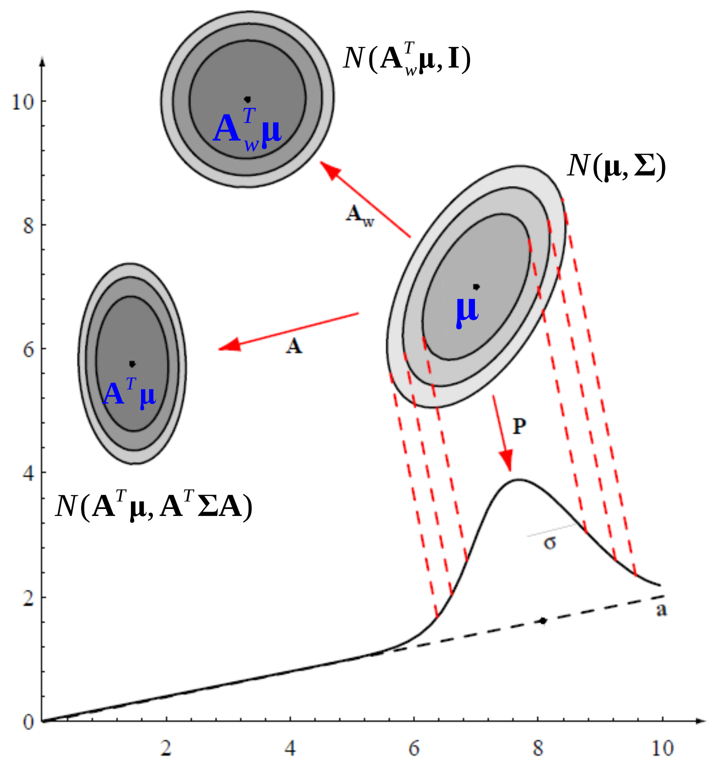
$$\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_d) \in R^{d\times d}$$

  - **Linear Transformation y = A^{T}x**

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

  - **Whitening transformation**

$$\mathbf{A}_w = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2}, \quad \Rightarrow \quad \mathbf{y} \sim N(\mathbf{A}_w^T\boldsymbol{\mu}, \mathbf{I})$$



$$\mathbf{A}_w^T\boldsymbol{\Sigma}\mathbf{A}_w = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\boldsymbol{\Sigma}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}$$

2.5 高斯密度下的判别函数

- **最小错误率贝叶斯决策**
  - **对于c类问题，假定 各类条件概率密度函数为多元正态分布：**

$$p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad i = 1, 2, \ldots, c$$

  - **判别函数(**Quadratic discriminant function (QDF))**:**

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

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

$$(i = 1, 2, \ldots, c)$$

  - **决策面方程**

$$g_i(\mathbf{x}) = g_j(\mathbf{x})$$

$$-\frac{1}{2}\left((\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - (\mathbf{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right) - \frac{1}{2}\ln\left(\frac{|\boldsymbol{\Sigma}_i|}{|\boldsymbol{\Sigma}_j|}\right) + \ln\left(\frac{|P(\omega_i)|}{|P(\omega_j)|}\right) = 0$$

- **第一种情形：** $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I} \in R^{d\times d}, \quad i = 1, 2, \ldots, c$
  - **协方差矩阵**

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}, \quad |\boldsymbol{\Sigma}_i| = \sigma^{2d}, \quad \boldsymbol{\Sigma}_i^{-1} = \frac{1}{\sigma^2}\mathbf{I}$$

  - **判别函数**(Quadratic discriminant function (QDF)):

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

等价

$$\begin{aligned} g_i(\mathbf{x}) &= -\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i) + \ln\left(P(\omega_i)\right) \\ &= -\frac{1}{2\sigma^2}\left\|\mathbf{x}-\boldsymbol{\mu}_i\right\|_2^2 + \ln\left(P(\omega_i)\right) \end{aligned}$$

欧氏距离

2.5 高斯密度下的判别函数

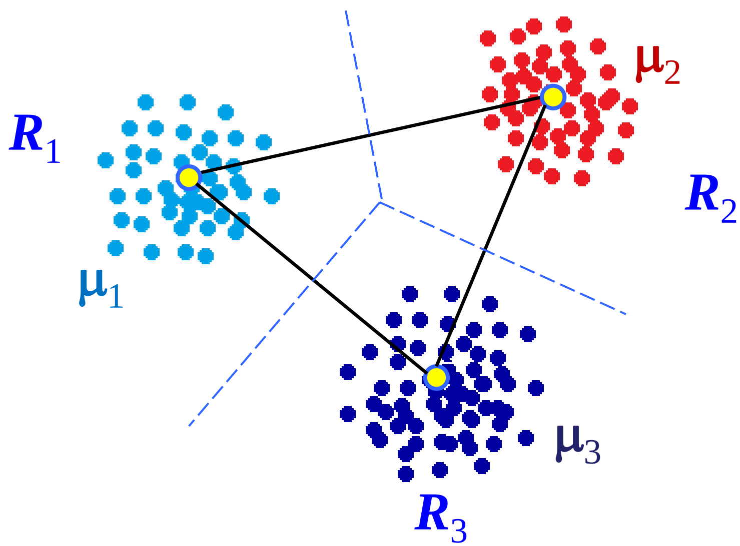
- **第一种情形：** $\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}, \quad i = 1, 2, \ldots, c$
  - 先验概率相等：$P(\omega_i) = P(\omega_j)$
  - 此时，判别函数可进一步简单化为 $g_i(\mathbf{x}) = -\dfrac{1}{2\sigma^2}\|\mathbf{x} - \boldsymbol{\mu}_i\|_2^2$
  - 因此，最小错误率贝叶斯规则相当简单：
    - 若要对样本x进行分类，只需要计算x到各类均值向量的欧氏距离平方，然后将归于距离最短的一类：

$$\arg\min_{i=1,2,\ldots,c} \|\mathbf{x} - \boldsymbol{\mu}_i\|_2^2$$

    - 这种分类器称为最小距离分类器

2.5 高斯密度下的判别函数

- **第一种情形：** $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - 先验概率相等：$P(\omega_i) = P(\omega_j)$
  - 最小距离分类器：



先验概率相等时的最小距离分类器

2.5 高斯密度下的判别函数

- **第一种情形：** $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - 先验概率不相等：$P(\omega_i) \neq P(\omega_j)$

$$
\begin{aligned}
g_i(\mathbf{x}) &= -\frac{1}{2\sigma^2}(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i) + \ln\left(P(\omega_i)\right) \\
&= -\frac{1}{2\sigma^2}\left(\mathbf{x}^T\mathbf{x} - 2\boldsymbol{\mu}_i^T\mathbf{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i\right) + \ln\left(P(\omega_i)\right)
\end{aligned}
$$

由于每一类的判别函数均包含$\mathbf{x}^T\mathbf{x}$，与下标i无关，因此可以进一步简化为线性判别函数：

$$
\begin{aligned}
g_i(\mathbf{x}) &= \frac{1}{\sigma^2}\boldsymbol{\mu}_i^T\mathbf{x} - \frac{1}{2\sigma^2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i + \ln\left(P(\omega_i)\right) \\
&= \mathbf{w}_i^T\mathbf{x} + w_{i0}
\end{aligned}
$$

$$
\mathbf{w}_i = \frac{1}{\sigma^2}\boldsymbol{\mu}_i
$$

$$
w_{i0} = \ln\left(P(\omega_i)\right) - \frac{1}{2\sigma^2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i
$$

2.5 高斯密度下的判别函数

- **第一种情形：** $\mathbf{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - 先验概率不相等：$P(\omega_i) \neq P(\omega_j)$
  - 决策规则:

    $$\text{if } \ g_k(\mathbf{x}) = \max_i g_k(\mathbf{x}), \ \text{ then } \ \mathbf{x} \in \omega_i$$

  - 判别函数g_i(x)是x的线性函数。
    - 判别函数为线性函数的分类器称为线性分类器。
    - 线性分类器的决策面方程为： g_i(x)- g_j(x)=0 所确定的一个超平面。

2.5 高斯密度下的判别函数

- **第一种情形：** $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - 先验概率不相等：$P(\omega_i) \neq P(\omega_j)$
  - 决策面方程：

$$g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0 \quad \Rightarrow \quad \mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0$$

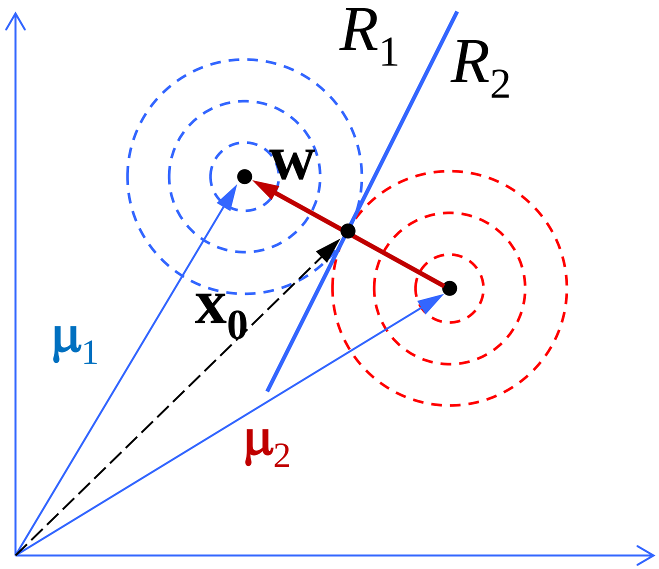
$$\Rightarrow \quad \mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \boxed{\frac{\sigma^2}{\left\| \boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right\|^2} \ln\left(\frac{P(\omega_i)}{P(\omega_j)} \right)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

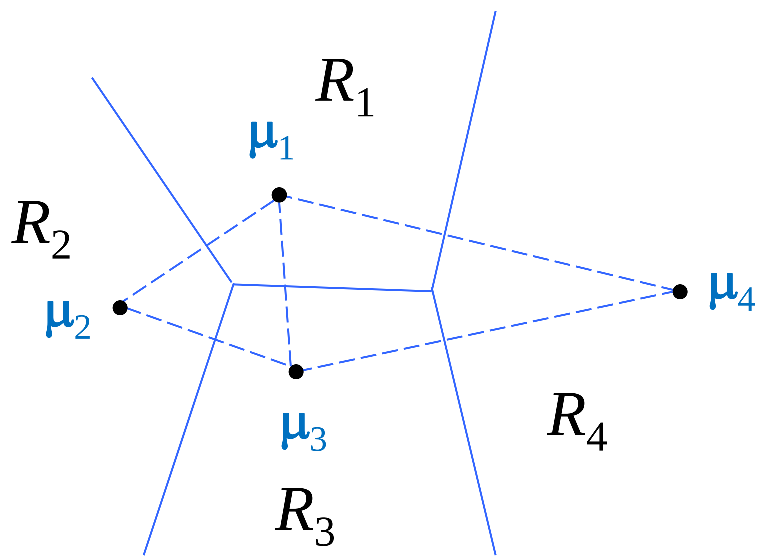
$$= \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - s_{ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

2.5 高斯密度下的判别函数

- **第一种情形：**$\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - **进一步分析先验概率相等：*P*(*ω*$_i$)= *P*(*ω*$_j$)**
  - 决策面方程：$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0, \quad \mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j, \quad \mathbf{x}_0 = \frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right)$



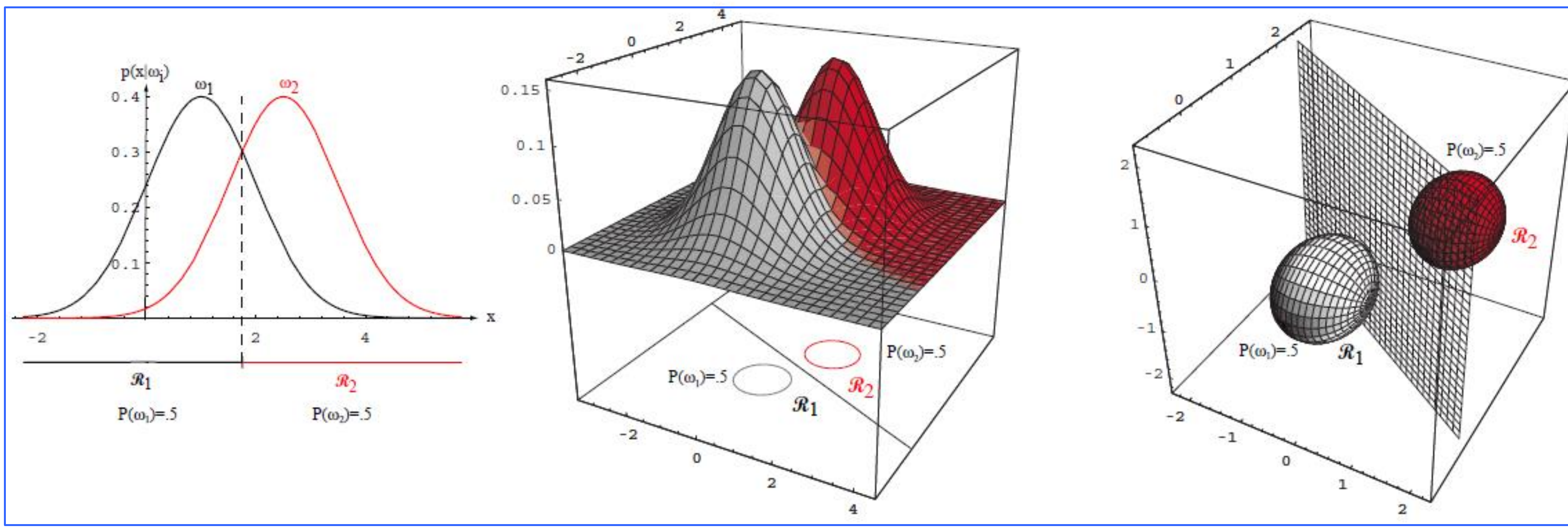
两类情形



多类情形

2.5 高斯密度下的判别函数

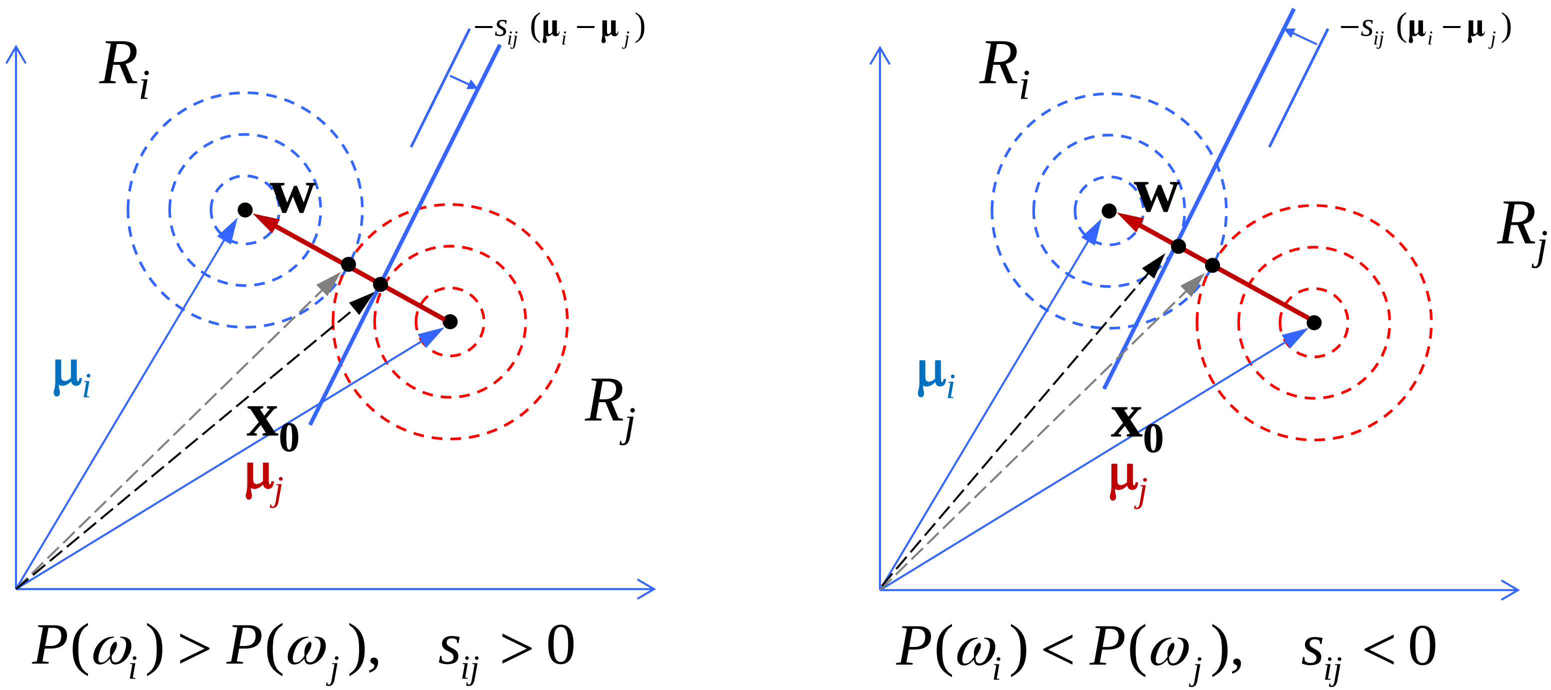
- 第一种情形：$\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}, \quad i = 1, 2, \ldots, c$
  - 进一步分析先验概率相等：$P(\omega_i) = P(\omega_j)$
  - 决策面方程：$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0, \quad \mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j, \quad \mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$



线性判别函数

- **第一种情形**：$\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - **进一步分析先验概率不相等**：$P(\omega_i) \neq P(\omega_j)$
  - 决策面方程：$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0,$

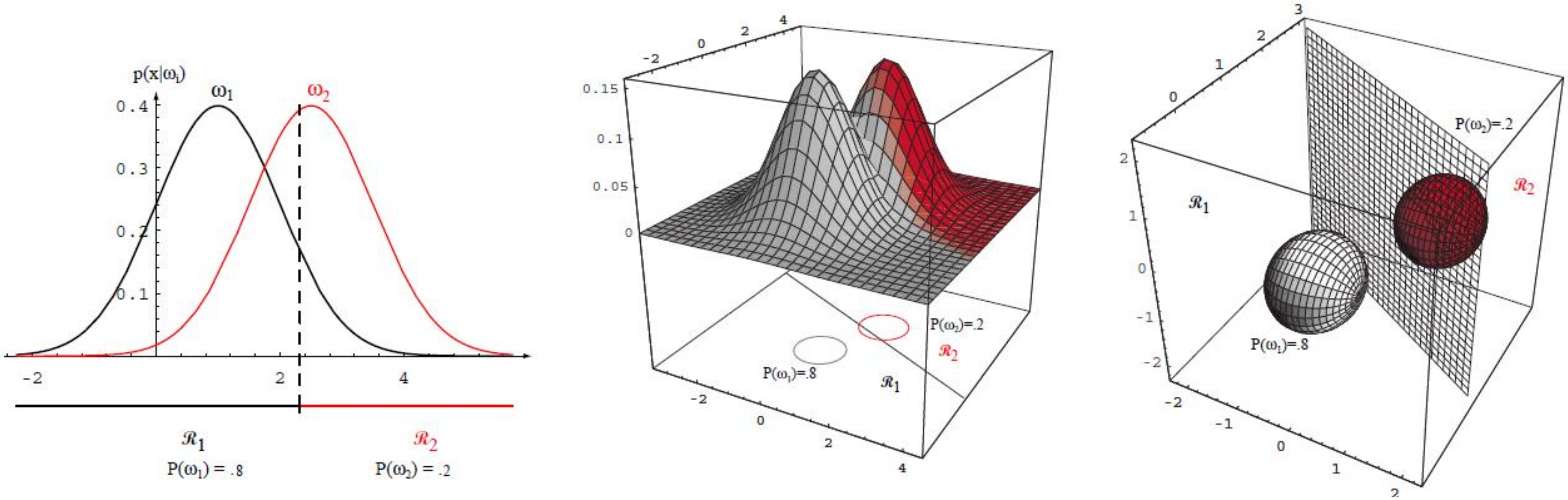
$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j; \quad \mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - s_{ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$



线性决策面

- **第一种情形：** $\boldsymbol{\Sigma}_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - **进一步分析先验概率不相等：$P(\omega_i) \neq P(\omega_j)$**
  - 决策面方程：$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0,$

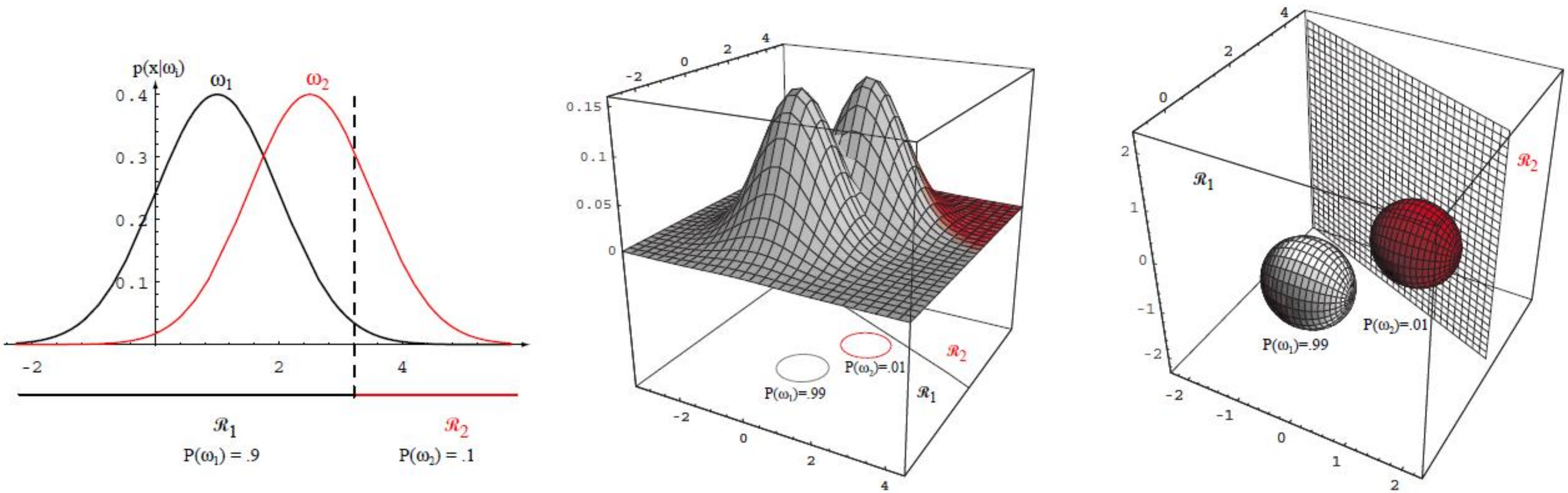
$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j; \quad \mathbf{x}_0 = \frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) - s_{ij}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)$$



例子1

- **第一种情形**：$\Sigma_i = \sigma^2 \mathbf{I}, \quad i = 1, 2, \ldots, c$
  - **进一步分析先验概率不相等**：$P(\omega_i) \neq P(\omega_j)$
  - 决策面方程：$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0,$

$$\mathbf{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j; \quad \mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - s_{ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$



例子2

2.5 高斯密度下的判别函数

- **第二种情形：** $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - 各类的协方差矩阵均相等。从几何上看，相当于各类样本集中于以该类均值μ_i为中心但大小和形状相同的椭球内。
  - 判别函数 (Quadratic discriminant function (QDF))：

$$g_i(\mathbf{x}) = \ln\left(p(\mathbf{x} \mid \omega_i)\right) + \ln\left(P(\omega_i)\right)$$

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

简化：

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

2.5 高斯密度下的判别函数

- **第二种情形：** $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - 先验概率相等：$P(\omega_i) = P(\omega_j)$
  - 判别函数：

$$g_i(\mathbf{x}) = r^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$$

  - 决策规则：
    - 若要对样本x进行分类，只需要计算x到各类均值向量的马氏距离平方，然后将归于距离最短的一类：

$$\arg \min_{i=1,2,\ldots,c} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$$

2.5 高斯密度下的判别函数

- 第二种情形：$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - 先验概率不相等：$P(\omega_i) \neq P(\omega_j)$
  - 判别函数：

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

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

忽略与i无关的项，简化可得：

$$g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{i0}$$

$$\mathbf{w}_i = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$$

$$w_{i0} = \ln\left(P(\omega_i)\right) - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i$$

2.5 高斯密度下的判别函数

- **第二种情形：** $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - **先验概率不相等：** $P(\omega_i) \neq P(\omega_j)$
  - 决策面方程：$g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$

  展开可得：$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0$，　线性判别函数

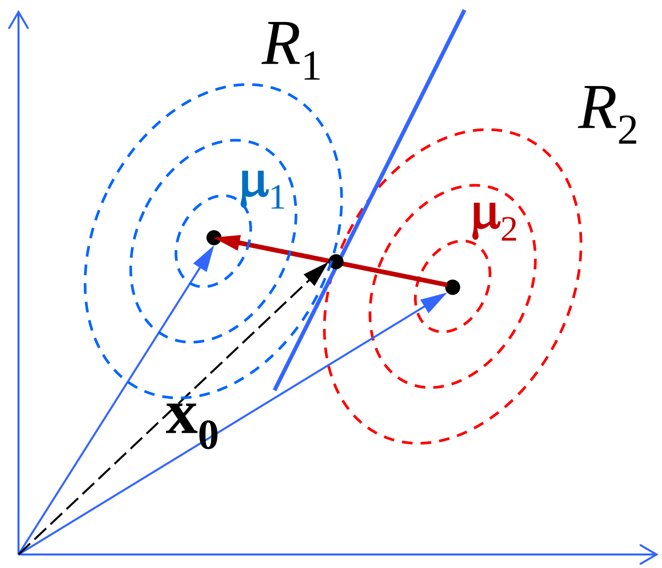
$$
\begin{cases}
\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \\
\mathbf{x}_0 = \dfrac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \boxed{\dfrac{\sigma^2}{(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)} \ln\left(\dfrac{P(\omega_i)}{P(\omega_j)}\right)} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)
\end{cases}
$$

$$
= \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - s_{ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)
$$

2.5 高斯密度下的判别函数

- **第二种情形：** $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - **进一步考察先验概率相等：** $P(\omega_i) = P(\omega_j)$

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0: \quad \mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \quad \mathbf{x}_0 = \frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right)$$



两类情形：$P(\omega_1) = P(\omega_2)$

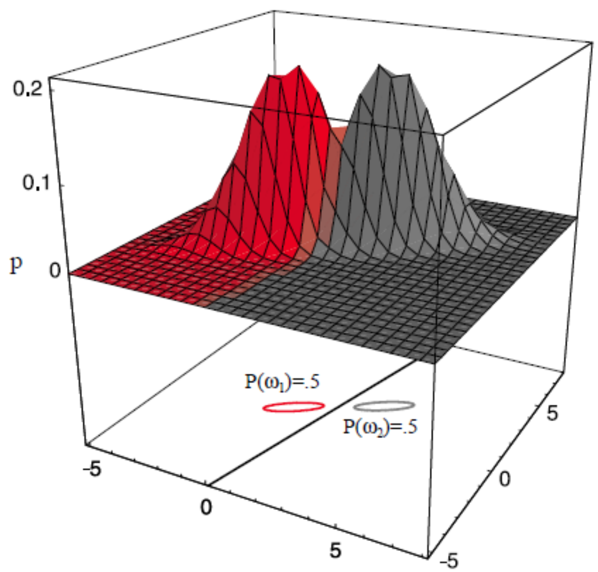
✓ $\mathbf{w}$通常不在$\boldsymbol{\mu}_i - \boldsymbol{\mu}_j$的方向上。决策面通过$\mathbf{x}_0$点，但不与$\boldsymbol{\mu}_i - \boldsymbol{\mu}_j$正交。

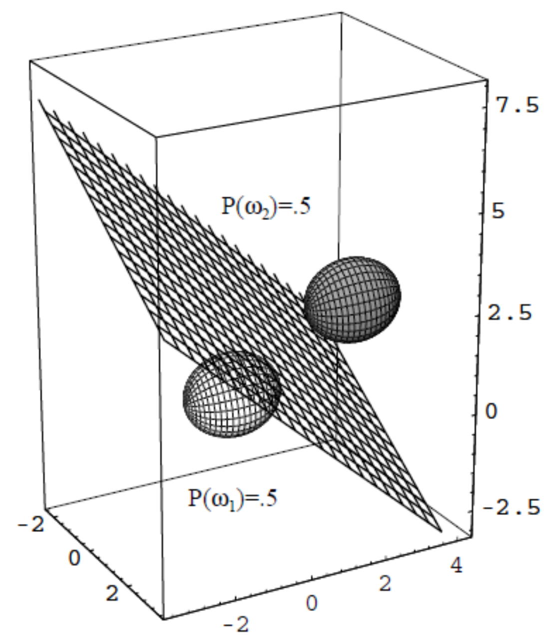
✓ 当各类先验概率相等时，$\mathbf{x}_0$为$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$的中点。

2.5 高斯密度下的判别函数

- 第二种情形：$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - 进一步考察先验概率相等：$P(\omega_i) = P(\omega_j)$

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0: \quad \mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), \quad \mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$

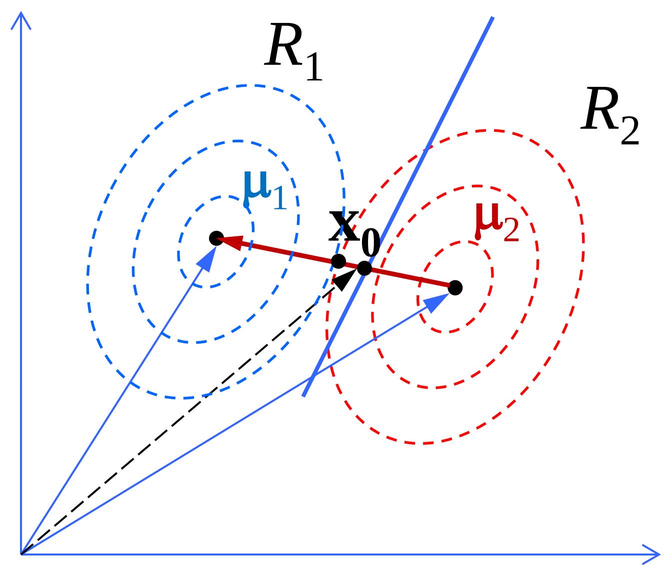




2.5 高斯密度下的判别函数

- **第二种情形**：$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - **进一步考察先验概率不相等**：$P(\omega_i) \neq P(\omega_j)$

$$\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)=0: \quad \mathbf{w}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j), \quad \mathbf{x}_0=\frac{1}{2}\left(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j\right)-s_{ij}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right)$$



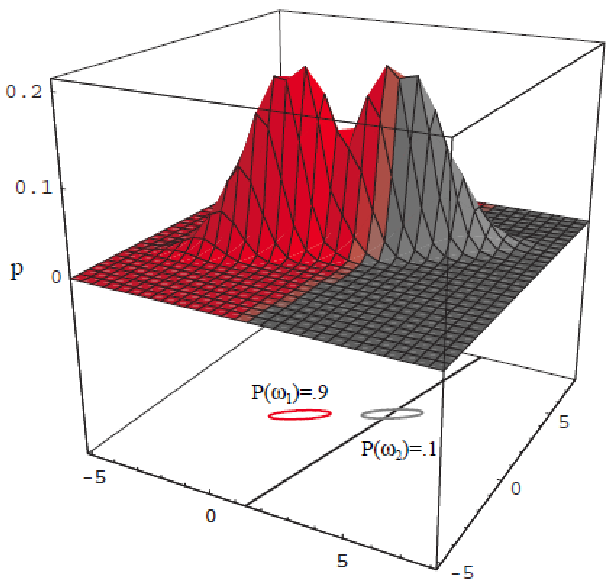
两类情形：$P(\omega_i) > P(\omega_j), \quad s_{ij} > 0$

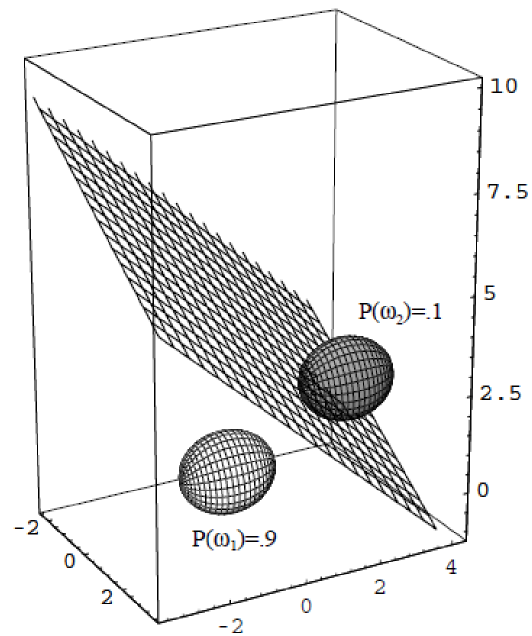
- ✓ $\mathbf{w}$通常不在$\boldsymbol{\mu}_i-\boldsymbol{\mu}_j$的方向上。决策面通过$\mathbf{x}_0$点，但不与$\boldsymbol{\mu}_i-\boldsymbol{\mu}_j$正交。
- ✓ 当各类先验概率不相等时，$\mathbf{x}_0$不在$\boldsymbol{\mu}_i$和$\boldsymbol{\mu}_j$的中点上，而是在连线上向先验概率较小的均值点偏移。

2.5 高斯密度下的判别函数

- 第二种情形： $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \quad i = 1, 2, \ldots, c$
  - 进一步考察先验概率不相等：$P(\omega_i) \neq P(\omega_j)$

$$\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0)=0: \quad \mathbf{w}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j), \quad \mathbf{x}_0=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-s_{ij}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$$





2.5 高斯密度下的判别函数

- **第三种情形：** $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j, \quad i, j = 1, 2, \ldots, c$
  - 判别函数:

$$g_i(\mathbf{x}) = \ln\left(p(\mathbf{x} \mid \omega_i)\right) + \ln\left(P(\omega_i)\right)$$

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

忽略常数项：

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

$$= \mathbf{x}^T \mathbf{W}_i \mathbf{x} + \mathbf{w}_i^T \mathbf{x} + w_{i0}$$

$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}, \quad \mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i, \quad w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln(|\boldsymbol{\Sigma}_i|) + \ln\left(P(\omega_i)\right)$$

2.5 高斯密度下的判别函数

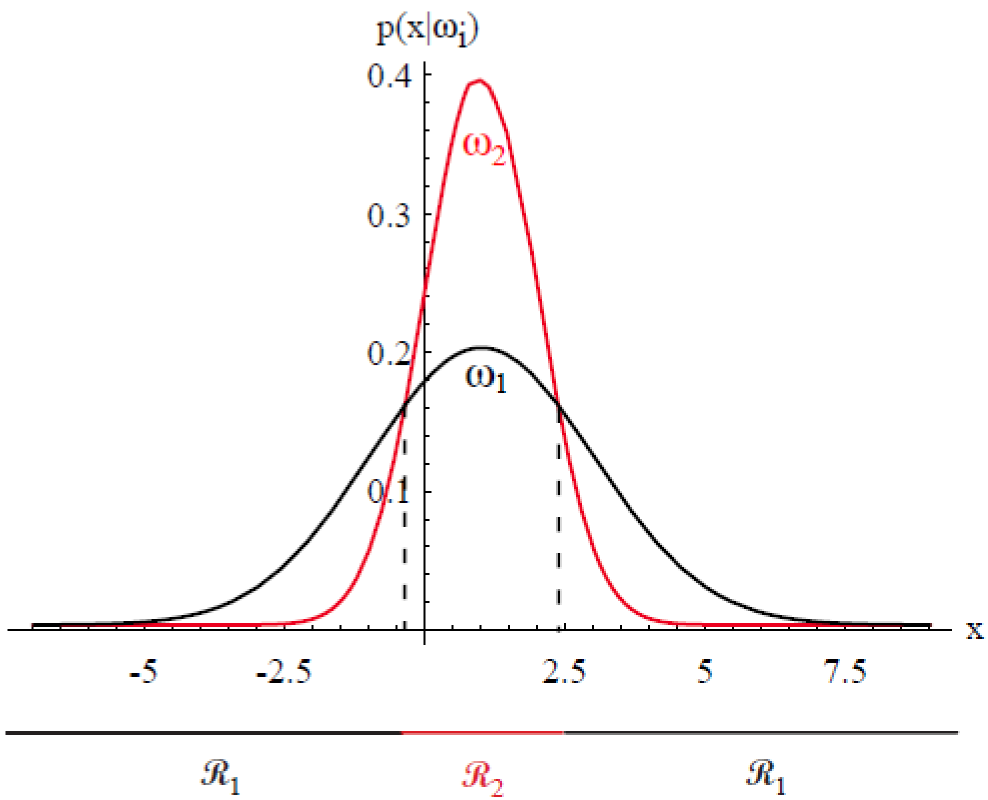
- **第三种情形：** $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j, \quad i, j = 1, 2, \ldots, c$
  - 决策方程：$g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$

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

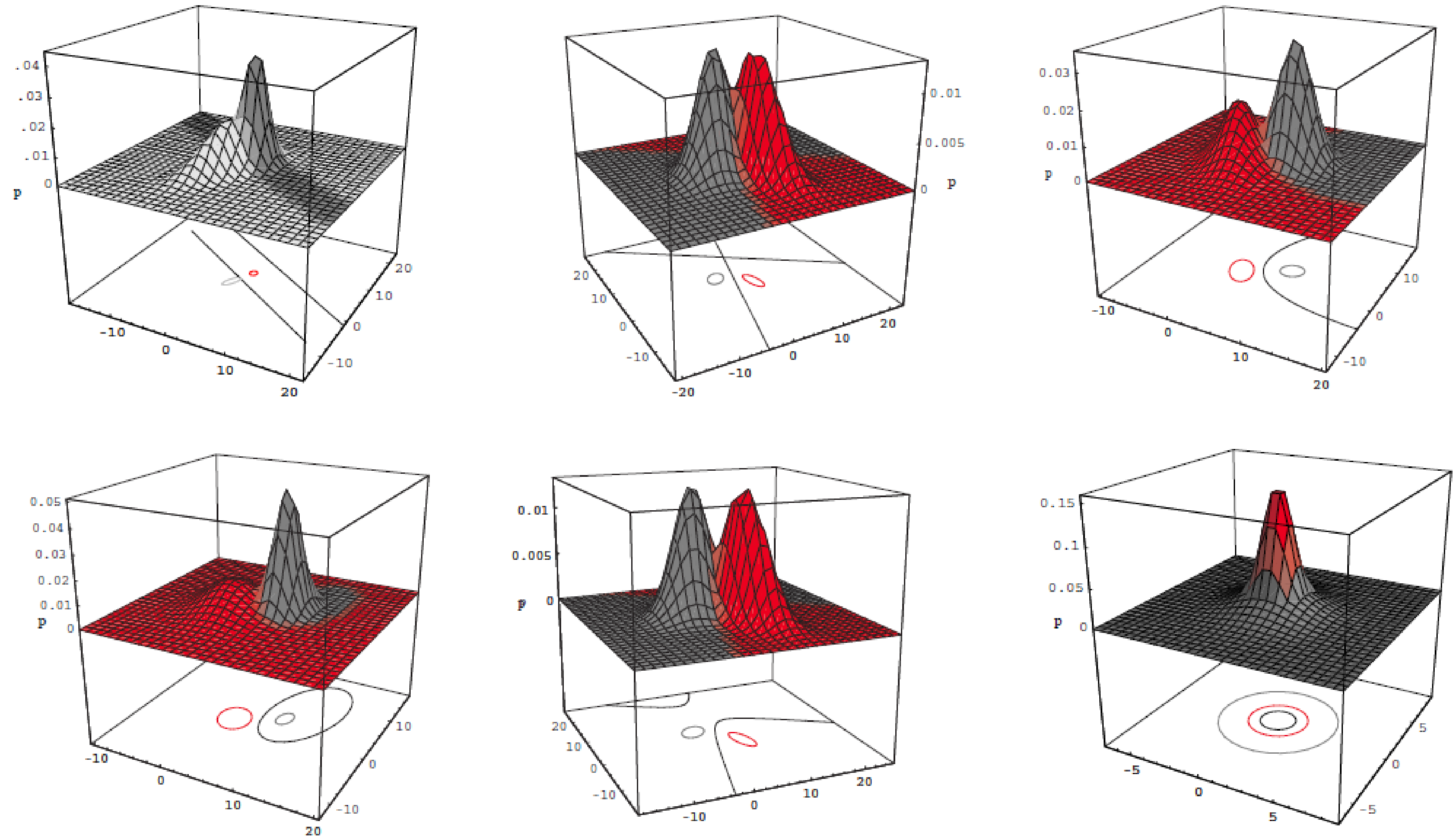
决策面为一个超二次曲面。随着$\boldsymbol{\Sigma}_i$，$\boldsymbol{\mu}_i$，$P(\omega_i)$ 等的不同而呈现出超球面、超椭球面、超双曲面或超平面等不同的情形。

2.5 高斯密度下的判别函数

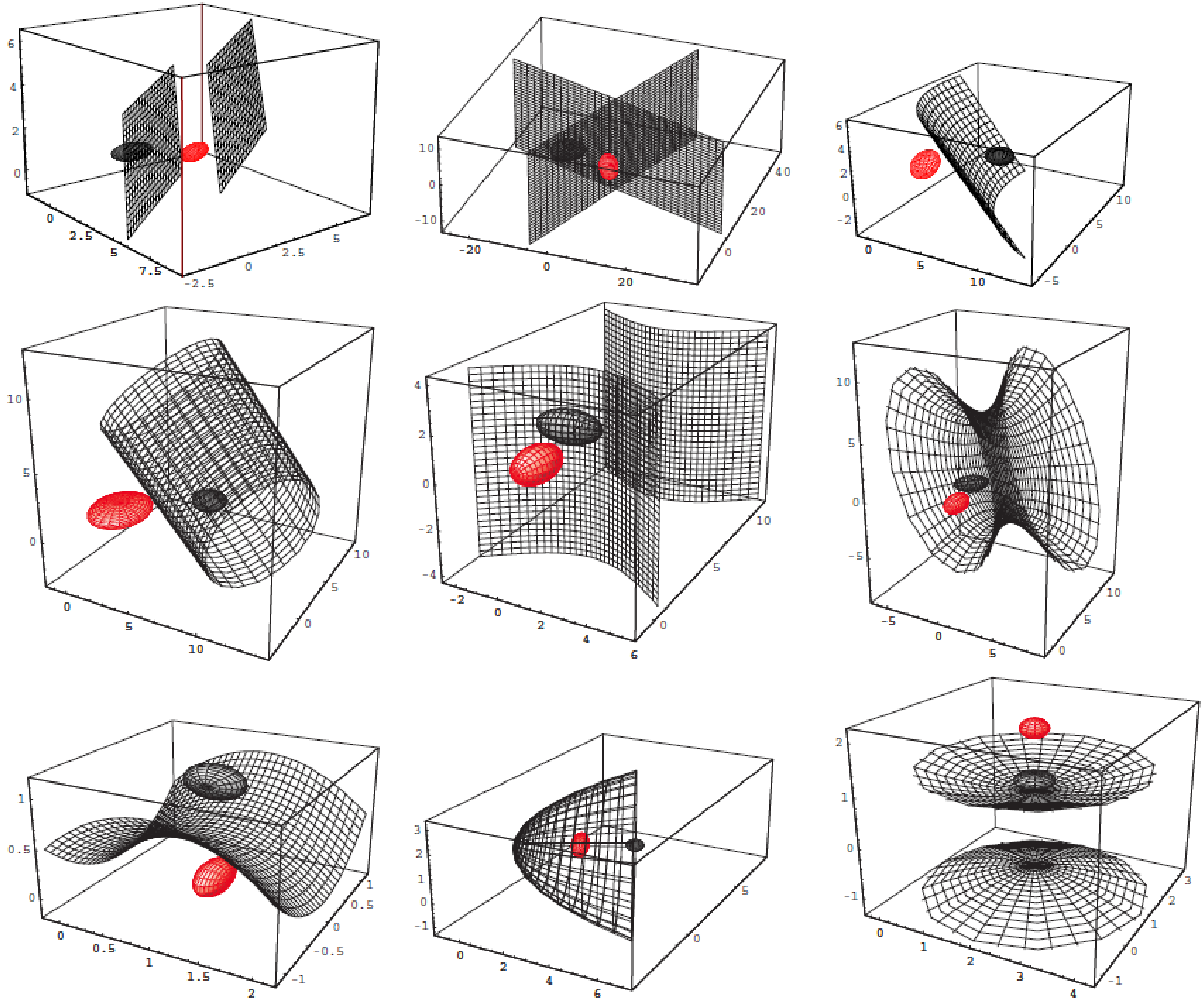
- **第三种情形：** $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_j, \quad i, j = 1, 2, \ldots, c$
  - 一些例子：一维情形，等均值但具有不同协方差



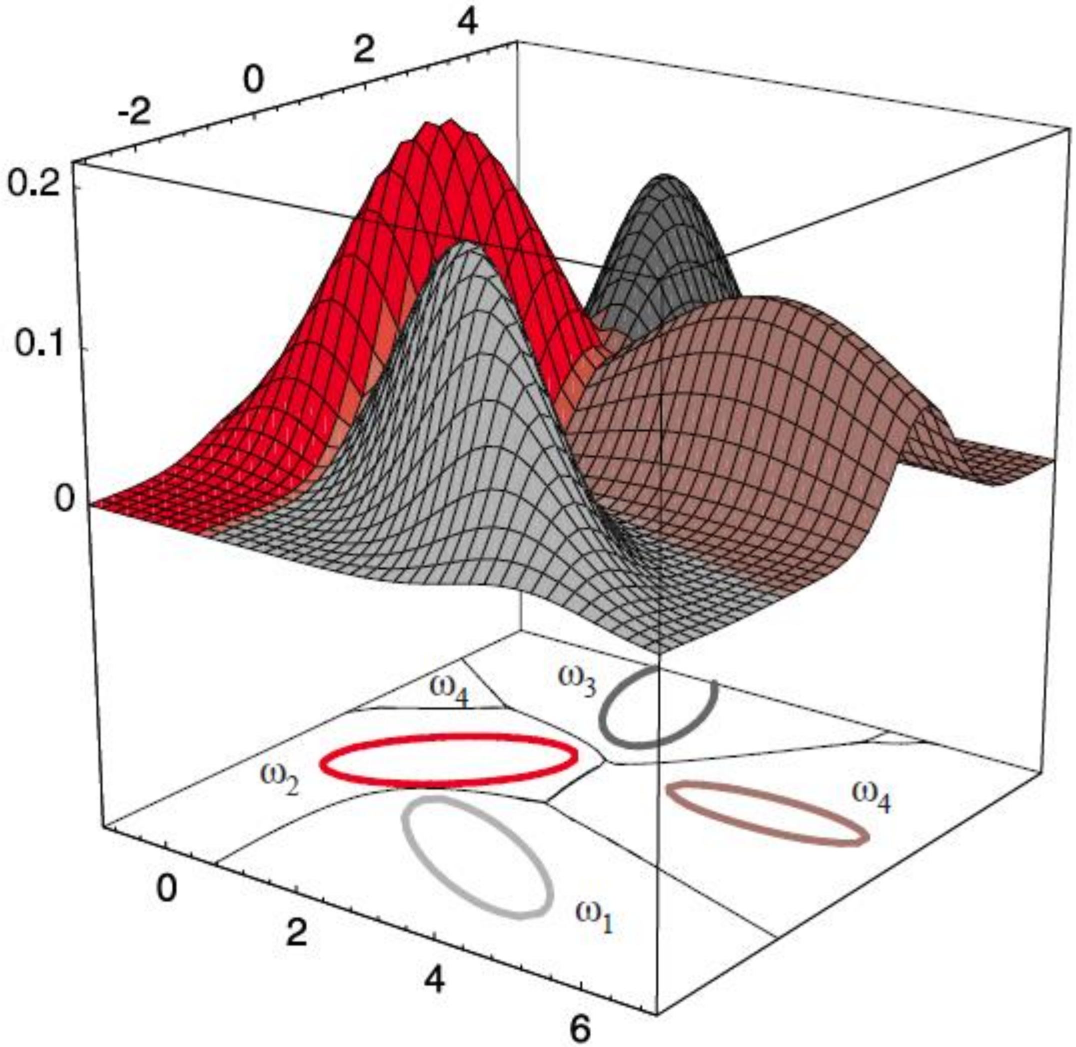
2.5 高斯密度下的判别函数



二维的一些例子


6 4 2 0
0 2.5 5 7.5
-2.5 0 2.5 5
10 0 -10
-20 0 20
0 20 40
6 4 2 0 -2
0 5 10
-5 0 5 10
10 5 0
0 5 10
0 5 10
4 2 0 -2 -4
-2 0 2 4 6
0 5 10
10 5 0 -5
-5 0 5
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
2 1 0 -1
0 1 2 3 4
-1 0 1 2 3

三维的一些例子



二维特征空间、四个类别（例子）

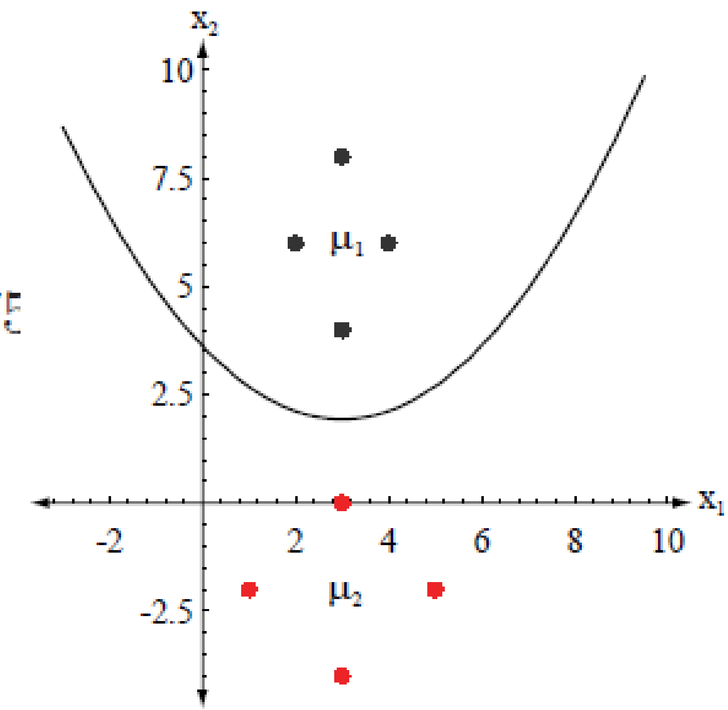
- **一个具体例子**
  - 2类，2D $P(\omega_1) = P(\omega_2) = 0.5$

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

$$\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \qquad \Sigma_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

  - 决策面 $g_1(\mathbf{x})=g_2(\mathbf{x})$

$$x_2 = 3.514 - 1.125x_1 + 0.1875$$



2.6 分类错误率

- **最小错误率贝叶斯决策**
  - **样本x的错误率：**
    - 任一决策都可能会有错误。

即：当我们将样本x判定为第二类时，这个判定失误的概率为$P(\omega_1|\mathbf{x})$。（因为样本以该概率属于第一类）

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

即：当我们将样本x判定为第一类时，这个判定失误的概率为$P(\omega_2|\mathbf{x})$。（因为样本以该概率属于第二类）

  - **样本x的最小错误率：**

$$P(error\,|\,\mathbf{x}) = \min\left(P(\omega_1\,|\,\mathbf{x}), P(\omega_2\,|\,\mathbf{x})\right)$$

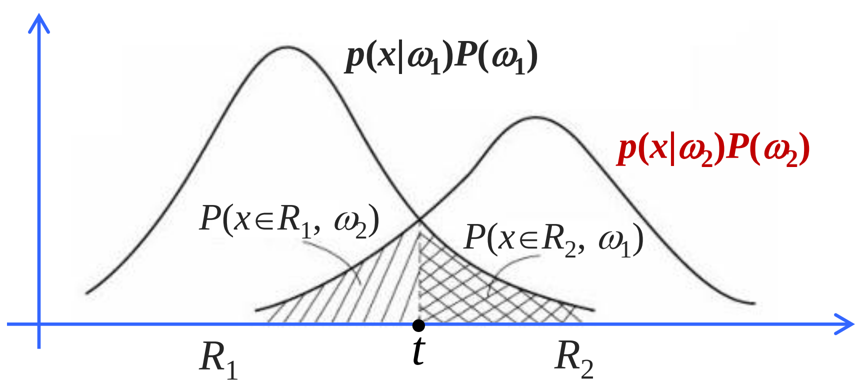
  - **贝叶斯决策的错误率：**
    - 贝叶斯决策的错误率定义为所有服从独立同分布的样本上的错误率的期望：

$$P(error) = \int P(error\,|\,\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

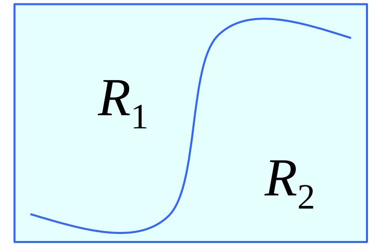
- **关于错误率，以一维为例举例说明：**
  - 考虑一个有关一维样本的两类分类问题。假设决策边界 t 将x轴分成两个区域 R_1和R_2。 R_1为$(-\infty, t)$,R_2为(t, ∞)。
  - 错误情形：样本在R_1中，但属于第二类的概率是存在的，即$P(\omega_2|x)$；样本在R_2中,但属于第一类的概率也是存在的，即$P(\omega_1|x)$；这两种情形就是决策一个给定样本x可能出现错误的概率。
  - 考虑样本自身的分布后的平均错误率计算如下：

$$
\begin{aligned}
P(error) &= \int_{-\infty}^{t} P(\omega_2 \mid x) p(x) dx + \int_{t}^{\infty} P(\omega_1 \mid x) p(x) dx \\
&= \int_{-\infty}^{t} P(x \mid \omega_2) P(\omega_2) dx + \int_{t}^{\infty} P(x \mid \omega_1) P(\omega_1) dx \\
&= P(x \in R_1, \omega_2) + P(x \in R_2, \omega_1)
\end{aligned}
$$

记为



2.6 分类错误率



- **两类情形**
  - 平均错分概率：

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

从联合概率来理解

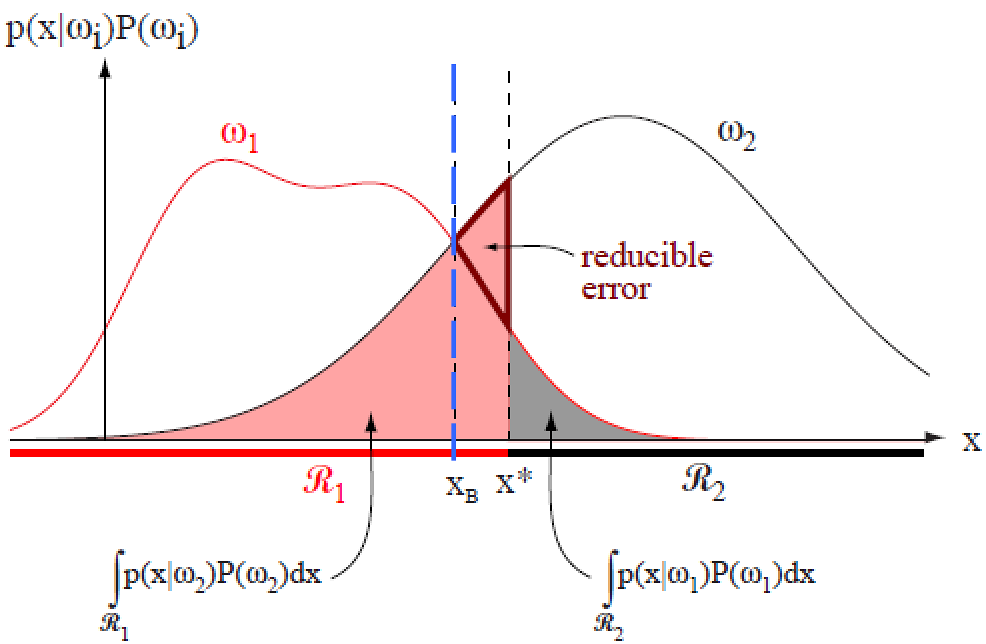
- **多类情形**
  - **平均错分概率：** $$\boxed{P(error) = \sum_{i=1}^{c} \sum_{j \neq i} P(\mathbf{x} \in R_j, \omega_i)}$$ 对样本$\mathbf{x}$
  - **平均分类精度：**

$$P(correct) = \sum_{i=1}^{c} P(\mathbf{x} \in R_i, \omega_i)$$
$$= \sum_{i=1}^{c} P(\mathbf{x} \in R_i \mid \omega_i) P(\omega_i)$$
$$= \int_{R_i} p(\mathbf{x} \mid \omega_i) P(\omega_i) d\mathbf{x}$$

2.6 分类错误率

- **两类的例子----平均错分概率：**

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



X*不是一个最优的决策点，X_B是一个最优的决策点

讨 论

- 贝叶斯分类器(基于贝叶斯决策的分类器)是最优的吗?
  - 最小风险、最大后验概率决策
  - 最优的条件:概率密度、风险能准确估计
  - 具体的参数法、非参数法是贝叶斯分类器的近似,实际中难以达到最优
  - 判别模型:回避了概率密度估计,以较小复杂度估计后验概率或判别函数
  - 什么方法能胜过贝叶斯分类器:在不同的特征空间才有可能!

下次课内容

– 最大似然参数估计
– 贝叶斯估计
– 。。。

Thank All of You!

(Questions?)

向世明

smxiang@nlpr.ia.ac.cn

http://people.ucas.ac.cn/~xiangshiming

时空数据分析与学习课题组(STDAL)

中科院自动化研究所· 模式识别国家重点实验室