数据挖掘与机器学习

1

潘斌
panbin@nankai.edu.cn
范孙楼227

上节回顾

- 图像特征提取
  - 灰度共生矩
  - 局部二值模式LBP
- 感知准则
  - $a^Ty_i > 0$——正确分类
  - $a^Ty_i < 0$——错误分类
  - 梯度下降求解
- 最小错分、最小平方误差准则
- 贝叶斯派和频率派

贝叶斯公式

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)}{P(X)}$$

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

4.1 基本概念

- **Bayes**分类器设计思路
  - 寻求概率意义上的最小错误率的分类器
  - 即具有最小错分概率的分类器——分类器设计的最优解

4.2.1 问题（P代表概率，p代表概率密度函数）

- **Bayes决策已知条件和问题（先考虑C = 2分类问题）**
  - 两类问题：ω_1和ω_2
  - 先验概率：$P(\omega_1)$ 和$P(\omega_2)$
  - 类条件概率密度函数：$p(x|\omega_1)$ 和$p(x|\omega_2)$
  - 发生了一个随机事件，其观察值为：特征向量x

  - **求最小错误率分类器**

4.2.2 后验概率

- **Bayes（条件概率）公式**
  - $p(x|\omega_1)\,P(\omega_1) = p(x)\,P(\omega_1|x)$
  - $p(x|\omega_2)\,P(\omega_2) = p(x)\,P(\omega_2|x)$

- 后验概率
  - $P(\omega_1|x) = p(x|\omega_1)\,P(\omega_1)\,/\,p(x)$
  - $P(\omega_2|x) = p(x|\omega_2)\,P(\omega_2)\,/\,p(x)$
  - $p(X) = p(x|\omega_1)\,P(\omega_1) + p(x|\omega_2)\,P(\omega_2)$

4.2.3 决策规则

- **决策规则**
  - 比较后验概率，取最大值进行类别判断
  - 对于未知样本x，若$P(\omega_1|x) > P(\omega_2|x)$，则$x \in \omega_1$
  - 若$P(\omega_1|x) < P(\omega_2|x)$，则$x \in \omega_2$

4.2.3 决策规则

- **等价规则一 后验概率分子**
  - 比较分子$p(x|\omega_1)\,P(\omega_1)$ 和 $p(x|\omega_2)\,P(\omega_2)$，取最大
  - 对于未知样本x，若$p(x|\omega_1)\,P(\omega_1) > p(x|\omega_2)\,P(\omega_2)$，则$x \in \omega_1$
  - 若$p(x|\omega_1)\,P(\omega_1) < p(x|\omega_2)\,P(\omega_2)$，则$x \in \omega_2$

4.2.3 决策规则

- **等价规则二 似然比**
  - 定义似然比函数$l(x) = p(x|\omega_1) / p(x|\omega_2)$
  - 对于未知样本x，若$l(x) > P(\omega_2) / P(\omega_1)$，则$x \in \omega_1$
  - 若$l(x) < P(\omega_2) / P(\omega_1)$，则$x \in \omega_2$

4.2.3 决策规则

- **等价规则三 负对数似然比**
  - 定义负对数似然比函数$h(x) = -\ln l(x) = -\ln p(x|\omega_1) + \ln p(x|\omega_2)$
  - 对于未知样本x，若$h(x) < \ln [P(\omega_1) / P(\omega_2)]$，则$x \in \omega_1$
  - 若$h(x) > \ln [P(\omega_1) / P(\omega_2)]$，则$x \in \omega_2$

4.2.4 实例

- **已知**
  - 癌细胞图像识别：正常和异常两类（即C = 2）
  - 已知未知样本特征观察值：x = 0.5
  - 又已知$P(\omega_1) = 0.9$和$P(\omega_2) = 0.1$
  - 查函数曲线得$p(0.5|\omega_1) = 0.2$和$p(0.5|\omega_2) = 0.4$
（可以是先验或学习得到）

  - 试对未知样本x = 0.5进行分类

4.2.4 实例

- 已知
  - 癌细胞图像识别：正常和异常两类（即C = 2）
  - 已知未知样本特征观察值：x = 0.5
  - 又已知$P(\omega_1) = 0.9$和$P(\omega_2) = 0.1$
  - 查函数曲线得$p(0.5|\omega_1) = 0.2$和$p(0.5|\omega_2) = 0.4$
  - 试对未知样本x = 0.5进行分类

解：利用贝叶斯公式，分别计算出 ω_1 及 ω_2 的后验概率。

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

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

根据贝叶斯决策规则式(2-2)，有

$$P(\omega_1|x)=0.818>P(\omega_2|x)=0.182$$

所以合理的决策是把 x 归类于正常状态。

4.2.5 最小错误率的说明

- **最小错误率的说明（设C=2，D=1）**
  - 错误率P(e)的定义

首先应指出所谓错误率是指平均错误率，以 $P(e)$ 来表示，其定义为

$$P(e)=\int_{-\infty}^{\infty}P(e,\mathbf{x})\mathrm{d}\mathbf{x}=\int_{-\infty}^{\infty}P(e|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2-6}$$

其中 $\int_{-\infty}^{\infty}(\quad)\mathrm{d}\mathbf{x}$ 表示在整个 d 维特征空间上的积分。

对两类别问题，从式(2-2)的决策规则可知，如果 $P(\omega_2|\mathbf{x})>P(\omega_1|\mathbf{x})$，则决策应为 ω_2。显然在作出决策 ω_2 时，$\mathbf{x}$ 的条件错误概率为 $P(\omega_1|\mathbf{x})$；反之，则应为 $P(\omega_2|\mathbf{x})$。可表示为

$$P(e|\mathbf{x})=\begin{cases}P(\omega_1|\mathbf{x}), & \text{当 } P(\omega_2|\mathbf{x})>P(\omega_1|\mathbf{x})\\ P(\omega_2|\mathbf{x}), & \text{当 } P(\omega_1|\mathbf{x})>P(\omega_2|\mathbf{x})\end{cases} \tag{2-7}$$

4.2.5 最小错误率的说明

- **最小错误率的说明（设C=2，D=1）**
  - 错误率P(e)的推导
    - 做ω_1判别时的错误概率

$$P(e_{12})=\int_{\Re_1} P(\omega_2 \mid x)p(x)dx=\int_{\Re_1} P(\omega_2)\, p(x \mid \omega_2)dx$$

    - 做ω_2判别时的错误概率

$$P(e_{21})=\int_{\Re_2} P(\omega_1 \mid x)p(x)dx=\int_{\Re_2} P(\omega_1)p(x \mid \omega_1)dx$$

    - 总错误概率

$$P(e)=\int_{\Re_1} P(\omega_2)\, p(x \mid \omega_2)dx+\int_{\Re_2} P(\omega_1)p(x \mid \omega_1)dx$$

4.2.5 最小错误率的说明

- **最小错误率的说明（设C=2，D=1）**
  - 再假设t为唯一分界点（**C=2，D=1**）
    - 做ω_1判别时的错误概率

$$P(e_{12}) = \int_{-\infty}^{t} P(\omega_2)\, p(x \mid \omega_2) dx$$

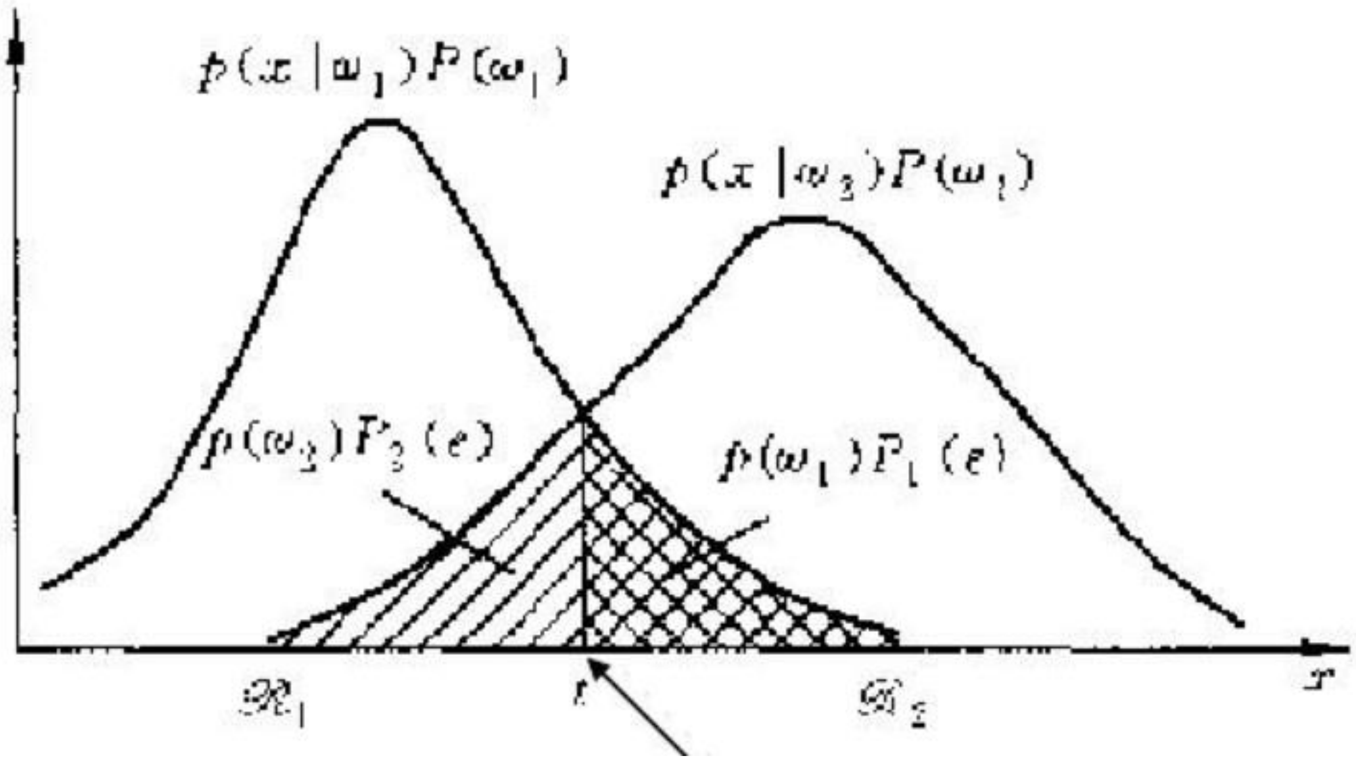
    - 做ω_2判别时的错误概率

$$P(e_{21}) = \int_{t}^{\infty} P(\omega_1)\, p(x \mid \omega_1) dx$$

    - 总错误概率

$$P(e) = \int_{-\infty}^{t} P(\omega_2)\, p(x \mid \omega_2) dx + \int_{t}^{\infty} P(\omega_1)\, p(x \mid \omega_1) dx$$

4.2.5 最小错误率的说明



$$P_1(e)=\int_{R_2} p(\mathbf{x}|\omega_1)d\mathbf{x}$$

$$P_2(e)=\int_{R_1} p(\mathbf{x}|\omega_2)d\mathbf{x}$$

4.2.5 最小错误率的说明

- **推广到多类（任意C类）**
  - 已知条件和问题

  - C类D维问题：ω_1， ω_2， …， ω_C
  - 先验概率：$P(\omega_1)$， $P(\omega_2)$， …， $P(\omega_C)$
  - 类条件概率密度函数：$p(x \mid \omega_1)$， $p(x \mid \omega_2)$， …， $p(x \mid \omega_C)$
  - 发生了一个随机事件，其观察值为：特征向量x

  - **求最小错误率分类器**

4.2.5 最小错误率的说明

- **推广到多类（任意C类）**
  - 判别函数
    - $P(\omega_i|x) = p(x|\omega_i)\, P(\omega_i) / p(x)$ 　　$i = 1, \ldots, C$
  - 决策规则
    - 对于未知样本x，若$P(\omega_j|x) = \max P(\omega_i|x)$，则$x \in \omega_j$

4.2.6 特点

- **最小错误率Bayes决策的特点**
  - 已知条件多——各类概率分布
  - 最小错误率——概率意义上最优
  - 非线性分类器
  - 设计过程复杂

4.3 最小风险Bayes决策

- **4.3.1 问题的提出**
- **4.3.2 决策规则**
- **4.3.3 其它说明**
- **4.3.4 特点**

4.3.1 问题的提出

- **最小错误率Bayes决策**
  - 最小错误率——概率意义上最优
  - 工程上是否最优?
- 错误分类的结果、代价或风险会是怎样的?
  - 考虑癌细胞图像识别的例子
- 出错的可能情况
  - 正常细胞ω_1错分为异常ω_2
  - 异常细胞ω_2错分为正常ω_1

4.3.1 问题的提出

- **区别状态和决策概念**
  - 状态：识别的目的是分类，把样本归类于其可能的自然状态（即类别）之一，将这种自然状态简称为状态，记为ω
  - 状态空间：所有可能的状态的集合构成状态空间，记为**Ω**

4.3.1 问题的提出

- **区别状态和决策概念**
  - 决策：把样本归类于某个状态，或不能进行归类，都是决策，记为α
  - 决策空间：所有可能的决策（包括拒绝决策）的集合构成决策空间，记为A

4.3.1 问题的提出

- **已知条件（类别C = 2，决策A = 2）**
  - ω_1、$P(\omega_1)$、 $p(x \mid \omega_1)$和ω_2 、$P(\omega_2)$、 $p(x \mid \omega_2)$
  - $\Omega = \{\omega_1，\omega_2\}$
  - $A = \{\alpha_1，\alpha_2\}$
  - 定义损失函数$\lambda(\alpha_i, \omega_j)$，简记$\lambda_{ij}$
  - 发生了一个随机事件，其观察值为特征向量x

  - **求最小风险分类器**

4. 3. 2 决策规则

- **判别函数**
  - $P(\omega_j | x) = p(x | \omega_j) P(\omega_j) / p(x)$ $j = 1, 2$
  - $R(\alpha_i | x) = E[\lambda(\alpha_i, \omega_j) | x] = \lambda_{i1} P(\omega_1 | x) + \lambda_{i2} P(\omega_2 | x)$ $i = 1, 2$

- **决策规则**
  - 对于未知样本x，若$R(\alpha_k | x) = \min R(\alpha_i | x)$，则$x \in \omega_k$，**即决策**$\alpha_k$

4.3.2 决策规则

- **推广到任意情况（C个类别，A个决策）**
  - ω_j、$P(\omega_j)$、 $p(x|\omega_j)$ $j = 1, \ldots, C$
  - $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{C.}\}$
  - $A = \{\alpha_1, \alpha_2, \ldots, \alpha_A\}$
  - 定义损失函数$\lambda(\alpha_i, \omega_j)$，简记$\lambda_{ij}$
  - 发生了一个随机事件，其观察值为特征向量x

  - **求最小风险分类器**

4.3.2 决策规则

- **判别函数**
  - $P(\omega_j | x) = p(x | \omega_j) P(\omega_j) / p(x)$ $\quad j = 1, 2, \ldots, C$
  - $R(\alpha_i, x) = E\left[\lambda(\alpha_i, \omega_j)\right] = \sum_{j=1} \lambda(\alpha_i, \omega_j) P(\omega_j | x)$ $\quad i = 1, 2, \ldots, A$
- **决策规则**
  - 对于未知样本x，若$R(\alpha_k | x) = \min R(\alpha_i | x)$，则$x \in \omega_k$，**即决策**$\alpha_k$

4.3.3 其它说明

- **最小风险与最小错误率的关系**
  - 0-1损失函数

$\lambda(\alpha_i, \omega_j) = 0 \qquad i = j$

$\lambda(\alpha_i, \omega_j) = 1 \qquad i \neq j$

$$R(\alpha_i, x) = E\left[\lambda(\alpha_i, \omega_j)\right] = \sum_{j=1} \lambda(\alpha_i, \omega_j) P(\omega_j \mid x)$$

$$R(\alpha_i, x) = \sum_{\substack{j=1 \\ j \neq i}} P(\omega_j \mid x) = 1 - P(\omega_i \mid x)$$

4.3.3 其它说明

- **最小风险与最小错误率的关系**

$$R(\alpha_i, x) = \sum_{\substack{j=1 \\ j \neq i}}^{L} P(\omega_j \mid x) = 1 - P(\omega_i \mid x)$$

  - 对于未知样本x，若$R(\alpha_k | x) = \min R(\alpha_i | x)$，则$x \in \omega_k$
  - 对于未知样本x，若$P(\omega_k|x) = \max P(\omega_i|x)$，则$x \in \omega_k$

  - **结论：最小错误率Bayes决策，等价于0-1损失函数的最小风险Bayes决策**

4.3.4 特点

- **最小风险Bayes决策的特点**
  - 已知条件多——各类概率分布及风险系数
  - 最小错误风险——概率意义上最优
  - 非线性分类器
  - 设计过程复杂

例题

- **1、已知**
  - 甲类：$P(\omega_1) = 0.7$和类条件概率密度函数$p(x|\omega_1)$
  - 乙类：$P(\omega_2) = 0.3$和类条件概率密度函数$p(x|\omega_2)$
  - 今有待分类样本特征观察值$x = 10$，且由函数曲线查得$p(10|\omega_1) = 0.2$，$p(10|\omega_2) = 0.5$
  - ⑴试用最小错误率Bayes决策对样本$x = 10$进行分类
  - ⑵试用最小风险Bayes决策对该样本进行分类，设$\lambda_{11}=\lambda_{22}=0$，$\lambda_{12}=2$，$\lambda_{21}=1$

4.4 最小最大决策

- **4.4.1 问题的提出**
- **4.4.2 期望损失**
- **4.4.3 最小最大风险**
- **4.4.4 特点**

4.4.1 问题的提出

- **已知条件和问题（C = 2情况）**
  - 先验概率：考虑$P(\omega_1)$ 和$P(\omega_2)$未知或不确定的情况
  - 此时绝对意义的最小风险不存在
  - 如何求Bayes分类器

  - 思路
    - 假设$P(\omega_1)$ 和$P(\omega_2)$确定
    - 设计系列最小风险Bayes分类器
    - 取其中最大风险为最小的一个来用
    - 目的是控制最大风险

4.4.1 问题的提出

- **问题（类别C = 2，决策A = 2）**
  - ω_1、$p(x \mid \omega_1)$和ω_2、$p(x \mid \omega_2)$
  - $\Omega = \{\omega_1, \omega_2\}$
  - $A = \{\alpha_1, \alpha_2\}$
  - 损失函数$\lambda(\alpha_i, \omega_j)$，简记$\lambda_{ij}$
  - 发生了一个随机事件，其观察值为特征向量x

  - **求最小最大风险分类器**

4.4.2 期望损失

- 期望损失的推导（回顾最小错误率证明）

$$P(e)=\int_{\Re_1} P(\omega_2)\,p(x\mid\omega_2)dx+\int_{\Re_2} P(\omega_1)\,p(x\mid\omega_1)dx$$

$$P(e_{12})=\int_{\Re_1} P(\omega_2)\,p(x\mid\omega_2)dx$$

$$P(e_{21})=\int_{\Re_2} P(\omega_1)\,p(x\mid\omega_1)dx$$

4.4.2 期望损失

- **期望损失的推导**

$$R=\int R(\alpha(\boldsymbol{x})|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=\int_{\mathcal{R}_1}R(\alpha_1|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}+\int_{\mathcal{R}_2}R(\alpha_2|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$=\int_{\mathcal{R}_1}[\lambda_{11}P(\omega_1)p(\boldsymbol{x}|\omega_1)+\lambda_{12}P(\omega_2)p(\boldsymbol{x}|\omega_2)]\mathrm{d}\boldsymbol{x}+\int_{\mathcal{R}_2}[\lambda_{21}P(\omega_1)p(\boldsymbol{x}|\omega_1)$$

$$+\lambda_{22}P(\omega_2)p(\boldsymbol{x}|\omega_2)]\mathrm{d}\boldsymbol{x}$$

$$P(\omega_2)=1-P(\omega_1)$$

$$\int_{\Re_2}p(x|\omega_1)dx=1-\int_{\Re_1}p(x|\omega_1)dx$$

4.4.2 期望损失

- 期望损失的推导

$$R=\lambda_{22}+(\lambda_{12}-\lambda_{22})\int_{\mathcal{R}_1}p(\boldsymbol{x}|\omega_2)\mathrm{d}\boldsymbol{x}+P(\omega_1)[(\lambda_{11}-\lambda_{22})$$

$$+(\lambda_{21}-\lambda_{11})\int_{\mathcal{R}_2}p(\boldsymbol{x}|\omega_1)\mathrm{d}\boldsymbol{x}-(\lambda_{12}-\lambda_{22})\int_{\mathcal{R}_1}p(\boldsymbol{x}|\omega_2)\mathrm{d}\boldsymbol{x}]$$

$$a=\lambda_{22}+(\lambda_{12}-\lambda_{22})\int_{\mathcal{R}_1}p(\boldsymbol{x}|\omega_2)\mathrm{d}\boldsymbol{x}$$

$$b=(\lambda_{11}-\lambda_{22})+(\lambda_{21}-\lambda_{11})\int_{\mathcal{R}_2}p(\boldsymbol{x}|\omega_1)\mathrm{d}\boldsymbol{x}-(\lambda_{12}-\lambda_{22})\int_{\mathcal{R}_1}p(\boldsymbol{x}|\omega_2)\mathrm{d}\boldsymbol{x}$$

推导思路：替换后改成只有P(w2)有关的形式

4.4.3 最小最大风险

- **最小最大风险的解**
  - 任给$P(\omega_1)^*$的值，$P(\omega_2)^* = 1 - P(\omega_1)^*$
  - ↓
  - 求对应的最小风险Bayes分类器，设决策域为R_1和R_2
  - ↓
  - 计算对应的a^*和b^*，得到$R^* = a^* – b^* P(\omega_1)^*$
  - ↓
  - 重复上述计算步骤，得到一系列最小风险Bayes分类器，并 可得$P(\omega_1)^*$——R^* 关系曲线
  - ↓
  - 比较所有最大风险，取其中最小的一个，作为最终的分类器

4.4.4 最小最大决策

- **最小最大决策的特点**
  - 已知条件多——各类概率分布及风险系数
  - 最小最大风险——概率意义上最优
  - 非线性分类器
  - 设计过程很复杂

4. 5 Bayes分类器设计

- **4.5.1 原理设计**
- **4.5.2 Bayes决策面**
- **4.5.3 错误率估计**
- **4.5.4 其它**

4.5.1 原理设计

- **两类情况（C=2）设计方法一**
  - 定义判别函数g(x)

    ① $g(\boldsymbol{x})=P(\omega_1|\boldsymbol{x})-P(\omega_2|\boldsymbol{x})$

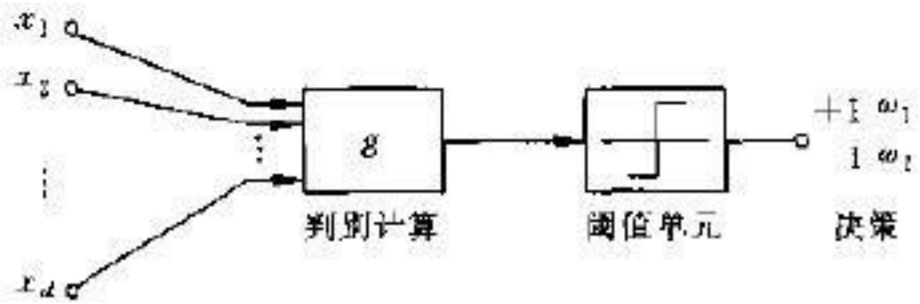
    ② $g(\boldsymbol{x})=p(\boldsymbol{x}|\omega_1)P(\omega_1)-p(\boldsymbol{x}|\omega_2)P(\omega_2)$

    ③ $g(\boldsymbol{x})=\ln\dfrac{p(\boldsymbol{x}|\omega_1)}{p(\boldsymbol{x}|\omega_2)}+\ln\dfrac{P(\omega_1)}{P(\omega_2)}$

  - 决策面H方程
    - g(x) = 0

4.5.1 原理设计

- **两类情况（C=2）设计方法一**
  - 决策规则
    - 对于未知样本x，若g(x) > 0，则x决策为ω_1类
    - 若g(x) < 0，则x决策为ω_2类
  - 原理图



4.5.1 原理设计

- **两类情况（C=2）设计方法二**
  - 定义判别函数$g_1(x)$和$g_2(x)$

    ① $g_i(\boldsymbol{x})=P(\omega_i|\boldsymbol{x})$

    ② $g_i(\boldsymbol{x})=p(\boldsymbol{x}|\omega_i)P(\omega_i)$

    ③ $g_i(\boldsymbol{x})=\ln p(\boldsymbol{x}|\omega_i)+\ln P(\omega_i)$

  - 决策面H方程
    - $g_1(x) = g_2(x)$

4.5.1 原理设计

- **两类情况（C=2）设计方法二**
  - 决策规则
    - 对于未知样本x，若$g_1(x) > g_2(x)$，则x决策为ω_1类
    - 若$g_1(x) < g_2(x)$，则x决策为ω_2类
  - 原理图

4. 5. 1 原理设计

- **多类情况（C任意）设计方法**
  - 定义判别函数$g_i(x)$　　$i = 1, 2, \ldots, C$

    ① $g_i(\boldsymbol{x}) = P(\omega_i \mid \boldsymbol{x})$

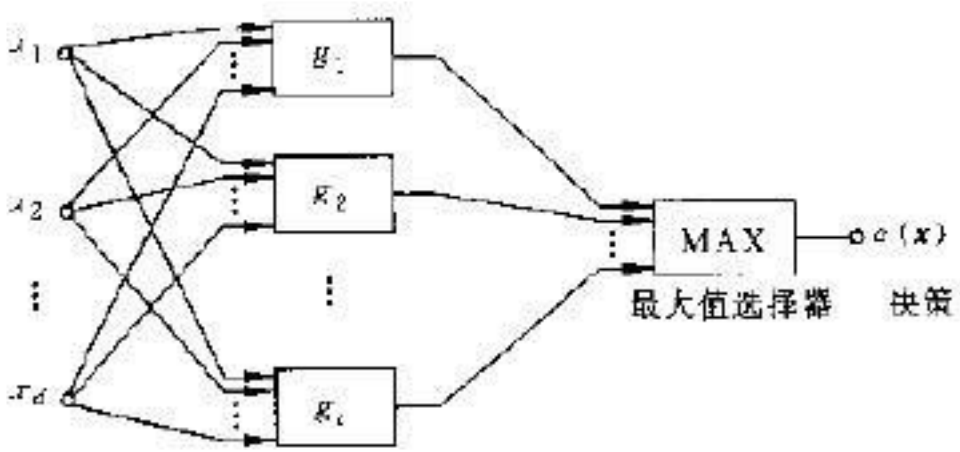
    ② $g_i(\boldsymbol{x}) = p(\boldsymbol{x} \mid \omega_i) P(\omega_i)$

    ③ $g_i(\boldsymbol{x}) = \ln p(\boldsymbol{x} \mid \omega_i) + \ln P(\omega_i)$

  - 决策面H方程
    - $g_i(x) = g_j(x)$　　$i, j = 1, 2, \ldots, C$

4.5.1 原理设计

- **多类情况（C任意）设计方法**
  - 决策规则
    - 对于未知样本x，若$g_j(x)$ = MAX $g_i(x)$，则x决策为ω_j类
  - 原理图



4.5.2 正态分布决策面

- 单变量正态分布 / 一元正态分布

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

4.5.2 正态分布决策面

- 多变量正态分布 / 多元正态分布

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

$$\boldsymbol{\mu}=E\{\boldsymbol{x}\}$$

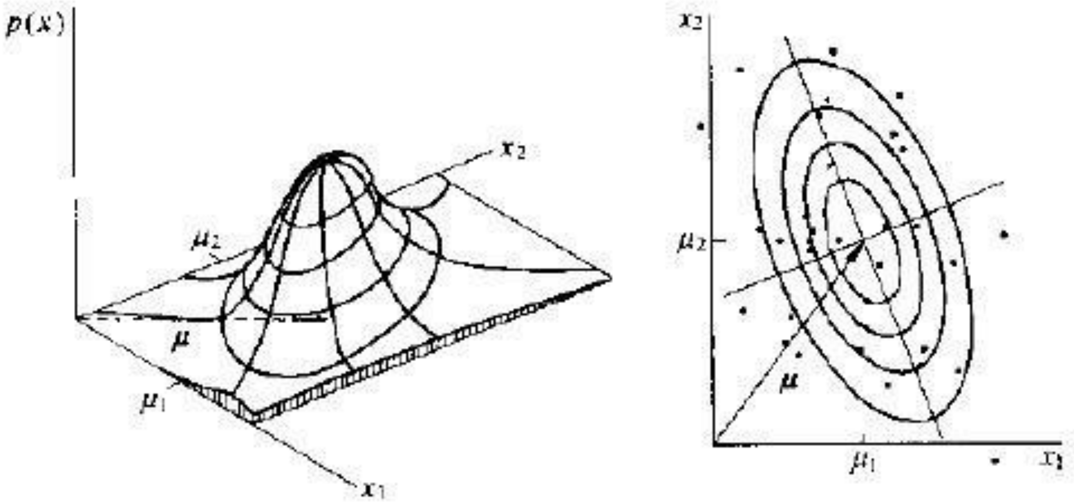
$$\Sigma=E\{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\}$$

4.5.2 正态分布决策面

- 多元正态分布的性质
  - 均值与方差
  - 等概率密度点的轨迹

正态分布的等密度点的轨迹为超椭球面

$$(x-\mu)^t \Sigma^{-1} (x-\mu) = \text{常数}$$



$$\gamma^2 = (x-\mu)^t \Sigma^{-1} (x-\mu)$$

4.5.2 正态分布决策面

- **多元正态分布的性质**
  - 不相关性与独立性：　　　不相关性 = 独立性

$$
\begin{aligned}
\sigma_{ij}^2 &= E[(x_i-\mu_i)(x_j-\mu_j)], \quad i,j=1,2,\cdots,d; i\neq j \\
&= E(x_i-\mu_i)\cdot E(x_j-\mu_j) \\
&= 0
\end{aligned}
$$

$$
\Sigma=\begin{bmatrix} \sigma_{11}^2 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \sigma_{dd}^2 \end{bmatrix}
$$

$$
\Sigma^{-1}=\begin{bmatrix} 1/\sigma_{11}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_{dd}^2 \end{bmatrix}
$$

4.5.2 正态分布决策面

- **多元正态分布的性质**
  - 边缘分布：　正态分布
  - 条件分布：　正态分布
  - 线性组合：　正态分布

4.5.2 正态分布决策面

- **正态分布时分类器的决策面方程**
  - 判别函数

③ $g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$

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

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

$$-\frac{1}{2}[(\mathbf{x}-\boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - (\mathbf{x}-\boldsymbol{\mu}_j)^T \Sigma_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)] - \frac{1}{2}\ln\frac{|\Sigma_i|}{|\Sigma_j|} + \ln\frac{P(\omega_i)}{P(\omega_j)} = 0$$

  - 决策面方程

4.5.2 正态分布决策面

- 正态分布时分类器的决策面方程
  - 二次型判别函数

展开

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

$$= \boldsymbol{x}^T W_i \boldsymbol{x} + \boldsymbol{w}_i^T \boldsymbol{x} + w_{i0}$$

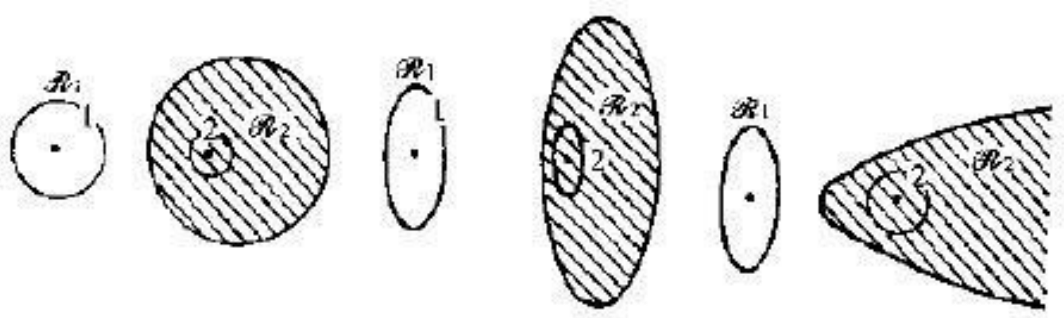
$$W_i = -\frac{1}{2}\Sigma_i^{-1} \qquad (d\times d \text{ 矩阵})$$

$$\boldsymbol{w}_i = \Sigma_i^{-1}\boldsymbol{\mu}_i \qquad (d \text{ 维列向量})$$

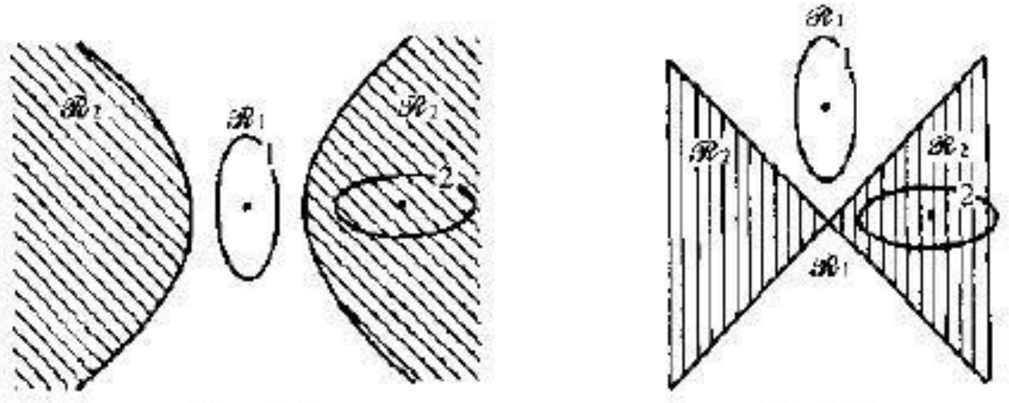
$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T\Sigma_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

4.5.2 正态分布决策面

- 决策面示例



(a) 圆 (b) 椭圆 (c) 抛物线



(d) 双曲线 (e) 直线

4.5.2 正态分布决策面

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

- 决策面特例一
  - $\Sigma_i = \sigma^2 I$　　　　　$i = 1, 2, \ldots, C$

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

朴素贝叶斯分类器

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

准垂直平分

$$\boldsymbol{w}^T(\boldsymbol{x}-\boldsymbol{x}_0) = 0$$

$$\boldsymbol{w} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j$$

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

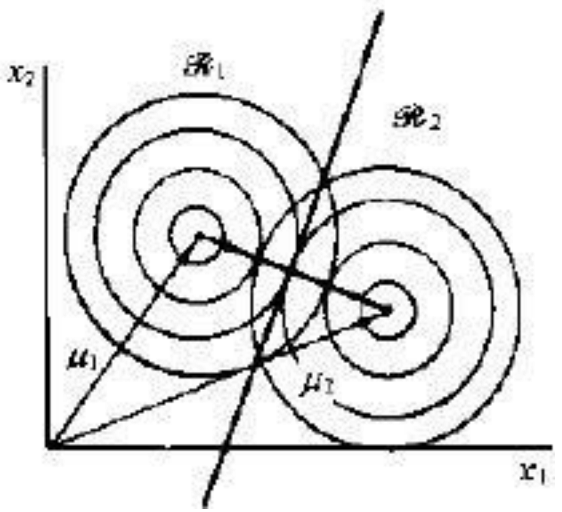
4.5.2 正态分布决策面

- 决策面特例一

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

$$w=\mu_i-\mu_j$$

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



4.5.2 正态分布决策面

- 决策面特例二
  - $\Sigma_i = \Sigma$ $\qquad$ i = 1, 2, …, C

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

$$\mathbf{w}^T(\mathbf{x}-\mathbf{x}_0) = 0$$

$$\mathbf{w} = \Sigma^{-1}(\mu_i - \mu_j)$$

$$\mathbf{x}_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln \frac{P(\omega_i)}{P(\omega_j)}}{(\mu_i - \mu_j)^T\Sigma^{-1}(\mu_i - \mu_j)}(\mu_i - \mu_j)$$

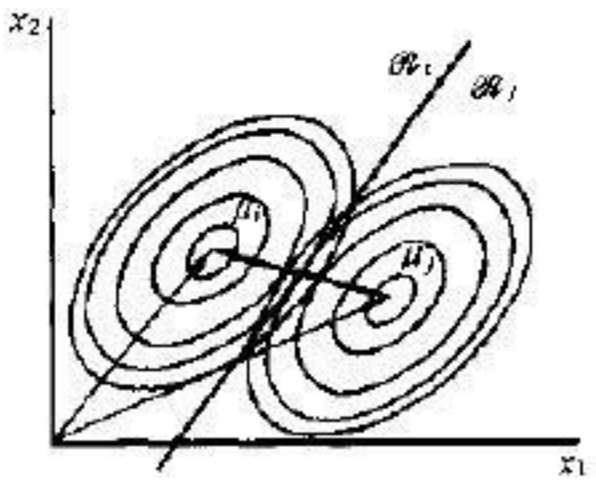
4.5.2 正态分布决策面

- 决策面特例二

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

$$w = \Sigma^{-1}(\mu_i - \mu_j)$$

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



4.5.3 错误率估计

- **Bayes分类器错误率的估计方法**
  - 按理论公式求解计算*
  - 计算错误率上界
  - 实验估计错误率

  - *注：直接计算十分困难，只能在特例情况下进行。

4.5.3 错误率估计

- **错误率的计算**
  - 两类正态分布等协方差阵情况下

$$P(e)=\int_{\Re_1} P(\omega_2)\,p(x\mid\omega_2)dx+\int_{\Re_2} P(\omega_1)\,p(x\mid\omega_1)dx$$

$$P(e_{12})=\int_{\Re_1} P(\omega_2)\,p(x\mid\omega_2)dx=P(\omega_2)\int_{\Re_1} p(x\mid\omega_2)dx=P(\omega_2)P_2(e)$$

$$P(e_{21})=\int_{\Re_2} P(\omega_1)\,p(x\mid\omega_1)dx=P(\omega_1)\int_{\Re_2} p(x\mid\omega_1)dx=P(\omega_1)P_1(e)$$

4.5.3 错误率估计

- 错误率的计算

回顾最小错误率贝叶斯决策规则的负对数似然比形式。

$$h(x)=-\ln l(x)=-\ln p(x|\omega_1)+\ln p(x|\omega_2)\lessgtr\ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right]\to x\in\begin{cases}\omega_1\\\omega_2\end{cases}$$

$h(x)$是x的函数，x是随机向量，因此$h(x)$是随机变量。我们记它的分布密度函数为$p(h|\omega_i)$。由于它是一维密度函数，因此易于积分，所以用它计算错误率有时较为方便。这样式(2-113)可表示为

$$P_1(e)=\int_{\mathscr{R}_2}p(x|\omega_1)dx=\int_t^{\infty}p(h|\omega_1)dh \tag{2-114}$$

$$P_2(e)=\int_{\mathscr{R}_1}p(x|\omega_2)dx=\int_{-\infty}^{t}p(h|\omega_2)dh \tag{2-115}$$

其中 $t=\ln[P(\omega_1)/P(\omega_2)]$ (2-116)

4.5.3 错误率估计

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

$$\mu=E\{\boldsymbol{x}\}$$

$$\Sigma=E\{(\boldsymbol{x}-\mu)(\boldsymbol{x}-\mu)^T\}$$

- 错误率的计算

$$\begin{aligned}
h(\boldsymbol{x}) &= -\ln l(\boldsymbol{x}) = -\ln p(\boldsymbol{x}|\omega_1)+\ln P(\boldsymbol{x}|\omega_2)\\
&= -\left[-\frac{1}{2}(\boldsymbol{x}-\mu_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\mu_1)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_1|\right]\\
&\quad+\left[-\frac{1}{2}(\boldsymbol{x}-\mu_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\mu_2)-\frac{d}{2}\ln 2\pi-\frac{1}{2}\ln|\Sigma_2|\right]\\
&= \frac{1}{2}(\boldsymbol{x}-\mu_1)^T\Sigma_1^{-1}(\boldsymbol{x}-\mu_1)-\frac{1}{2}(\boldsymbol{x}-\mu_2)^T\Sigma_2^{-1}(\boldsymbol{x}-\mu_2)+\frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|}\\
&\lessgtr \ln\frac{P(\omega_1)}{P(\omega_2)} \rightarrow \boldsymbol{x}\in\begin{cases}\omega_1\\ \omega_2\end{cases}
\end{aligned}$$

4.5.3 错误率估计

- **错误率的计算**
  - 两类正态分布等协方差阵情况下

$$h(\boldsymbol{x}) = (\mu_2 - \mu_1)^T \Sigma^{-1} \boldsymbol{x} + \frac{1}{2}(\mu_1^T \Sigma^{-1} \mu_1 - \mu_2^T \Sigma^{-1} \mu_2)$$

$$\lessgtr \ln \frac{P(\omega_1)}{P(\omega_2)} \rightarrow \boldsymbol{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

4.5.3 错误率估计

$$h(\boldsymbol{x})=(\mu_2-\mu_1)^T\Sigma^{-1}\boldsymbol{x}+\frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1-\mu_2^T\Sigma^{-1}\mu_2)$$
$$\lessgtr \ln\frac{P(\omega_1)}{P(\omega_2)}\rightarrow \boldsymbol{x}\in\begin{cases}\omega_1\\ \omega_2\end{cases}$$

- **错误率的计算**
  - 线性组合仍然服从正态分布：通过均值和方差来估计错误率

$$\eta_1=E[h(\boldsymbol{x})|\omega_1]=(\mu_2-\mu_1)^T\Sigma^{-1}\mu_1+\frac{1}{2}(\mu_1^T\Sigma^{-1}\mu_1-\mu_2^T\Sigma^{-1}\mu_2)$$
$$=-\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)$$

$$\eta=\frac{1}{2}[(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)]$$

$$\eta_1=-\eta$$
$$\sigma_1^2=E\{[h(x)-\eta]^2|\omega_1\}=(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)=2\eta$$

$$\eta_2=\frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)=\eta$$
$$\sigma_2^2=(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)=2\eta$$

4.5.3 错误率估计

- **错误率的计算**
  - 线性组合仍然服从正态分布

$$
\begin{aligned}
P_1(e) &= \int_t^{\infty} p(h \mid \omega_1)\,\mathrm{d}h \\
&= \int_t^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{h+\eta}{\sigma}\right)^2\right\}\mathrm{d}h \\
&= \int_t^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{h+\eta}{\sigma}\right)^2\right\}\mathrm{d}\left(\frac{h+\eta}{\sigma}\right) \\
&= \int_{\left(\frac{t+\eta}{\sigma}\right)}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\xi^2\right)\mathrm{d}\xi
\end{aligned}
$$

4.5.3 错误率估计

- **错误率的计算**
  - 线性组合仍然服从正态分布

$$
\begin{aligned}
P_2(e) &= \int_{-\infty}^{t} p(h|\omega_2)\mathrm{d}h \\
&= \int_{-\infty}^{t} (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\frac{h-\eta}{\sigma}\right)^2\right\}\mathrm{d}\left(\frac{h-\eta}{\sigma}\right) \\
&= \int_{-\infty}^{\frac{t-\eta}{\sigma}} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\xi^2\right)\mathrm{d}\xi
\end{aligned}
$$

$$
t = \ln\left[\frac{P(\omega_1)}{P(\omega_2)}\right], \sigma = \sqrt{2\eta}
$$

4.5.3 错误率上界

- **两类情况下最小错误率的上界**
  - Chernoff切尔诺夫上界

$$P(e_{12}) \le P(\omega_2)e^{[-\mu(s)+(1-s)t]}$$

$$P(e_{21}) \le P(\omega_1)e^{[-\mu(s)-st]}$$

  - Bhattacharyya巴氏系数上界

$$J_B = -\ln\left[\int \sqrt{p(x|\omega_1)p(x|\omega_2)}dx\right]$$

$$P(e) \le \sqrt{P(\omega_1)P(\omega_2)}\exp\{-J_B\}$$

4.5.3 错误率估计

- **错误率的实验估计**
  - 测试样本集（区别于训练样本集）
  - 测试结果（有限样本情况，混淆矩阵）
  - 错误率评价

4.5.4 讨论

- **Bayes决策的先决条件**
  - 类条件概率密度的估计
  - 例：密度函数估计的简化
    - 假设特征向量各分量相互独立

4.5.4 讨论

- **Bayes决策的最小错误概率**
  - 设计难度大（类条件概率密度估计）
  - 寻求高“性价比”的分类器
    - 错误概率逼近
    - 决策面逼近

5 非线性分类器

- 5.1 多类问题概述
- 5.2 最小距离分类器
- 5.3 分段线性分类器概述
- 5.4 决策树
- 5.5 近邻法分类器
- 5.6 Adaboost
- 5.7 SVM
- 5.8 人工神经网络

5.1 多类问题概述

- 5.1.1 多类问题
- 5.1.2 解决方案

5.1.1 多类问题

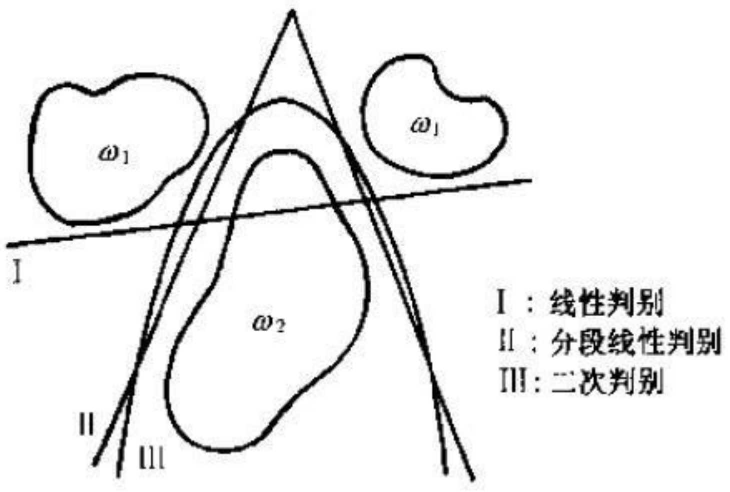
- **多类问题包括**
  - 两类情况（C = 2）
    - 样本集具有多峰分布

5.1.1 多类问题概述

- **多类问题包括**
  - 多类情况（类别数C > 2）
    - 单峰分布
    - 多峰分布

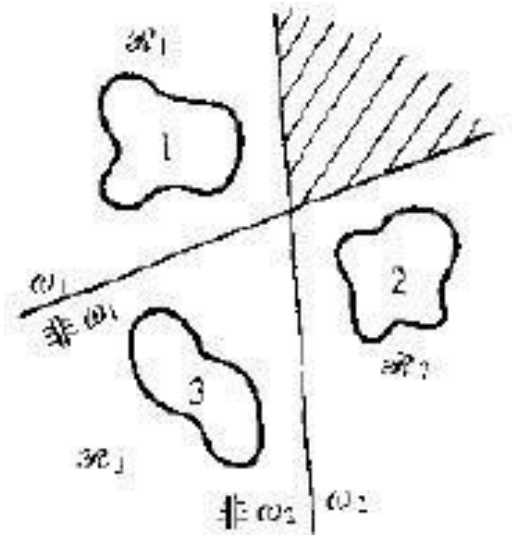
5.1.2 解决方案

- 如何解多类问题
  - Bayes分类器
    - 二次型判别函数
  - 线性分类器
  - 其它分类器



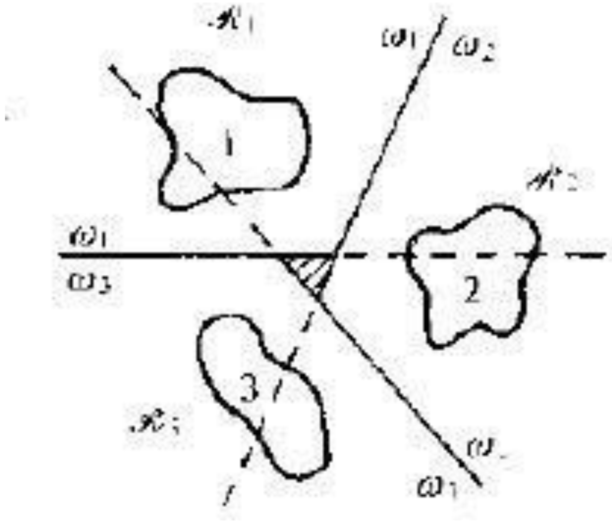
5.1.2 解决方案

- **多类问题是否可以用线性分类器解？**
  - 思路一
    - 对“类与类的非”进行线性分类
    - 只需要C – 1个线性分类器就可以



5.1.2 解决方案

- **多类问题是否可以用线性分类器解？**
  - 思路二
    - “两两分类”进行线性分类
    - 需要C(C – 1)/2个线性分类器就可以

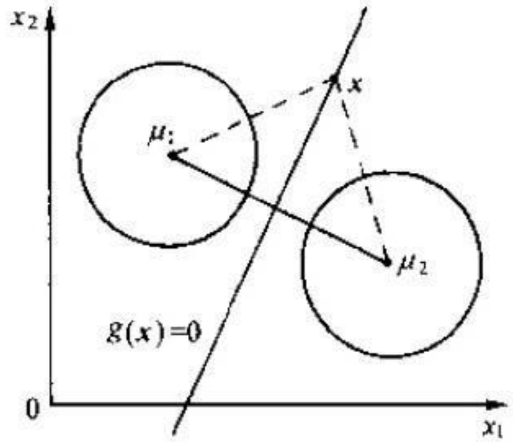


5.2 最小距离分类器

- 5.2.1 最小距离分类器原理
- 5.2.2 分段最小距离分类器
- 5.2.3 特点

5.2.1 最小距离分类器原理

- **回顾两类单峰线性分类器**
  - 垂直平分 / 最小距离分类器
  - 基于两类样本均值点作垂直平分线

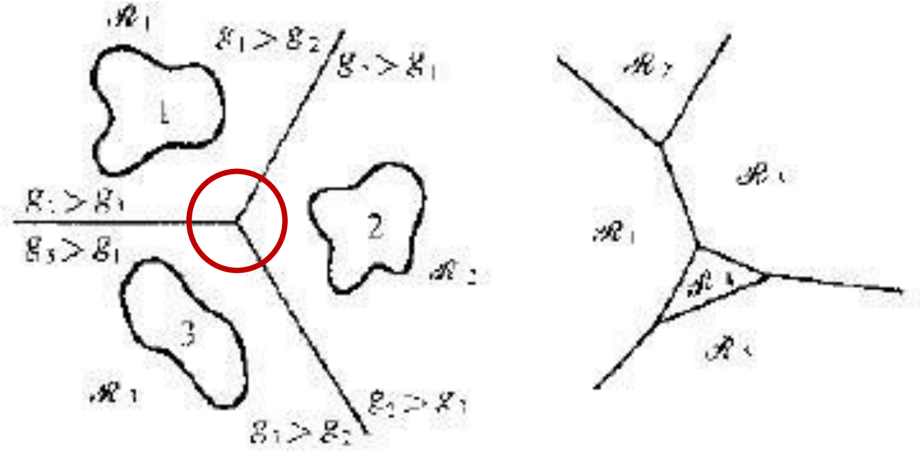


5.2.1 最小距离分类器原理

- **其最小距离形式**
  - 判别函数
    - $G_1(x) = d_1(x) = \|x - m_1\|$
    - $G_2(x) = d_2(x) = \|x - m_2\|$
  - 决策规则
    - 对于未知样本x，若$d_1(x) < d_2(x)$， 则x决策为ω_1类
    - 若$d_1(x) > d_2(x)$， 则x决策为ω_2类

5.2.1 最小距离分类器原理

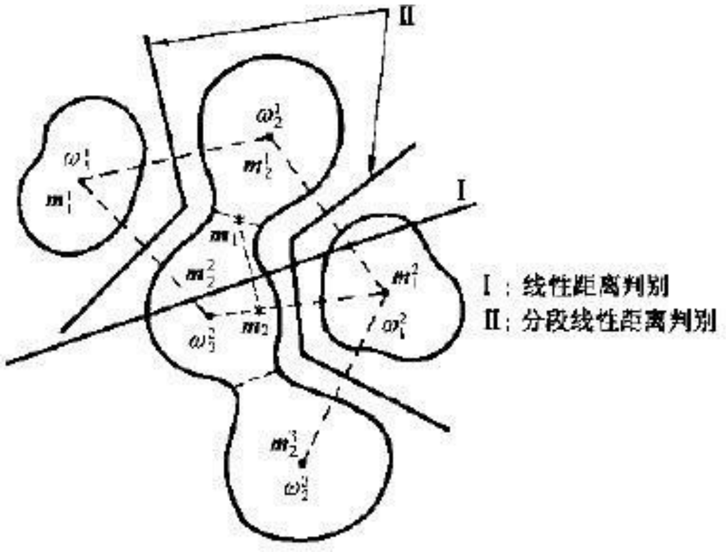
- **直接使用可以解决多类问题**
  - 解决C类单峰问题



(a) 三类　(b) 五类

5.2.1 最小距离分类器原理

- 直接使用可以解决多类问题
  - 解决两类多峰问题



5.2.2 分段最小距离分类器

- **问题**
  - 已知各类及其子类
  - 求分段最小距离分类器

5.2.2 分段最小距离分类器

- **分类器设计**
  - 先求各子类均值
    - m_{ij} （ω_i类的第j子类）
  - 定义各类判别函数
    - $G_i(x) = \min_j \| x - m_{ij} \|$
  - 决策规则
    - 对于未知样本x，若$G_k(x) = \min_i G_i(x)$，则x决策为ω_k类

5.2.3 特点

- **分类器特点**
  - 解决两类多峰或多类问题的分段线性分类器
  - 可以解决几乎所有分类问题但要已知各类子类
  - 概念直观简单，未经优化
  - 分类器设计简单容易
  - （无重叠区或空白区）

THE END !