数据挖掘与机器学习

1

潘斌
panbin@nankai.edu.cn
范孙楼227

上节回顾

- 线性分类器
  - 垂直平分分类器
  - 感知准则
  - 最小错分样本数分类器
  - 最小平方误差分类器
- 贝叶斯派和频率派

本节提要

- 贝叶斯分类器
  - 最小错误率Bayes决策
  - 最小风险Bayes决策
  - 最大最小Bayes决策
  - 贝叶斯分类器的设计

贝叶斯公式

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

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

4.2 最小错误率Bayes决策

- 4.2.1 问题
- 4.2.2 后验概率
- 4.2.3 决策规则
- 4.2.4 实例
- 4.2.5 其它问题
- 4.2.6 特点

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

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

其中 $\int_{-\infty}^{\infty}(\quad)\mathrm{d}x$ 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 $P(\omega_2|x)>P(\omega_1|x)$，则决策应为 ω_2。显然在作出决策 ω_2 时，x 的条件错误概率为 $P(\omega_1|x)$；反之，则应为 $P(\omega_2|x)$。可表示为

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

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

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

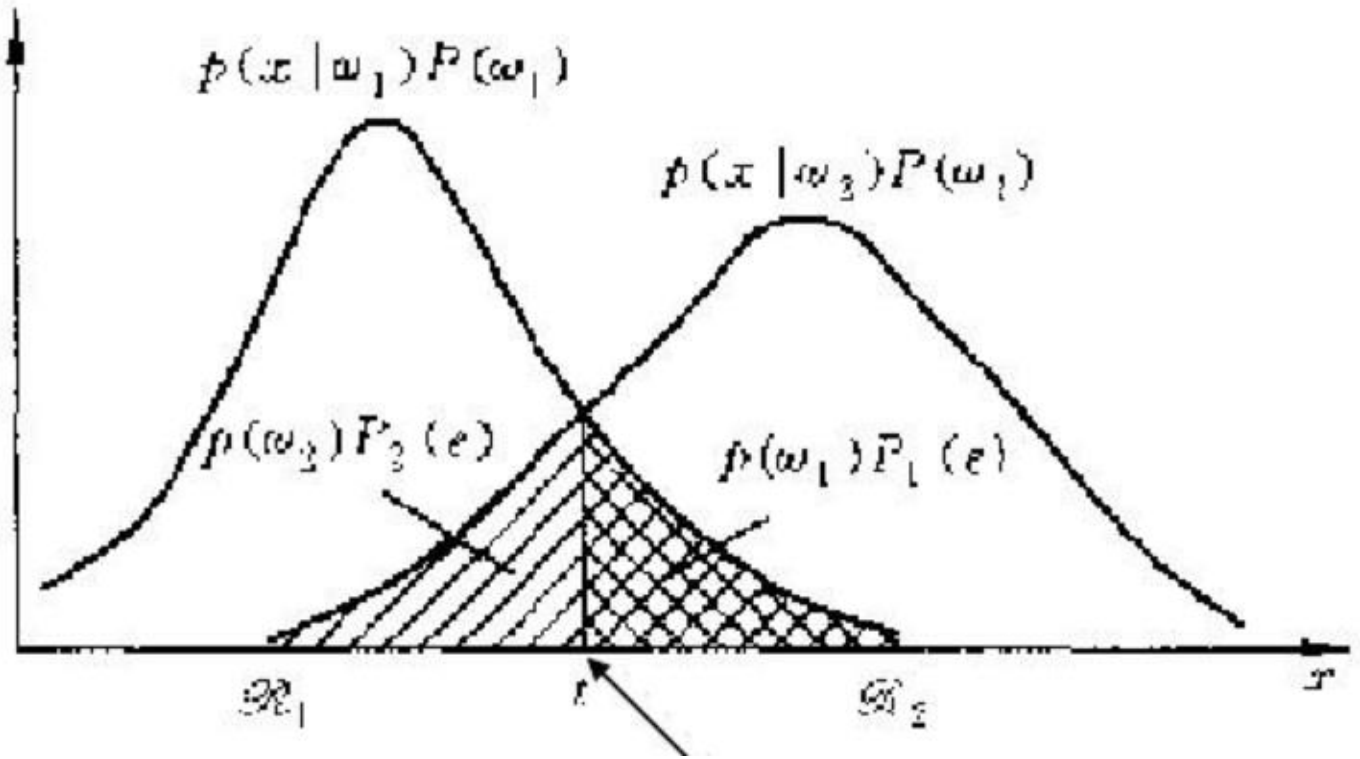
    - 做ω_2判别时的错误概率

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

    - 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
    - $P(\omega_i|x) = p(x|\omega_i)\,P(\omega_i) / p(x) \qquad i = 1, \dots, C$
  - 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**
  - $P(\omega_j \mid x) = p(x \mid \omega_j)\, P(\omega_j) / p(x)$ $\quad j = 1, 2$
  - $R(\alpha_i \mid x) = E[\lambda(\alpha_i, \omega_j) \mid x] = \lambda_{i1} P(\omega_1 \mid x) + \lambda_{i2} P(\omega_2 \mid x)$ $\quad i = 1, 2$

- **决策规则**
  - 对于未知样本x，若$R(\alpha_k \mid x) = \min R(\alpha_i \mid x)$，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**
  - $P(\omega_j \mid x) = p(x \mid \omega_j)\, P(\omega_j) / p(x)$　　　　　$j = 1, 2, \ldots, C$
  - $R(\alpha_i \mid x) = E[\lambda(\alpha_i, \omega_j) \mid x] = \sum_{j=1} \lambda(\alpha_i, \omega_j) P(\omega_j \mid x)$　　$i = 1, 2, \ldots, A$
- **决策规则**
  - 对于未知样本x，若$R(\alpha_k \mid x) = \min R(\alpha_i \mid x)$，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

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

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

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

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

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$R(\alpha_k \mid x) = \min R(\alpha_i \mid x)$，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
    - 设计一系列最小风险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**

联合或条件的
形式均可

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

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

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

$P(\omega_2)=1-P(\omega_1)$，代入

4.4.2 期望损失

- 期望损失的推导

$$
\begin{aligned}
R = \lambda_{22} &+ (\lambda_{12} - \lambda_{22})\int_{\mathcal{R}_1} p(\boldsymbol{x}|\omega_2)\mathrm{d}\boldsymbol{x} + P(\omega_1)[(\lambda_{11} - \lambda_{22}) \\
&+ (\lambda_{21} - \lambda_{11})\int_{\mathcal{R}_2} p(\boldsymbol{x}|\omega_1)\mathrm{d}\boldsymbol{x} - (\lambda_{12} - \lambda_{22})\int_{\mathcal{R}_1} p(\boldsymbol{x}|\omega_2)\mathrm{d}\boldsymbol{x}]
\end{aligned}
$$

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

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

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$R^* = a^* - b^* P(\omega_1)^*$ （注意！**R1R2**随**P(w)**变，即决策会根据先验改变。故**R**和**P(w)**不是线性）
  - ↓
  - 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 $g(\boldsymbol{x})=P(\omega_1|\boldsymbol{x})-P(\omega_2|\boldsymbol{x})$

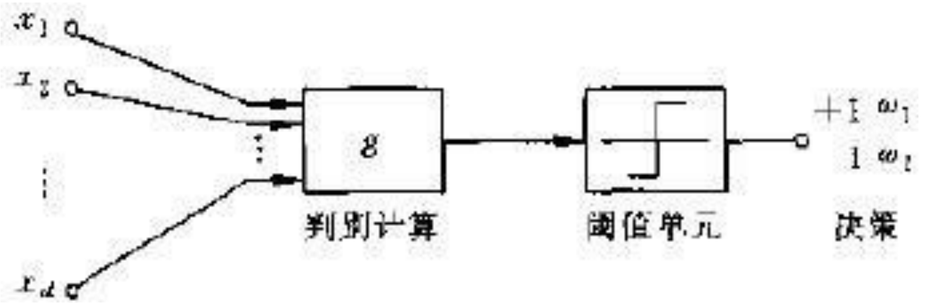
② $g(\boldsymbol{x})=p(\boldsymbol{x}|\omega_1)P(\omega_1)-p(\boldsymbol{x}|\omega_2)P(\omega_2)$

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

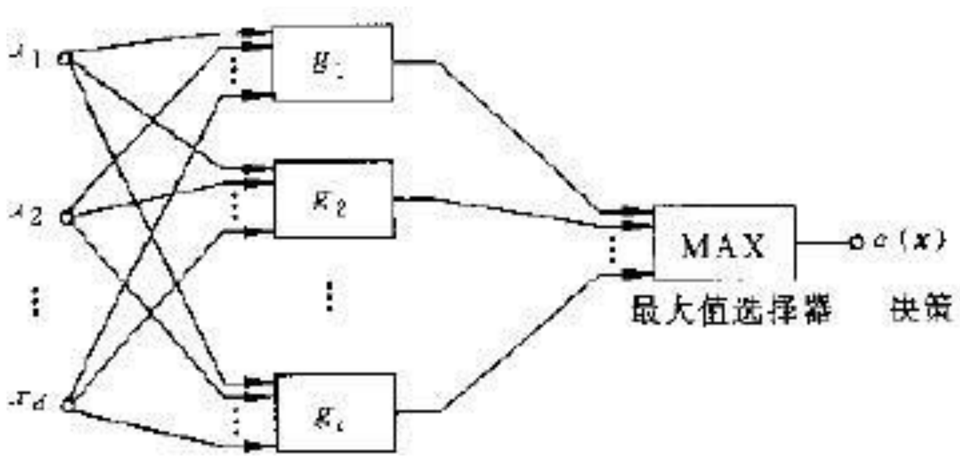
    ③ $g_i(\boldsymbol{x}) = \ln p(\boldsymbol{x} \mid \omega_i) + \ln P(\omega_i)$

  - 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

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \triangleq \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

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

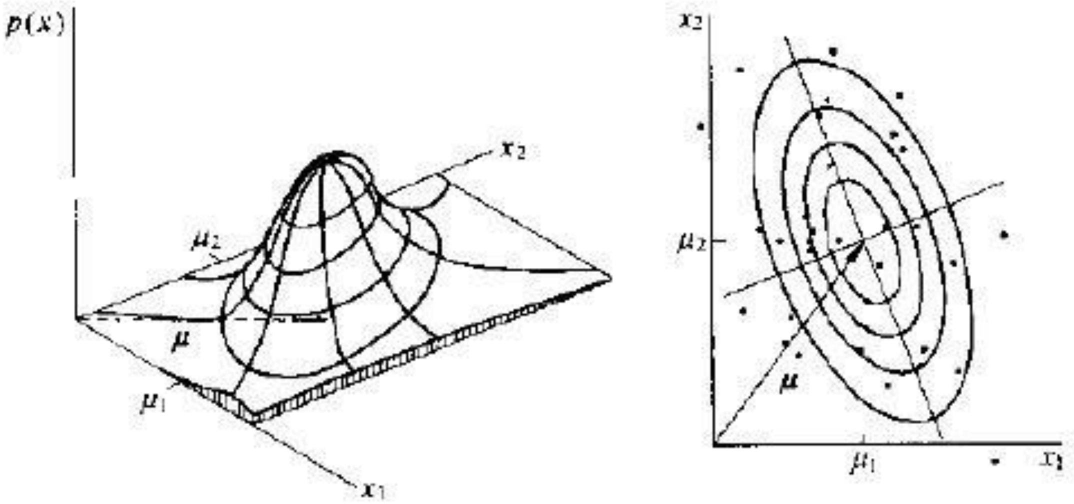
$$\Sigma = E\{(x-\mu)(x-\mu)^T\}$$

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

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

- 正态分布时分类器的决策面方程
  - 判别函数

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

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

  - 决策面方程

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

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

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

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

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

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

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

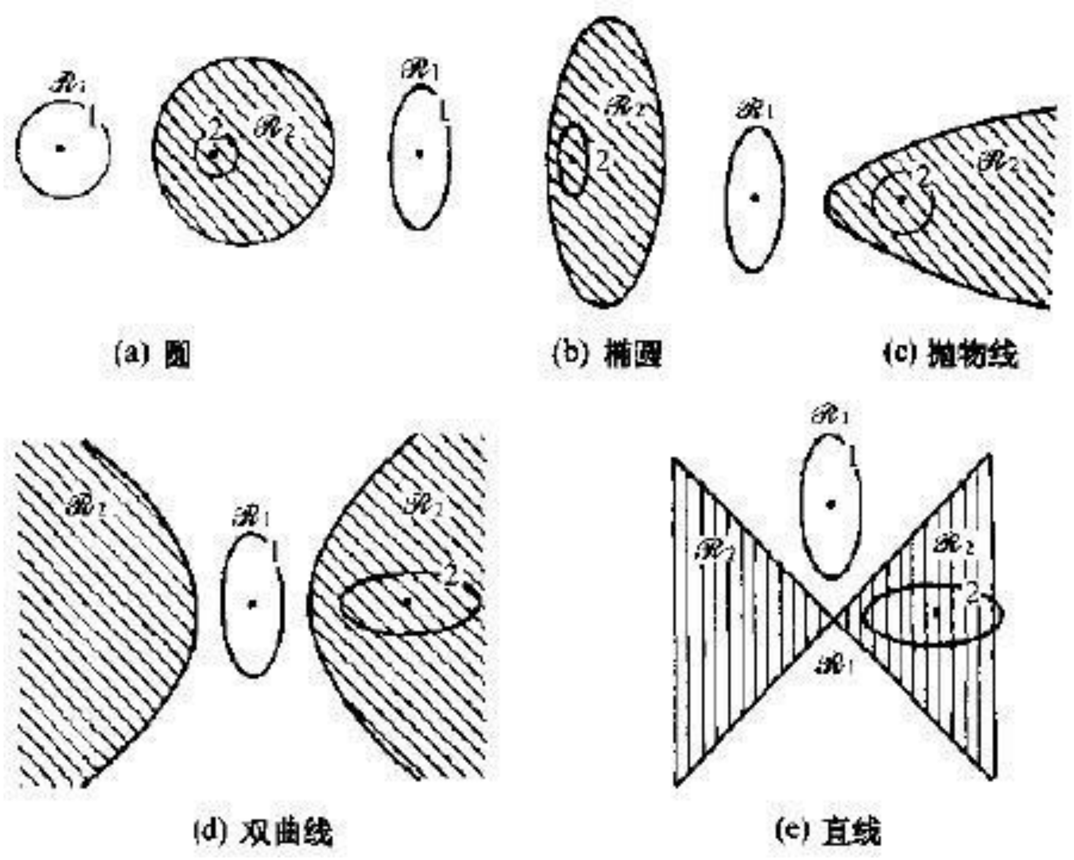
$$w_{i0} = -\frac{1}{2}\boldsymbol{\mu}_i^T\Sigma_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

  - 决策面方程

相减

4.5.2 正态分布决策面

- 决策面示例



THE END !