模式识别

张志飞
zhifeizhang@tongji.edu.cn

第二章 贝叶斯决策论

2.1 引例：鱼分拣
2.2 最小错误率贝叶斯决策
2.3 最小风险贝叶斯决策
2.4 分类器、判别函数
2.5 正态分布下的贝叶斯决策
2.6 离散变量的贝叶斯决策

总括

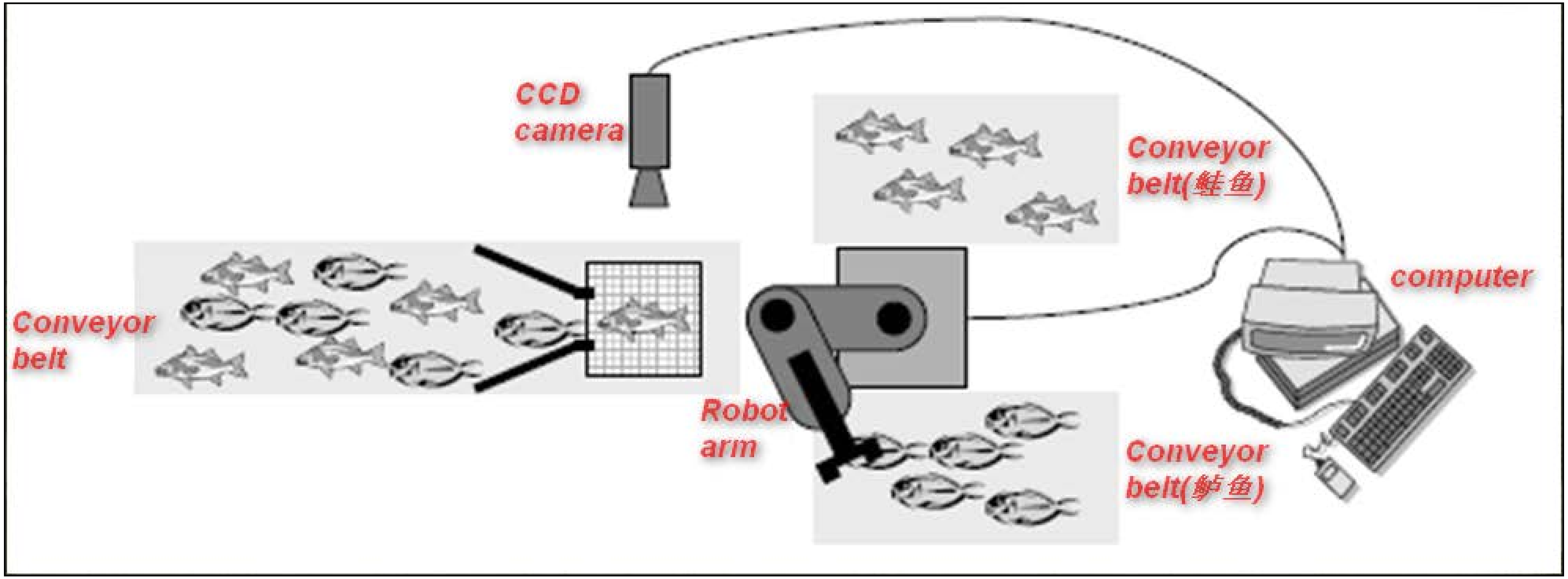
- 贝叶斯决策论是解决模式分类问题的一种基本统计途径。
- 出发点是利用概率决策与相应代价之间的折中。

它做了如下假设：

1)决策问题可以用概率形式来描述；

2)所有有关的概率结构均已知。

2.1 引例　鱼分拣



实现自动分拣的关键在于：

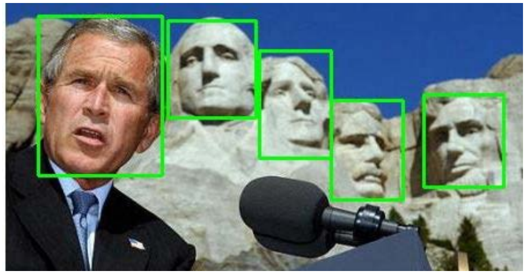
设计一个能分开两类鱼的分类器。

模式和模式识别

- 模式的两个层次：
  - 样本(sample,instance)：观测对象
  - 类别(class)：彼此相似的观测对象构成集合。

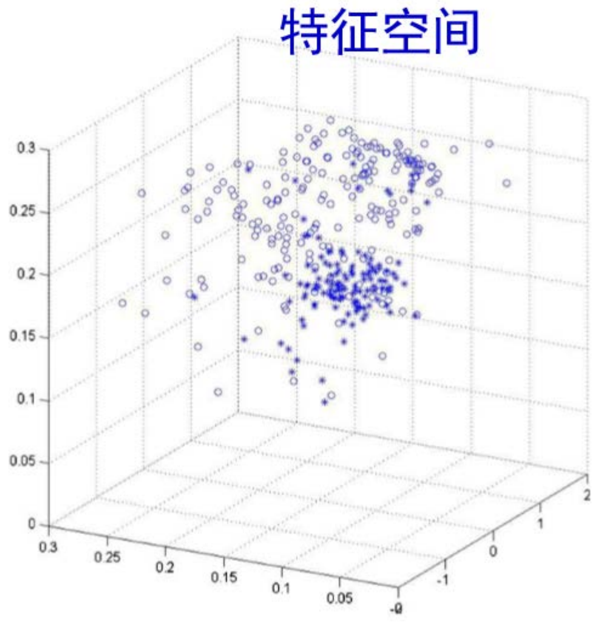
8个样本，2个类别

- 模式识别核心技术：**模式分类**
  - 检测：2-class
  - 判别：2-class, multi-class
  - 分类器设计：机器学习
  - 相关问题：特征提取、特征选择

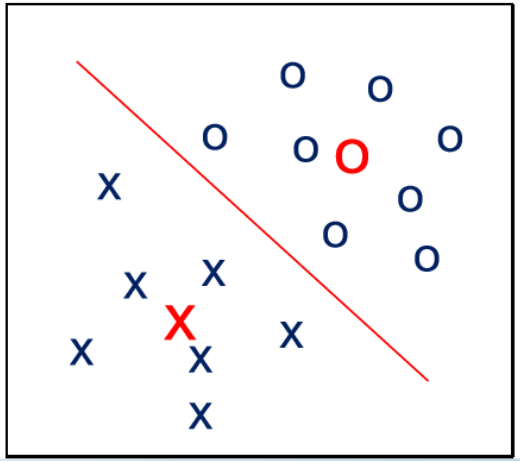


模式表示：两个方面

- 识别对象的表示：
  - 特征向量：$x=[x_1, x_2, \cdots, x_n]^T$　　x_1亮度、 x_2宽度
  - 特征空间：一个模式(样本)对应空间中的一点
  - 好处：容易计算样本之间的距离/相似度
  - 问题：特征提取、特征选择

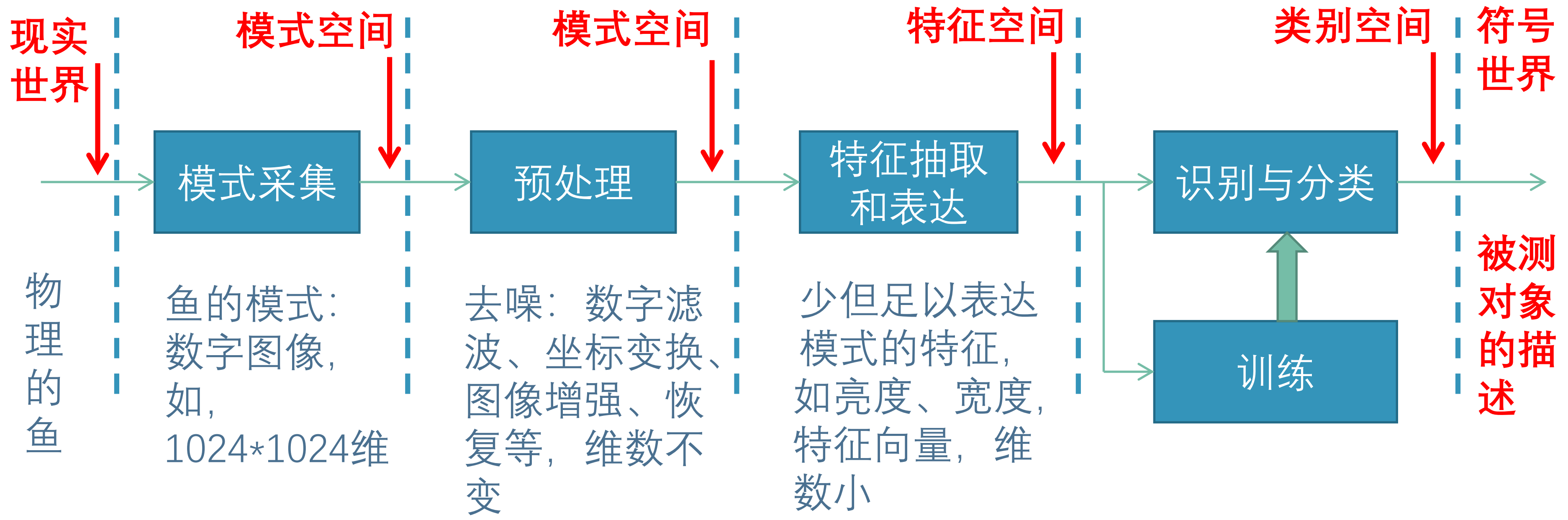


分类问题 → 特征空间划分问题



- 分类器的表示：
  - 类别模型：$M_i=M(x,\theta_i)$
  - 判别函数：$y_i=f(x,w_i)$
  - 决策面：$f(x,w_i) = f(x,w_j)$

模式识别系统的基本构成



根据特征向量在特征空间中的分布，划分特征空间，使得每一区域尽可能只包含来自同一个类别的样本。建立特征空间中的一个分割区域与一个类别之间的关联。获取这种关联关系的过程称为**训练**。根据训练结果，由新样本落在特征空间的区域决定其类别，这称为**分类**或**识别**。

规范化描述

- 类别：$\omega_1, \cdots, \omega_c$表示 c 个类别的有限集
- 特征向量： $\mathbf{x} = (x_1, x_2, \ldots, x_d)^{\mathrm{T}} \in \mathbf{R}^d$
- 先验概率： $p(\omega_j)$ $\sum_{i=1}^{c} P(\omega_i) = 1$
- 类(条件)概率密度：$p(\mathbf{x}|\omega_j)$
- 后验概率，可由**贝叶斯公式**计算，

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

2.2 最小错误率决策：鱼分拣

- 鲈鱼ω_1，鲑鱼ω_2
- 若仅知道两种鱼先验概率，如何判定一新到鱼x的类别？判错概率？

$$\text{如果 } P(\omega_1) > P(\omega_2)\text{，则 } x \in \omega_1\text{；否则，} x \in \omega_2 \qquad (2\text{-}1)$$

$$\because \quad P(\omega_1) + P(\omega_2) = 1,$$

$$\therefore \quad P(error) = \begin{cases} P(\omega_2), & \text{若决策} x \in \omega_1 \\ P(\omega_1), & \text{若决策} x \in \omega_2 \end{cases}$$

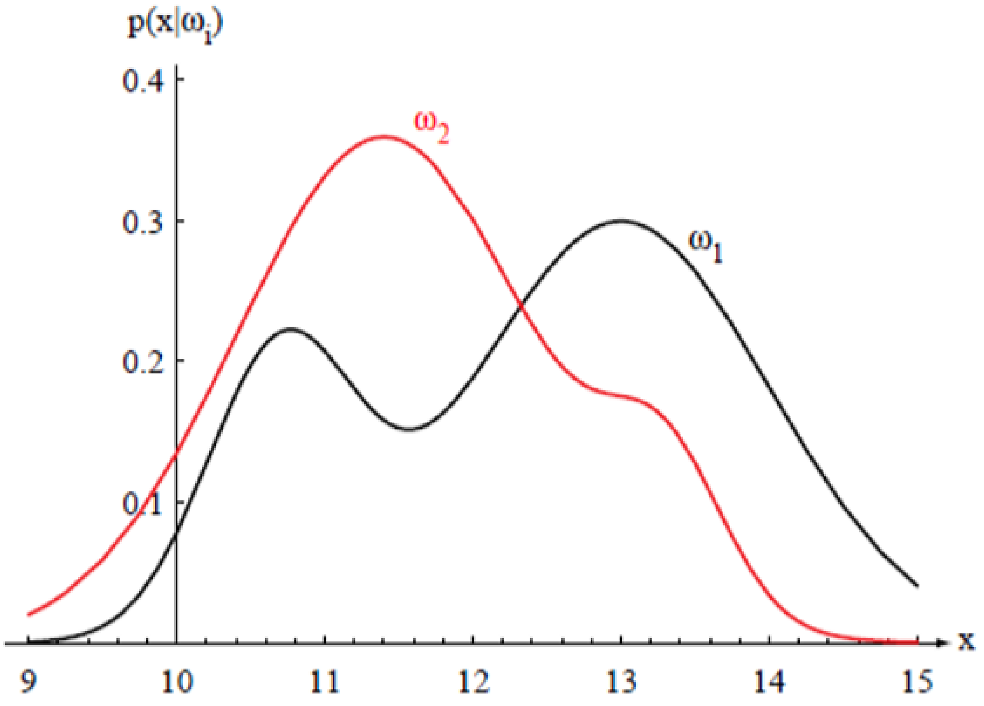
显然，采用式(2-1)的决策犯错误的概率最小。 $\min\{P(\omega_1), P(\omega_2)\}$

- 在所有可能出现的样本上类别决策错误的概率被称为**错误率**。
- 式(2-1)的准则实际上是最小错误率准则。

同济大学 TONGJI UNIVERSITY

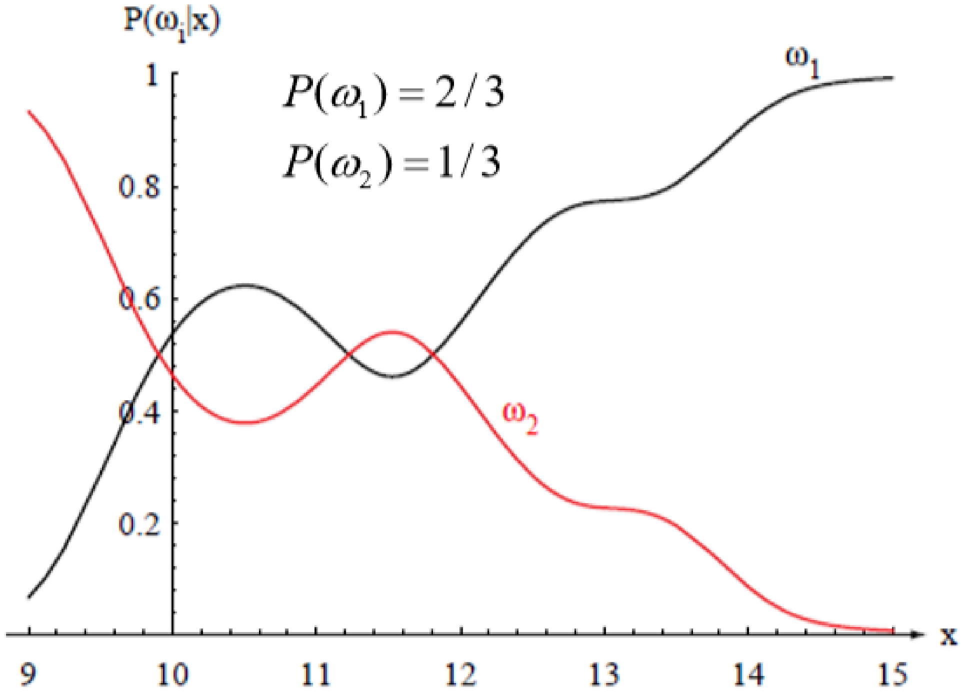
若观测到鱼的光泽度x?

观测量分布$p(x|\omega_i)$



x轴：一维特征空间

计算后验概率



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

贝叶斯公式

- $p(x) = \sum_j p(x|\omega_j) P(\omega_j)$是归一化因子(也称**证据因子**)，**与类别状态无关**。
- $p(x|\omega_j)$为ω_j关于x的**似然函数**，表明在其他条件都相等的情况下，使得$p(x|\omega_j)$较大的ω_j更有可能是真实的类别。

$$P(\boldsymbol{x}, \boldsymbol{\omega}_j) = P(\boldsymbol{\omega}_j|\boldsymbol{x})p(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{\omega}_j)P(\boldsymbol{\omega}_j)$$

$$P(\boldsymbol{\omega}_j|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\omega}_j)P(\boldsymbol{\omega}_j)}{p(\boldsymbol{x})}$$

$$\text{后验} = \frac{\text{似然} \times \text{先验}}{\text{证据}}$$

基于后验概率决策

- 对某一观测 x,
  - 如果 $P(\omega_1|\boldsymbol{x}) > P(\omega_2|\boldsymbol{x})$，则判为$\omega_1$；否则，判为$\omega_2$。 (2-2)
  - 判错概率: $P(error|\boldsymbol{x}) = \begin{cases} P(\omega_1|\boldsymbol{x}) & 如果决策\ \boldsymbol{x} \in \omega_2 \\ P(\omega_2|\boldsymbol{x}) & 如果决策\ \boldsymbol{x} \in \omega_1 \end{cases}$ (2-3)

- 错误率：平均错误概率，定义为独立同分布样本错误概率的期望，即

$$P(error) = \int_{-\infty}^{\infty} P(error,\boldsymbol{x})d\boldsymbol{x} = \int_{-\infty}^{\infty} P(error|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \qquad (2\text{-}4)$$

- 式(2-2)仍然是最小错误率的决策，因为它保证了对每一个x, $P(error|\boldsymbol{x})$ 最小。

最小错误率贝叶斯决策：两类情况

- 从最小错误率的要求出发，利用贝叶斯公式，就能得出使错误率最小的分类决策，称之为**最小错误率贝叶斯决策**。
- 根据式(2-3)可知，使**错误率最小的决策**就是**使后验概率最大的决策**。

对两类问题，最小错误率贝叶斯决策规则：

如果 $P(\omega_1|x) > P(\omega_2|x)$, 则 $x \in \omega_1$; 反之，则 $x \in \omega_2$。

等价形式

(1) 若 $P(\omega_i|x)=\max\limits_{j=1,2} P(\omega_j|x)$, 则 $x \in \omega_i$

(2) 若 $p(x|\omega_i)P(\omega_i) = \max\limits_{j=1,2} P(x|\omega_j)p(\omega_j)$，则$x \in \omega_i$

(3) 似然比 $l(x)$，若 $l(x) = \dfrac{p(x|\omega_1)}{p(x|\omega_2)} \gtrless \dfrac{P(\omega_2)}{P(\omega_1)}$， 则$x \in \begin{cases}\omega_1 \\ \omega_2\end{cases}$

(4) 对数似然比，取 $h(x) = \ln\dfrac{P(\mathrm{X}|\omega_1)}{P(\mathrm{X}|\omega_2)} + \ln\dfrac{P(\omega_1)}{P(\omega_2)}$

决策边界

判决门限为 t
Ω_1 为(-∞,t)
Ω_2 为(t, ∞)

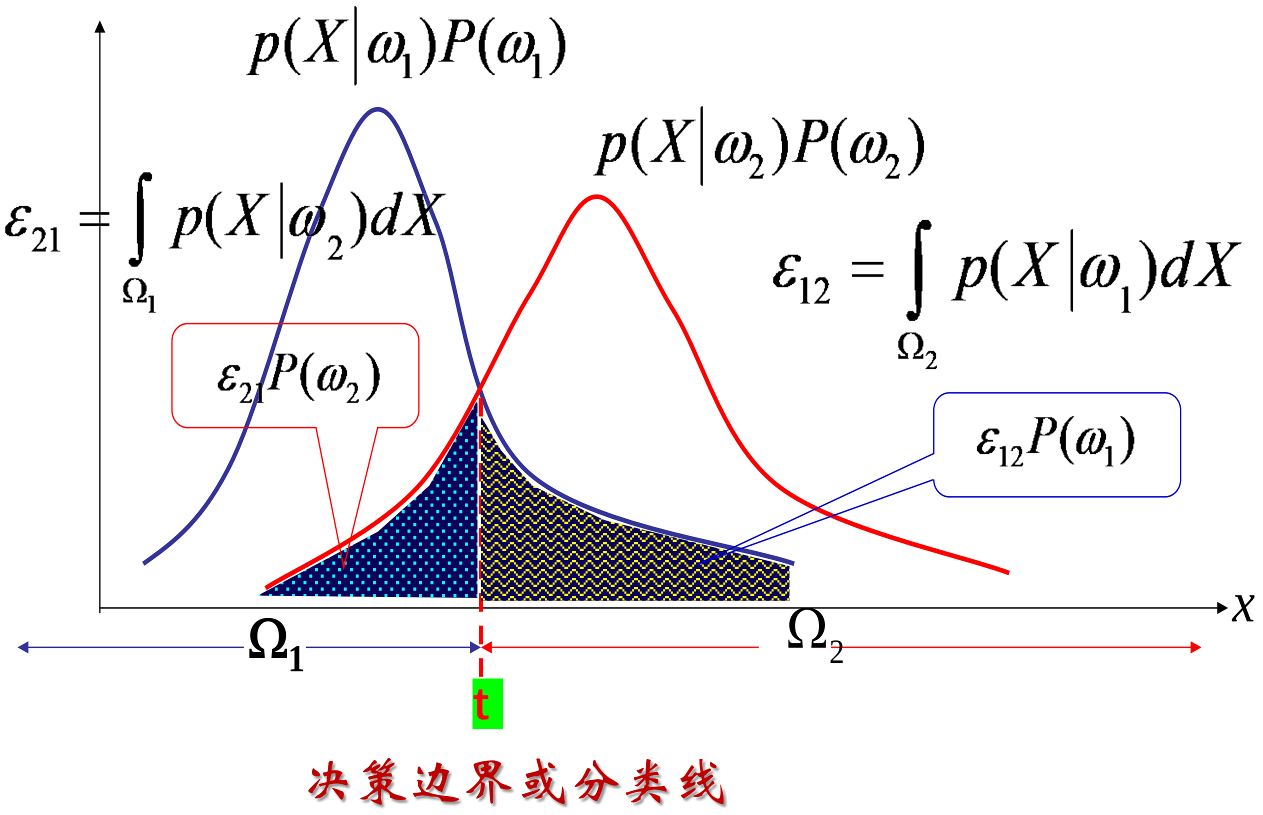


图2-1基于最小错误率的贝叶斯决策

计算错误率

写成

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

$$P(e) = P(x \in \Omega_1, \omega_2) + P(x \in \Omega_2, \omega_1) = P(x \in \Omega_1|\omega_2)P(\omega_2) + P(x \in \Omega_2|\omega_1)P(\omega_1) = P(\omega_2)\int_{\Omega_1} p(x|\omega_2)\,dx + P(\omega_1)\int_{\Omega_2} p(x|\omega_1)\,dx = \boldsymbol{P(\omega_2)\mathcal{E}_{21} + P(\omega_1)\,\mathcal{E}_{12}}$$

其中，$\mathcal{E}_{21} = \int_{\Omega_1} p(x|\omega_2)\,dx$，把第二类样本决策为第一类的错误率；

$\mathcal{E}_{12} = \int_{\Omega_2} p(x|\omega_1)\,dx$，把第一类样本决策为第二类的错误率。

两种错误率用相应类别的先验概率加权，就是总的平均错误率。

例2.1 垃圾邮件过滤

一封电子邮件不是垃圾邮件就是正常邮件，我们用ω_1、ω_2分别代表垃圾和正常邮件，各自取值的概率分别为$P(\omega_1)$，$P(\omega_2)$。

设$\{w_1, \cdots w_n\}$代表一些特殊的词(或词的组合)形成的集合，它们出现后就表示邮件是垃圾的。对每个**i(i=1,…,n)**, X_i是伯努利随机变量，用来定义词汇w_i是否出现在信息中，即当w_i出现时，$X_i = 1$,否则$X_i = 0$。

假设条件概率$p(x_i|\omega_1)$ 和$p(x_i|\omega_2)$已知，$x_i = 0,1$。简单起见，假设在给定类别ω_j的条件下随机变量$X_1, \cdots, X_n$是相互独立的。

现在想根据响应向量$(x_1, \cdots, x_n)$，判断一封邮件是垃圾还是正常邮件。

例2.1（续）

用贝叶斯公式，计算两类邮件的后验概率，

$$P(\omega_m|X_1=x_1,\cdots X_n=x_n)=\frac{P(\omega_m)\prod_{i=1}^{n}p\,(x_i|\omega_m)}{\sum_{j=1}^{2}p\,(\omega_j)\prod_{i=1}^{n}p\,(x_i|\omega_j)}\;，\quad m=1,2$$

若下式成立，则判定该邮件为垃圾邮件：

$$P(\omega_1|X_1=x_1,\cdots X_n=x_n)>P(\omega_2|X_1=x_1,\cdots X_n=x_n)$$

或等价地

$$P(\omega_1)\prod_{i=1}^{n}p\,(x_i|\omega_1)>P(\omega_2)\prod_{i=1}^{n}p\,(x_i|\omega_2)$$

最小错误率贝叶斯决策：多类情况

决策规则：

如果 $P(\omega_i|x) = \max\limits_{j=1,\dots,c} P(\omega_j|x)$, 则 $\boldsymbol{x} \in \boldsymbol{\omega_i}$

等价形式

若 $\boldsymbol{p(x|\omega_i)P(\omega_i) = \max\limits_{j=1,\dots,c} P(x|\omega_j)p(\omega_j)}$，则$\boldsymbol{x \in \omega_i}$

错误率：

- **直接计算的计算量大。因为特征空间被分割成c个区域，可能错误的情况c(c-1)种。**
- **通过计算平均正确率*P*(c)来计算错误率。**

$$\boldsymbol{P(c) = \sum_{j=1}^{c} P(x \in \Omega_j|\omega_j)P(\omega_j) = \sum_{j=1}^{c} \int_{\Omega_j} p(x|\omega_j)P(\omega_j)dx}$$

$$\boldsymbol{P(e) = 1 - P(c) = 1 - \sum_{j=1}^{c} P(\omega_j) \int_{\Omega_j} p(x|\omega_j)dx}$$

2.3 最小风险贝叶斯决策

- 很多场合，不同的错误决策所造成的损失不同
  - 癌症误判、鱼分拣
- 引入更一般的损失函数来替代错误率
  - 计算各种不同误判的风险；采取总风险最小的决策。**最小风险贝叶斯决策**
- 允许更多行动(action)而不仅仅是类别判定
  - 当决策风险大于不决策时，可以拒绝做决策。

判决的风险

- 决策代价(loss)
  - 对某样本x，真实类别状态为ω_j而决策为α_i 的损失函数 $\lambda(\alpha_i|\omega_j)$
- 条件风险(Condition risk)，某样本x属于各个状态的后验概率是$P(\omega_j|\mathbf{x})$，j=1,⋯,c，对它采取决策α_i的期望损失，

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

- 总的风险， $\boldsymbol{R} = \int R(\alpha(\mathbf{x})|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ ⟵ 最小化
  - $p(\mathbf{x})$已知且与决策无关，因此，要使积分和最小，每个x都使$\boldsymbol{R}(\boldsymbol{\alpha}(\mathbf{x})|\mathbf{x})$最小。
- 最小风险贝叶斯决策就是

若 $\boldsymbol{R}(\boldsymbol{\alpha}_i|\mathbf{x}) = \min_{j=1\ldots c} \boldsymbol{R}(\boldsymbol{\alpha}_j|\mathbf{x})$，则采取决策$\boldsymbol{\alpha}_i$

最小风险决策：两类问题

- **计算条件风险：**

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

其中 $\lambda_{ij}=\lambda(\alpha_i|\omega_j)$

- **最小风险贝叶斯决策：**

若 $R(\alpha_1|\mathrm{x}) < R(\alpha_2|\mathrm{x})$, 则判为$\omega_1$。

等价地，

若 $(\lambda_{21}-\lambda_{11})P(\omega_1|\mathrm{x}) > (\lambda_{12}-\lambda_{22})P(\omega_2|\mathrm{x})$, 则判为$\omega_1$。

或写成“似然比”，

若 $\frac{p(\mathrm{x}|\omega_1)}{p(\mathrm{x}|\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}}\frac{P(\omega_2)}{P(\omega_1)}$，则判为$\omega_1$.

最小风险贝叶斯决策与最小错误率贝叶斯决策

- 当$\lambda_{11} = \lambda_{22} = 0, \lambda_{12} = \lambda_{21} = 1$时，最小风险贝叶斯决策就转化成最小错误率贝叶斯决策。
- 可以把最小错误率贝叶斯决策看作最小风险贝叶斯决策的特例，是0-1损失时的最小风险贝叶斯决策。

例2.2　比较两种贝叶斯决策

假设在某个地区细胞识别中正常(ω_1)和异常(ω_2)两类的先验概率分别为，

正常状态　$P(\omega_1)=0.9$

异常状态　$P(\omega_2)=0.1$

已知正确决策没有损失，错误决策损失分别为，

对真实状态为正常(ω_1)的细胞，决策为α_2的损失为1，即λ_{21} =1

对真实状态为异常(ω_2) 的细胞，决策为α_1的损失为6，即$\lambda_{12}=6$

现有一待识别的细胞，其观测值为x，从类条件概率曲线上分别查得，

$p(x|\omega_1)$=0.2 ，　$p(x|\omega_2)$=0.4

试分别用<u>最小错误率贝叶斯决策</u>和<u>最小风险贝叶斯决策</u>对该细胞 x 进行分类。

例2.2 （续）

最小错误率贝叶斯决策：

利用贝叶斯公式计算后验概率，

$$P(\omega_1|x) = \frac{0.2\times0.9}{0.2\times0.9+0.4\times0.1} = 0.818$$

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

根据最小错误率贝叶斯决策规则，因

$$P(\omega_1|x) > P(\omega_2|x)$$

把 x 归为正常状态，属于ω_1 类。

最小风险贝叶斯决策：

利用贝叶斯公式计算后验概率，

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

计算条件风险，

$$R(\alpha_1|x) = \sum_{j=1}^{2} \lambda_{1j} P(\omega_j|x) = \lambda_{12} P(\omega_2|x) = 1.092$$

$$R(\alpha_2|x) = \lambda_{21} P(\omega_1|x) = 0.818$$

由于 $R(\alpha_1|x) > R(\alpha_2|x)$，根据最小风险贝叶斯决策，我们把 x 归为异常，属于 ω_2类。

带拒识的决策

- 根据最小风险贝叶斯决策规则，若分类的风险大于不作决策的风险，可考虑拒绝分类。
- 定义**拒绝域**，$\mathcal{R}=\{\mathbf{x}|\min_i R(\omega_i|\mathbf{x})>t\}$，其中t为阈值。
- 决策规则，

若 $R(\omega_i|\mathbf{x})=\min_j R(\omega_j|\mathbf{x})\le t$，则接受 x 并将其归入ω_i；

否则，拒绝对 x 做出决策。

2. 4 分类器、判别函数和决策面

- **分类器**的表示方式有多种，最常用的是一组判别函数。
- 设类别ω_i对应**判别函数** $g_i(\mathbf{x}), i = 1,2,\cdots,c$ 。给定一特征向量$\mathbf{x}$，在最大值判决下，判决规则为，

$$\text{若 } \mathbf{g}_i(\mathbf{x}) > \mathbf{g}_j(\mathbf{x}),\ \forall j \neq i,\ j = 1,2,\cdots,c，\quad \text{则} \quad \mathbf{x} \in \boldsymbol{\omega}_i \text{ 。}$$

- 判决规则将特征空间分成c个判决区域，$\mathcal{R}_1, \cdots, \mathcal{R}_c$。若对$\forall j \neq i$ 有 $g_i(\mathbf{x}) > g_j(\mathbf{x})$，那么x属于$\mathcal{R}_i$。此区域由判决边界来分割。
- **判别边界**即判决空间中使判别函数值最大的曲面(**决策面**)。

分类器可视为一个网络或机器，对每个输入 $\mathbf{x} = (x_1, x_2, \cdots x_d)^T$，通过计算c个判别函数，选择与最大判别值相对应的类别。如图2-2

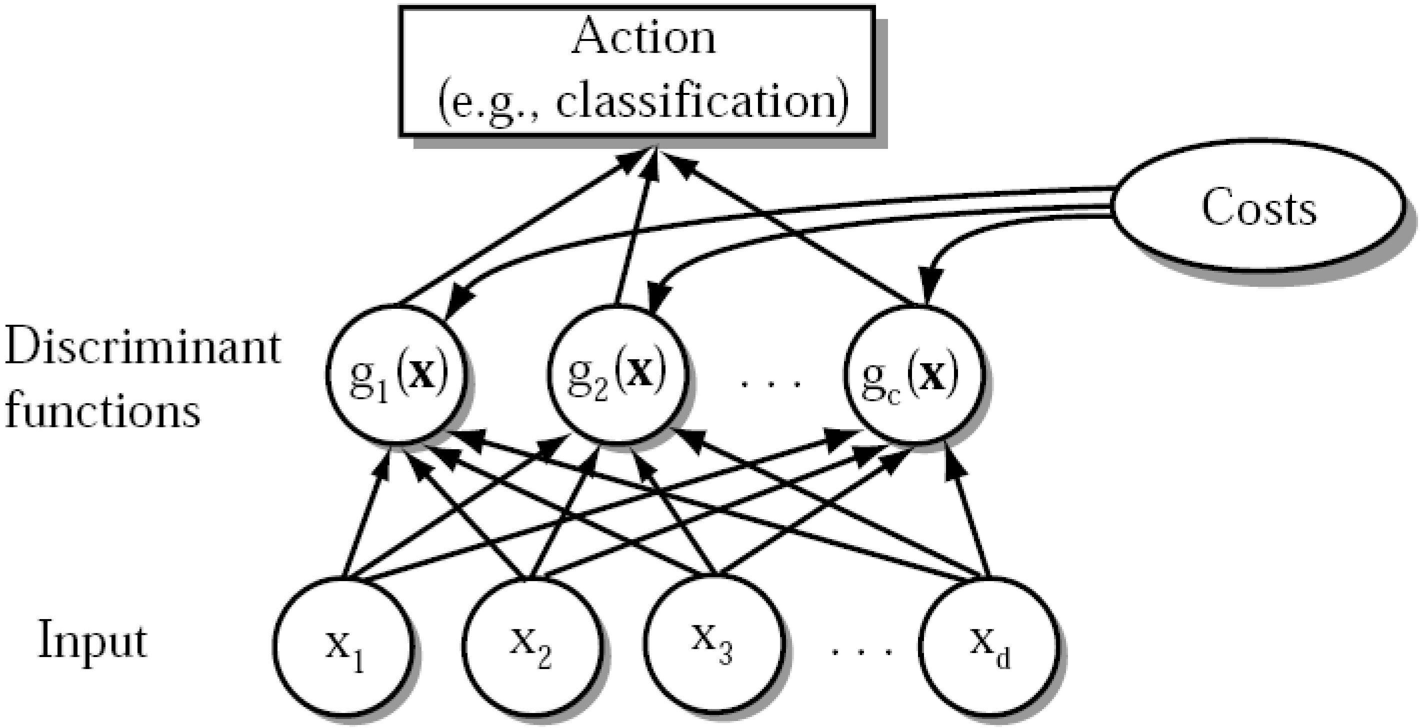


图2-2 统计模式分类器的一般结构

贝叶斯分类器的判别函数

- 对于**最小错误率**的情况，取

$$\boxed{g_i(\mathbf{x}) = P(\omega_i|\mathbf{x})}$$

这样，最大判别函数即为最大后验概率。

- 一般**有风险**的情况下，取

$$\boxed{g_i(\mathbf{x}) = -R(\alpha_i|\mathbf{x})}$$

因为，最大判别函数等价于最小条件风险。

判别函数的性质

- 判别函数的单调增函数不改变分类结果。即

  若f(*)是一个单调增函数，则$f\left(g_i(\mathbf{x})\right)$与$g_i(\mathbf{x})$分类结果相同。

  可利用该性质构造更简单的判别函数。如最小错误率分类判别函数，

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

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

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

2.5 正态分布下的贝叶斯决策

- 正态 (高斯) 分布及其性质回顾
  - 单变量正态密度函数

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

式中μ、σ^2、σ分别是随机变量x的期望、方差和标准差。

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

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

单变量高斯概率密度函数由参数μ和σ^2完全确定，常简记$p(x)$为$N(\mu,\sigma^2)$

– 多元正态密度函数

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

式中 $\mathbf{x} = (x_1, x_2, \cdots, x_d)^T$ 是d维随机向量，　$\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_d)^T$ 是$\mathbf{x}$的均值向量，

$\boldsymbol{\Sigma}$是$\mathbf{x}$的$d \times d$维协方差矩阵，$\boldsymbol{\Sigma}^{-1}$和$|\boldsymbol{\Sigma}|$分别是$\boldsymbol{\Sigma}$的逆矩阵和行列式。

$$\mu_i = E\{x_i\} = \int_{-\infty}^{\infty} x_i p(x_i) dx_i$$

其中 $p(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{x}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_d$

$$\Sigma = E\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T\} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1d}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2d}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1}^2 & \sigma_{d2}^2 & \cdots & \sigma_{dd}^2 \end{bmatrix}$$

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

其中 $p(x_i, x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{x}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_{j-1} dx_{j+1} \cdots dx_d$

正态分布下的最小错误率判别函数

- 取最小错误率贝叶斯判别函数如下，

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

- 对多元正态概型 $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$，$i = 1, \cdots, c$，**判别函数通式**如下，

$$\boxed{g_i(\mathbf{x}) = -\underbrace{\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x}-\boldsymbol{\mu}_i)}_{\text{第1项}} - \underbrace{\frac{d}{2}\ln 2\pi}_{\text{第2项}} - \underbrace{\frac{1}{2}\ln|\boldsymbol{\Sigma}_i|}_{\text{第3项}} + \underbrace{\ln P(\omega_i)}_{\text{第4项}}}$$

情况1：$\Sigma_i = \sigma^2 I$

- 协方差阵是对角阵
  - 最简单情况。发生在各特征独立且具有相同方差 σ^2 时。样本呈球状分布。
  - 有$|\Sigma_i| = \sigma^{2d}$, $\Sigma_i^{-1} = (1/\sigma^2)I$
- <u>判别函数</u>：

与i无关，故可省略

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

其中‖·‖是欧氏范数，$\|\mathbf{x}-\boldsymbol{\mu}_i\|^2 = (\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)$

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

- 将二次型$(\mathbf{x}-\boldsymbol{\mu}_i)^T(\mathbf{x}-\boldsymbol{\mu}_i)$展开，得

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

由于所有类别的判别函数中均包含$\mathbf{x}^T\mathbf{x}$项，可省略。**最终的判别函数**可写成，

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

判别函数 $\mathrm{g}_i(x)$是x的**线性函数**，

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

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

称ω_{i0}为第i个方向上的偏置

- 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是由线性方程

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

所确定的一个**超平面**（如果决策域 $\mathcal{R}_i$ 和 $\mathcal{R}_j$ 相邻)。

- 对于$\Sigma_i = \sigma^2 I$的正态密度情况，**决策面方程**可改写为

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

一个过点x_0且与权向量w正交的超平面

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

将$\mathcal{R}_i$与$\mathcal{R}_j$分开的超平面与两均值向量的连线垂直

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

若$P(\omega_i) = P(\omega_j)$,平分；若不等，则判定边界远离先验概率大的类别。

对高斯分布，**情况1：$\Sigma_i = \sigma^2 I$**

若所有 c类的先验概率$P(\omega_i)$相等，则判别函数简化为，

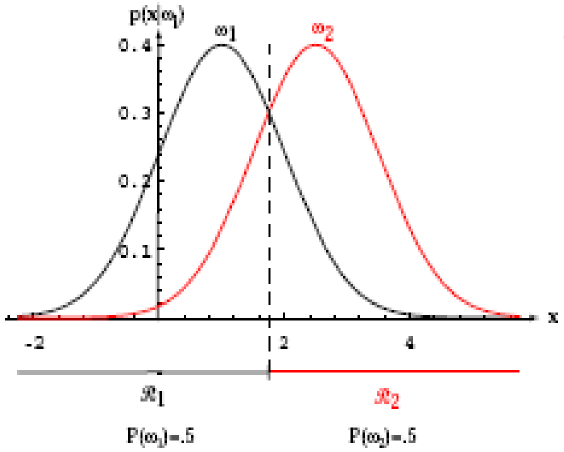
$$g_i(\mathbf{x}) = -\|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

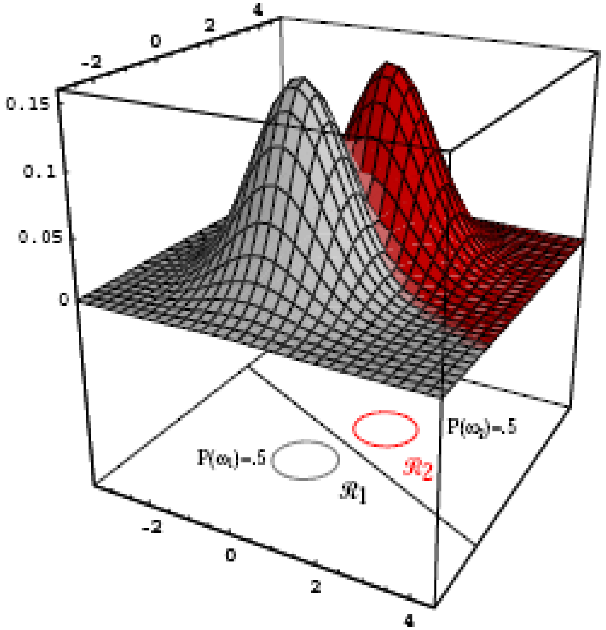
判决规则为：

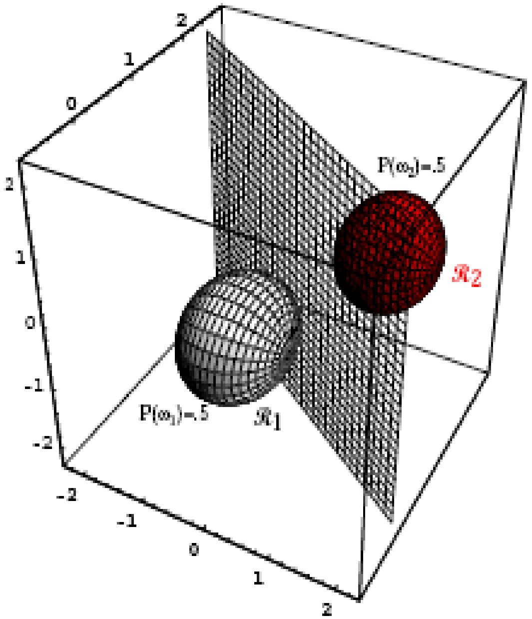
为将某特征向量x归类，通过测量每一x到c个均值向量中的每一个的欧氏距离，将x归为离它最近的那一类中。

最近邻算法

一维 二维 三维



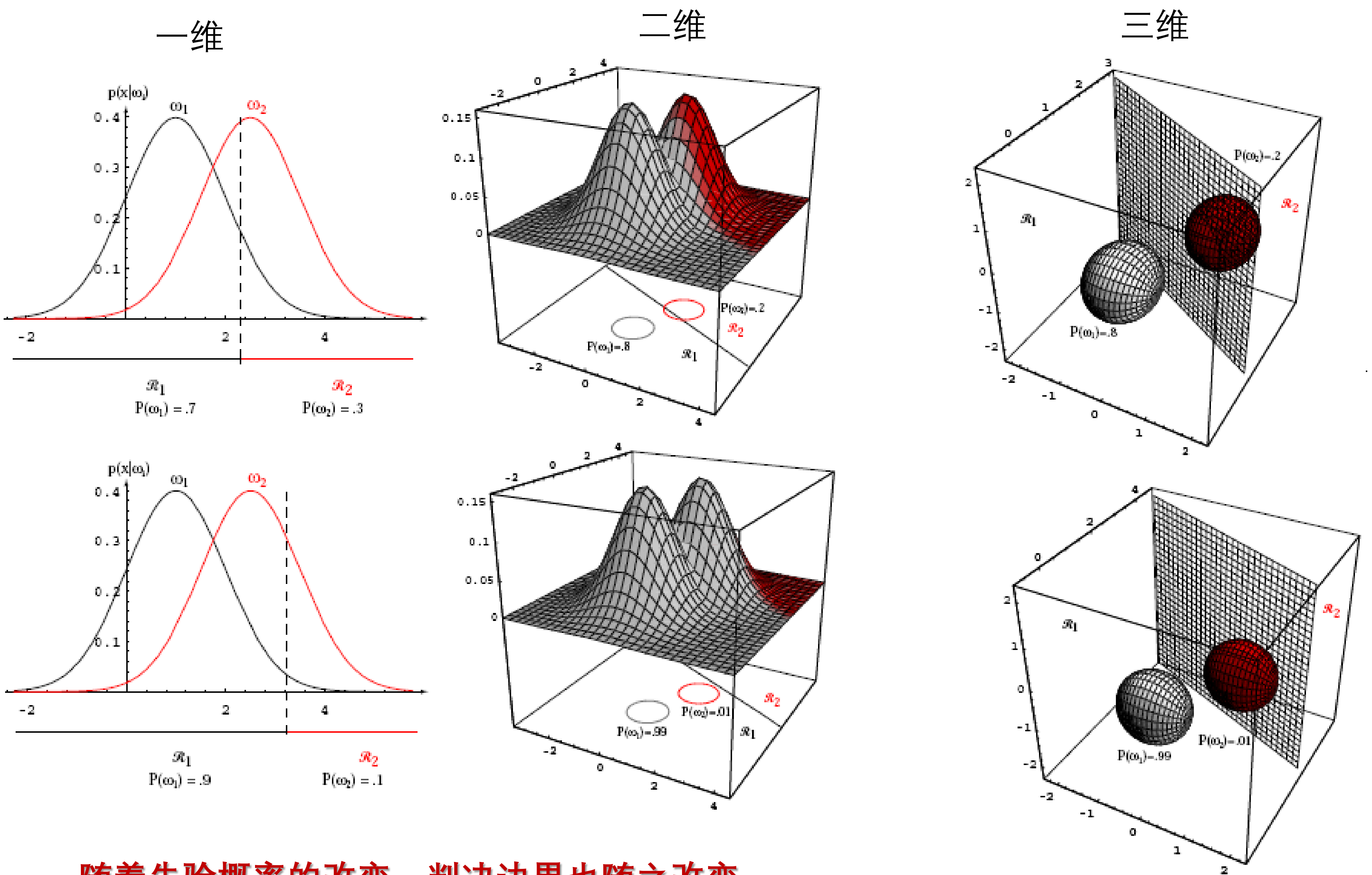




如果两种分布的协方差矩阵相等，且与单位阵成比例，那么它们呈d维球状分布，其判别边界是一个d-1维归一化超平面，垂直于两个中心的连线。

在这些一维、二维、三维的例子中，假设$P(\omega_1)=P(\omega_2)$

边界垂直平分两类中心(均值向量)



随着先验概率的改变，判决边界也随之改变。

对于差别较大的离散先验概率，判决边界不会落于球状高斯分布的中心点之间。

情况2：$\Sigma_i = \Sigma \neq$ 对角阵的情况

- **判别函数**：省略正态概型判别函数通式中的第1、2项，得

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

- 二次型展开后，去掉与i无关项$\mathbf{x}^T\Sigma^{-1}\mathbf{x}$,得如下**线性判别函数**,

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

其中，

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

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

- **决策面方程**：**超平面**

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

其中，$\mathbf{w}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$

将$\mathcal{R}_i$与$\mathcal{R}_j$分开的超平面与两均值向量连线未必垂直

$$\mathbf{x_0}=\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j)-\frac{\ln[P(\omega_i)/P(\omega_j)]}{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\Sigma^{-1}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)$$

若$P(\omega_i)=P(\omega_j)$,则决策面过两均值向量连续的中点，即平分；若不等，则判定边界向先验概率小的均值点偏移。

对高斯分布，**情况2：$\Sigma_i = \Sigma$**

若c类先验概率都相等，则判别函数简化为，

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

此时**决策规则**：

为对某样本x进行分类，通过计算x到每类的均值向量的**马氏距离** γ^2，将x归于γ^2最小的类别。

最近邻算法

情况3：Σ_i =任意

- 一般的多元正态分布情况，判别函数通式中只有 $\frac{d}{2}ln2\pi$与i无关，可省略。简化后，

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

- 判别函数是二次函数：

$$g_i(\mathbf{x}) = \mathbf{x}^T\mathbf{W}_i\mathbf{x} + \mathbf{w}_i^T\mathbf{x} + w_{i0} \qquad (1)$$

其中，

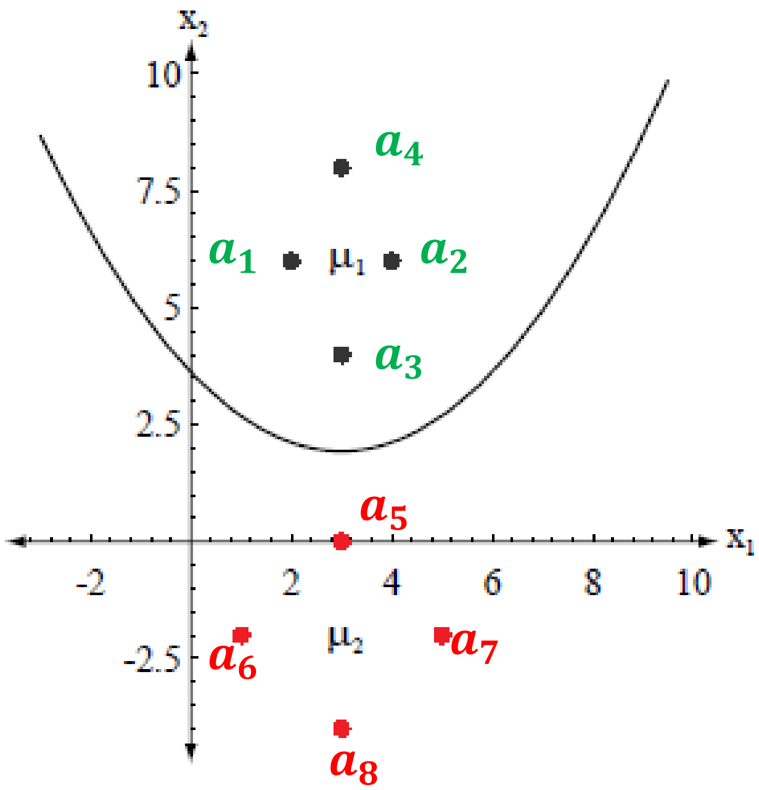
$$\mathbf{W}_i = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1} \qquad (d\times d\ \text{矩阵}) \qquad (2)$$

$$\mathbf{w}_i = \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i \qquad (d\text{维列向量}) \qquad (3)$$

$$w_{i0} = \frac{-1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i) \qquad (4)$$

- 判别边界：超二次曲面

例2.3 二维高斯分布数据的判决边界



已知

- ω_1表示黑点的集合，

$$\boldsymbol{a}_1=\begin{bmatrix}2\\6\end{bmatrix},\ \boldsymbol{a}_2=\begin{bmatrix}4\\6\end{bmatrix},\ \boldsymbol{a}_3=\begin{bmatrix}3\\4\end{bmatrix},\ \boldsymbol{a}_4=\begin{bmatrix}3\\8\end{bmatrix}$$

- ω_2是红点的集合,

$$\boldsymbol{a}_5=\begin{bmatrix}3\\0\end{bmatrix},\ \boldsymbol{a}_6=\begin{bmatrix}1\\-2\end{bmatrix},\ \boldsymbol{a}_7=\begin{bmatrix}5\\-2\end{bmatrix},\ \boldsymbol{a}_8=\begin{bmatrix}3\\-4\end{bmatrix}$$

(1) 假设两类的先验概率相等，求决策面方程。

(2) 判别点$(2,2)^T$所属类别。

$g_i(\mathbf{x})=\mathbf{x}^T\mathbf{W}_i\mathbf{x}+\mathbf{w}_i^T\mathbf{x}+w_{i0}$	(1)
$\mathbf{W}_i=-\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}$ $(d\times d$ 矩阵$)$	(2)
$\mathbf{w}_i=\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i$ $(d$维列向量$)$	(3)
$w_{i0}=\frac{-1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i-\frac{1}{2}ln\lvert\Sigma_i\rvert+lnP(\omega_i)$	(4)

同济大学 TONGJI UNIVERSITY

- **均值$\boldsymbol{\mu}_i$和协方差$\boldsymbol{\Sigma}_i$, i=1,2, 用下式计算,**

$$\boldsymbol{\mu}_1 = \frac{1}{4}\sum_{k=1}^{4} \boldsymbol{a}_k \quad , \quad \boldsymbol{\Sigma}_1 = \frac{1}{4}\sum_{k=1}^{4}(\boldsymbol{a}_k - \boldsymbol{\mu}_1)(\boldsymbol{a}_k - \boldsymbol{\mu}_1)^T$$

$$\boldsymbol{\mu}_2 = \frac{1}{4}\sum_{k=5}^{8} \boldsymbol{a}_k \quad , \quad \boldsymbol{\Sigma}_2 = \frac{1}{4}\sum_{k=58}(\boldsymbol{a}_k - \boldsymbol{\mu}_2)(\boldsymbol{a}_k - \boldsymbol{\mu}_2)^T$$

得,

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

$$\mu_2 = \begin{pmatrix} 3 \\ -2 \end{pmatrix} \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \quad \Sigma_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad |\Sigma_2| = 4$$

- **把$P(\omega_1) = P(\omega_2) = 0.5$代入一般形式的判别式(1)~(4),**

$$W_1 = \begin{pmatrix} -1 & 0 \\ 0 & -0.25 \end{pmatrix} \quad W_2 = \begin{pmatrix} -0.25 & 0 \\ 0 & -0.25 \end{pmatrix} \quad \begin{matrix} w_1 = (6 \quad 3) \\ w_2 = (1.5 \quad -1) \end{matrix} \quad \begin{matrix} w_{10} = -18.6931 \\ w_{20} = -4.6363 \end{matrix}$$

- 得

```
g1 = -1.0*x1**2 + 6.0*x1 - 0.25*x2**2 + 3.0*x2 - 18.6931471805599
g2 = -0.25*x1**2 + 1.5*x1 - 0.25*x2**2 - 1.0*x2 - 4.63629436111989
```

$$g_1(\mathbf{x}) = -x_1^2 + 6x_1 - 0.25x_2^2 + 3x_2 - 18.6931$$

$$g_2(\mathbf{x}) = -0.25x_1^2 + 1.6x_1 - 0.25x_2^2 - x_2 - 4.6363$$

- 令$g_1(\mathbf{x}) = g_2(\mathbf{x})$，得判决边界为，

$$\boxed{x_2 = 3.514 - 1.125x_1 + 0.1875x_1^2}$$

此方程描述了一个顶点位于(3 ， 1.83)^{T}的抛物线。

- 当 $\mathbf{x} = (2，2)^T$ 时，$\mathbf{g_1 < g_2}$， 故判定$\mathbf{x}$属于$\boldsymbol{\omega}_2$.

连续型朴素贝叶斯算法

输入： 训练数据集$D=\{(x_1,y_1),\dots,(x_N,y_N)\}$ 各分量为连续型数据

过程：

（1）计算先验概率$p(y=c_k)$的极大似然估计：

$$\hat{p}(y=c_k)=\frac{\sum_{i=1}^{N}I(y_i=c_k)+\lambda}{N+K\lambda},k=1,2,\dots,K$$

其中 I 是指示性函数，若$y_i=c_k$ ，则 $I(y_i=c_k)=1$ ，否则$I(y_i=c_k)=0$

λ是平滑项，$\lambda=1$时则是拉普拉斯平滑，这里$\lambda=1$。

（2）计算条件概率密度$p(X^{(j)}=a_{jl}|y=c_k)$ ：

$$p(X^{(j)}=a_{jl}\,|y=c_k)=\frac{1}{\sqrt{2\pi}\sigma_{jk}}e^{-\frac{(a_{il}-u_{jk})^2}{2\sigma_{jk}^2}}$$

这里有两个参数u_{jk}、σ_{jk}，它们可以用极大似然估计法定出：

$$\hat{u}_{jk}=\frac{1}{N_k}\sum_{i=1}^{N}x_i^{(j)}I(y_i=c_k)$$

$$\hat{\sigma}_{jk}^2=\frac{1}{N_k}\sum_{i=1}^{N}(x_i^{(j)}-u_{jk})^2I(y_i=c_k)$$

其中，$N_k=\sum_{i=1}^{N}I(y_i=c_k)$是类别$c_k$的样本数。

（3）构造决策函数 f(x) （根据贝叶斯决策，数据 $x^*=(x^{*(1)},\dots,x^{*(n)})^T$ 的类别应判定为后验概率最大的类别）：

$$f(x^*)=arg\max_{c_k}\hat{p}(y=c_k)\prod_{j=1}^{n}p(X^{(j)}=x^{*(j)}|y=c_k)$$

输出： 新数据 $x^*=(x^{*(1)},\dots,x^{*(n)})^T$ 所属类别，$y^*=f(x^*)$

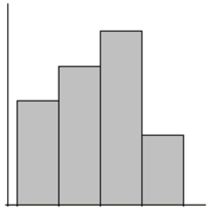
2.6 离散变量的贝叶斯决策

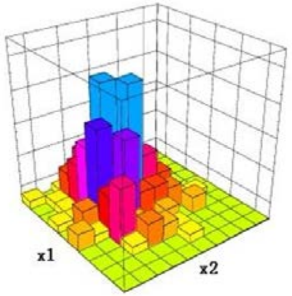
- 贝叶斯决策
  - 最小风险决策准则： $\alpha^* = \arg min_i\, R(\alpha_i|\mathbf{x})$

    即，选择行为α_i使条件风险$\boldsymbol{R}(\boldsymbol{\alpha}_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$最小。
  - 最小错误率决策准则：max($\boldsymbol{P}(\boldsymbol{\omega}_i|\mathbf{x})$)

    即，通过最大后验概率(MAP)来最小化错误率。
- 离散特征变量：
  - x是m个离散值$\mathbf{v}_1, \ldots, \mathbf{v}_m$中的一个
  - 如，医疗诊断(是否有某症状)，问卷调查(每个问题有2或多个选项)
  - 概率密度函数$\boldsymbol{p}(\mathbf{x}|\boldsymbol{\omega}_i) = \boldsymbol{p}(x_1 x_2 \ldots x_d|\boldsymbol{\omega}_i)$

    (非参数、直方图表示)





独立的二值特征

考虑**两类问题**，其特征向量中的元素是二值的，且条件独立。

设 x=$(x_1,\cdots,x_d)^T$, $x_i = 1$ 或 0, 且

$$p_i = P(x_i = 1|\omega_1), \qquad \Rightarrow P(x_i|\omega_1) = p_i^{x_i}(1-p_i)^{1-x_i}$$

$$q_i = P(x_i = 1|\omega_2) \qquad \Rightarrow P(x_i|\omega_2) = q_i^{x_i}(1-q_i)^{1-x_i}$$

因条件独立，则类条件概率密度为

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

- **似然比：** $\frac{P(\mathbf{x}|\omega_1)}{P(\mathbf{x}|\omega_2)} = \prod_{i=1}^{d}\left(\frac{p_i}{q_i}\right)^{x_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-x_i}$
- **判别函数：** $\mathrm{g}(\mathbf{x}) = \log\frac{P(\mathrm{X}|\omega_1)P(\omega_1)}{P(\mathrm{X}|\omega_2)P(\omega_2)} = \sum_{i=1}^{d}\left[x_i \ln\frac{p_i}{q_i} + (1-x_i)\ln\frac{1-p_i}{1-q_i}\right] + \ln\frac{P(\omega_1)}{P(\omega_2)}$

  该判别函数关于x_i是线性的，写成，

$$\boxed{\mathrm{g}(\mathbf{x}) = \sum_{i=1}^{d} w_i x_i + w_0 \quad , i = 1,2,\cdots,d}$$

w_i表征每个特征的判别性

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

- **判决规则：** 如果 $\mathrm{g}(\mathbf{x}) > 0$，将$\mathbf{x}$归为$\omega_1$；否则，将$\mathbf{x}$归为$\omega_2$。
- **判决面：** $\mathrm{g}(\mathbf{x}) = 0$

  所定义的是一个将ω_1的顶点同ω_2的顶点分割开的超平面。

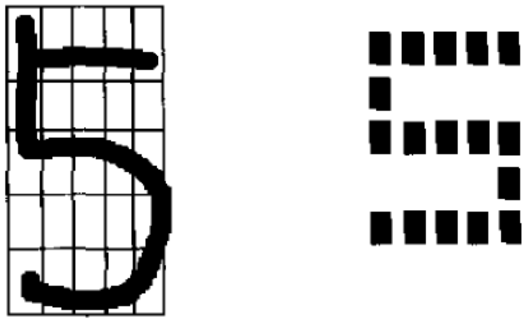
例2.3 手写字符识别

基于二值数据的贝叶斯分类实现

1. 理论基础

所谓**二值数据**，即各样本的**每一特征只取1或0**。对于手写数字的分类问题，在对每个数字图形提取特征时，定义N×N模板；在本例中N=5，将每个样本的长度和宽度5等分，构成一个5×5的模板，相当于将模板罩在对应的样本上，如下图所示，对于每一份内的像素个数进行累加统计，除以该模板每一份的面积总数。设阈值T=0.05，模板所对应的元素黑像素占有率大于T，则特征值取1；否则取0。

25 features

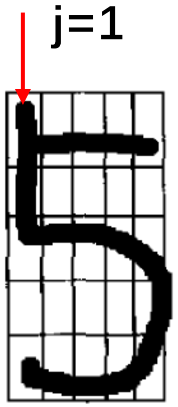


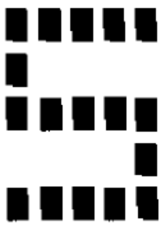
实现步骤：

①计算先验概率$P(\omega_i)$，先验概率可由各类的样本数和样本总数近似计算：

$$P(\omega_i) \approx {}^{N_i}/_{N} \qquad i = 0,1,2,\dots,9$$

其中$P(\omega_i)$为类别是数字i 的先验概率；**N_i为数字i的样本数**；**N为样本总数**。





② 先计算x各分量的类条件概率$P_j(\omega_i)$，再计算x的类条件概率$P(\mathbf{x}|\omega_i)$ ：

$P_j(\omega_i)$表示样本$\mathbf{x}(x_0, x_1, x_2, \cdots, x_{24})$属于$\omega_i$类，$\mathbf{x}$的第j个分量为1（$x_j = 1$）的概率估计值。

$$\boxed{P_j(\omega_i) = \Big(\sum_{\substack{k=1 \\ \mathbf{x}\in\omega_i}}^{N_i} x_j^{(k)} + 1\Big)/(N_i + 2)} \qquad j = 0,1,\cdots,24$$

离散概型估计

所有属于第i类的样本，第j个特征取1的总数

由此可计算，

$$P\big(x_j = 1\big|\mathbf{x} \in \omega_i\big) = P_j(\omega_i) \qquad i = 0,1,2,\cdots,9$$

$$P\big(x_j = 0\big|\mathbf{x} \in \omega_i\big) = 1 - P_j(\omega_i) \qquad j = 0,1,2,\cdots,24$$

样本x的类条件概率为，

$$P(\mathbf{x}|\omega_i) = P[\mathbf{x} = (x_0, x_1, x_2, \cdots, x_{24})|\mathbf{x} \in \omega_i]$$

$$P(\mathbf{x}|\omega_i) = P[\mathbf{x} = (x_0, x_1, x_2, \cdots, x_{24})|\mathbf{x} \in \omega_i] \quad = \prod_{j=0}^{24} P\left(x_j = \alpha \middle| \mathbf{x} \in \omega_i\right)$$

$i = 0,1,\cdots,9$　　其中，$\alpha = 0$或1。

③应用贝叶斯公式求后验概率

$$P(\omega_i|\mathbf{x}) = \frac{P(\omega_i)P(\mathbf{x}|\omega_i)}{P(\omega_0)P(\mathbf{x}|\omega_0) + P(\omega_1)P(\mathbf{x}|\omega_1) + \cdots + P(\omega_9)P(\mathbf{x}|\omega_9)}$$

$i = 0,1,\cdots,9$

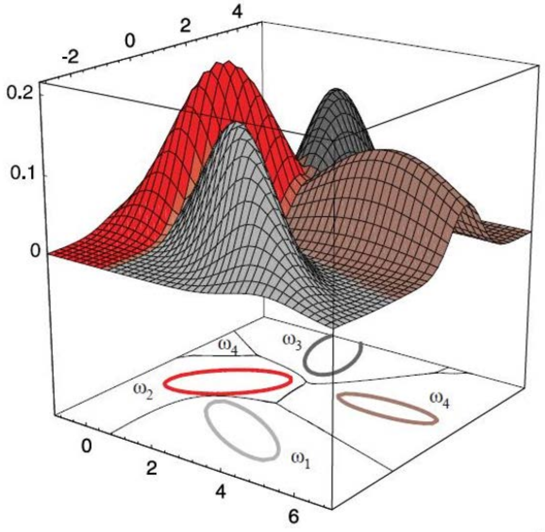
④后验概率的最大值的类别(0~9)就是手写数字的所属类别。

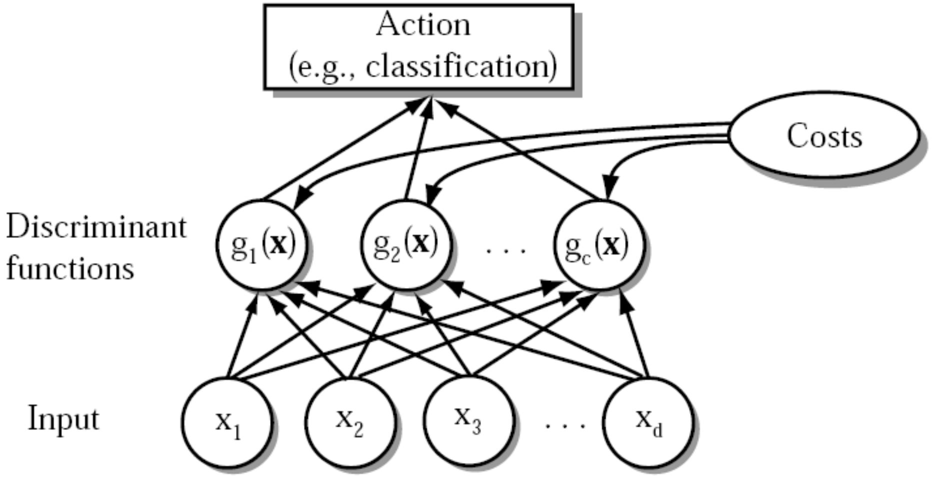
离散型朴素贝叶斯算法

输入： 训练数据集$D=\{(x_1,y_1),\ldots,(x_N,y_N)\}$ 各分量为离散型数据
过程： （1）计算先验概率$p(y=c_k)$的极大似然估计： $\hat{p}(y=c_k)=\dfrac{\sum_{i=1}^{N}I(y_i=c_k)+\lambda}{N+K\lambda},k=1,2,\ldots,K$ 其中 I 是指示性函数，若$y_i=c_k$，则 $I(y_i=c_k)=1$，否则$I(y_i=c_k)=0$ λ是平滑项，$\lambda=1$时则是拉普拉斯平滑，这里$\lambda=1$。 （2）计算条件概率$p(X^{(j)}=a_{jl}\|y=c_k)$的极大似然估计（设每个单独输入的 n 维向量x_i的第 j 维特征$x^{(j)}$可能的取值集合为$\{a_{j1},\cdots,a_{jM_j}\}$）： $\hat{p}(X^{(j)}=a_{jl}\|y=c_k)=\dfrac{\sum_{i=1}^{N}I\left(x_i^{(j)}=a_{jl},y_i=c_k\right)+\lambda}{\sum_{i=1}^{N}I(y_i=c_k)+M_j\lambda}$ 其中 $l=1,2,\ldots,M_j$；$k=1,2,\ldots,K$；$\lambda=1$ （3）构造决策函数 f(x) （根据贝叶斯决策，数据 $x^*=(x^{*(1)},\ldots,x^{*(n)})^T$ 的类别应判定为后验概率最大的类别）： $f(x^*)=arg\max_{c_k}\hat{p}(y=c_k)\prod_{j=1}^{n}\hat{p}(X^{(j)}=x^{*(j)}\|y=c_k)$
输出： 新数据 $x^*=(x^{*(1)},\ldots,x^{*(n)})^T$ 所属类别， $y^*=f(x^*)$
输出： 新数据 $x^*=(x^{*(1)},\ldots,x^{*(n)})^T$ 所属类别，$y^*=f(x^*)$

小结：统计模式分类的基本框架

- 特征空间划分
  - 判别函数(Discriminant function)、决策面(Decision surface)
  - 生成模型(Generative model)： $\mathbf{x} \rightarrow p(\mathbf{x}|\omega_i) \rightarrow g_i(\mathbf{x})$,
  - 判别模型(Discriminative model): $\mathbf{x} \rightarrow g_i(\mathbf{x})$





生成模型

- 从一定的概率模型出发，把模式识别问题转化成概率模型估计问题
- 分类器设计实是对概率模型的估计
- 又称为基于(概率)模型的模式识别方法

判别模型

- 从要解决的问题和训练样本出发，直接求出判别函数
- 有些方法可事先确定判别函数的形式，通过训练样本确定其中的参数。

  如，SVM，神经网络
- 有的甚至不用事先设计分类器，只确定分类原则，根据已知样本直接对未知样本进行分类。如 近邻法
- 也称为基于数据的模式识别方法

模式识别相关杂志和主流国际会议

- 模式识别相关杂志
  - IEEE Trans. On PAMI TPAMI
  - Journal of Machine Learning Research JMLR
  - Artificial Intelligence AI
  - International Journal of Computer Vision IJCV
  - Pattern Recognition
- 主流国际会议
  - IJCAI (International Joint Conference on Artificial Intelligence)
  - CVPR (IEEE Conference on Computer Vision & Pattern Recognition)
  - ICML (International Conference on Machine Learning)
  - AAAI (AAAI Conference on Artificial Intelligence)
  - ICCV (International Conference on Computer Vision)
  - COLT / NIPS /KDD /ICLR

作业1

设计离散型或连续型朴素贝叶斯分类器。

二选一