

## 第二章 统计判别

## 2.1 作为统计判别问题的 模式分类

- 模式识别的目的就是要确定某一个给定的模式样本属于哪一类。
- 可以通过对被识别对象的多次观察和测量，构成特征向量，并将其作为某一个判决规则的输入，按此规则来对样本进行分类。

## 2.1 作为统计判别问题的模式分类

- 在获取模式的观测值时，有些事物具有确定的因果关系，即在一定的条件下，它必然会发生或必然不发生。
  - 例如识别一块模板是不是直角三角形，只要凭“三条直线边闭合连线和一个直角”这个特征，测量它是否有三条直线边的闭合连线并有一个直角，就完全可以确定它是不是直角三角形。
  - 这种现象是确定性的现象

## 2.1 作为统计判别问题的模式分类

- 但在现实世界中，由许多客观现象的发生，就每一次观察和测量来说，即使在基本条件保持不变的情况下也具有不确定性。
- 只有在大量重复的观察下，其结果才能呈现出某种规律性，即对它们观察到的特征具有统计特性。
- 特征值不再是一个确定的向量，而是一个随机向量。
- 此时，只能利用模式集的统计特性来分类，以使分类器发生错误的概率最小。

# 2.1 作为统计判别问题的模式分类

## 2.1.1 贝叶斯判别原则

- 两类模式集的分类
  - 目的：要确定 $x$ 是属于 $\omega_1$ 类还是 $\omega_2$ 类，要看 $x$ 是来自于 $\omega_1$ 类的概率大还是来自 $\omega_2$ 类的概率大。
- 贝叶斯判别

# 2.1 作为统计判别问题的模式分类

## 2.1.1 贝叶斯判别原则

- 例子（地震预测）

- 对某一地震高发区进行统计，地震以 $\omega_1$ 类表示，正常以 $\omega_2$ 类表示
- 统计的时间区间内，每周发生地震的概率为20%，即 $P(\omega_1)=0.2$ ，当然 $P(\omega_2)=1-0.2=0.8$
- 在任意一周，要判断该地区是否会有地震发生。显然，因为 $P(\omega_2) > P(\omega_1)$ ，只能说是正常的可能性大。如要进行判断，只能其它观察现象来实现。

# 2.1 作为统计判别问题的模式分类

## 2.1.1 贝叶斯判别原则

- 例子

- 通常地震与生物异常反应之间有一定的联系。
- 若用生物是否有异常反应这一观察现象来对地震进行预测，生物是否异常这一结果以模式 $x$ 代表，这里 $x$ 为一维特征，且只有 $x = \text{“异常”}$ 和 $x = \text{“正常”}$ 两种结果。

# 2.1 作为统计判别问题的模式分类

## 2.1.1 贝叶斯判别原则

- 例子

- 假设根据观测记录，发现这种方法有以下统计结果：

- 地震前一周内出现生物异常反应的概率=0.6，即 $p(x=\text{异常} | \omega_1)=0.6$
    - 地震前一周内出现生物正常反应的概率=0.4，即 $p(x=\text{正常} | \omega_1)=0.4$
    - 一周内没有发生地震但也出现了生物异常的概率=0.1，即 $p(x=\text{异常} | \omega_2)=0.1$
    - 一周内没有发生地震时，生物正常的概率=0.9，即 $p(x=\text{正常} | \omega_2)=0.9$



# 2.1 作为统计判别问题的模式分类

## 2.1.1 贝叶斯判别原则

- 问题

- 若某日观察到明显的生物异常反应现象，一周内发生地震的概率为多少，即求 $P(\omega_1 | x=\text{异常})=?$
- 这里 $P(\omega_1)$ 是根据以往的统计资料得到的地震发生的先验概率。现在经过观察，需要求出 $P(\omega_1 | x=\text{异常})$ ，即观测到生物异常时一周内发生地震的概率，称为后验概率。

- [计算]

## 2.1 作为统计判别问题的模式分类

### 2.1.2 贝叶斯最小风险判别

- 当考虑到对于某一类的错误判决要比对另一类的判决更为关键时，就需要把最小错误概率的贝叶斯判别做一些修正，提出条件平均风险 $r_j(x)$ 。
- M类分类问题的条件平均风险 $r_j(x)$ 
  - 对M类问题，如果观察样本被判定属于 $\omega_j$ 类，则条件平均风险为：
$$r_{\hat{y}}(x) = \sum_{i=1}^M L_{ij} P(\omega_i | x)$$
  - $L_{ij}$ 称为将本应属于 $\omega_i$ 类的模式判别成属于 $\omega_j$ 类的是非代价。

# 2.1 作为统计判别问题的模式分类

## 2.1.2 贝叶斯最小风险判别

- 意义

- 对于自然属性是属于 $\omega_i$ 类的模式 $x$ 来说，它来自 $\omega_i$ 类的概率应为 $P(\omega_i | x)$ 。
- 如果分类器判别 $x$ 是属于 $\omega_j$ 类，但它实际上来自 $\omega_i$ 类，也就是说分类器失败，这时 $L_{ij}$ 为失分，对应的条件风险为后验概率进行 $L_{ij}$ 的加权运算。
- 由于模式 $x$ 的自然属性可能来自 $M$ 类中的任一类，因此可将观察样本指定为 $\omega_j$ 类的条件平均风险用 $r_j(x)$ 的公式运算。

# 2.1 作为统计判别问题的模式分类

## 2.1.2 贝叶斯最小风险判别

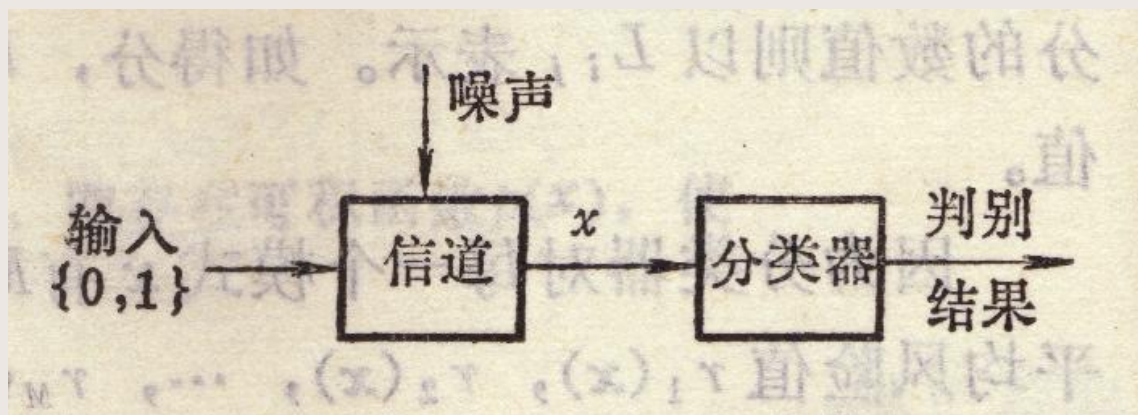
- $L_{ij}$ 的取值
  - 若 $i=j$ ，即判别正确，得分， $L_{ij}$ 可以取负值或零，表示不失分。
  - 若 $i \neq j$ ，即判别错误，失分， $L_{ij}$ 应取正值。
- 最小平均条件风险分类器
  - 分类器对每一个模式 $x$ 有 $M$ 种可能的类别可供选择。
  - 若对每一个 $x$ 计算出全部类别的平均风险值 $r_1(x)$ ,  $r_2(x), \dots, r_M(x)$ ，并且将 $x$ 指定为是具有最小风险值的那一类，则这种分类器称为最小平均条件风险分类器。
  - 表达式

## 2.1 作为统计判别问题的模式分类

### 2.1.2 贝叶斯最小风险判别

- 两类 ( $M=2$ ) 的情况

- [例子]



- 一般多类 ( $M$ 类) 的情况

## 2.2 正态分布模式的贝叶斯分类器

- 出发点
  - 当已知或者有理由设想类概率密度函数  $P(\mathbf{x}|\omega_i)$  是多变量的正态分布时，上一节介绍的贝叶斯分类器可以导出一些简单的判别函数。
  - 由于正态密度函数易于分析，且对许多重要的实际应用又是一种合适的模型，因此受到很大的重视。

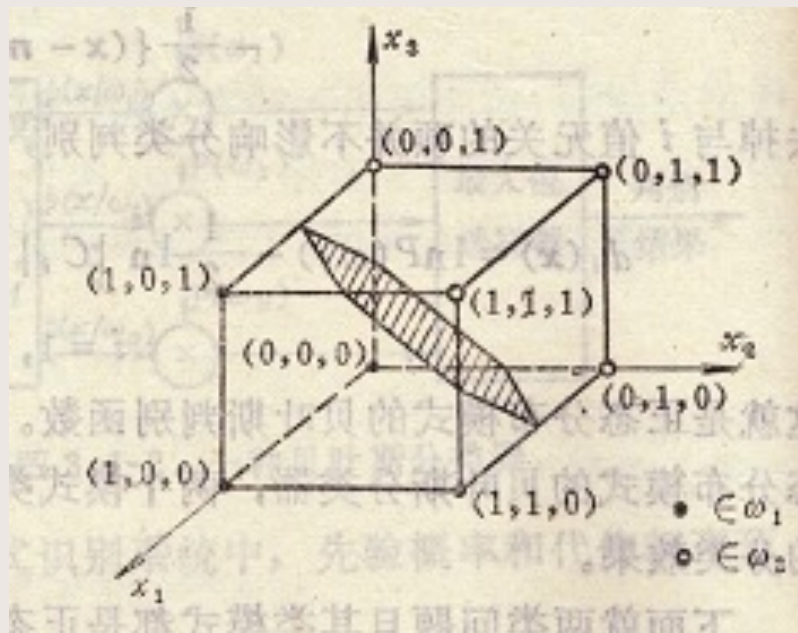
## 2.2 正态分布模式的贝叶斯分类器

- M种模式类别的多变量正态类密度函数
  - 判别函数是一个超二次曲面。
  - 对于正态分布模式的贝叶斯分类器，两个模式类别之间用一个二次判别界面分开，就可以求得最优的分类效果。
- 两类问题且其类模式都是正态分布的特殊情况
  - 当 $C_1 \neq C_2$ 时的情况
    - 显然，判别界面 $d_1(x) - d_2(x) = 0$ 是 $x$ 的二次型方程，即 $\omega_1$ 和 $\omega_2$ 两类模式可用二次判别界面分开。
    - 当 $x$ 是二维模式时，判别界面为二次曲线，如椭圆，圆，抛物线或双曲线等。
  - 当 $C_1 = C_2 = C$ 时的情况
    - 判别界面为 $x$ 的线性函数，为一超平面。
    - 当 $x$ 是二维时，判别界面为一直线。



## 2.2 正态分布模式的贝叶斯分类器

- [例子]



- 讨论

- 贝叶斯分类规则是基于统计概念的。
- 如果只有少数模式样本，一般较难获得最优的结果。



# 作业及编程（编程可选）

- 设以下模式类别具有正态概率密度函数：

$$\omega_1: \{(0\ 0)^T, (2\ 0)^T, (2\ 2)^T, (0\ 2)^T\}$$

$$\omega_2: \{(4\ 4)^T, (6\ 4)^T, (6\ 6)^T, (4\ 6)^T\}$$

（1）设 $P(\omega_1) = P(\omega_2) = 1/2$ ，求这两类模式之间的贝叶斯判别界面的方程式。

（2）绘出判别界面。

- 编写两类正态分布模式的贝叶斯分类程序。  
（可选例题或上述作业题为分类模式）

## 2.3 均值向量和协方差矩阵的参数估计

- 在贝叶斯分类器中，构造分类器需要知道类概率密度函数 $p(\mathbf{x}|\omega_i)$ 。
- 如果按先验知识已知其分布，则只需知道分布的参数即可。
  - 例如：类概率密度是正态分布，它完全由其均值向量和协方差矩阵所确定。
- 对均值向量和协方差矩阵的估计即为贝叶斯分类器中的一种参数估计问题。

## 2.3 均值向量和协方差矩阵的参数估计

- 参数估计的两种方式
  - 一种是将参数作为非随机变量来处理，例如矩估计就是一种非随机参数的估计。
  - 另一种是随机参数的估计，即把这些参数看成是随机变量，例如贝叶斯参数估计。

## 2.3 均值向量和协方差矩阵的参数估计

- 均值和协方差矩阵的非随机参数的估计
  - 均值和协方差矩阵的估计量定义
  - 均值和协方差矩阵估计量的迭代运算

## 2.3 均值向量和协方差矩阵的参数估计

- 均值向量和协方差矩阵的贝叶斯学习
  - 将概率密度函数的参数估计量看成是随机变量  $\theta$ ，它可以是纯量、向量或矩阵。
  - 按这些估计量统计特性的先验知识，可以先粗略地预选出它们的密度函数。
  - 通过训练模式样本集  $\{x_i\}$ ，利用贝叶斯公式设计一个迭代运算过程求出参数的后验概率密度  $p(\theta | x_i)$ 。
  - 当后验概率密度函数中的随机变量  $\theta$  的确定性提高时，可获得较准确的估计量。

## 2.3 均值向量和协方差矩阵的参数估计

- 均值向量和协方差矩阵的贝叶斯学习
  - 一般概念
  - 单变量正态密度函数的均值学习