

模式识别(Pattern Recognition): 确定一个样本的类别属性（模式类）的过程，即把某一样本归属于多个类型中的某个类型。

样本 (Sample): 一个具体的研究（客观）对象。如患者，某人写的一个汉字，一幅图片等。

模式(Pattern): 对客体（研究对象）特征的描述（定量的或结构的描述），是取自客观世界的某一样本的测量值的集合（或综合）。

特征(Features): 能描述模式特性的量（某一模式的测量值集合中的同一量）。在统计模式识别方法中，通常用一个矢量 \vec{x} 表示，称之为特征矢量，记为 $\vec{x}=(x_1, x_2, \dots, x_n)'$ 。（一个特征矢量描述一种模式）

模式类(Class): 具有某些共同特性的模式的集合。

模式识别的三大任务

模式采集: 从客观世界（对象空间）到模式空间的过程称为模式采集。

特征提取和特征选择: 由模式空间到特征空间的变换和选择。

类型判别: 特征空间到类型空间所作的操作。

模式识别系统的主要环节

特征提取: 符号表示，如长度、波形、。。。。

特征选择: 选择有代表性的特征，能够正确分类

学习和训练: 利用已知样本建立分类和识别规则

分类识别: 对所获得样本按建立的分类规则进行分类识别

一、统计模式识别

模式描述方法:

特征向量 $\vec{x}=(x_1, x_2, \dots, x_n)'$

模式判定:

模式类用条件概率分布 $P(X/w_i)$ 表示, m 类就有 m 个分布，然后判定未知模式属于哪一个分布。

主要方法: 线性、非线性分类、Bayes 决策、聚类分析

主要优点:

- 1) 比较成熟
- 2) 能考虑干扰噪声等影响
- 3) 识别模式基元能力强

主要缺点:

- 1) 对结构复杂的模式抽取特征困难
- 2) 不能反映模式的结构特征，难以描述模式的性质
- 3) 难以从整体角度考虑识别问题

二、句法模式识别

模式描述方法:

符号串，树，图

模式判定:

是一种语言，用一个文法表示一个类， m 类就有 m 个文法，然后判定未知模式遵循哪一个文法。

主要方法: 自动机技术、CYK 剖析算法、Early 算法、转移图法

主要优点:

- 1) 识别方便, 可以从简单的基元开始, 由简至繁。
- 2) 能反映模式的结构特征, 能描述模式的性质。
- 3) 对图象畸变的抗干扰能力较强。

主要缺点:

当存在干扰及噪声时, 抽取特征基元困难, 且易失误

三、模糊模式识别

模式描述方法:

模糊集合 $A = \{(ua, a), (ub, b), \dots (un, n)\}$

模式判定:

是一种集合运算。用隶属度将模糊集合划分为若干子集, m 类就有 m 个子集, 然后根据择近原则分类。

主要方法: 模糊统计法、二元对比排序法、推理法、模糊集运算规则、模糊矩阵

主要优点:

由于隶属度函数作为样本与模板间相似程度的度量, 故往往能反映整体的与主体的特征, 从而允许样本有相当程度的干扰与畸变。

主要缺点:

准确合理的隶属度函数往往难以建立, 故限制了它的应用。

四、人工神经网络法

模式描述方法:

以不同活跃度表示的输入节点集 (神经元)

模式判定:

是一个非线性动态系统。通过对样本的学习建立起记忆, 然后将未知模式判决为其最接近的记忆。

主要方法: BP 模型、HOP 模型、高阶网

主要优点:

可处理一些环境信息十分复杂, 背景知识不清楚, 推理规则不明确的问题。允许样本有较大的缺损、畸变。

主要缺点:

模型在不断丰富与完善中, 目前能识别的模式类还不够多。

五、逻辑推理法 (人工智能法)

模式描述方法:

字符串表示的事实

模式判定:

是一种布尔运算。从事实出发运用一系列规则, 推理得到不同结果, m 个类就有 m 个结果。

主要方法: 产生式推理、语义网推理、框架推理

主要优点:

已建立了关于知识表示及组织, 目标搜索及匹配的完整体系。对需要众多规则的推理达到识别目标确认的问题, 有很好的效果。

主要缺点:

当样本有缺损, 背景不清晰, 规则不明确甚至有歧义时, 效果不好。

特征矢量:

设一个研究对象的 N 个特征量测量值分别为 x_1, x_2, \dots, x_n , 将它们作为一个整体来考虑, 让他们构成一个 N 维特征向量 $\vec{x} = (x_1, x_2, \dots, x_n)'$

特征空间:

各种不同的特征矢量构成了 N 维特征空间。特征矢量是特征空间中的一点,

随机矢量:

对许多对象而言, 各个特征分量都是随机变量, 即许多对象的特征向量在 n 维空间中呈随机性分布, 称为随机矢量。

设 $\vec{X} = (X_1, X_2, \dots, X_n)'$ 为随机矢量, $\vec{x} = (x_1, x_2, \dots, x_n)'$ 为确定性矢量。

随机矢量的联合概率分布函数定义为:

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

$$F(\vec{x}) = P(\vec{X} \leq \vec{x})$$

随机矢量 \vec{X} 的联合概率密度函数定义为:

$$p(x_1, x_2, \dots, x_n) \triangleq p(\vec{x}) = \partial^n F(x_1, x_2, \dots, x_n) / \partial x_1 \partial x_2 \dots \partial x_n$$

当第 i 类为 w_i 类的情况下, 则他的概率分布和概率密度函数如下

$$F(\vec{x} | w_i) = P(\vec{X} \leq \vec{x} | w_i)$$

$$p(\vec{x} | w_i) = \partial^n F(x_1, x_2, \dots, x_n | w_i) / \partial x_1 \partial x_2 \dots \partial x_n$$

在模式识别中, 经常以类别 w_i 作为条件, 在这种情况下随机矢量 \vec{x} 的条件期望矢量定义为

$$\bar{\mu}_{w_i} = E[\vec{X} | w_i] = \int_{X^n} \vec{x} p(\vec{x} | w_i) d\vec{x}$$

随机矢量 \vec{x} 的自协方差矩阵表征各分量围绕其均值的散布情况及各分量间的相关关系, 其定义为:

$$\Sigma = E[(\vec{X} - \bar{\vec{X}})(\vec{X} - \bar{\vec{X}})'] = \int_{X^n} (\vec{x} - \bar{\mu})(\vec{x} - \bar{\mu})' p(\vec{x}) d\vec{x} = (\sigma_{ij}^2)_{n \times n}$$

式中 σ_{ij}^2 是第 i 个分量与第 j 个分量的协方差

$$\sigma_{ij}^2 = E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)'] = \int \int_{-\infty}^{+\infty} (x_i - \mu_i)(x_j - \mu_j) p(x_i, x_j) dx_i dx_j$$

正态分布随机矢量 $\vec{X} = (X_1, X_2, \dots, X_n)'$ 的概率密度函数定义为:

$$p(x_1, x_2, \dots, x_n) = p(\vec{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\vec{x} - \bar{\mu})' \Sigma^{-1} (\vec{x} - \bar{\mu})\right]$$

正态分布随机矢量性质:

分布函数完全由 μ 和 Σ 确定

等概率密度点的轨迹为一超球面

对正态分布, 不相关等价于独立

其边缘密度和条件密度仍然是正态分布

正态分布随机矢量的线性变化仍然是正态分布随机矢量

聚类分析的基本思想

相似的归为一类。

模式相似性的度量和聚类算法。

无监督分类 (Unsupervised)。

特征量的类型

物理量----(重量、长度、速度)

次序量----(等级、技能、学识)

名义量----(性别、状态、种类)

分类无效时的情况

1. 特征选取不当使分类无效
2. 特征选取不足可能使不同类别的模式判为一类
3. 特征选取过多可能无益反而有害,增加分析负担并使分析效果变差
4. 量纲选取不当

对分类结果会产生极大影响的问题

选择什么特征?

选择多少个特征?

选择什么样的量纲?

选择什么样的距离测度?

聚类过程的基本步骤

- 一、特征选择(feature selection)
尽可能多地包含任务关心的信息
- 二、近邻测度(proximity measure)
定量测定两特征如何“相似”或“不相似”
- 三、聚类准则 (clustering criterion)
以蕴涵在数据集中类的类型为基础
- 四、聚类算法 (clustering algorithm)
按近邻测度和聚类准则揭示数据集的聚类结构
- 五、结果验证 (validation of the results)
常用逼近检验验证聚类结果的正确性
- 六、结果判定 (interpretation of the results)
由专家用其他方法判定结果的正确性

聚类的应用 (四个基本方向)

- 一、减少数据
- 二、假说生成
- 三、假说检验
- 四、基于分组的预测

模式相似性测度

用于描述各模式之间特征的相似程度

- 一、**距离测度** (差值测度)

测度基础: 两个矢量矢端的距离

测度数值: 两矢量各相应分量之差的函数。

常用的距离测度有:

1. 欧氏(Euclidean)距离

$$d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\| = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2}$$

2. 绝对值距离(街坊距离或 Manhattan 距离)

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^n |x_i - y_i|$$

3. 切氏(Chebyshev)距离

$$d(\vec{x}, \vec{y}) = \max_i |x_i - y_i|$$

4. 明氏(Minkowski)距离

$$d(\vec{x}, \vec{y}) = \left[\sum_{i=1}^n |x_i - y_i|^m \right]^{1/m}$$

5. 马氏距离

$$d^2(\vec{x}, \vec{y}) = (\vec{x} - \vec{y})' V^{-1} (\vec{x} - \vec{y})$$

$$V = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y}) D^2 \hat{1}$$

马氏距离对一切非奇异线性变换都是不变的,这说明它不受特征量纲选择的影响,并且是平移不变的。 V 的含义是这个矢量集的协方差阵的统计量,故马氏距离加入了对特征的相关性的考虑。

二、相似测度

测度基础: 以两矢量的方向是否相近作为考虑的基础,矢量长度并不重要。

1. 角度相似系数(夹角余弦)

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x}' \vec{y}}{\|\vec{x}\| \|\vec{y}\|} = \frac{\vec{x}' \vec{y}}{(\vec{x}' \vec{x} \vec{y}' \vec{y})^{1/2}}$$

2. 相关系数

$$r(\vec{x}, \vec{y}) = \frac{\text{cov}(\vec{x}, \vec{y})}{(D(\vec{x}) D(\vec{y}))^{1/2}} = \frac{(\vec{x} - \bar{\vec{x}})' (\vec{y} - \bar{\vec{y}})}{[(\vec{x} - \bar{\vec{x}})' (\vec{x} - \bar{\vec{x}}) (\vec{y} - \bar{\vec{y}})' (\vec{y} - \bar{\vec{y}})]^{1/2}}$$

3. 指数相似系数

$$e(\vec{x}, \vec{y}) = \frac{1}{n} \sum_{i=1}^n \exp \left[-\frac{3}{4} \frac{(x_i - y_i)^2}{\sigma_i^2} \right]$$

三、匹配测度

当特征只有两个状态 (0, 1) 时, 常用匹配测度。

0 表示无此特征 1 表示有此特征。故称之为二值特征。

对于给定的 x 和 y 中的某两个相应分量 x_i 与 y_j

若 $x_i=1, y_j=1$, 则称 x_i 与 y_j 是 (1-1)匹配; $a = \sum_i x_i y_i$; a 代表 1-1 匹配特征

数目

若 $x_i=0, y_j=1$, 则称 x_i 与 y_j 是 (0-1)匹配; $b = \sum_i (1-x_i)y_i$; b 代表 0-1 匹配

特征数目

若 $x_i=1, y_j=0$, 则称 x_i 与 y_j 是 (1-0)匹配; $c = \sum_i x_i(1-y_i)$; c 代表 1-0 匹配

特征数目

若 $x_i=0, y_j=0$, 则称 x_i 与 y_j 是 (0-0)匹配。 $a = \sum_i (1-x_i)(1-y_i)$; e 代表 0-0

匹配特征数目

1. Tanimoto 测度

$$s(\vec{x}, \vec{y}) = \frac{a}{a+b+c} = \frac{\vec{x}'\vec{y}}{\vec{x}'\vec{x} + \vec{y}'\vec{y} - \vec{x}'\vec{y}}$$

等于**共同具有的特征数目**与分别具有的**特征种类总数之比**。这里只考虑(1-1)匹配而**不考虑(0-0)匹配**。

2. Rao 测度

$$s(\vec{x}, \vec{y}) = \frac{a}{a+b+c+e} = \frac{\vec{x}'\vec{y}}{n}$$

(1-1)匹配特征数目和所选用的特征数目之比。

3. 简单匹配系数

$$m(\vec{x}, \vec{y}) = \frac{a+e}{n}$$

(1-1)匹配特征数目与(0-0)匹配特征数目之和与特征总数目之比。

4. Dice 系数

$$m(\vec{x}, \vec{y}) = \frac{a}{2a+b+c} = \frac{\vec{x}'\vec{y}}{\vec{x}'\vec{x} + \vec{y}'\vec{y}}$$

(1-1) 匹配总个数与 x, y 矢量中 1 的总个数之比。

5. Kulzinsky 系数

$$m(\vec{x}, \vec{y}) = \frac{a}{b+c} = \frac{\vec{x}'\vec{y}}{\vec{x}'\vec{x} + \vec{y}'\vec{y} - 2\vec{x}'\vec{y}}$$

(1-1) 匹配总个数与 (1-0) 匹配和 (0-1) 匹配总个数之比。

模式类

定义之 1 设特征矢量集合 S 中任意元素 \vec{x}_i 与 \vec{x}_j 间的距离 d_{ij} (距离测度) 有

$$d_{ij} \leq h$$

其中 h 为给定的阈值, 称集合 S 对于阈值 h 组成一模式类。

类间距离测度方法

两个**模式类的类间距离**。

1. 最近距离法

$$D_{kl} = \min_{i,j} [d_{ij}]$$

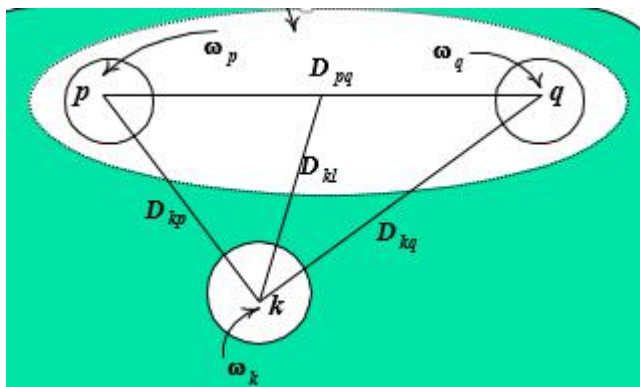
d_{ij} 表示 $\vec{x}_i \in w_k$ 与 $\vec{x}_j \in w_l$ 之间的距离。表示**两模式类中模式距离测度的最小值**。

2. 最远距离法

$$D_{kl} = \max[d_{ij}]$$

d_{ij} 表示 $\vec{x}_i \in w_k$ 与 $\vec{x}_j \in w_l$ 之间的距离。表示两模式类中模式距离测度的最大值。

3. 中间距离法



$$D_{kl}^2 = \frac{1}{2} D_{kp}^2 + \frac{1}{2} D_{kq}^2 - \frac{1}{4} D_{pq}^2$$

其中 $w_l = w_p \cup w_q$ (模式类 L 中的特征矢量是模式类 p 和模式类 q 中特征矢量的集合)。

4. 重心距离法

$$D_{kl}^2 = \frac{n_p}{n_p + n_q} D_{kp}^2 + \frac{n_q}{n_p + n_q} D_{kq}^2 - \frac{n_p n_q}{(n_p + n_q)^2} D_{pq}^2$$

其中 $w_l = w_p \cup w_q$ (模式类 L 中的特征矢量是模式类 p 和模式类 q 中特征矢量的集合)。 n_p, n_q 分别为模式类 w_p, w_q 中特征矢量的个数。

5. 平均距离法

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{i,j} d_{ij}^2$$

d_{ij} 表示 $\vec{x}_i \in w_p$ 与 $\vec{x}_j \in w_q$ 之间的距离。

6. 离差平方和法

类内离差平方和法

$$s_t = \sum_{\vec{x}_i \in w_t} (\vec{x}_i - \vec{x}_t)' (\vec{x}_i - \vec{x}_t)$$

其中 \vec{x}_t 是模式类 w_t 的重心。

$$D_{pq}^2 = s_l - s_p - s_q$$

其中 $w_l = w_p \cup w_q$ 。则

$$D_{pq}^2 = \frac{n_p n_q}{n_p + n_q} (\vec{x}_p - \vec{x}_q)' (\vec{x}_p - \vec{x}_q)$$

其中 \vec{x}_p, \vec{x}_q 分别为模式类 w_p, w_q 的重心。

$$D_{kl}^2 = \frac{n_k + n_p}{n_k + n_l} D_{kp}^2 + \frac{n_k + n_q}{n_k + n_l} D_{kq}^2 - \frac{n_k}{n_k + n_l} D_{pq}^2$$

聚类的准则函数

判别**分类结果好坏的一般标准**：

类内距离小，类间距离大。

如果聚类准则函数选择得好，聚类质量就会高。**聚类准则往往是和类的定义有关的**，是类的定义的某种体现。

一、类内距离准则

类内距离准则函数定义为：

$$J_w = \sum_{j=1}^c \sum_{i=1}^{n_j} \left\| \vec{x}_i^{(j)} - \vec{m}_j \right\|^2 \Rightarrow \min$$

其中 $\{\vec{x}_i^{(j)}; j=1, 2, \dots, c; i=1, 2, \dots, n_j\}$ 为待分类的模式集合， $\vec{x}_i^{(j)}$ 表示将该模式 \vec{x}_i

划分在 w_j 类中。 \vec{m}_j 是分入模式类 w_j 的均值矢量。

分类的目标是使得 J_w 最小，**这种方法也称作误差平方和准则**。

加权类内距离准则 J_{ww} ：

$$J_{ww} = \sum_{j=1}^c \frac{n_j}{N} \bar{d}_j^2$$

$$\bar{d}_j^2 = \frac{2}{n_j(n_j-1)} \sum_{\substack{\vec{x}_k^{(j)} \in \omega_j \\ \vec{x}_i^{(j)} \in \omega_j}} \left\| \vec{x}_i^{(j)} - \vec{x}_k^{(j)} \right\|^2$$

其中 N 为待分类模式的总数。 n_j 为分入模式类 w_j 的模式的个数。 \bar{d}_j^2 表示类内两模式间的均方距离。

二、类间距离准则

$$J_B = \sum_{j=1}^c (\vec{m}_j - \vec{m})' (\vec{m}_j - \vec{m}) \Rightarrow \max$$

其中 $\vec{m}_j = \frac{1}{n_j} \sum_{\vec{x}_i \in w_j} \vec{x}_i$ 为分入模式类 w_j 的模式的均值。 $\vec{m} = \frac{1}{N} \sum_{i=1}^n \vec{x}_i$ 是待分模式的均值。

加权类间距离准则：

$$J_{wB} = \sum_{j=1}^c \frac{n_j}{N} (\vec{m}_j - \vec{m})' (\vec{m}_j - \vec{m}) \Rightarrow \max$$

三、基于类内距离类间距离的准则函数

分入类 w_j 的类内离差阵定义为

$$S_w^{(j)} = \frac{1}{n_j} \sum_{i=1}^{n_j} (\vec{x}_i^{(j)} - \vec{m}_j)(\vec{x}_i^{(j)} - \vec{m}_j)' \quad (j=1, 2, \dots, c)$$

总的类内离差阵定义为

$$S_w = \sum_{j=1}^c \frac{n_j}{N} S_w^{(j)} \quad (j=1, 2, \dots, c)$$

类间离差阵定义为

$$S_B = \sum_{j=1}^c \frac{n_j}{N} (\vec{m}_j - \vec{m})(\vec{m}_j - \vec{m})'$$

类内类间离差阵则为

$$S_T = \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})'$$

可以证明 $S_T = S_w + S_B$

$$\begin{aligned} S_T &= \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \vec{m})(\vec{x}_i - \vec{m})' = \frac{1}{N} \sum_{j=1}^c \sum_{i=1}^{n_j} (\vec{x}_i^{(j)} - \vec{m})(\vec{x}_i^{(j)} - \vec{m})' \\ &= \sum_{j=1}^c \frac{n_j}{N} \frac{1}{n_j} \sum_{i=1}^{n_j} (\vec{x}_i^{(j)} - \vec{m} + \vec{m}_j - \vec{m}_j)(\vec{x}_i^{(j)} - \vec{m} + \vec{m}_j - \vec{m}_j)' \\ &= \sum_{j=1}^c \frac{n_j}{N} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} (\vec{x}_i^{(j)} - \vec{m}_j)(\vec{x}_i^{(j)} - \vec{m}_j)' + (\vec{m}_j - \vec{m})(\vec{m}_j - \vec{m})' + 0 + 0 \right] \\ &= S_w + S_B \end{aligned}$$

四个聚类准则

$$J_1 = \text{Tr}[S_w^{-1} S_B], \quad J_2 = |S_w^{-1} S_B|, \quad J_3 = \text{Tr}[S_w^{-1} S_T], \quad J_4 = |S_w^{-1} S_T|$$

为了得到更好的聚类效果，应该使他们的值足够大。

聚类分析的算法归纳起来就是三大类：

1、按最小距离原则简单聚类方法

针对具体问题确定相似性阈值，将模式到各聚类中心间的距离与阈值比较，当大于阈值时该模式就作为另一类的类心，小于阈值时按最小距离原则将其分划到某一类中。这类算法运行中模式的类别及类的中心一旦确定将不会改变。

2、按最小距离原则进行两类合并的方法

首先视各模式自成一类，然后将距离最小的两类合并成一类，不断地重复这个过程，直到成为两类为止。这类算法运行中，类心不断地修正，但模式类别

一旦指定后就不再改变，就是**模式一旦划为一类后就不再被分划开**，这类算法也称为谱系聚类法。

3、依据准则函数动态聚类方法

设定一些**分类的控制参数**，定义一个能**表征聚类结果优劣的准则函数**，聚类过程就是**使准则函数取极值的优化过程**。算法运行中，**类心不断地修正**，各模式的类别的指定也不断地更改。这类方法有一C 均值法、ISODATA 法等。

聚类的算法

一、简单聚类算法

- (1) 取任意一个模式特征矢量作为第一个聚类中心。例如令 w_1 类的中心

$$\vec{z}_1 = \vec{x}_1$$

- (2) 计算下一个模式特征矢量 \vec{x}_2 到 \vec{z}_1 的距离 d_{21} ，若 $d_{21} > T$ 则建立新类 w_2 ，

类的中心 $\vec{z}_2 = \vec{x}_2$ ；否则，将 \vec{x}_2 归入 w_1 类

- (3) 假设已有类 $\{w_1, w_2, \dots, w_k\}$ ，其中心分别为 $\{\vec{z}_1, \vec{z}_2, \dots, \vec{z}_k\}$ ，计算尚未确

定模式 \vec{x}_i 到各类 w_j 中心 \vec{z}_j 的距离 $d_{ij}, j=1, 2, \dots, k$ ，如果有 $d_{ij} > T$ 则

建立新类 w_{k+1} ，类的中心 $\vec{z}_{k+1} = \vec{x}_i$ ；否则如果 $d_{il} = \min_j [d_{ij}]$ ，则模式 \vec{x}_i

归入类 w_l 中。检查所有模式是否划分结束，如果没有结束，重复步骤 3。

算法特点：

这类算法的突出优点是**算法简单**。但聚类过程中，**类的中心一旦确定将不会改变，模式一旦指定类后也不再改变**。

二、最大最小距离法

- (1) 取任意一个模式特征矢量作为第一个聚类中心。例如令 w_1 类的中心

$$\vec{z}_1 = \vec{x}_1$$

- (2) 从待分类模式特征矢量集中选距离 \vec{z}_1 最远的特征矢量作为第二个类的中心 \vec{z}_2

- (3) 计算未被作为聚类中心的各模式特征矢量与 \vec{z}_1, \vec{z}_2 的距离

$$d_{ij} = \|\vec{x}_i - \vec{z}_j\|, j=1, 2$$

- (4) 若 $d_l = \max_i [\min(d_{i1}, d_{i2})] > \theta \|\bar{z}_1 - \bar{z}_2\|$, 则模式特征矢量 \bar{x}_l 作为第三个聚类中心 $\bar{z}_3 = \bar{x}_l$; 否则, 转至最后一步 (6)
- (5) 假设已有类 $\{w_1, w_2, \dots, w_k\}$, 其中心分别为 $\{\bar{z}_1, \bar{z}_2, \dots, \bar{z}_k\}$, 计算尚未确定模式 \bar{x}_i 到各类 w_j 中心 \bar{z}_j 的距离 $d_{ij}, j=1, 2, \dots, k$, 并计算出 $d_i = \max_j [\min(d_{i1}, d_{i2}, \dots, d_{ik})]$, 若 $d_i \Rightarrow \theta \|\bar{z}_1 - \bar{z}_2\|$, 则模式特征矢量 \bar{x}_i 作为第 $k+1$ 个聚类中心 $\bar{z}_{k+1} = \bar{x}_i$, 然后转至步骤 (5); 否则转至步骤 (6)
- (6) 当判断出不再有新的聚类中心产生之后, 将未分类的模式特征矢量按照最小距离原则分到各类中去。计算 $d_{ij} = \|\bar{x}_i - \bar{z}_j\| (j=1, 2, \dots; i=1, 2, \dots, N)$, 当 $d_{il} = \min_j [d_{ij}]$ 则把特征矢量 \bar{x}_i 描述的模式划分到 w_l 类中。

这种算法的聚类结果与参数 θ 以及第一个聚类心的选取有关。

三、谱系聚类法

- (1) 初始分类。摸个模式自成一类 $G_i^{(0)} = \{\bar{x}_i\}, i=1, 2, \dots, N$, 其中 $G_i^{(k)}$ 表示第 k 次合并时的第 i 类。
- (2) 计算各类间的距离 D_{ij} , 生成一个对称的距离矩阵 $D^{(k)} = (D_{ij})_{m \times m}$, m 为类的个数 (初始 $m=N$)。
- (3) 找出矩阵 $D^{(k)}$ 中的最小元素, 假设是 $G_i^{(k)}$ 与 $G_j^{(k)}$ 间的距离, 将 $G_i^{(k)}$ 与 $G_j^{(k)}$ 合并成一类, 令 $k = k + 1, m = m - 1$
- (4) 检查类的个数, 如果个数大于 2, 则转至步骤 (2); 否则停止。

四、动态聚类算法

动态聚类算法要点

- 1、 确定模式和聚类的距离测度。当采用欧氏距离时, 是计算此模式和该类中心的欧氏距离; 为能反映出类的模式分布结构, 应采用马氏距离
- 2、 确定评估聚类质量的准则函数。
- 3、 确定模式分划及聚类合并或分裂的规则。

动态聚类算法基本步骤

- 1、 建立初始聚类中心, 进行**初始聚类**;
- 2、 计算模式和类的距离, **调整模式的类别**;
- 3、 计算各**聚类的参数**, **删除、合并或分裂一些聚类**;
- 4、 从初始聚类开始, **运用迭代算法动态地改变模式的类别和聚类的中心**使**准则函数取得极值**或**设定的参数达到设计要求**时停止。

(1) C-均值法

第一步：选取 C 个初始聚类中心 $\bar{z}_j^{(0)} = \bar{x}_j, j=1,2,\dots,C$

第二步：计算各个模式 \bar{x}_i 到各个类 $G_j^{(k)}$ 的中心 $\bar{z}_j^{(k)}$ 的距离测度 $d_{ij}, i=1,2,\dots,N, j=1,2,\dots,C$ ，选取 $d_{ii} = \min_j [d_{ij}]$ 将模式 \bar{x}_i 归入 $G_i^{(k+1)}$ 类中。

第三步：根据新建的类更新聚类中心。 $\bar{z}_j^{(k+1)} = \frac{1}{n_j} \sum_{\bar{x}_i \in G_j^{(k+1)}} \bar{x}_i$ ，若存在

$z_j^{(k)} \neq z_j^{(k+1)}$ ，则重复第二步；否则结束。

当模式分布呈现类内团聚状，**C-均值算法**是能达到很好的聚类结果，故应用较多。**C-均值算法**是能使各模式到其所判属的类别中心距离(平方)之和为最小的最佳聚类。

C 的调整：作一条 $C-J$ 曲线，其曲率变化的最大点对应的类数是比较接近最优的类数。在类别数未知的情况下，可使类数 C 由较小值逐步增加，对于每个选定的 C 分别使用该算法。

初始聚类中心的选取：① 凭经验选择初始类心。② 将模式随机地分成 C 类，计算每类中心，以其作为初始类心。③ (最大密度)，求以每个特征点为球心、某一正数 d_0 为半径的球形域中特征点个数，这个数称为该点的密度。选取密度最大的特征点作为第一个初始类心 z_1 ，然后在与 z_1 大于某个距离 d 的那些特征点中选取具有“最大”密度的特征点作为第二个初始类心 z_2 ，如此进行，选取 C 个初始聚类中心。④ 用相距最远的 C 个特征点作为初始类心。具体地讲，是按前述的最大最小距离算法求取 C 个初始聚类中心。⑤ 当 N 较大时，先随机地从 N 个模式中取出一部分模式用谱系聚类法聚成 C 类，以每类的重心作为初始类心。⑥ 设已标准化的待分类模式集为 $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$ 希望将它们分为 C 类。 $\bar{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})'$ ，

$$sum(i) = \sum_{k=1}^n x_{ki} \quad , \quad MAX = \max_i [sum(i)] \quad , \quad MIN = \min_i [sum(i)] \quad ,$$

$$\alpha_i = \frac{(c-1)[sum(i) - MIN]}{MAX - MIN} + 1, (i=1,2,\dots,N), \text{ 若 } \alpha_i \text{ 最接近整数 } j, \text{ 则把}$$

x_i 分划至类 w_j 中，通过初始划分的类计算类中心。

用类核代替类中心：当类的分布不是球状或近似球状时，这种算法很难有较好的效果。类核可以是一个函数、一个点集或其他适当的模型。比如马式距离。

(2) IOSDATA 法

第一步：设定聚类分析控制参数： C - 预期的类数， N_c - 初始聚类中

心个数(可以不等于 C), θ_n -每一类中允许的最少模式数目和分裂时判定, θ_s -类内各分量分布的距离标准差上界(分裂用), θ_d -两类中心间的最小距离下界(合并用), L -在每次迭代中可以合并的类的最多对数, I -允许的最多迭代次数。选定初始聚类中心 $\bar{z}_j, j=1,2,\dots,N_c$: 可随机选择待分类模式集合中的 N_c 个模式为类中心。

第二步: 按照最小距离原则将模式集 $\{\bar{x}_i\}$ 中的每一个模式分到某一个类中。 $d_{il} = \min_j [\bar{x}_i - \bar{z}_j]$, \bar{x}_i 模式分入 w_l 类中。

第三步: 依据 θ_n 判断合并。如果类 w_j 中的模式个数 $n_j < \theta_n$, 则取消该类, 取消类中心 \bar{z}_j , $N_c = N_c - 1$, 转至第二步

第四步: 计算分类后的参数: 各类中心、类内平均距离及总体平均距离。① 计算各类的中心: $\bar{z}_j = \frac{1}{n_j} \sum_{\bar{x}_i \in \omega_j} \bar{x}_i, j=1,2,\dots,N_c$ ② 计算各类中

模式到类心的平均距离 $\bar{d}_j = \frac{1}{n_j} \sum_{\bar{x}_i \in \omega_j} \|\bar{x}_i - \bar{z}_j\|, j=1,2,\dots,N_c$ ③ 计算各

个模式到其类内中心的总体平均距离 $\bar{d} = \frac{1}{N} \sum_{j=1}^{N_c} n_j \bar{d}_j$

第五步: 依据 I_p, N_c 判断停止、分裂或合并。① 若迭代次数 I_p 达到 I , 则转第十一步 (**准备停止**) ② 若 $N_c \leq \frac{C}{2}$, 则转到第六步 (**进行分裂**) ③ 若 $N_c \geq 2C$, 则转到第九步 (**进行合并**) ④ 若 $\frac{C}{2} < N_c < 2C$, 当 I_p 为奇数时转至第六步 (**进行分裂**), 当 I_p 为偶数时转至第九步 (**进行合并**)

第六步: 计算各类类内距离的标准差矢量 $\bar{\sigma}_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{mj})'$ 。

$\sigma_{kj} = \left[\frac{1}{n_j} \sum_{\bar{x}_i \in \omega_j} (x_{ki} - z_{kj})^2 \right]^{\frac{1}{2}}, k=1,2,\dots,n, j=1,2,\dots,N_c$ 。 σ_{kj} 代表 w_j 类的类

内距离标准差的第 k 分量。

第七步：对每一聚类**求出类内距离标准差矢量** $\vec{\sigma}_j$ **的最大分量**

$$\vec{\sigma}_{ij} = \max_k [\sigma_{kj}], k=1,2,\dots,n, j=1,2,\dots,N_c$$

第八步：在 $\{\sigma_{ij}\}$ 中，对每一个 $\sigma_{ij} > \theta_s$ ，同时又满足下列条件之一：1)

$\bar{d}_j > \bar{d}$ （类中模式到类心的平均距离大于所有类的模式到类心的总体平均距离）并且 $n_j > 2(\theta_n + 1)$ （类中模式超过阈值）2) $N_c \leq \frac{C}{2}$ 。则

将该类 w_j 分裂为两类， $N_c = N_c + 1$ ，**新类的中心** \vec{z}_j^+, \vec{z}_j^- **计算规则：**

在**原来的类中心** \vec{z}_j **的** l **分量上对应加减** $k\sigma_{ij}$ ， $0 < k < 1$ ，并且使得新的

的类中心仍然在老的 w_j 类域空间且，新的类中心与 w_j 类域模式近，

与其他类域模式较远。当所有 σ_{ij} 进行判断过，并且进行过类分裂，

则 $I_p = I_p + 1$ ，然后转至第二步；否则转至第九步。

第九步：**计算各类间距离** $D_{ij} = \|\vec{z}_i - \vec{z}_j\|, i=1,2,\dots,N_c, j=i+1,\dots,N_c$

第十步：依据 θ_D 进行**合并判断**。取出 D_{ij} 中小于 θ_D 的值，进行递增排

序，**若个数超过** L **则取前** L **个值**。对每个值，从最小的 D_{ij} 开始合并，

若 w_i, w_j 没有进行过合并，则将相应的两类 w_i, w_j 合并（**注意一个类**

只能被合并一次，假如存在 $D_{12} < D_{24} < \theta_D$ ，则 w_4 不能与 w_2 合并），合

并后，新的类中心 $\vec{z} = \frac{1}{n_i + n_j} [n_i \vec{z}_i + n_j \vec{z}_j]$ ， $N_c = N_c - 1$ 。

第十一步：迭代次数 $I_p \geq I$ ，则停止计算。否则， $I_p = I_p + 1$ ，若需要调整参数，转至第一步；不需要调整参数，转至第二步。

判别域代数界面方程法（有监督分类）

用判别域界面方程分类的概念

分类原理：不同模式对应不同特征点在空间中散部。运用已知类别的**训练样本进行学习**，产生若干**代数界面** $d(\vec{x}) = 0$ ，将代数空间划分成一些互不重叠的子区域。

$d(\vec{x}) = 0$ 又称判别函数。

线性可分的定义：对于来自两类的一组模式 $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ ，如果能用一个线性判别函数将其分为两类，则称他们是线性可分的。

线性判别函数

$$\vec{x} = (x_1, x_2, \dots, x_n)'$$

$$d(\vec{x}) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + w_{n+1}$$

$$d(\vec{x}_1) = w' \vec{x}_1$$

$$w = (w_1, w_2, \dots, w_n, w_{n+1})', \vec{x}_1 = (x_1, x_2, \dots, x_n, 1)'$$

w 称为增广权矢量， \vec{x}_1 称为增广特征矢量

两类处理

设判别函数为 $d(\vec{x})$ ，判别准则如下

$$d(\vec{x}) = w' \vec{x} \begin{cases} > 0, \vec{x} \in \omega_1 \\ < 0, \vec{x} \in \omega_2 \\ = 0, \vec{x} \in \omega_i \end{cases}$$

多类处理

方法一：

w_i / \bar{w}_i 两分法

将属于 w_i 与不属于 w_i 的模式分划开。将划分成 C 个类的问题分解为划分成 $C-1$ 个类的问题。可以建立 C 个判别函数 $d_i(\vec{x}) = w_i' \vec{x}, i=1, 2, \dots, C$ 。经过训练，使判别函数具有以下性质

$$d_i(\vec{x}) = w_i' \vec{x} \begin{cases} > 0, \vec{x} \in \omega_i \\ < 0, \vec{x} \notin \omega_i \end{cases}$$

判决规则：对于 C 个判别函数如果有 $d_k(\vec{x}_i) > 0, d_m(\vec{x}_i) < 0, m=1, 2, \dots, C, m \neq k$ ，则

模式 $\vec{x}_i \in \omega_k$ 。

注意：如果存在两个及以上判别函数 $d_i(\vec{x}_j) > 0$ ，那么无法判断 \vec{x}_j 模式属于哪个类。如果所有判别函数 $d_i(\vec{x}_j) < 0$ ，那么无法判断 \vec{x}_j 模式属于哪个类。

方法二：

存在不确定区的 w_i / w_j 两分法

对 C 个类中的两个类 w_i, w_j 建立一个判别函数。此函数不提供其他类的信息，需要提供 $C(C-1)/2$ 个判别函数 $d_{ij}(\vec{x}) = w'_{ij} \vec{x}, i=1,2,\dots,C, j=i+1,\dots,C$ 。经过训练，

得到区分两类 w_i, w_j 的判别函数具有以下性质

$$d_{ij}(\vec{x}) = w'_{ij} \vec{x} \begin{cases} > 0, \vec{x} \in \omega_i \\ < 0, \vec{x} \in \omega_j \end{cases}$$

判决规则：对于 $C(C-1)/2$ 个判别函数如果有 $d_{ij}(\vec{x}) > 0, j=i+1,\dots,C, d_{ki}(\vec{x}) < 0, k=1,2,\dots,i-1$ ，则模式 $\vec{x} \in w_i$ （所有含有 w_i 类判别的判别函数都判别为 w_i ）。

注意：当不存在一个 i 能够满足上面的条件，则为不确定区域。

方法三：

没有不确定区的 w_i / w_j 两分法

取判定函数 $d_{ij}(\vec{x}) = d_i(\vec{x}) - d_j(\vec{x}) = (w'_i - w'_j) \vec{x}$ 。当 $d_{ij}(\vec{x}) > 0$ 则 $d_i(\vec{x}) > d_j(\vec{x})$ ，此时模式属于 w_i 类。因此只需要对每一个类 w_i 建立一个判别函数 $d_i(\vec{x})$ 。**C 类问题有**

C 个判别函数。 $d_i(\vec{x}) = w'_i \vec{x}, i=1,2,\dots,C$

判决规则：若 $d_k(\vec{x}) = \max_i [d_i(\vec{x})], i=1,2,\dots,C$ ，则该模式属于 w_k 类。

判别函数值的大小、正负的数学意义

线性判别界面方程 $d(\vec{x}) = \vec{w}' \vec{x} + w_{n+1}, \vec{x} = (x_0, x_1, \dots, x_n), \vec{w} = (w_1, w_2, \dots, w_n)$ 是一个超平面 π 。那么其存在一下性质：

1) **系数矢量 \vec{w} 为其法向量。**

2) **判别函数 $d(\vec{x})$ 的值正比于 \vec{x} 到超平面 $d(\vec{x}) = 0$ 的距离 $d_{\vec{x}}$ 。** $d_{\vec{x}} = \frac{1}{\|\vec{w}\|} |d(\vec{x})|$ 。（定

理： $d(\vec{u}) = \vec{m}' \vec{u} + n = 0$ 所确定的 \vec{u} 空间中的平面， \vec{u} 空间中任意一点 \vec{u}_i 到界面

的距离为 $d_{u_i} = \frac{1}{\|\vec{m}\|} |d(\vec{u}_i)|$ ，同时 \vec{m} 为平面的法向量)

3) 若 $\vec{x} \in \mathcal{W}_1$ 等价于 $d(\vec{x}) > 0$ ，则 \vec{x} 位于法向量 \vec{w} 指向的半空间；否则， \vec{x} 位于法向量 \vec{w} 背向的半空间

权空间

增广特征矢量与增广权矢量是相对的。当进行样本训练时， $\vec{x} = (x_1, x_2, \dots, x_n, 1)$ 应

视为 $\vec{w} = (w_1, w_2, \dots, w_n, w_{n+1})$ 的权，判别函数记为 $d(\vec{x}) = \vec{x}'\vec{w}$ 。若 $\vec{x} \in \mathcal{W}_1$ 等价于

$d(\vec{x}) > 0$ ，则 \vec{w} 位于法向量 \vec{x} 指向的半空间；否则， \vec{w} 位于法向量 \vec{x} 背向的半空

间。权空间为以 $(w_1, w_2, \dots, w_n, w_{n+1})$ 为坐标系的 $n+1$ 维空间。权空间中任意一

点 \vec{w}_i 到界面 $\vec{x}'\vec{w} = 0$ 的距离为 $d_{w_i} = \frac{1}{\|\vec{x}\|} |d(\vec{w}_i)|$ 。

解矢量

当模式根据已知类，进行训练时，判别函数记为 $d(\vec{w}) = \vec{x}'\vec{w}$ ，

$\vec{x} = (\vec{x}_j, 1) = (x_1, x_2, \dots, x_n, 1)$ ， $\vec{w} = (w_1, w_2, \dots, w_n, w_{n+1})$ 。当训练模式 $\vec{x}_j \in \mathcal{W}_1$ 时，

$\vec{x}'\vec{w} > 0$ ；当训练模式 $\vec{x}_j \in \mathcal{W}_2$ 时， $\vec{x}'\vec{w} < 0$ ，这时的 \vec{w} 称为解矢量。

解空间

N 个训练模式将有 N 个界面（解矢量界面），每一个界面将权控件分成两个半空间， N 个正的半空间的交空间是一个以原点为顶点的凸多面锥。满足

$\vec{x}_i'\vec{w} > 0 | \vec{x}_i'\vec{w} < 0$ (\vec{x}_i 为矢量 (各训练模式, 1)) 的所有 \vec{w} 都在该锥内部。解矢量

\vec{w} 不是唯一的，凸多面锥包含了全部解矢量，称其为解区，解空间，解锥。每一个训练模式都对解区提供一个约束，训练模式越多，解区的限制就越多，解区就越小，就越靠近解区的中心，解矢量 \vec{w} 就越可靠，由它构造的判别函数错分的可能性就越小。

余量

为了使解矢量更可靠，可以通过增加训练模式数以及引入余量 b 。使得 $\vec{x}'\vec{w} \geq b > 0$ ，从而达到更好的效果。引入余量后界面方程变为 $\vec{x}'\vec{w} - b = 0$ ，此时权空间中任意

一点 \vec{w}_i 到界面 $\vec{x}'\vec{w} - b = 0$ 的距离为 $d_{w_i} = \frac{1}{\|\vec{x}\|} |\vec{x}\vec{w}_i - b|$ ，而该点 \vec{w}_i 到界面 $\vec{x}'\vec{w} = 0$ 的

距离为 $d_{w_i} = \frac{1}{\|\vec{x}\|} |\vec{x}\vec{w}_i|$ ，因此由 $\vec{x}'\vec{w} \geq b > 0$ 确定的凸面锥在 $\vec{x}'\vec{w} > 0$ 确定的凸面锥内

部，边界距离为 $\frac{b}{\|\vec{x}\|}$ 。引入了余量可有效地避免量测的误差、引入的误差以及某些算法求得的解矢量收敛于解区的边界上，从而提高了解的可靠性。

Fisher 线性判别

思想：将 N 维模式投影到一维空间

方法：根据训练确定最佳投影方向（根据使准则函数获取最大值），确定判别规则

假设给定已经分类的训练模式集 $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ ，分属于 c 个类 $\{w_1, w_2, \dots, w_c\}$ ，每个类中的模式集合为 $\{x_j^{(i)}\}, i=1, 2, \dots, C$ ， i 代表某个类， $j \in 1, 2, \dots, N$ 代表模式集中的某个模式。

则各类的模式均值矢量 $\vec{m}_i = \frac{1}{N_i} \sum_j x_j^{(i)}$ ，全部模式均值矢量

$\vec{m} = \frac{1}{N} \sum_{j=1}^n \vec{x}_j$ ，总的类内离差阵 $S_w = \frac{1}{N} \sum_{i=1}^c \sum_j (\vec{x}_j^{(i)} - \vec{m}_i)(\vec{x}_j^{(i)} - \vec{m}_i)'$ ，类间离差阵

$$S_B = \sum_{i=1}^c \frac{n_i}{N} (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})'。$$

(1) 确定最佳投影方向

模式经过投影后 $y_j^{(i)} = \vec{u}' x_j^{(i)}$ 。经过投影变换后的各类的模式均值矢量 $\tilde{m}_i = \vec{u}' \vec{m}_i$ ，

全部模式均值矢量 $\tilde{m} = \vec{u}' \vec{m}$ ，总的类内离差度 $\tilde{S}_w = \vec{u}' S_w \vec{u}$ ，类间离差度 $\tilde{S}_B = \vec{u}' S_B \vec{u}$ 。

定义目标准则函数 $J_F(\vec{u}) = \frac{\tilde{S}_B}{\tilde{S}_w} = \frac{\vec{u}' S_B \vec{u}}{\vec{u}' S_w \vec{u}}$ ，确定能够使得目标准则函数取得最大值的 \vec{u} ，即为投影方向。

$\frac{\partial J_F}{\partial \vec{u}} = \frac{2(\vec{u}' S_w \vec{u}) S_B \vec{u} - 2(\vec{u}' S_B \vec{u}) S_w \vec{u}}{(\vec{u}' S_w \vec{u})^2} = \vec{0}$ ，此时 $\frac{\vec{u}' S_B \vec{u}}{\vec{u}' S_w \vec{u}} = \frac{S_B \vec{u}}{S_w \vec{u}} = \lambda$ ，

$\lambda \vec{u} = S_w^{-1} S_B \vec{u} = S_w^{-1} \sum_{i=1}^c \frac{n_i}{N} (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})' \vec{u}$ ，由于 $\lambda, \frac{n_i}{N}, (\vec{m}_i - \vec{m})' \vec{u}$ 都为标量，我们只

关心 \vec{u} 的方向，因此将标量都置 1，得 $\vec{u} = S_w^{-1} \sum_{i=1}^c (\vec{m}_i - \vec{m})$ ，因此 Fisher 变换函数

$$为 y = \sum_{i=1}^c (\vec{m}_i - \vec{m})' S_w^{-1} \vec{x}$$

(2) 确定判别规则

根据训练模式确定阈值 y_i , $\begin{cases} w_i, \vec{u}'\vec{x} = y > y_i \\ w_j, \vec{u}'\vec{x} = y < y_i \end{cases}$, 判别阈值可取两个类心在 \vec{u} 方向上

轴的投影连线的中点作为阈值 $y_i = \frac{\tilde{m}_i + \tilde{m}_j}{2}$ 。

两类问题中确定 \vec{w} 权值的感知器算法

给定符合规范化 (若 $\vec{x} \in w_1$, 则保持不变, 若 $\vec{x} \in w_2$, 则 \vec{x} 取反, 保证规范化后的训练集中所有模式属于同一个类) 后的增广训练集 $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$, 属于两个类 w_1, w_2

- (1) 置步数 $k = 1$, 令增量 ρ 为一正的常数, 赋值权值 \vec{w} 分量为任一较小值
- (2) 输入训练模式 \vec{x}_k , 计算判别函数值 $\vec{w}(k)' \vec{x}_k$
- (3) 调整权值 \vec{w} , 规则如下

若 $\vec{w}(k)' \vec{x}_k \leq 0$, 则 $\vec{w}(k+1) = \vec{w}(k) + \rho \vec{x}_k$ (增大判别函数值)

若 $\vec{w}(k)' \vec{x}_k > 0$, 则 $\vec{w}(k+1) = \vec{w}(k)$
- (4) 若 $k < N$, 则 $k = k + 1$, 转 (2)。否则检验判别函数 $\vec{w}'\vec{x}$ 对所有 $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ 能否正确分类。若正确则停止; 否则令 $k = 1$, 转 (2)。

两类问题中确定 \vec{w} 权值的梯度下降法

构造准则函数 $J(\vec{w}) = k(|\vec{w}'\vec{x}| - \vec{w}'\vec{x})$, J 的最小值 0, 当且仅当 $\vec{w}'\vec{x} \geq 0$ 。我们的目

标是, 对于已经符号规范化后的增广训练集 $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$, 寻找 \vec{w} 使得所有 $\vec{w}'\vec{x} \geq 0$, 即对构造的准则函数 J , 寻找使 J 取得最小值的 \vec{w} 。

令 $k = 1/2$, 求得准则函数梯度 $\nabla J(\vec{w}) = \frac{\partial J}{\partial \vec{w}} = \frac{1}{2} [\vec{x} \operatorname{sgn}(\vec{w}'\vec{x}) - \vec{x}]$ 。增广权矢量的修

正迭代公式为 (最速下降法: 为时性能指标函数 $J(x)$ 达到最小值, 迭代时

$$\begin{aligned} \vec{w}(k+1) &= \vec{w}(k) - \rho_k \nabla J(\vec{w}(k)) \\ x(k+1) = x(k) - \alpha \nabla J &: = \vec{w}(k) - \frac{\rho_k}{2} [\vec{x}_k \operatorname{sgn}[\vec{w}'(k)\vec{x}_k] - \vec{x}_k] \\ &= \begin{cases} \vec{w}(k), & \vec{w}'(k)\vec{x}_k > 0 \\ \vec{w}(k) + \rho_k \vec{x}_k, & \vec{w}'(k)\vec{x}_k \leq 0, \rho_k > 0 \end{cases} \end{aligned}$$

当 ρ_k 为正常数时，梯度下降法与感知器算法相同。若 ρ_k 较小，则收敛慢，若 ρ_k 较大，则会引起震荡。可使 ρ_k 随 k 变化而变化。

多类问题中感知器算法

判决规则：若 $d_k(\bar{x}) = \max_i [d_i(\bar{x})], i=1,2,\dots,C$ ，则该模式属于 w_k 类。

算法步骤：

- (1) 赋初值,分别给 C 个权矢量 $\bar{w}_i, i=1,2,\dots,C$ 赋任意的初值, ρ 选择正常数, 置步数 $k=1$ 。
- (2) 输入已知类别的增广训练模式 \bar{x}_k ，计算 C 个判别函数
$$d_i(\bar{x}_k) = \bar{w}_i'(k) \bar{x}_k, i=1,2,\dots,C$$
- (3) 若 $\bar{x}_k \in w_l$ ，将每个判别函数 $d_i(\bar{x}_k), i \neq l$ 与 $d_l(\bar{x}_k)$ 进行比较：

若 $d_l > d_i$ ，则 $\bar{w}_l(k+1) = \bar{w}_l(k)$ ， $\bar{w}_i(k+1) = \bar{w}_i(k)$ ，继续比较

若 $d_l \leq d_i$ ，则 $\bar{w}_l(k+1) = \bar{w}_l(k) - \rho \bar{x}_k$ ， $\bar{w}_i(k+1) = \bar{w}_i(k) + \rho \bar{x}_k$ ，跳出比较
- (4) 若 $k < N$ ，则 $k = k+1$ ，转 (2)。否则检验判别函数 $\bar{w}'\bar{x}$ 对所有 $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ 能否正确分类。若正确则停止；否则令 $k=1$ ，转 (2)。

二次准则函数

一次准则函数及其算法（如感知器算法）只适用于线性可分的情况，如果是线性不可分的，分类过程将不收敛！二次准则函数能够测试出模式样本集是否线性可分，并且对线性不可分的情况也能给出“次最优”的解。

最小错分模式数目准则：

对于两类问题，设已经符号规范化后的 $N+1$ 维增广训练集 $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ ，如果训练模式是线性可分的，则存在权矢量 \bar{w} 使不等式组 $\bar{w}\bar{x}_i > 0, i=1,2,\dots,N$ 成立，若线性可分，不等式组有解，线性不可分，不等式无解。引入 N 维余量 $\bar{b} = (b_1, b_2, \dots, b_n)'$ ，则不等式方程组变为 $X\bar{w} \geq \bar{b} > \bar{0}$ ，

其中 $X = \begin{pmatrix} \bar{x}_1' \\ \bar{x}_2' \\ \vdots \\ \bar{x}_N' \end{pmatrix} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)'$ 。对 $X\bar{w} = \bar{b}$ 构造方差准则函数

$J(\vec{w}) = (X\vec{w} - \vec{b})'(X\vec{w} - \vec{b}) = \sum_{i=1}^n (\vec{x}_i' \vec{w} - b_i)^2$ ，当 $J(\vec{w})$ 取最小时， $\vec{x}_i' \vec{w}$ 都在 b_i 附近，因此，被错分的模式数目可达到最小。

求 $J(\vec{w})$ 对 \vec{w} 的梯度，

$$\nabla J(\vec{w}) = 2X'(X\vec{w} - \vec{b})。$$

1) 伪逆法求解 \vec{w}

令 $\nabla J(\vec{w}) = 0$ ，可得 $X'X\vec{w} = X'\vec{b}$ ，当 $(X'X)^{-1}$ 存在的时候 $\vec{w} = (X'X)^{-1}X'\vec{b}$ ，

当 $(X'X)^{-1}$ 不存在的时候， $\vec{w} = (X'X)^+X'\vec{b}$ ， $(X'X)^+$ 为的广义逆矩阵。

2) 梯度法求解 \vec{w}

由于广义逆矩阵计算量比较大，因此常采用梯度下降法。由梯度 $\nabla J(\vec{w}) = 2X'(X\vec{w} - \vec{b})$ 可得，对于任意初值 $\vec{w}(0)$ 的增广权矢量的修正迭代

公式为： $\vec{w}(k+1) = \vec{w}(k) - \rho_k \nabla J(\vec{w}(k)) = \vec{w}(k) - \rho_k X'(X\vec{w}(k) - \vec{b})$ ，当 $\rho_k = \frac{\rho}{k}$

时，该算法使得权矢量序列 $\{\vec{w}(k)\}$ 收敛于 \vec{w}^* ，该 \vec{w}^* 称为 MES 解。为了减少计算量和存储量，可以仿照单样本修正法：由于

$X'(X\vec{w} - \vec{b}) = \sum_{k=1}^N (\vec{w}'\vec{x}_k - b_k)\vec{x}_k$ ，迭代式可以修正为

$\vec{w}(k+1) = \vec{w}(k) - \rho_k (\vec{w}'(k)\vec{x}_k - b_k)\vec{x}_k$ 。此算法通常称为 W-H(Widrow-Hoff)

算法。W-H 算法有两个性质：

1、当 $\vec{b} = \left\{ \underbrace{\frac{N}{N_1}, \dots, \frac{N}{N_1}}_{N_1}, \underbrace{\frac{N}{N_2}, \dots, \frac{N}{N_2}}_{N_2} \right\}$ 时，MES 解等价于 Fisher 解。

2、当 $\vec{b} = \{1, 1, \dots, 1\}$ 时，当样本数 $N \rightarrow \infty$ ，MES 解以最小均方误差逼近贝

叶斯判决函数 $d_B(\vec{x}) = P(w_1 / \vec{x}) - P(w_2 / \vec{x})$ 。

3) HK 算法

在之前的算法中， \vec{b} 为取定的常量，这将影响 \vec{w} 的求取。HK 算法通

过求解使 $J(\vec{b})$ 趋向于最小值的 \vec{b} ，来获取最佳分类器。 $\vec{b}(k)$ 的迭代

公式为 $\vec{b}(k+1) = \vec{b}(k) + \beta(k)$ 。记误差矢量 $\vec{e}(k) = X\vec{w}(k) - \vec{b}(k)$ ，若

$\vec{e}(k) \leq 0$ ，则 $\beta(k) = 0$ （当 $\vec{x}_i' \vec{w}$ 小于 b_i ，此时 b_i 足够大，不需要调整）；

若 $\bar{e}(k) > 0$ ，则 $\beta(k) = -\rho \nabla J(\vec{b}) = 2\rho \bar{e}(k)$ （当 $\vec{x}_i' \vec{w}$ 小于 b_i ，此时 b_i 太小，余量不够，加大余量 b_i ）。 $\beta(k)$ 可统一记为

$\vec{b}(k+1) = \vec{b}(k) + \rho(\bar{e}(k) + |e(k)|)$ 。由 $\vec{w} = (X'X)^{-1}X'\vec{b} = X^+\vec{b}$ 得到

$\vec{w}(k+1) = X^+\vec{b}(k+1) = \vec{w}(k) + \rho X^+(\bar{e}(k) + |e(k)|)$ 。

HK 算法步骤：

1、将训练样本规范化，求 $X^+ = (X'X)^{-1}X'$

2、置初值 $\vec{b}(0) > \vec{0}$ ， $\rho = \frac{1}{2}$ ， $k = 1$

3、计算 $\vec{w}(k) = X^+\vec{b}(k)$ ， $\bar{e}(k) = X\vec{w}(k) - \vec{b}(k)$

4、执行判断：

若 $\vec{b}(k)$ 各分量连续两次都为负值，没有可行解，退出；

若 $\bar{e}(k) \Rightarrow \vec{0}$ ，则输出可行解 $\vec{w}(k)$ ，退出；

否则，执行下一步

5、更新 $\vec{b}(k)$ 。 $\vec{b}(k+1) = \vec{b}(k) + \rho(\bar{e}(k) + |e(k)|)$

6、 $k = k + 1$ ，转至步骤 3

广义线性判别函数

原理：将 n 维模式特征集 $\{\vec{x}_i\}$ 映射到 m 维模式特征级 $\{\vec{y}_i\}$ ，其中 $\vec{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})'$ ， \vec{y}_i 为对应 \vec{x}_i 的映射 $\vec{y}_i = (f_1(\vec{x}_i), f_2(\vec{x}_i), \dots, f_m(\vec{x}_i))'$ ， f_i 为不同的映射函数，经过映射之后的 $\{\vec{y}_i\}$ 能够线性可分。

经变化后的线性判别函数 $\tilde{d}(\vec{y}) = \vec{w}'\vec{y} = w_1f_1(\vec{x}) + \dots + w_mf_m(\vec{x}) + w_{m+1} = d(\vec{x})$ ，其中

$\vec{w} = (w_1, w_2, \dots, w_m, w_{m+1})$ ， $\vec{y} = (f_1(\vec{x}), f_2(\vec{x}), \dots, f_m(\vec{x}), 1)$ 。

典型的 $f_i(\vec{x})$

(1) $d(\vec{x})$ 为二次多项式

$$d(\vec{x}) = \sum_{i=1}^n w_{ii}x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij}x_ix_j + \sum_{i=1}^n w_ix_i + w_{n+1}$$

变换后的特征 \vec{y}_i 的维数为 $\frac{n(n+3)}{2}$

(2) $d(\vec{x})$ 为 r 次多项式

$$d(\vec{x}) = d^r(\vec{x}) + d^{r-1}(\vec{x}) + \dots + d^0(\vec{x}) = d^{(r)}(\vec{x})$$

$$d^0(\vec{x}) = w_{n+1} = d^{(0)}(\vec{x}), \quad 1 \text{ 项}$$

$$d^1(\vec{x}) = \sum_{i=1}^n w_i x_i, \quad C_n^1 \text{ 项}$$

$$d^2(\vec{x}) = \sum_{i_1=1}^n \sum_{i_2=i_1}^n w_{i_1 i_2} x_{i_1} x_{i_2}, \quad C_{n+1}^2 \text{ 项}$$

$$d^r(\vec{x}) = \sum_{i_1=1}^n \sum_{i_2=i_1}^n \dots \sum_{i_r=i_{r-1}}^n w_{i_1 i_2 \dots i_r} x_{i_1} x_{i_2} \dots x_{i_r}, \quad C_{n+r-1}^r \text{ 项}$$

$$d^{(r)}(\vec{x}) \text{ 的项数为: } \sum_{k=1}^r C_{n+k-1}^k + 1 = (n+r)!/n!r!$$

变换后的特征 \vec{y}_i 的维数为 $(n+r)!/n!r! - 1$

二次判别函数

在 n 维特征空间中，二次判别函数的一般表示式为：

$$d(\vec{x}) = \vec{x}' W \vec{x} + \vec{w}' \vec{x} + w_{n+1} = \sum_{k=1}^n w_{kk} x_k^2 + 2 \sum_{j=1}^{n-1} \sum_{k=j+1}^n w_{jk} x_j x_k + \sum_{j=1}^n w_j x_j + w_{n+1}$$

一般的判别规则是：

计算训练模式 $\vec{x}_j \in \omega_l (j=1, 2, \dots, N_l)$,

$$\hat{m}_l = \frac{1}{N_l} \sum_{j=1}^{N_l} \vec{x}_j, \quad \hat{C}_l = \frac{1}{N_l - 1} \sum_{j=1}^{N_l} (\vec{x}_j - \hat{m}_l)(\vec{x}_j - \hat{m}_l)'$$

构造判别函数： $d(\vec{x}) = K^2 - (\vec{x} - \hat{m}_l)' \hat{C}_l^{-1} (\vec{x} - \hat{m}_l)$

$$\text{对未知模式: } \begin{cases} d(x) > 0, x \in \omega_1 \\ d(x) < 0, x \in \omega_2 \end{cases}$$

位势函数分类法

特点：(1) 可直接确定判决函数（位势函数）

(2) 适用于非线性和线性可分的情况

对于两类问题认为：如果 $\vec{x} \in w_1$ ，则 \vec{x} 带正电荷 $K(\vec{x}) > 0$ ；如果 $\vec{x} \in w_2$ ，则 \vec{x} 带负电荷 $K(\vec{x}) < 0$ 。

定义模式位势函数 $K(\vec{x}, \vec{x}_j)$ ，表示模式 \vec{x} 在 \vec{x}_j 的势场中的位势或模式 \vec{x}_j 在 \vec{x} 的势场中的位势。 $K(\vec{x}, \vec{x}_j)$ 有如下性质： $K(\vec{x}, \vec{x}_j) = K(\vec{x}_j, \vec{x})$ ； $K(\vec{x}, \vec{x}_j)$ 连续光滑； \vec{x} 与 \vec{x}_j 相距越远，位势越趋近于 0，当 $\vec{x} = \vec{x}_j$ ，位势最高。

两种模式位势函数

(1) 设 $\{\varphi_i(\vec{x})\}$ 是一组正交函数集，则 $K(\vec{x}, \vec{x}_j) = \sum_{i=1}^n \varphi_i(\vec{x}) \varphi_i(\vec{x}_j)$

(2) 取关于 \vec{x}, \vec{x}_j 距离的对称函数，例如

$$K(\vec{x}, \vec{x}_j) = \exp(-\alpha \|\vec{x} - \vec{x}_j\|^2), K(\vec{x}, \vec{x}_j) = \frac{1}{1 + \alpha \|\vec{x} - \vec{x}_j\|^2}$$

确定判别函数步骤

设训练集 $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n\}$ ，分属于 w_1, w_2 ，定义一个模式位势函数 $K(\vec{x}, \vec{x}_j)$ 。通过以下

步骤确定位势函数 $K(\vec{x}) = \sum_{j=1}^n \alpha_j K(\vec{x}, \vec{x}_j)$ 中 $\alpha_j, j=1, 2, \dots, n$ 的值。

(1) 令初始位势函数 $K_0(\vec{x}) = 0$ ，判错标记 $flag = 0$ ，步数 $j = 0$

(2) $j = j + 1$ ，调整位势函数：

若 $\vec{x}_j \in w_1$ ， $K_{j-1}(\vec{x}_j) > 0$ ，则 $K_j(\vec{x}) = K_{j-1}(\vec{x})$

若 $\vec{x}_j \in w_2$ ， $K_{j-1}(\vec{x}_j) < 0$ ，则 $K_j(\vec{x}) = K_{j-1}(\vec{x})$

若 $\vec{x}_j \in w_1$ ， $K_{j-1}(\vec{x}_j) \leq 0$ ，则 $K_j(\vec{x}) = K_{j-1}(\vec{x}) + K(\vec{x}, \vec{x}_j)$ ， $flag = 1$

若 $\vec{x}_j \in w_2$ ， $K_{j-1}(\vec{x}_j) \geq 0$ ，则 $K_j(\vec{x}) = K_{j-1}(\vec{x}) - K(\vec{x}, \vec{x}_j)$ ， $flag = 1$

(3) 若 $j < n$ ，转 (2)。否则判断判错标记 $flag$ 是否为 0， $flag$ 为 0，退出；

否则， $flag = 0$ ， $j = 0$ ，转 (2)。

位势函数分类法与感知器算法有其相似点：在训练时，当判别函数错误时，对判别函数进行修正（若 $\vec{x}_i \in w_1$ ，但 $d(\vec{x}_i) \leq 0$ ，则向 $d(\vec{x})$ 增加正值；若 $\vec{x}_i \in w_2$ ，但

$d(\vec{x}_i) \geq 0$ ，则向 $d(\vec{x})$ 增加负值)。

对于多类问题，位势函数分类法确定判别函数步骤：

- (1) 令初始位势函数 $K_0^i(\vec{x}) = 0, i = 1, 2, \dots, n$ ，判错标记 $flag = 0$ ，步数 $j = 0$
- (2) $j = j + 1$ ，假设 $\vec{x}_j \in w_l$ ，对每一个位势函数 $K_{j-1}^{(i)}(\vec{x}_j), i = 1, 2, \dots, n, i \neq l$ 与 $K_{j-1}^{(l)}(\vec{x}_j)$ 进行比较
若 $K_{j-1}^{(l)}(\vec{x}_j) > K_{j-1}^{(i)}(\vec{x}_j)$ ，则 $K_j^{(l)}(\vec{x}) = K_{j-1}^{(l)}(\vec{x})$ ， $K_j^{(i)}(\vec{x}) = K_{j-1}^{(i)}(\vec{x})$ ，继续比较
若 $K_{j-1}^{(l)}(\vec{x}_j) \leq K_{j-1}^{(i)}(\vec{x}_j)$ ，则 $K_j^{(l)}(\vec{x}) = K_{j-1}^{(l)}(\vec{x}) + K(\vec{x}, \vec{x}_j)$ ，
 $K_j^{(i)}(\vec{x}) = K_{j-1}^{(i)}(\vec{x}) - K(\vec{x}, \vec{x}_j)$ ， $flag = 1$ ，跳出比较
- (3) 若 $j < n$ ，转 (2)。否则判断判错标记 $flag$ 是否为 0， $flag$ 为 0，退出；
否则， $flag = 0$ ， $j = 0$ ，转 (2)。

统计判决

随机模式分类识别，通常称为 **Bayes(贝叶斯)判决**。主要依据类的概率、概密，按照某种准则使分类结果从统计上讲是最佳的。准则函数不同，所导出的判决规则就不同，分类结果也不同。

基础知识:

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

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

先验概率: $P(w_i)$ 表示类 w_i 出现的先验概率, 简称类 w_i 的概率。

后验概率: $P(w_i|x)$ 表示在 x 出现条件下类 w_i 出现的概率, 称其为类别的后验概率,

对于模式识别来讲可理解为 x_i 来自类 w_i 的概率。

类概密: $p(\vec{x}|w_i)$ 表示在类 w_i 条件下的概率密度, 即类 w_i 模式 \vec{x} 的概率分布密度, 简称为类概密。(由于 \vec{x} 可以在空间中取任意值, 因此 $p(\vec{x}|w_i)$ 代表其概率)。

根据后验概率进行判决

若 $P(w_1|x) > P(w_2|x)$, 则 $x \in w_1$

若 $P(w_1|x) < P(w_2|x)$, 则 $x \in w_2$

根据 Bayes 公式, 后验概率 $P(w_i|x)$ 可表示为:

$$P(\omega_i | \vec{x}) = \frac{p(\vec{x} | \omega_i)P(\omega_i)}{p(\vec{x})} = \frac{p(\vec{x} | \omega_i)P(\omega_i)}{\sum_{i=1}^2 p(\vec{x} | \omega_i)P(\omega_i)}$$

因此判决也为:

若 $\frac{p(\vec{x} | w_1)}{p(\vec{x} | w_2)} > \frac{P(w_2)}{P(w_1)}$, 则 $x \in w_1$

若 $\frac{p(\vec{x} | w_1)}{p(\vec{x} | w_2)} < \frac{P(w_2)}{P(w_1)}$, 则 $x \in w_2$

最小误判概率准则判决

根据类的概率 $P(w_i)$ 和类概密 $p(\vec{x}|w_i)$ 将空间划分为 Ω_1, Ω_2 两个区域

$\Omega_1 \cup \Omega_2 = \Omega, \Omega_1 \cap \Omega_2 = \emptyset$, 此时正确的分类: $\vec{x} \in \Omega_1$ 时 $\vec{x} \in w_1$; $\vec{x} \in \Omega_2$ 时 $\vec{x} \in w_2$ 。

因此问题为**如何划分 Ω_1, Ω_2 可使得误判概率最小。**

错误的分类:

$\vec{x} \in \Omega_2$ 时判断其 $\in w_1$, 错误分类概率为 $\xi_{12} = \int_{\Omega_2} p(\vec{x}|w_1) d\vec{x}$ (在类 w_1 出现的条件下

各模式 $\in w_1$ 的总概率)。

$\vec{x} \in \Omega_1$ 时判断其 $\in w_2$, 错误分类概率为 $\xi_{21} = \int_{\Omega_1} p(\vec{x}|w_2) d\vec{x}$ (在类 w_2 出现的条件下

各模式 $\in w_2$ 的总概率)。

总的错误划分概率为:

$$P_e = \xi_{12}P(w_1) + \xi_{21}P(w_2) = \int_{\Omega_1} P(w_2)p(\vec{x}|w_2) + \int_{\Omega_2} P(w_1)p(\vec{x}|w_1)$$

使 P_e 最小的划分: $\forall \vec{x} \in \Omega_1$ 满足 $P(w_1)p(\vec{x}|w_1) > P(w_2)p(\vec{x}|w_2)$, $\forall \vec{x} \in \Omega_2$ 满足

$$P(w_1)p(\vec{x}|w_1) < P(w_2)p(\vec{x}|w_2)。$$

因此**判决准则**为:

若 $P(w_1)p(\vec{x}|w_1) > P(w_2)p(\vec{x}|w_2)$, 则 $\vec{x} \in w_1$

若 $P(w_1)p(\vec{x}|w_1) < P(w_2)p(\vec{x}|w_2)$, 则 $\vec{x} \in w_2$

正态模式最小误判概率准则

在 C 类问题中, w_i 类条件下的类概密为:

$$p(\vec{x}|w_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp[-\frac{1}{2}(\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)] \text{ 其中 } \Sigma_i \text{ 为协方差矩阵, } \vec{\mu}_i \text{ 为均值矢量。}$$

w_i 类的判决函数可以变为:

$$d_i(\vec{x}) = p(\vec{x}|w_i)P(w_i) = \ln P(w_i) - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)$$

由于其中的常量不改变判决的结果, 因此**判决函数**变为:

$$d_i(\vec{x}) = \ln P(w_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i), i = 1, 2, 3, \dots, C$$

判决规则：若 $d_k(\bar{x}) > d_i(\bar{x}), \forall i = 1, 2, \dots, C; i \neq k$ ，则该模式属于 w_k 类。

正态模式分类的误判概率

考虑两类问题，**设两类模式为协方差阵相等的多变量正态分布。**
总得误判概率为

$$\begin{aligned} P(e) &= P(\omega_i)P(L_{ij} < \alpha | \omega_i) + P(\omega_j)P(L_{ij} > \alpha | \omega_j) \\ &= P(\omega_i)\Phi\left[\frac{\alpha - \frac{1}{2}r_{ij}^2}{r_{ij}}\right] + P(\omega_j)\left\{1 - \Phi\left[\frac{\alpha + \frac{1}{2}r_{ij}^2}{r_{ij}}\right]\right\} \\ r_{ij}^2 &= (\vec{\mu}_i - \vec{\mu}_j)' \Sigma^{-1} (\vec{\mu}_i - \vec{\mu}_j) \end{aligned}$$

$$\Phi(u) = \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy$$

当两类先验概率相等时，**误判概率与两类的马氏距离的关系：**

$P(e)$ 随 r_{ij} 的增大而单调递减，只要**两类马氏距离足够大**，其**误判概率可足够小**。

最小损失准则判决

设模式空间中存在 c 个类别： w_1, w_2, \dots, w_c

决策空间由 a 个决策： $\alpha_1, \alpha_2, \dots, \alpha_a$

决策 α_j 表示，将模式分入 w_j 类或者拒判。

对一个实属 w_i 类的模式采用了决策 α_j 所造成的单类损失记为： $\lambda(a_j | w_i) \Rightarrow \lambda_{ij}$
(也就是单类误判厉害程度)

采用决策 α_j 而产生的总损失为： $R_j(\bar{x}) = \sum_{i=1}^c \lambda_{ij} P(w_i | \bar{x})$

判决准则：

采用哪种决策可使产生的损失最小则 \bar{x} 使用该决策进行判断。

$R_j(\bar{x}) = \min_i \{R_i(\bar{x})\}, i = 1, 2, \dots, c$ ，则 $\bar{x} \in w_j$

当 $\lambda_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$ 时，最小损失则判决与最小误差准则判决一样。

含拒绝判决的最小损失判决

决策 α_{c+1} 表示：对模式拒绝进行判决。

对一个实属 w_i 类的模式拒绝判决所造成的损失记为 $\lambda(\alpha_{c+1} | w_i)$ 。

当 $R(\alpha_{c+1} | \bar{x}) < \min_j R(\alpha_j | \bar{x}), j=1,2,\dots,c$ ，则拒绝进行判决。

假设 $\lambda(\alpha_{c+1} | w_i) = \lambda_r, \forall i=1,2,\dots,c$ ，各类拒判损失相同

假设 $\lambda(\alpha_j | w_i) = \lambda_e, i \neq j, \forall i, j=1,2,\dots,c$ ，各类误判损失相同

假设 $\lambda(\alpha_j | w_i) = \lambda_c, i = j, \forall i=1,2,\dots,c$ ，各类正确判决损失相同

通常 $\lambda_c < \lambda_r < \lambda_e$ 。

$$\text{因此: } R(\alpha_j | \bar{x}) = \sum_{i=1}^c \lambda(\alpha_j | w_i) P(w_i | \bar{x}) = \lambda_e - (\lambda_e - \lambda_c) P(w_j | \bar{x})$$

由于 $\lambda_c, \lambda_r, \lambda_e$ 都为常数，且 $\lambda_e > \lambda_c$ ，因此 $\min_j R(\alpha_j | \bar{x}) \Rightarrow \max_j P(w_j | \bar{x})$ 。

$R(\alpha_{c+1} | \bar{x}) < \min_j R(\alpha_j | \bar{x})$ 可改写为

$$R(\alpha_{c+1} | \bar{x}) = \sum_{i=1}^c \lambda_r P(w_i | \bar{x}) = \lambda_r < \lambda_e - (\lambda_e - \lambda_c) \max_j P(w_j | \bar{x})$$

$$\max_j P(w_j | \bar{x}) < 1 - \frac{\lambda_r - \lambda_c}{\lambda_e - \lambda_c} = 1 - t, t = \frac{\lambda_r - \lambda_c}{\lambda_e - \lambda_c} \in (0, 1)$$

含拒判的判决准则:

若 $\max_j P(w_j | \bar{x}) < 1 - t$ ，则拒绝判决;

若 $\max_j P(w_j | \bar{x}) = P(w_k | \bar{x}) \geq 1 - t$ ，则判决 $\bar{x} \in w_k$

由于 $\max_j P(w_j | \bar{x}) \geq \frac{1}{c}, j=1,2,\dots,c$ ，因此当 $t > 1 - \frac{1}{c}$ ，不存在拒判。

最小最大损失准则

当类的先验概率未知时，可使用最小最大损失准则

一种分类识别决策将模式特征空间 Ω 分为 c 个 $\{\Omega_i\}$ 子空间，空间内所有特征的判决总损失为:

$$R = \sum_{i=1}^c \int_{\Omega_i} R(\alpha_i | \bar{x}) p(\bar{x}) d\bar{x} = \sum_{i=1}^c \int_{\Omega_i} \sum_{k=1}^c \lambda_{ki} p(\bar{x} | w_k) P(w_k) d\bar{x}$$

其中子空间的划分受 $P(w_j), j=1,2,\dots,c$ 的影响。由于 $P(w_j)$ 是一个未知值，在特定的 $P(w_j)$ 下通过最小损失准则可以确定子空间的划分，并得出空间内所有特征

的判决总损失 R 。通过选取不同的 $P(W)=[P(w_1), P(w_2), \dots, P(w_c)]'$ 可以得到不同的 R ，因此可以构造 $P(W)-R$ 的曲线图。通过曲线图可以得到使 R 最大的 $P(W)$ ，此时的 R 就是最小损失中的最大损失（最坏情况）。然后通过使 R 最大的 $P(W)$ 来进行模式的分类。

设计过程：

- (1) 给定一个 $P(W)=[P(w_1), P(w_2), \dots, P(w_c)]'$ ，通过最小损失准则划分出子空间
- (2) 计算判决总损失 R
- (3) 选取不同的 $P(W)$ ，构造 $P(W)-R$ 曲线，得到能使 R 取得最大值的 $P^*(W)$
- (4) 使用 $P^*(W)$ 来进行最小损失准则计算，对模式进行分类。

N-P(Neyman—Pearson)判决

N-P 判决是**严格限制较重要的一类错误概率令其等于某常数**而使另一类误判概率最小。

对于两类问题，错误划分为：

$$\xi_{12} = \int_{\Omega_2} p(\bar{x} | w_1) d\bar{x}, \quad \xi_{21} = \int_{\Omega_1} p(\bar{x} | w_2) d\bar{x} \quad (\text{意义同最小误差准则中的})$$

假设 ξ_{21} 错误分类比较重要，即原模式 $\bar{x} \in w_1$ 却错误分类为 w_2 ，我们令其为一个

较小的常数 ξ_0 ，即 $\xi_{21} = \xi_0$ 。在该条件下获取使得 ξ_{12} 取得最小值的 Ω_1, Ω_2 划分方法。

构造拉格朗日函数 $F = \xi_{12} + \lambda(\xi_{21} - \xi_0) = (1 - \lambda\xi_0) - \int_{\Omega_1} (p(\bar{x} | w_1) - \lambda p(\bar{x} | w_2)) d\bar{x}$ 。可以

得到，使得 ξ_{12} 取得最小值的 Ω_1, Ω_2 划分方法为：

$$\frac{p(\bar{x} | w_1)}{p(\bar{x} | w_2)} > \lambda, \quad \bar{x} \in \Omega_1;$$

$$\frac{p(\bar{x} | w_1)}{p(\bar{x} | w_2)} < \lambda, \quad \bar{x} \in \Omega_2。$$

$$\text{由 } \frac{p(\bar{x} | w_1)}{p(\bar{x} | w_2)} = \lambda \text{ 可得划分边界 } \bar{x} = f(\lambda)。 \text{由约束条件 } \xi_{21} = \xi_0 \Rightarrow \int_{f(\lambda)} p(\bar{x} | w_2) d\bar{x} = \xi_0$$

可以得到 λ 的值。

判决规则如下：

$$\frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} > \lambda \text{ , } \vec{x} \in w_1 \text{ ;}$$

$$\frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} < \lambda \text{ , } \vec{x} \in w_2 \text{ .}$$