IPL

计算机学院

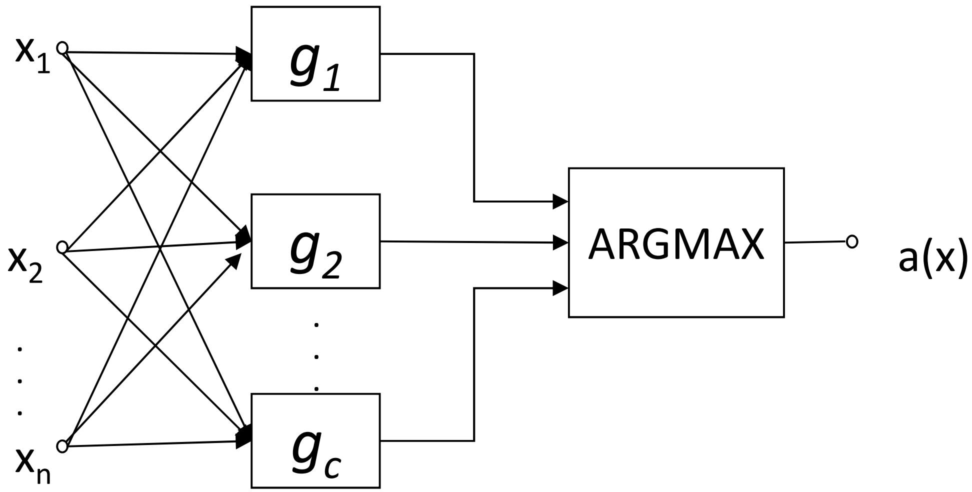
第四章 线性判别函数

孔万增 Kong Wanzeng, Ph.D
Tel: 15967146928
Email: kongwanzeng@hdu.edu.cn

Table of Contents

4.1 引言

分类器
功能结构



基于样本的Bayes分类器：通过估计类条件概率密度函数，设计相应的判别函数

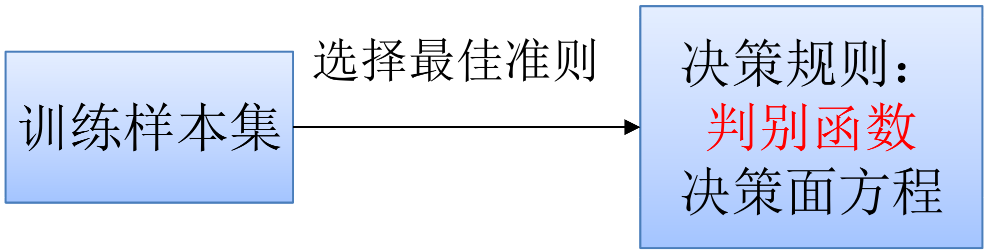
- 最一般情况下适用的“最优”分类器：错误率最小，对分类器设计在理论上有指导意义。
- 获取统计分布及其参数很困难，实际问题中并不一定具备获取准确统计分布的条件。

训练样本集 → 样本分布的统计特征：概率密度函数 → 决策规则：判别函数 决策面方程

直接确定判别函数

◆基于样本的直接确定判别函数方法：

- 设定判别函数形式，用样本集确定判别函数的参数。
- 定义准则函数，表达分类器应满足的要求。
- 这些准则的“最优”并不一定与错误率最小相一致：次优分类器。
- 实例：正态分布最小错误率贝叶斯分类器在特殊情况下，是线性判别函数$g(\boldsymbol{x})=\boldsymbol{w}^T\boldsymbol{x}$（决策面是超平面）。那么我们能否基于样本直接确定$\boldsymbol{w}$?



线性分类器设计步骤

引言

设计

- 线性分类器设计任务：给定样本集K，确定线性判别函数$g(\mathbf{x})=\mathbf{w}^T\mathbf{x}$的各项系数$\mathbf{w}$。
- 步骤：

1. 收集一组已知类别的样本$K=\{\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N\}$
2. 按需要确定一准则函数$J(K,\mathbf{w})$，其值反映分类器的性能，其极值解对应于“最好”分类。
3. 用最优化技术求准则函数J的极值解$\mathbf{w}^*$，从而确定判别函数，完成分类器设计。

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmax}}\, J(K,\mathbf{w})$$

应用

- 对于未知样本$\mathbf{x}$，计算$g(\mathbf{x})$，判断其类别。

线性判别函数

- *d*维空间中的线性判别函数的一般形式：

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$$

- **x**是样本向量，即样本在*d*维特征空间中的描述， **w**是权向量，w_0是一个常数(阈值权)。

$$\mathbf{x} = \left[x_1, x_2, \ldots x_d\right]^T \quad \mathbf{w} = \left[w_1, w_2, \ldots w_d\right]^T$$

两类问题的分类决策规则

规则表达1

$$\text{如果}\begin{cases} g(\mathbf{x})>0, & \text{则决策}\mathbf{x}\in\omega_1 \\ g(\mathbf{x})<0, & \text{则决策}\mathbf{x}\in\omega_2 \\ g(\mathbf{x})=0, & \text{可将其任意分类或拒绝} \end{cases}$$

规则表达2

$$j=\underset{i}{\operatorname{argmax}}\, g_i(\mathbf{x})$$

线性判别函数的几何意义

引言

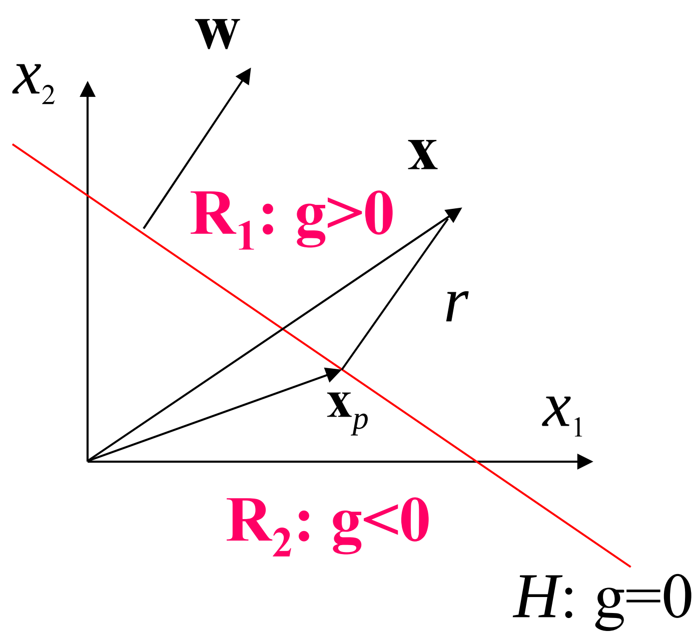
- 决策面(decision boundary)方程：$g(\mathbf{x})=0$
- 决策面将特征空间分成决策区域。
- 向量**w**是决策面H的法向量
- $g(\mathbf{x})$是点**x**到决策面H的距离的一种代数度量

$$\mathbf{x} = \mathbf{x}_p + r\frac{\mathbf{w}}{\|\mathbf{w}\|}, \quad g(\mathbf{x}) = r\|\mathbf{w}\|$$

r是**x**到H的垂直距离

$\mathbf{x}_p$是**x**在H上的投影向量

$$r_0 = \frac{w_0}{\|\mathbf{w}\|}$$

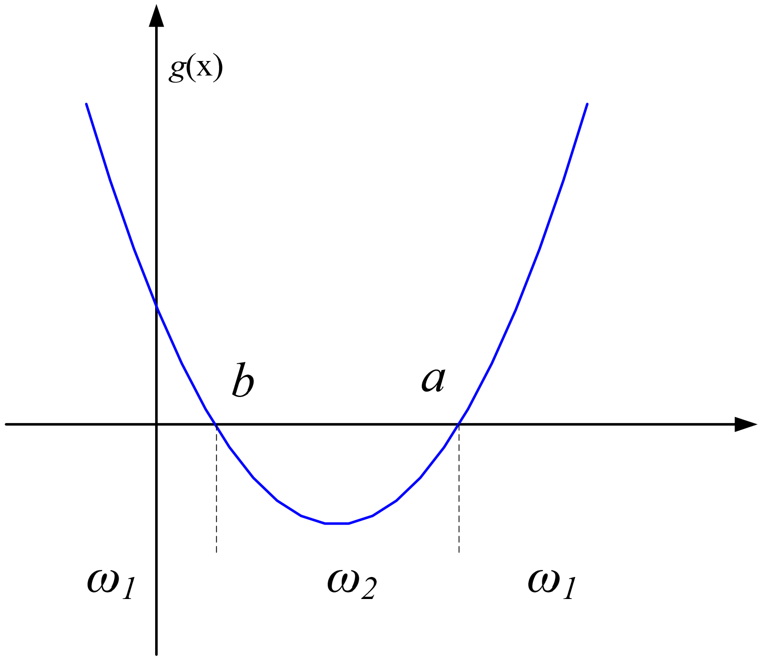


广义线性判别函数

引言

- 线性判别函数是形式最为简单的判别函数，但是它不能用于复杂情况。
  - 例：设计一个一维分类器，使其功能为：

$$\text{如果}\begin{cases} x<b \text{ 或 } x>a & \text{则决策} x\in\omega_1 \\ b\le x\le a & \text{则决策} x\in\omega_2 \end{cases}$$



判别函数：

$$g(x)=(x-a)(x-b)$$

广义线性判别函数(2)

引言

- 二次函数的一般形式： $g(x) = c_0 + c_1 x + c_2 x^2$

映射X→Y

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}$$

- g(x)又可表示成： $g(x) = \mathbf{a}^T \mathbf{y} = \sum_{i=1}^{3} a_i y_i$

广义线性判别函数(3)

- 按照上述方法，任何非线性函数*g*(**x**)用级数展开成高次多项式后，都可转化成线性来处理。
- 一种特殊映射方法：增广样本向量**y**与增广权向量**a**

$$\mathbf{y} = \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = [x_1, \ldots, x_d, 1]^T$$

$$\mathbf{a} = \begin{bmatrix} \mathbf{w} \\ w_0 \end{bmatrix} = [w_1, \ldots, w_d, w_0]^T$$

广义线性判别函数(4)

- 线性判别函数的齐次简化：

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = \mathbf{a}^T\mathbf{y}$$

- 增广样本向量使特征空间增加了一维，但保持了样本间的欧氏距离不变，对于分类效果也与原决策面相同，只是在Y空间中决策面是通过坐标原点的，这在分析某些问题时具有优点，因此经常用到。

广义线性判别函数举例

例1：设五维空间的线性方程为 $55x_1+68x_2+32x_3+16x_4+26x_5+10=0$，试求出其权向量与样本向量点积的表达式 $\mathbf{w}^T\mathbf{x}+w_0=0$ 中的 $\mathbf{w}$，$\mathbf{x}$ 以及增广权向量与增广样本向量形式 $\mathbf{a}^T\mathbf{y}$ 中的 $\mathbf{a}$ 与 $\mathbf{y}$。

答：

- 样本向量：$\mathbf{x}=(x_1, x_2, x_3, x_4, x_5)^T$
- 权向量：$\mathbf{w}=(55, 68, 32, 16, 26)^T$，$w_0=10$
- 增广样本向量：$\mathbf{y}=(1, x_1, x_2, x_3, x_4, x_5)^T$
- 增广权向量：$\mathbf{a}=(10, 55, 68, 32, 16, 26)^T$

广义线性判别函数举例(2)

例2：有一个三次判别函数：$z=g(x)=x^3+2x^2+3x+4$。试建立一映射$x \to \mathbf{y}$，使得z转化为$\mathbf{y}$的线性判别函数。

答：映射X→Y如下：

$$\mathbf{y}=\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}=\begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \end{bmatrix}, \quad \mathbf{a}=\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}=\begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

$$z=g(x)=h(\mathbf{y})=\mathbf{a}^T\mathbf{y}=\sum_{i=1}^{4}a_i y_i$$

广义线性判别函数举例(3)

引言

例3：设在三维空间中一个类别分类问题拟采用二次曲面。如欲采用广义线性方程求解，试问其广义样本向量与广义权向量的表达式，其维数是多少？

答：设次二次曲面为:

二次曲面

$$ax_1^2+bx_2^2+cx_3^2+dx_1x_2+ex_1x_3+fx_2x_3+gx_1+hx_2+lx_3+m=0$$

广义权向量

$$\mathbf{a}=(a,b,c,d,e,f,g,h,l,m)^T$$

广义样本向量

$$\mathbf{y}=(x_1{}^2,x_2{}^2,x_3{}^2,x_1x_2,x_1x_3,x_2x_3,x_1,x_2,x_3,1,)^T$$

维数为10

$$z=g(\mathbf{x})=h(\mathbf{y})=\mathbf{a}^T\mathbf{y}$$

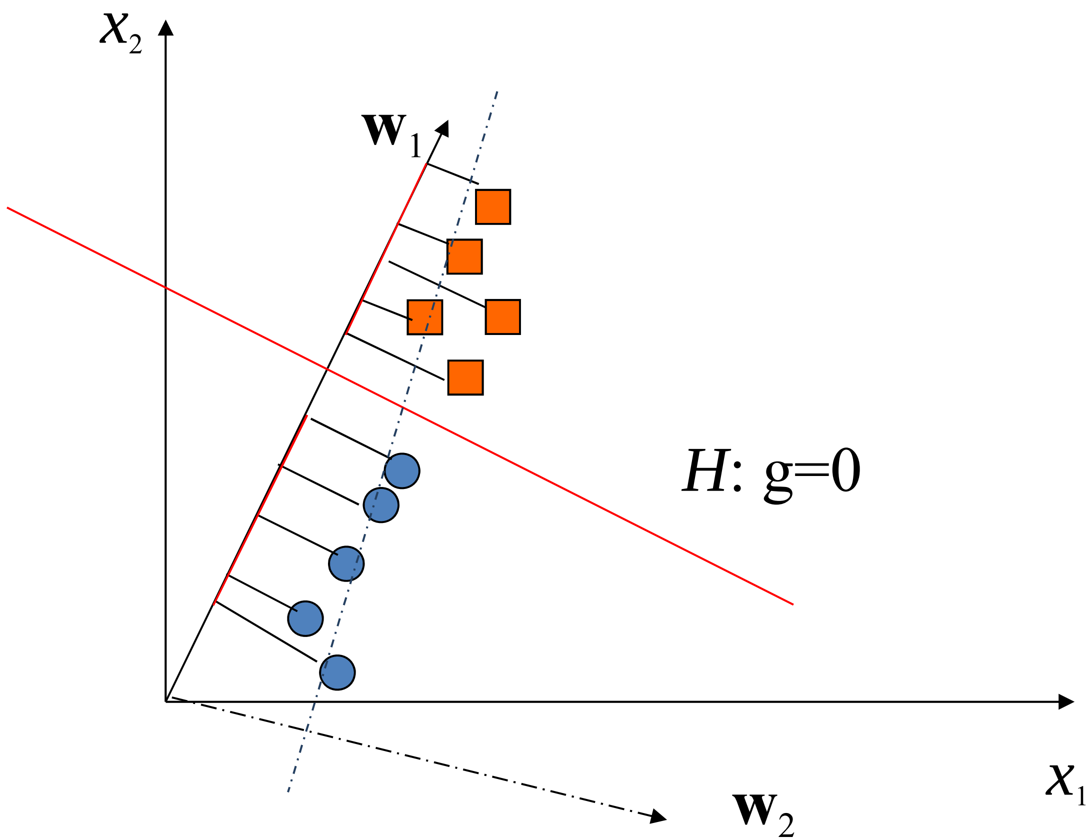
广义线性判别函数

4.2 Fisher线性判别

- 线性判别函数$y=g(\mathbf{x})=\mathbf{w}^T\mathbf{x}$:
  - 样本向量**x**各分量的线性加权
  - 样本向量**x**与权向量**w**的向量点积
  - 如果|| **w** ||=1，则视作向量**x**在**w**上的投影
- Fisher准则的基本原理：找到一个最合适的投影轴，使两类样本在该轴上投影之间的距离尽可能远，而每一类样本的投影尽可能紧凑，从而使分类效果为最佳。

Fisher线性判别图例

Fisher



Fisher准则的描述：用投影后数据的统计性质（均值和离散度的函数）作为判别优劣的标准。

d维空间样本分布的描述量

Fisher

- 各类样本均值向量$\mathbf{m}_i$ $\quad \mathbf{m}_i = \frac{1}{N_i}\sum_{x\in K_i}\mathbf{x} \quad i=1,2$
- 样本类内离散度矩阵$\mathbf{S}_i$与总类内离散度矩阵$\mathbf{S}_w$

$$\mathbf{S}_i = \sum_{\mathbf{x}\in\chi_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T, \quad i=1,2 \qquad \mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2$$

- 样本类间离散度矩阵$\mathbf{S}_b$： $\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$

离散度矩阵在形式上与协方差矩阵很相似，但协方差矩阵是一种总体期望值，而离散矩阵只是表示有限个样本在空间分布的离散程度

一维Y空间样本分布的描述量

Fisher

- 各类样本均值 $\tilde{m}_i = \frac{1}{N_i}\sum_{y\in\psi_i} y, \quad i=1,2$
- 样本类内离散度和总类内离散度

$$\tilde{S}_i = \sum_{y\in\psi_i}(y-\tilde{m}_i)^2, \quad i=1,2 \qquad \tilde{S}_w = \tilde{S}_1 + \tilde{S}_2$$

- 样本类间离散度 $\tilde{S}_b = (\tilde{m}_1 - \tilde{m}_2)^2$

以上定义描述*d*维空间样本点到一向量投影后的**分散情况**。样本离散度的定义与随机变量方差相类似

样本与其投影统计量间的关系

Fisher

◆样本$\mathbf{x}$与其投影y的统计量之间的关系：

$$\tilde{m}_i = \frac{1}{N_i}\sum_{y\in\psi_i} y = \frac{1}{N_i}\sum_{\mathbf{y}\in K_i} \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{m}_i, \quad i = 1,2$$

$$\tilde{S}_b = (\tilde{m}_1 - \tilde{m}_2)^2 = (\mathbf{w}^T\mathbf{m}_1 - \mathbf{w}^T\mathbf{m}_2)^2$$

$$= \mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} = \mathbf{w}^T S_b \mathbf{w}$$

样本与其投影统计量间的关系(2)

Fisher

$$\tilde{S}_i = \sum_{y\in\psi_i}(y-\tilde{m}_i)^2$$

$$= \sum_{x\in K_i}(\mathbf{w}^T\mathbf{x}-\mathbf{w}^T\mathbf{m}_i)^2$$

$$= \mathbf{w}^T\left[\sum_{x\in K_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T\right]\mathbf{w}$$

$$= \mathbf{w}^T S_i \mathbf{w}$$

$$\tilde{S}_w = \tilde{S}_1 + \tilde{S}_1 = \mathbf{w}^T(S_1+S_2)\mathbf{w} = \mathbf{w}^T S_w \mathbf{w}$$

Fisher准则函数

Fisher

- 评价投影方向**w**的原则，使原样本向量在该方向上的投影能兼顾类间分布尽可能分开，类内尽可能密集的要求
- Fisher准则函数的定义：

$$J_F(w) = \frac{\tilde{S}_b}{\tilde{S}_1 + \tilde{S}_2} = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}$$

- Fisher最佳投影方向的求解

$$\mathbf{w}^* = \underset{\mathbf{w}}{\arg\max}\, J_F(\mathbf{w})$$

Fisher最佳投影方向的求解

Fisher

◆采用拉格朗日乘子算法解决

$$\mathbf{w}^* = S_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

$\mathbf{m}_1\text{-}\mathbf{m}_2$是一向量，对与($\mathbf{m}_1\text{-}\mathbf{m}_2$)平行的向量投影可使两均值点的距离最远。但是如果从使类间分得较开，同时又使类内密集程度较高这样一个综合指标来看，则需根据两类样本的分布离散程度对投影方向作相应的调整，这就体现在对$\mathbf{m}_1\text{-}\mathbf{m}_2$ 向量按S_w^{-1}作一线性变换，从而使Fisher准则函数达到极值点

判别函数的确定

Fisher

- 前面讨论了使Fisher准则函数极大的*d*维向量**w***的计算方法，判别函数中的另一项w_0（阈值）可采用以下几种方法确定：

$$w_0 = -\frac{\tilde{m}_1 + \tilde{m}_2}{2} \qquad w_0 = -\frac{N_1\tilde{m}_1 + N_2\tilde{m}_2}{N_1 + N_2} = \tilde{m}$$

$$w_0 = \frac{\tilde{m}_1 + \tilde{m}_2}{2} - \frac{\ln\left[P(\omega_1)/P(\omega_2)\right]}{N_1 + N_2 - 2}$$

- 分类规则:

$$y = \mathbf{w}^T\mathbf{x} + w_0 > 0 \rightarrow \mathbf{x} \in \omega_1$$

$$y = \mathbf{w}^T\mathbf{x} - w_0 < 0 \rightarrow \mathbf{x} \in \omega_2$$

Fisher公式的推导

Fisher

$$J_F(w)=\frac{\tilde{S}_b}{\tilde{S}_1+\tilde{S}_2}=\frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}}$$

令 $\mathbf{w}^T S_w \mathbf{w}=\mathrm{c}\neq 0$

定义Lagrange函数: $L(\mathbf{w},\lambda)=\mathbf{w}^T S_b \mathbf{w}-\lambda(\mathbf{w}^T S_w \mathbf{w}-c)$

令: $$\frac{\partial L(\mathbf{w},\lambda)}{\partial \mathbf{w}}=S_b\mathbf{w}-\lambda S_w\mathbf{w}=0$$

$$S_w^{-1}S_b\mathbf{w}=\lambda\mathbf{w}$$

$$\lambda\mathbf{w}=S_w^{-1}S_b\mathbf{w}=S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)(\mathbf{m}_1-\mathbf{m}_2)^T\mathbf{w}$$
$$=S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)R$$

$$\mathbf{w}^*=\frac{R}{\lambda}S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)\simeq S_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)$$

Fisher准则举例

Fisher

例1：设两类样本的类内离散矩阵分别为$\mathbf{S}_1,\mathbf{S}_2$,各类样本均值分别为$\mathbf{m}_1=(2, 0)^t$, $\mathbf{m}_2=(2, 2)^t$, 试用Fisher准则求其决策面方程。

$$\mathbf{S}_1=\begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{bmatrix}$$

$$\mathbf{S}_2=\begin{bmatrix}1 & -\frac{1}{2}\\ -\frac{1}{2} & 1\end{bmatrix}$$

答：

$$\mathbf{S}_W=\mathbf{S}_1+\mathbf{S}_2=\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix}$$

Fisher准则最佳投影

$$\mathbf{w}^*=\mathbf{S}_w^{-1}(\mathbf{m}_1-\mathbf{m}_2)=\begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}\begin{bmatrix}0\\ -2\end{bmatrix}=\begin{bmatrix}0\\ -1\end{bmatrix}$$

由于两类样本分布形状是相同的（只是方向不同），因此$-w_0$应为（投影后）两类均值的中点

$$w_0=-\frac{\tilde{m}_1+\tilde{m}_2}{2}=-\frac{\mathbf{w}^{*T}(\mathbf{m}_1+\mathbf{m}_2)}{2}=1$$

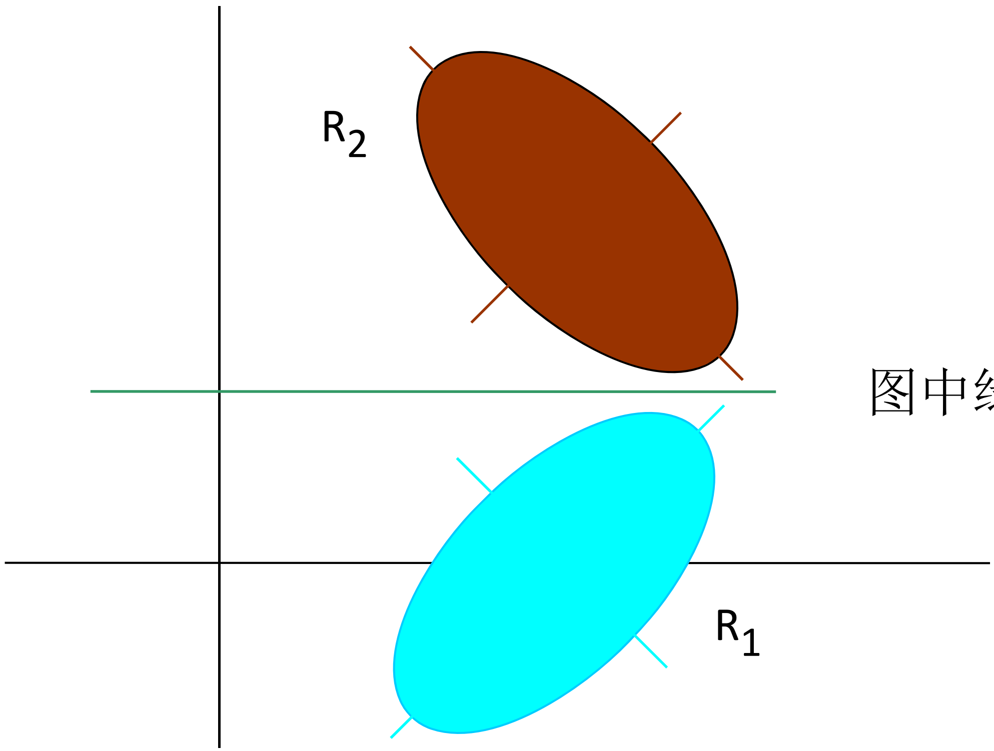
$$\mathbf{w}^{*T}\mathbf{x}+w_0=0$$

即：$x_2=1$

Fisher准则最佳分界面

Fisher最佳线性分界面

Fisher



图中绿线为最佳分界面

4.3 感知器准则

◆感知器准则是五十年代由Rosenblatt提出的一种自学习判别函数生成方法，由于Rosenblatt企图将其用于脑模型感知器(Perceptron)，因此被称为感知准则函数。其特点是随意确定的判别函数初始值，在对样本分类训练过程中逐步修正直至最终确定。

基本概念

感知器

- 感知器：Perceptron，Rosenblatt，50d/20thc
- 线性可分性：训练样本集中的两类样本在特征空间可以用一个线性分界面正确无误地分开。在线性可分条件下，对合适的(广义)权向量a应有：

$$\text{如果}\quad \mathbf{y}\in\omega_1,\ \text{则}\ \mathbf{a}^T\mathbf{y}>0$$

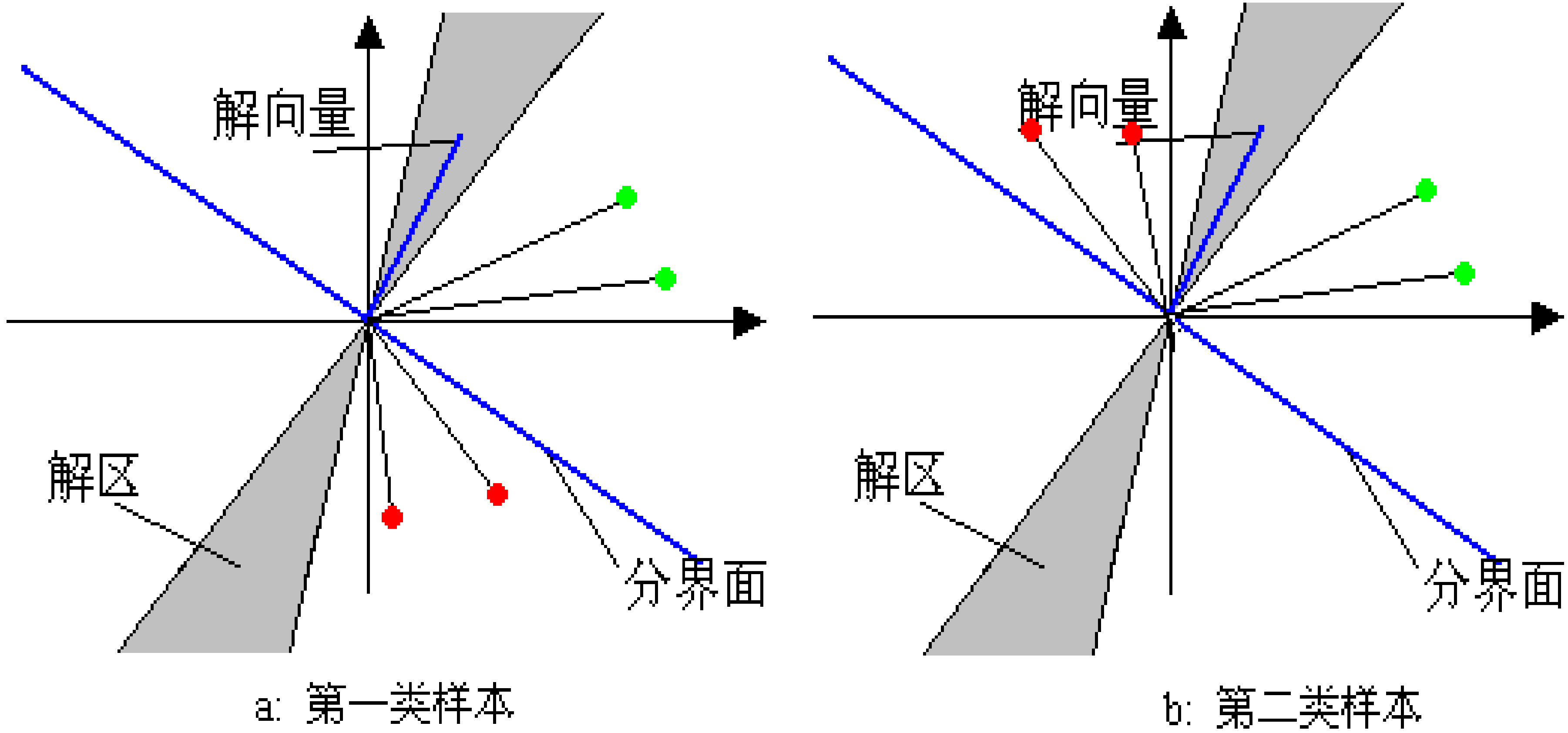
$$\text{如果}\quad \mathbf{y}\in\omega_2,\ \text{则}\ \mathbf{a}^T\mathbf{y}<0$$

- 规范化样本向量：将第二类样本取其反向向量

$$\mathbf{y}'=\begin{cases}\mathbf{y} & \text{如果}\ \mathbf{y}\in\omega_1\\ -\mathbf{y} & \text{如果}\ \mathbf{y}\in\omega_2\end{cases} \Rightarrow \mathbf{a}^T\mathbf{y}_i'>0\quad i=1,\ldots,N$$

解向量与解区

感知器



(a) 未规范化

(b) 规范化

感知器准则函数

- 对于任何一个增广权向量**a**，
  - 对样本**y**正确分类，则有：$\mathbf{a}^T\mathbf{y}>0$
  - 对样本**y**错误分类，则有：$\mathbf{a}^T\mathbf{y}<0$
- 定义一准则函数$J_P(\mathbf{a})$ (感知准则函数)：

$$J_P(\mathbf{a}) = \sum_{\mathbf{y}\in Y^k} (-\mathbf{a}^T\mathbf{y})$$

被错分类的规范化增广样本集

- 恒有$J_P(\mathbf{a})\geq 0$，且仅当**a**为解向量，Y^k为空集（不存在错分样本）时，$J_P(\mathbf{a})=0$，即达到极小值。确定向量**a**的问题变为对$J_P(\mathbf{a})$求极小值的问题。

梯度下降算法

- 梯度下降算法：对(迭代)向量沿某函数的负梯度方向修正，可较快到达该函数极小值。

$$\nabla J_p(\mathbf{a}) = \frac{\delta J_p(\mathbf{a})}{\delta \mathbf{a}} = \sum_{\mathbf{y} \in Y^k} (-\mathbf{y})$$

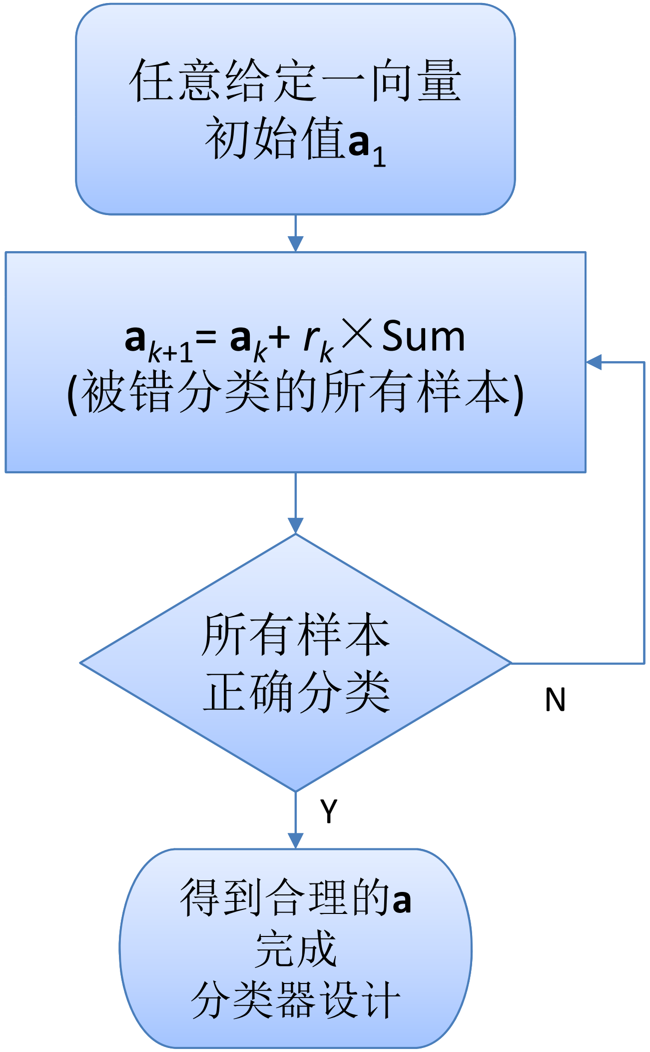
$$\mathbf{a}_{k+1} = \mathbf{a}_k - r_k \nabla J_p(\mathbf{a}) = \mathbf{a}_k + r_k \sum_{\mathbf{y} \in Y_k} \mathbf{y}$$

Y_k是被$\mathbf{a}_k$错分类的样本集

算法(step by step)

感知器

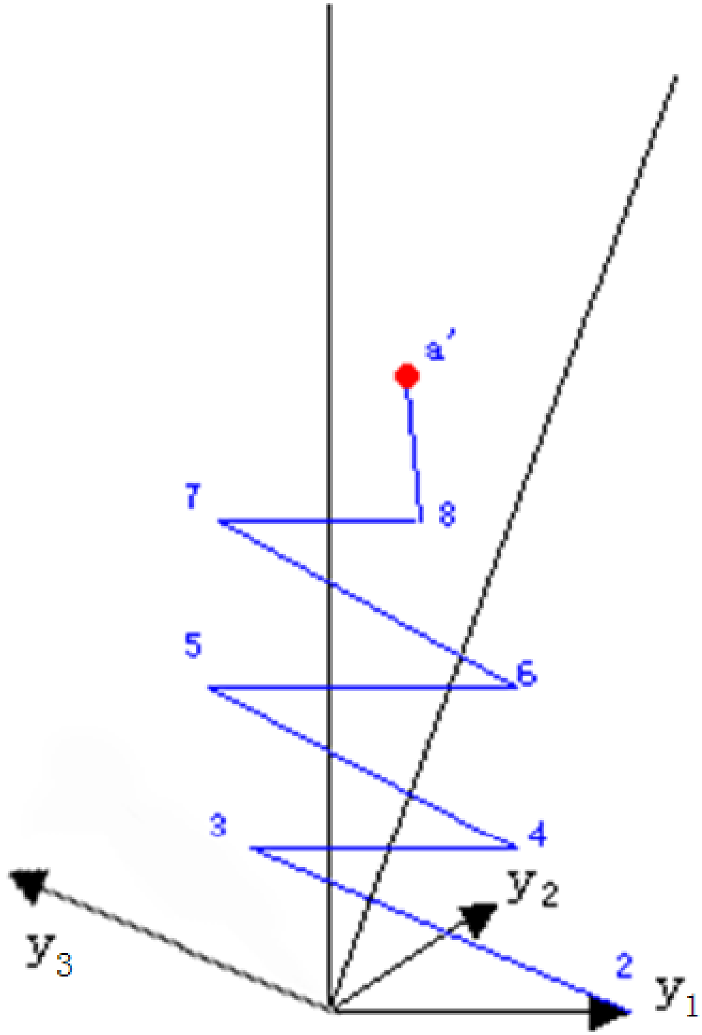
1. 初值：任意给定一向量初始值$\mathbf{a}_1$
2. 迭代：第k+1次迭代时的权向量$\mathbf{a}_{k+1}$等于第k次的权向量$\mathbf{a}_k$加上被错分类的所有样本之和与r_k的乘积
3. 终止：对所有样本正确分类



感知器方法例解

- 固定增量法与可变增量法
- 批量样本修正法与单样本修正法
  - 单样本修正法：样本集视为不断重复出现的序列，逐个样本检查，修正权向量
  - 批量样本修正法：样本成批或全部检查后，修正权向量

$$\mathbf{a}_{k+1} = \mathbf{a}_k + r_k \sum_{\mathbf{y} \in Y_k} \mathbf{y}$$



感知器方法小结

- 感知准则函数方法的思路是：先随意找一个初始向量$\mathbf{a}_1$，然后用训练样本集中的每个样本来计算。若发现一个$\mathbf{y}$出现$\mathbf{a}^T\mathbf{y}<0$，则只要$\mathbf{a}_{k+1} = \mathbf{a}_k + r_k\mathbf{y}$，$r_k$为正(步长系数)，则必有$\mathbf{a}_{k+1}{}^T\mathbf{y} = \mathbf{a}_k{}^T\mathbf{y} + r_k\mathbf{y}^T\mathbf{y}$，就有趋势做到使$\mathbf{a}_{k+1}{}^T\mathbf{y} >0$。当然，修改后的$\mathbf{a}_{k+1}$还可以使某些$\mathbf{y}$出现$\mathbf{a}_{k+1}{}^T\mathbf{y} <0$的情况，理论证明，只要训练样本集线性可分，无论$\mathbf{a}_1$的初值是什么，经过有限次叠代，都可收敛。

4.4 最小平方误差准则

- 规范化增广样本向量$\mathbf{y}_i$，增广权向量$\mathbf{a}$，正确分类要求：$\mathbf{a}^T\mathbf{y}_i>0,\ i=1,\ldots,N$
- 线性分类器设计⇔求一组N个线性不等式的解$\mathbf{a}^*$
- 样本集增广矩阵Y及一组N个线性不等式的的矩阵表示：

$$Y=\begin{bmatrix}\mathbf{y}_1^T\\ \mathbf{y}_1^T\\ \cdots\\ \mathbf{y}_N^T\end{bmatrix}=\begin{bmatrix}y_{11} & y_{12} & \cdots & y_{1\hat{d}}\\ y_{21} & y_{22} & \cdots & y_{2\hat{d}}\\ \cdots & \cdots & \cdots & \cdots\\ y_{N1} & y_{N2} & \cdots & y_{N\hat{d}}\end{bmatrix}$$

$$Y\mathbf{a}>0$$

- 引入余量(目标向量) $\mathbf{b}=[b_1, b_2, \ldots, b_N]^T$， b_i为任意给定正常数， $\mathbf{a}^T\mathbf{y}_i = b_i >0$
- N个线性方程的的矩阵表示：

$$Y\mathbf{a}=\mathbf{b}$$

矛盾方程组，没有精确解

平方误差准则函数

MSE

- 定义误差向量 **e**=*Y***a**-**b**：
- 定义平方误差准则函数$J_s(\mathbf{a})$:

$$J_s(\mathbf{a}) = \|\mathbf{e}\|^2 = \|Y\mathbf{a}-\mathbf{b}\|^2 = \sum_{i=1}^{N}(\mathbf{a}^T\mathbf{y}_i - b_i)^2$$

- 最小二乘近似解（MSE解）：

$$\mathbf{a}^* = \underset{\mathbf{a}}{\arg\min}\, J_s(\mathbf{a})$$

MSE方法的思想：对每个样本，设定一个“理想”的判别函数输出值，以最小平方误差为准则求最优权向量

MSE准则函数的伪逆解

MSE

$$\nabla J_s(\mathbf{a}) = \sum_{i=1}^{N} 2(\mathbf{a}^T\mathbf{y}_i - b_i)\mathbf{y}_i = 2Y^T(Y\mathbf{a} - \mathbf{b})$$

$$\nabla J_s(\mathbf{a}^*) = 0 \Leftrightarrow Y^T Y\mathbf{a}^* = Y^T\mathbf{b}$$

$$\mathbf{a}^* = (Y^T Y)^{-1} Y^T \mathbf{b} = Y^+\mathbf{b}$$

*Y*的伪逆矩阵

MSE方法的迭代解

MSE

- $\mathbf{a}^*=Y^+\mathbf{b}$, $Y^+=(Y^TY)^{-1}Y^T$，计算量大
- 实际中常用梯度下降法：

$$\nabla J_s(\mathbf{a})=\sum_{i=1}^{N}2(\mathbf{a}^T\mathbf{y}_i-b_i)\mathbf{y}_i=2Y^T(Y\mathbf{a}-\mathbf{b})$$

$$\begin{cases}\mathbf{a}_1,\text{任意初始化}\\ \mathbf{a}_{k+1}=\mathbf{a}_k-r_kY^T(Y\mathbf{a}-\mathbf{b})\end{cases}$$

批量样本修正法

$$\begin{cases}\mathbf{a}_1,\text{任意初始化}\\ \mathbf{a}_{k+1}=\mathbf{a}_k+r_k(b_k-\mathbf{a}_k^T\mathbf{y}_k)\mathbf{y}_k\end{cases}$$

单样本修正法

$$r_k=\frac{r_1}{k}$$

Widrow-Hoff

MSE方法与Fisher方法的关系

MSE

- 与Fisher方法的关系：当

$$\mathbf{b} = \begin{array}{l} \left[\begin{array}{c} N/N_1 \\ \cdots \\ N/N_1 \\ N/N_2 \\ \cdots \\ N/N_2 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} N \\ \cdots \\ N \end{array}}\right\} N_1\text{个} \\ \left.\vphantom{\begin{array}{c} N \\ \cdots \\ N \end{array}}\right\} N_2\text{个} \end{array} \end{array}$$

MSE解等价于Fisher解

MSE方法与Bayes方法的关系

MSE

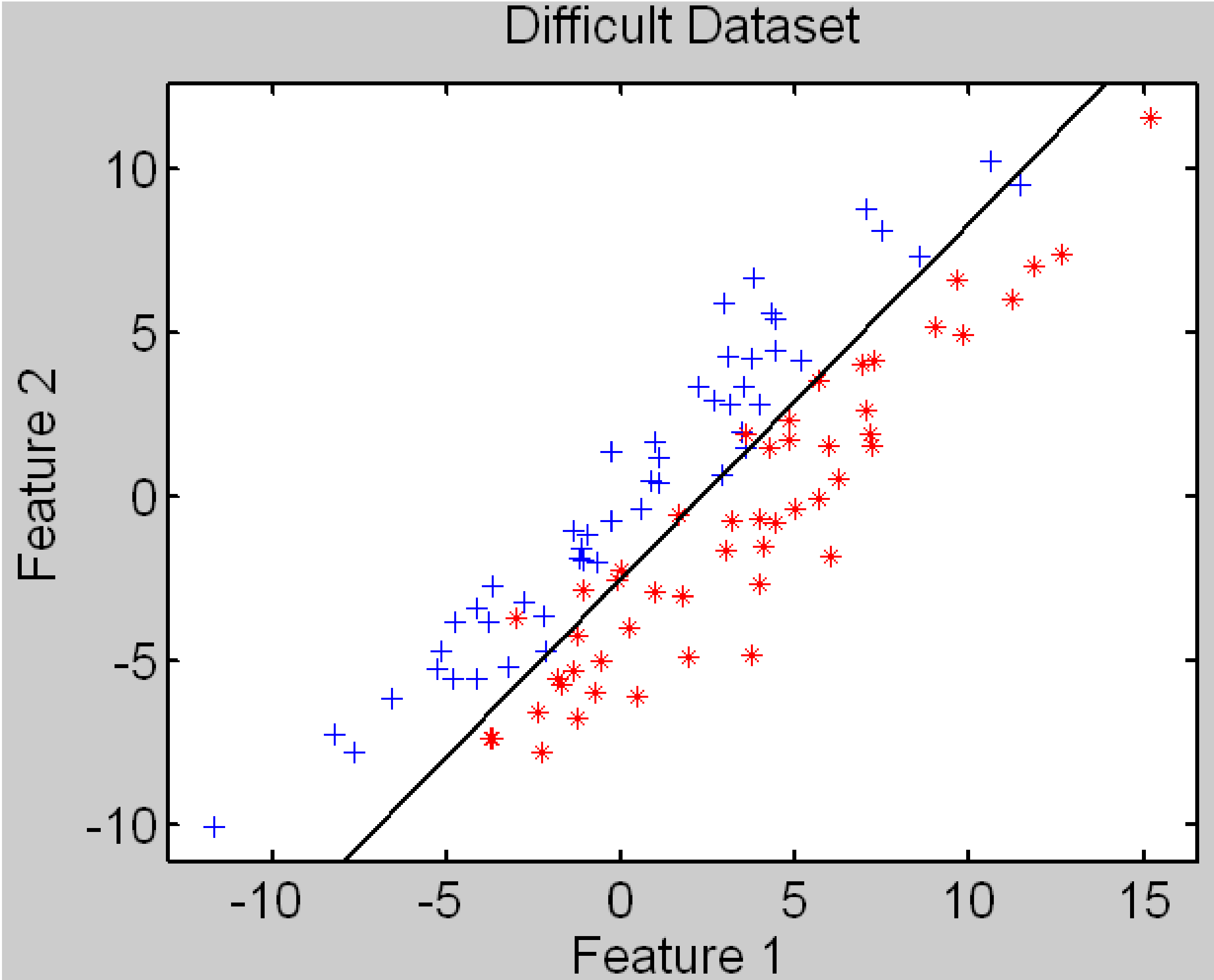
- 当N→∞，$\mathbf{b}=\mathbf{u}_N=[1,1,\ldots,1]^T$时，则它以最小均方误差逼近Bayes判别函数：

$$g(\mathbf{x})=P(\omega_1\mid\mathbf{x})-P(\omega_2\mid\mathbf{x})$$

$$\mathbf{a}^*=\arg\min_{\mathbf{a}} J_s(\mathbf{a})\Leftrightarrow\arg\min_{\mathbf{a}}\int[\mathbf{a}^T\mathbf{y}-g(\mathbf{x})]^2p(\mathbf{x})dx$$

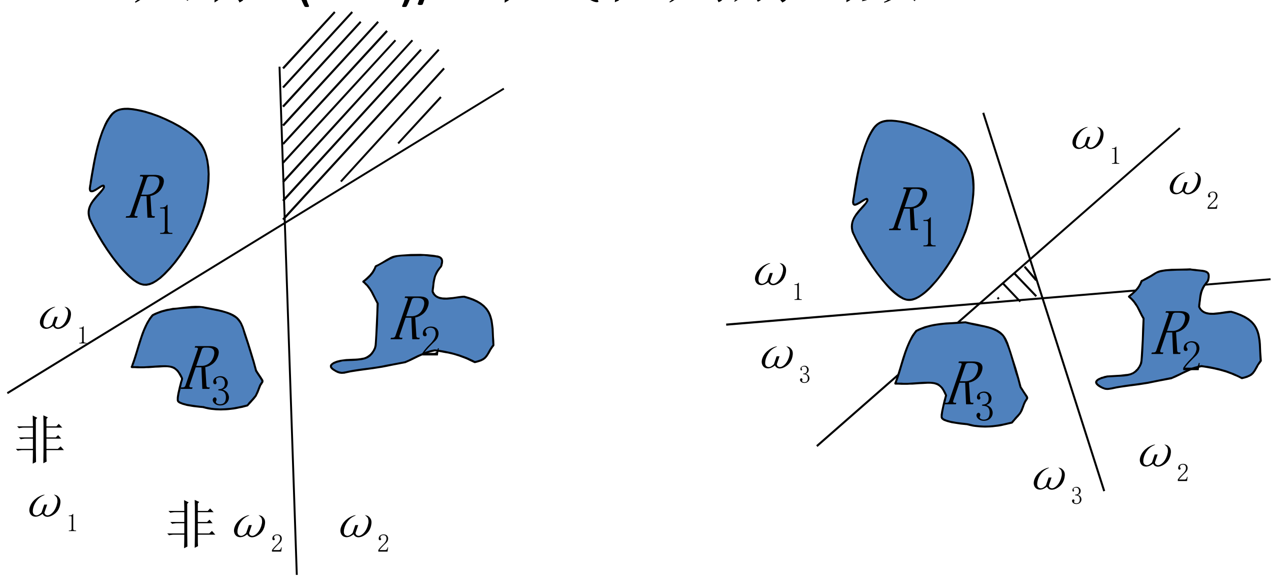
MSE方法应用举例

MSE



4.5 多类问题

- 两类别问题可以推广到多类别问题
  - $\omega_i/\sim\omega_i$ 法：将*C*类别问题化为(*C*-1)个两类（第*i*类与所有非*i*类）问题，按两类问题确定其判别函数与决策面方程 。
  - ω_i/ω_j 法：将*C*类中的每两类别单独设计其线性判别函数，因此总共有*C*(*C*-1)/2个线性判别函数。



多类线性判别函数

多类问题

- 将特征空间确实划分为c个决策域，共有c个判别函数

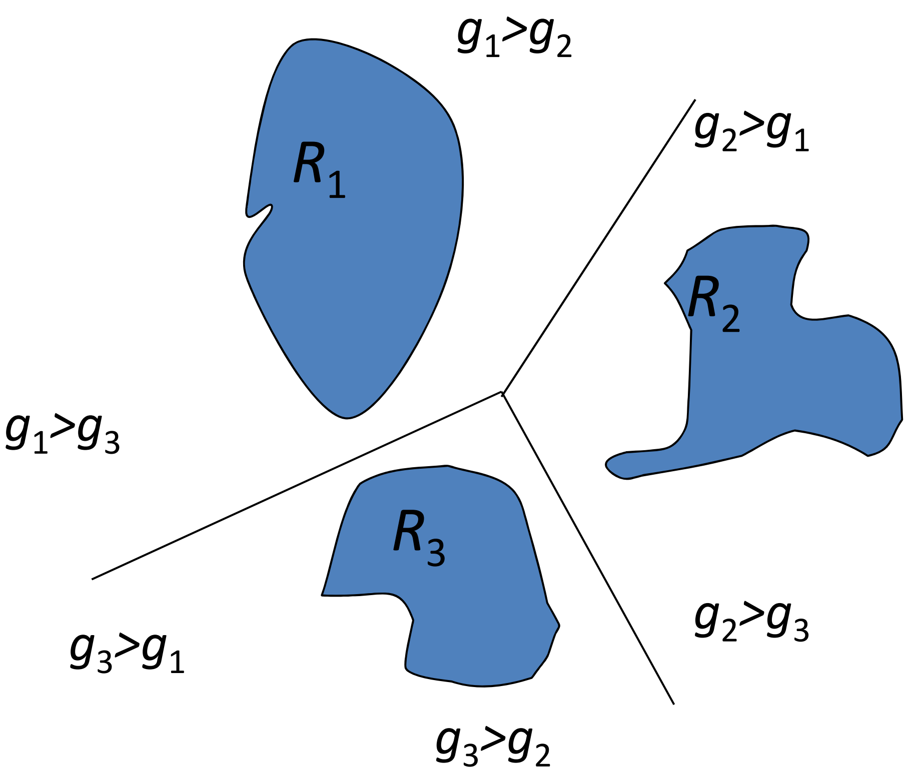
$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}, \quad i = 1, \ldots, c$$

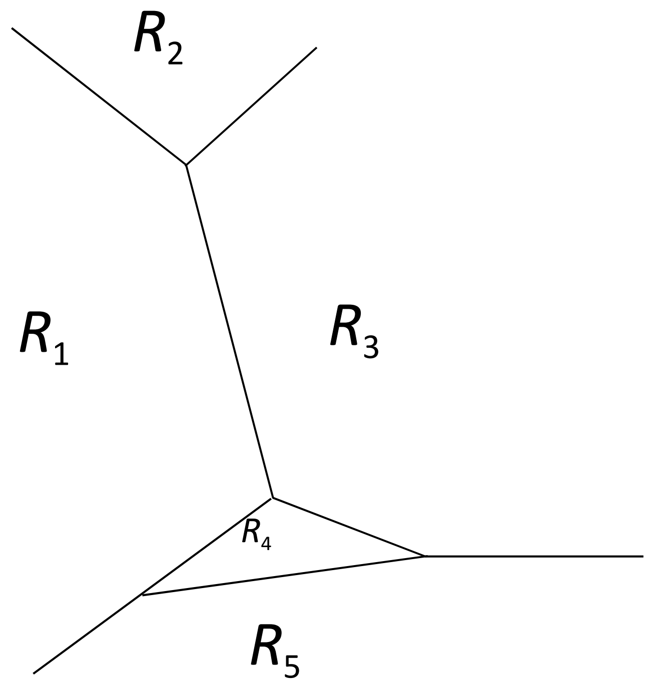
- 决策规则： $j = \underset{i}{\operatorname{argmax}}\, g_i(\mathbf{x}), \quad i = 1, \ldots, c$

- 决策域的边界由相邻决策域的判别函数共同决定，此时应有$g_i(\mathbf{x})=g_j(\mathbf{x})$
- 线性分类器的决策面是凸的，决策区域是单连通的
- 多类分类器的分界面是分段线性的

多类线性决策面图例

多类
问题

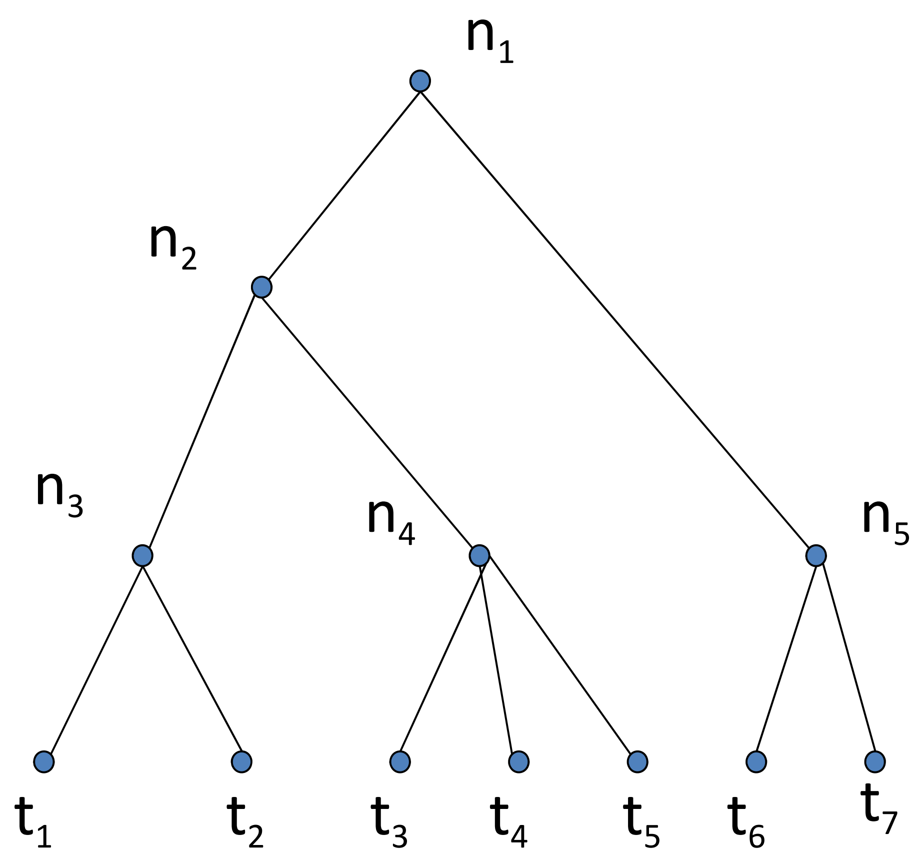




决策树简介

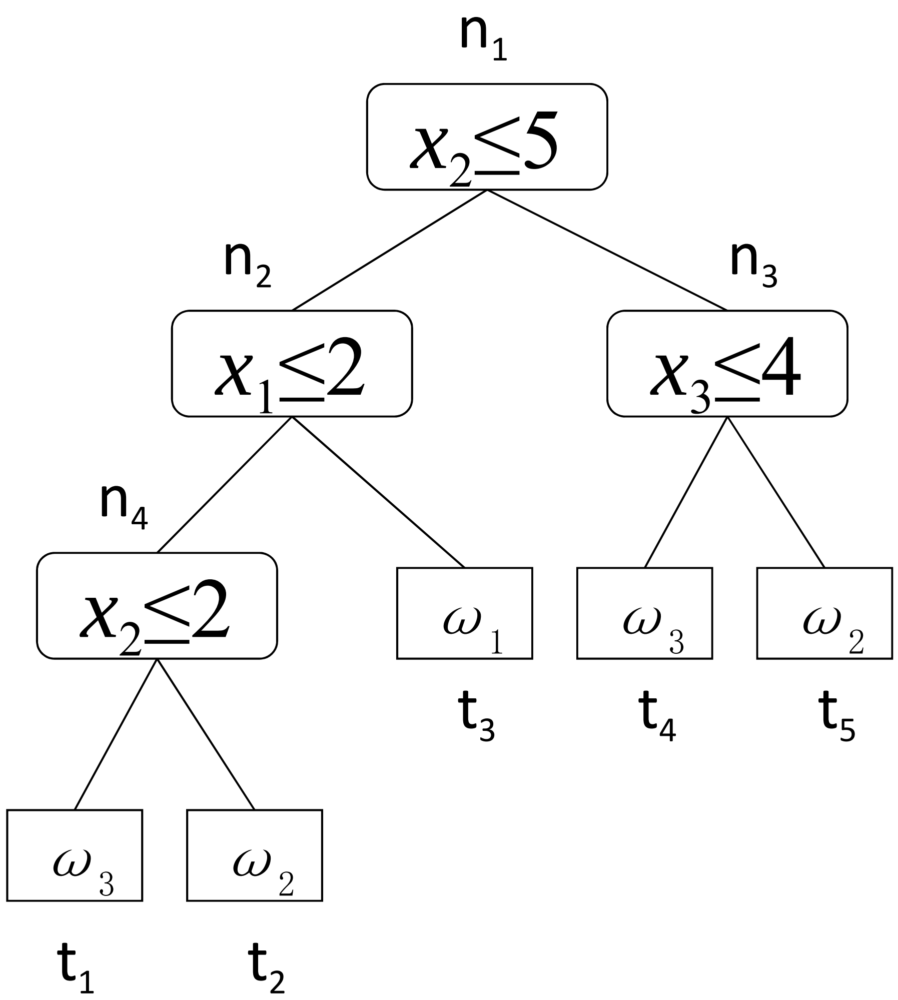
多类问题

◆决策树：一种多级分类器，它采用分级的形式，综合用多个决策规则，逐步把复杂的多类别分类问题转化为若干个简单的分类问题来解决



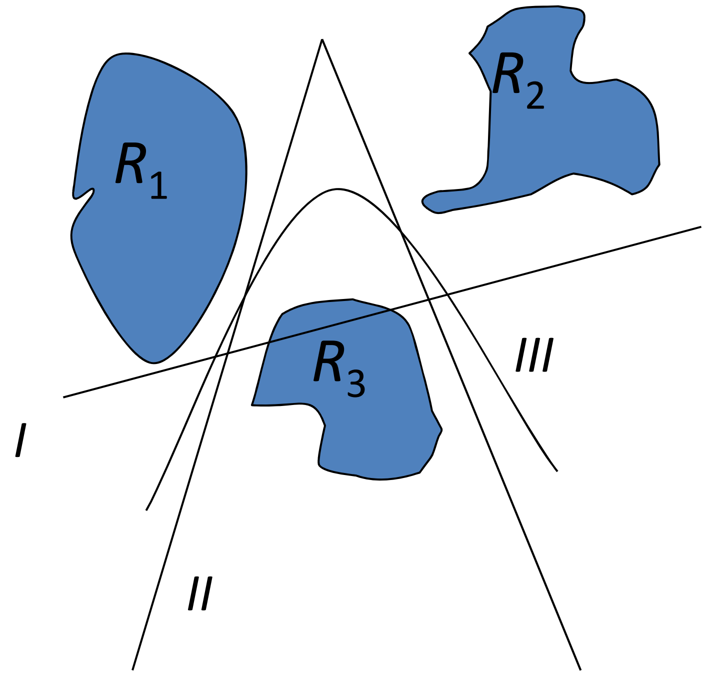
二叉决策树

二叉决策树：除叶节点外，决策树的每个节点n_i都有且只有两个子节点n_{il}和n_{ir}。二叉决策树把复杂的多类别分类问题转化为多级两类分类问题来解决。在每个节点n_i，都把样本集分成两个子集。每个子集可能仍包含多类别的样本，继续分直至仅包含单类别样本的叶节点



4.6 分段线性判别函数

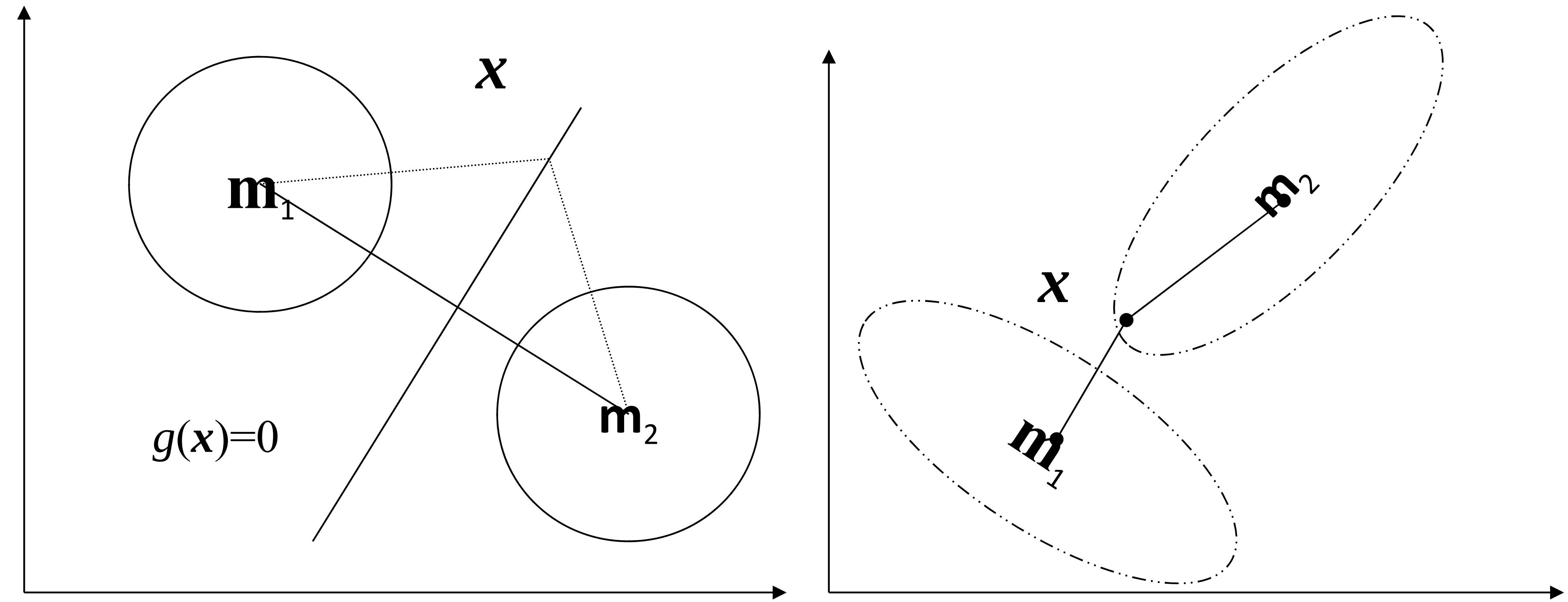
- 有些复杂模式识别问题不是线性可分的，需使用非线性的分类方法
- 分段线性判别函数：一种特殊的非线性判别函数，它的决策面是若干超平面
- 树分类器的各节点上采用线性判别规则，即构成分段线性分类器



I: 线性判别
II: 分段线性判别
III: 二次判别

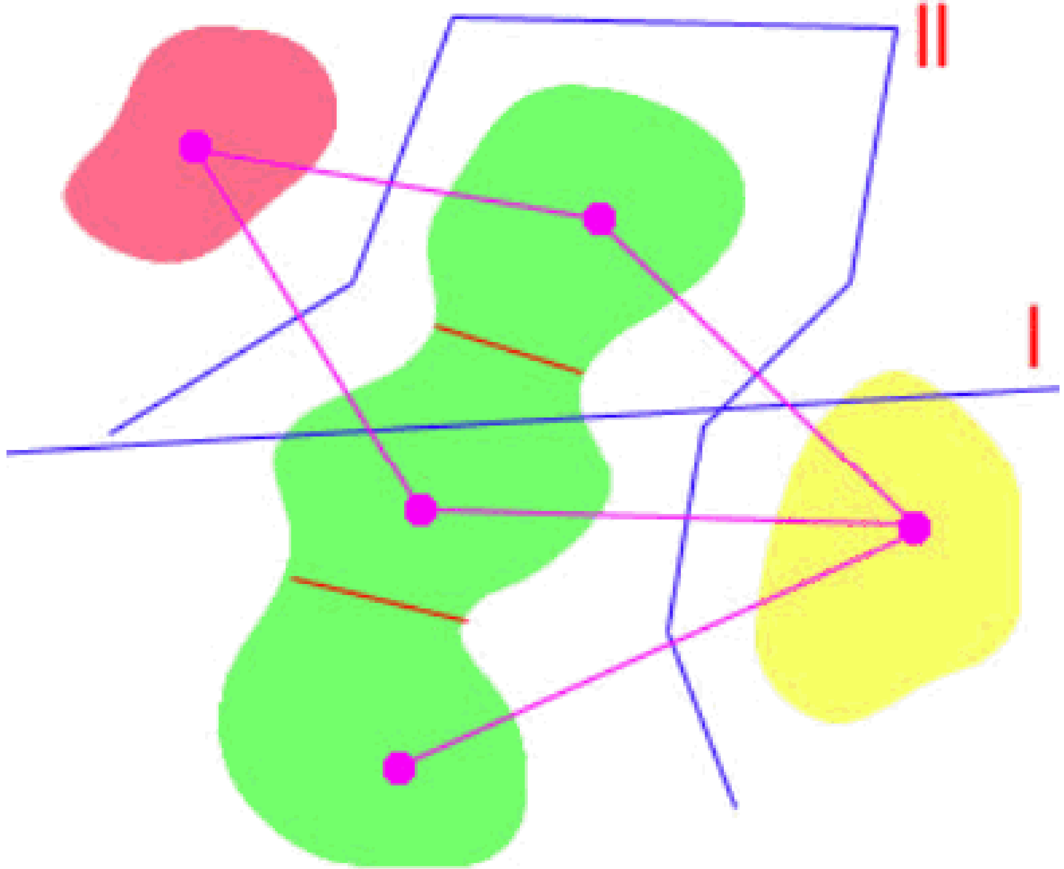
分段线性距离分类器

- 最小距离分类器：把各类别样本的均值向量作为各类的代表点(prototype)，根据待识样本到各类别代表点的最小距离判别其类别。决策面是两类别均值连线的垂直平分面。



分段线性距离分类器(2)

- 分段线性距离分类器：将各类别划分成相对密集的子类，每个子类以它们的均值作为代表点，然后按最小距离分类。



I： 线性距离判别

II： 分段线性距离判别

基于距离的分段线性判别函数

分段线性

- 判别函数定义：ω_i有l_i个子类，即属于ω_i的决策域R_i分成l_i个子域$R_i^1, R_i^2, \ldots, R_i^{l_i}$)，每个子区域用均值$m_i^k$代表点

$$g_i(\mathbf{x}) = \min_{k=1,\ldots,l_i} \left\| \mathbf{x} - \mathbf{m}_i^k \right\|$$

- 判别规则：
$$j = \operatorname*{argmin}_{i=1,\ldots,c} g_i(\mathbf{x})$$

or
$$\text{if } \; g_j(\mathbf{x}) = \min_{i=1,\ldots,c} g_i(\mathbf{x}) \;\; \text{then} \;\; \mathbf{x} \in \omega_j$$

分段线性判别函数

- 分段线性判别函数的形式: $g_i^k(\mathbf{x})$表示第i类第k段线性判别函数，l_i为i类所具有的判别函数个数，$\mathbf{w}_i^k$与w_{i0}^k分别是第k段的权向量与阈值

$$g_i^k(\mathbf{x}) = \mathbf{w}_i^{(k)T}\mathbf{x} + w_{i0}^k,\ \ k = 1,2,\ldots,l_i; i = 1,\ldots,c$$

- 第i类的判别函数:

$$g_i(\mathbf{x}) = \max_{k=1,\ldots,l_i} g_i^k(\mathbf{x})$$

分段线性判别函数(II)

- 判别规则：

$$\text{if } g_j(\mathbf{x}) = \max_{i=1,\ldots,c} g_i(\mathbf{x}) \text{ then } \mathbf{x} \in \omega_j$$

- 决策面取决于相邻的决策域，如第*i*类的第*n*个子类与第*j*类的第*m*个子类相邻，则由它们共同决定的决策面方程为

$$g_i^n(\mathbf{x}) = g_j^m(\mathbf{x})$$

4.7 讨论

- 基于样本的直接确定判别函数方法主要包含两个步骤：
  1. 确定使用的判别函数类型或决策面方程类型，如线性分类器，分段线性分类器等
  2. 在选定函数类型的条件下，确定相应的参数，从而完成整个分类器设计
- 线性判别函数计算简单，在一定条件下能实现最优分类，经常是一种“有限合理”的选择
- 分段线性分类器可以实现更复杂的分类面

习题

1. 有一个三次判别函数：$z=g(x)=x^3+2x^2+3x+4$。试建立一映射$x\rightarrow\mathbf{y}$，使得z转化为$\mathbf{y}$的线性判别函数。
2. 证明决策面H：$\mathbf{w}^T\mathbf{x}+w_0=0$的系数向量$\mathbf{w}$是决策面H的法向量
3. Ex-4.15
4. 设五维空间的线性方程为$55x_1+68x_2+32x_3+16x_4+26x_5+10=0$，试求出其权向量与样本向量点积的表达式$\mathbf{w}^T\mathbf{x}+w_0=0$中的$\mathbf{w}$，$\mathbf{x}$以及增广权向量与增广样本向量形式$\mathbf{a}^T\mathbf{y}$中的$\mathbf{a}$与$\mathbf{y}$

习题（续）

5. 设在三维空间中一个类别分类问题拟采用二次曲面。如欲采用广义线性方程求解，试问其广义样本向量与广义权向量的表达式，其维数是多少？