模式识别

第3讲 判别函数分类法（I）
（线性判别准则）

2018~2019学年

内容安排

一、绪论、数学基础（第1讲）

二、聚类分析（第2讲）

三、判别函数分类法（几何分类法）（第3、4讲）

四、统计决策分类法（概率分类法）（第5、6讲）

五、特征提取与选择（第7讲）

六、模糊模式识别（第8讲）

七、神经网络模式识别（第9讲）

期末考试（平时作业：40%，期末考试：60%）

三、判别函数分类法

3.1 判别函数

3.2 线性判别函数

3.3 广义线性判别函数

3.4 线性判别函数的几何性质

3.5 Fisher线性判别

3.6 感知器算法

3.7 梯度法

3.8 最小平方误差算法

3.9 非线性判别函数

3.1　判别函数

- **统计模式识别**：按任务类型划分
  - 聚类分析（Clustering Analysis）——简称：聚类
    - 简单聚类方法：最大最小距离法
    - 层次聚类方法：分裂式、凝聚式
    - 动态聚类方法：C-均值，ISODATA
  - 判别分析（Discriminatory Analysis）——简称：分类
    - 几何分类法（判别函数分类法）：线性、分段线性、二次、支持向量机
    - 概率分类法（统计决策分类法）：判别式 (Discriminative)、生成式 (Generative)
    - 近邻分类法（几何分类法和概率分类法的一种融合方法）
- 所谓 **几何分类法**，是指在特征空间中，利用矢量空间的直观概念，使用代数方程方法，对模式进行分类。因此也被称为：**代数界面方程法**。
- 所谓 **概率分类法**，是指把模式视为随机变量的抽样，利用统计决策理论（贝叶斯决策理论）成熟的判决准则与方法，对模式样本进行分类。

1. 判别函数的定义

直接用来对模式进行分类的决策函数。

若分属于ω_1，ω_2的两类模式在空间中的分布区域，可以用一代数方程$d(X)=0$来划分，那么称$d(X)$为判别函数，或称决策函数。显然，这一方程表示的是*n*维空间的*(n-1)*维超曲面（或超平面）。

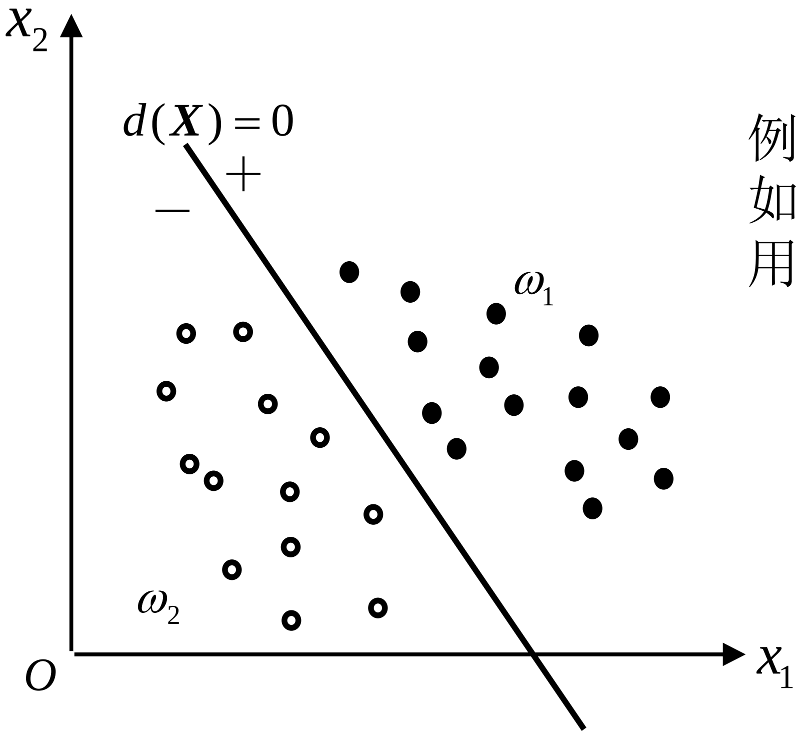


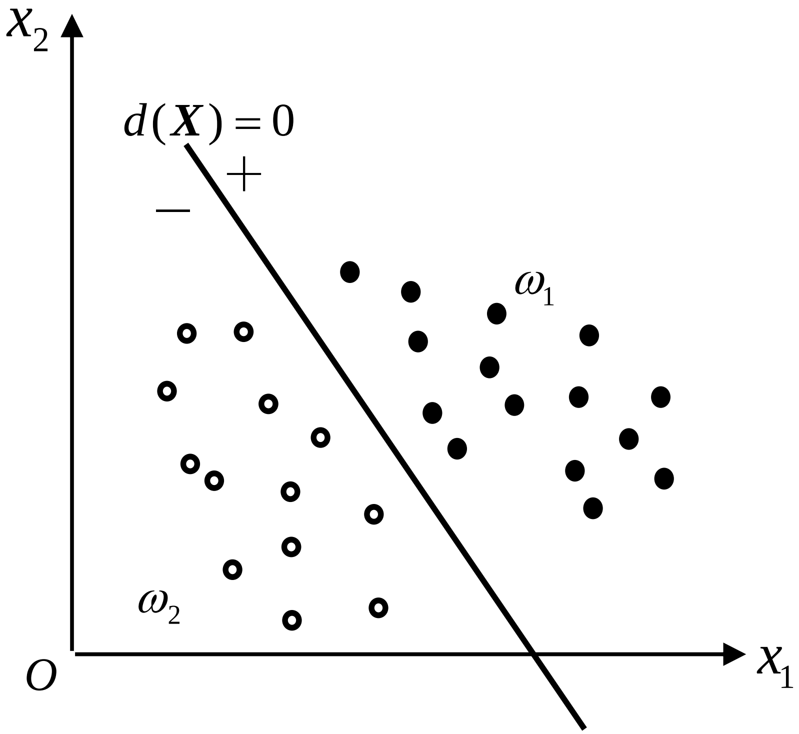
图3.2 两类二维模式的分布

例：一个二维的两类判别问题，模式分布如图示，这些分属于ω_1，ω_2两类的模式可用一直线方程$d(X)=0$来划分。

$$d(\boldsymbol{X})=w_1x_1+w_2x_2+w_3=0$$

式中：x_1,x_2为坐标变量，

w_1,w_2,w_3为方程参数。



示例：线性判别函数。

将某一未知模式 X 代入：

$$d(\boldsymbol{X}) = w_1x_1 + w_2x_2 + w_3$$

若$d(\boldsymbol{X})>0$，则 $\boldsymbol{X}\in\omega_1$类；

若$d(\boldsymbol{X})<0$，则 $\boldsymbol{X}\in\omega_2$类；

若$d(\boldsymbol{X})=0$，则$\boldsymbol{X}\in\omega_1$或$\boldsymbol{X}\in\omega_2$

或：拒绝分类

维数=3时：判别边界为一平面。
维数>3时：判别边界为一超平面。

注：为了清晰的了解*d(x)*的含义，应该画出**函数值这一轴**。

2．判别函数正负值的确定

判别界面的正负侧，是在训练判别函数的权值时**人为确定的**。
一般，令第1类样本的函数值大于零，第2类样本的函数值小于零。
——具体值的大小不重要（后面的学习算法会讲到）。

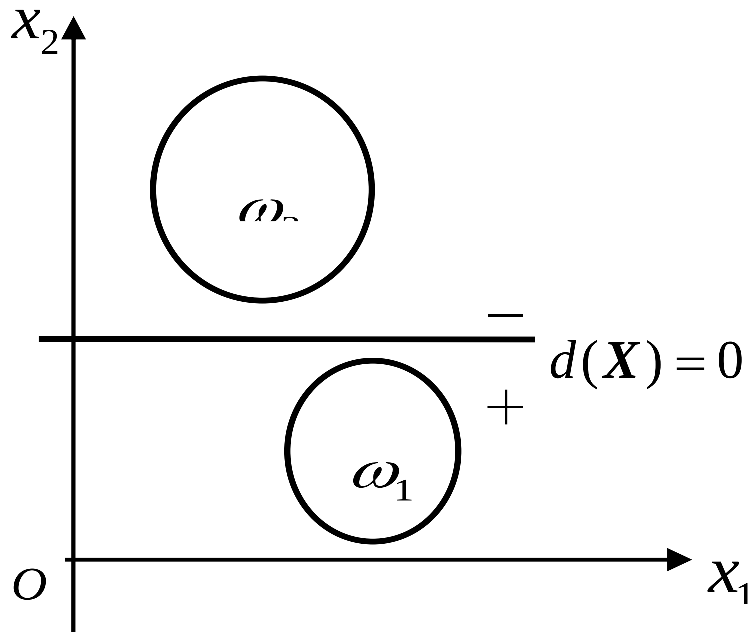


图3.3 判别函数正负的确定

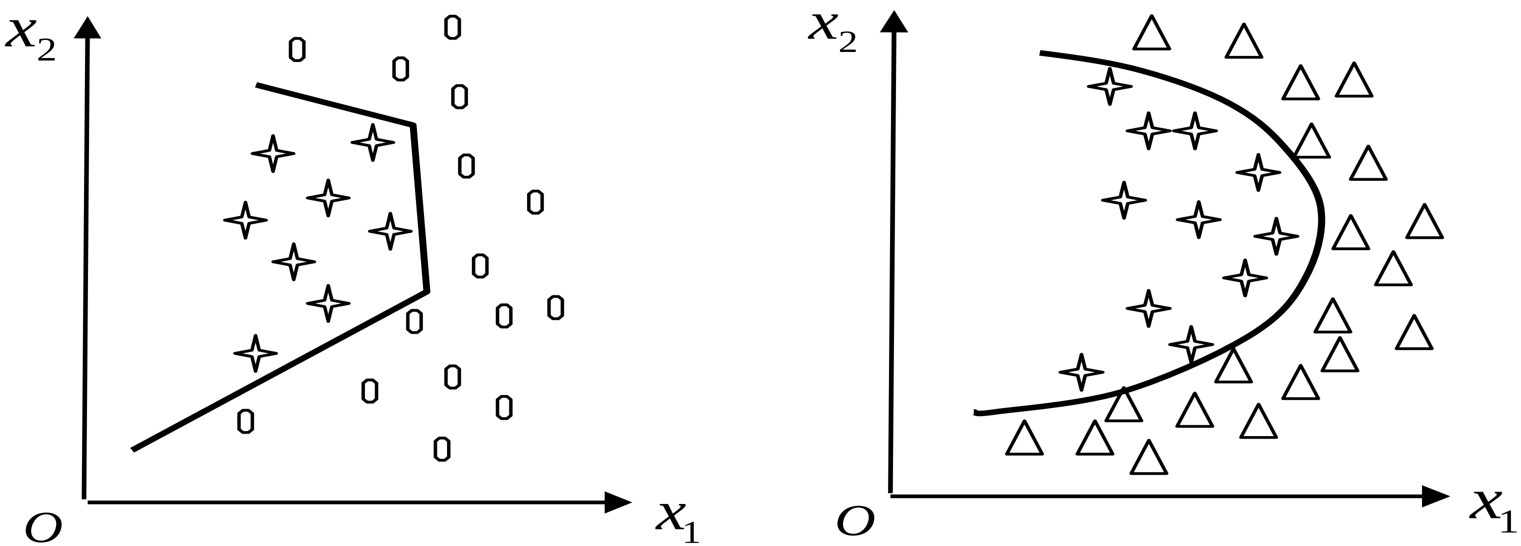
$d(\boldsymbol{X})$ 表示的是一种分类的标准，它可以是1、2、3维的，也可以是更高维的。

3. 确定判别函数的两个因素

1）判决函数$d(\boldsymbol{X})$的几何性质：它可以是线性的或非线性的函数，维数在特征提取时已经确定。

如：已知三维线性分类 —— 判决函数的性质就确定了判决函数的形式：$d(\boldsymbol{X}) = w_1x_1 + w_2x_2 + w_3x_3 + w_4$

例：非线性判决函数



2）判决函数$d(\boldsymbol{X})$的系数：用所给的模式样本确定。

3.2 线性判别函数

3.2.1 线性判别函数的一般形式

将二维模式推广到n维，线性判别函数的**一般形式**为：

$$d(\boldsymbol{X}) = w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_{n+1} = \boldsymbol{W}_0^{\mathrm{T}}\boldsymbol{X} + w_{n+1} \quad (3\text{-}2)$$

式中： $\boldsymbol{X} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$

$\boldsymbol{W}_0 = [w_1, w_2, \ldots, w_n]^{\mathrm{T}}$：权向量，即参数向量。

增广向量形式：

$$d(\boldsymbol{X}) = w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_{n+1} \cdot 1$$

$$= \begin{bmatrix} w_1 & w_2 & \cdots\cdots w_n & w_{n+1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}$$

式中：

$\boldsymbol{X} = [x_1, x_2, \ldots, x_n, 1]^{\mathrm{T}}$

为**增广模式向量**。

$\boldsymbol{W} = [w_1, w_2, \ldots, w_n, w_{n+1}]^{\mathrm{T}}$

为**增广权向量**，

3.2 线性判别函数

3.2.2 线性判别函数的性质

1. 两类情况

$$d(\boldsymbol{X})=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}\begin{cases}>0, & 若\boldsymbol{X}\in\omega_1\\<0, & 若\boldsymbol{X}\in\omega_2\end{cases}$$

$d(X)=0$：不可判别情况，可令 $\boldsymbol{X}\in\omega_1$ 或 $\boldsymbol{X}\in\omega_2$ 或拒绝分类

2. 多类情况

对M个线性可分模式类，ω_1， ω_2， … ω_M，有三种分类方式：

$\omega_i/\overline{\omega}_i$ 是非两分法

ω_i/ω_j 成对两分法

ω_i/ω_j 成对两分法特例（没有不确定区）

3.2 线性判别函数

（1）多类情况1: $\omega_i / \overline{\omega}_i$ 是非两分法

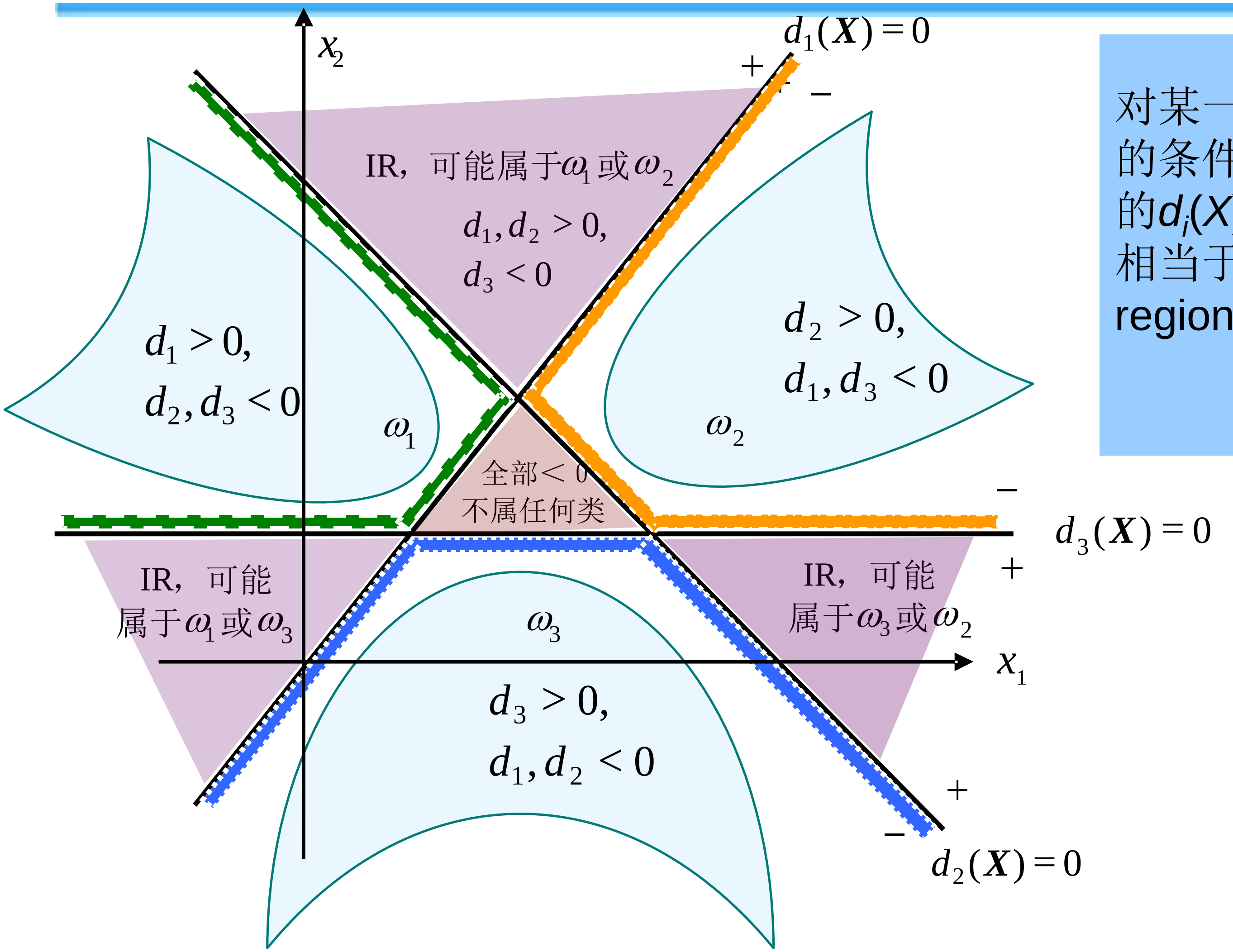
用线性判别函数将属于ω_i类的模式与其余不属于ω_i类的模式分开。能用本方法分类的模式集称为 **整体线性可分的**。

$$d_i(\boldsymbol{X}) = \boldsymbol{W}_i^{\mathrm{T}} \boldsymbol{X} \begin{cases} > 0, & 若\boldsymbol{X} \in \omega_i \\ < 0, & 若\boldsymbol{X} \in \overline{\omega}_i \end{cases} \qquad i = 1, \cdots, \quad M$$

识别分类时：

将某个待分类模式 $\boldsymbol{X}$ 分别代入 M 个类的$d(\boldsymbol{X})$中，若只有$d_i(\boldsymbol{X})>0$，其他$d_j(\boldsymbol{X})$均<0，则判为ω_i类。

3.2 线性判别函数



对某一模式区，$d_i(X)>0$的条件超过一个，或全部的$d_i(X)<0$，分类失效。相当于**不确定区**(indefinite region，**IR**)。

此法将 M 类问题分成M个两类问题，识别每一类均需M个判别函数。识别出所有的M类仍是这M个函数。

例3.1 设有一个二维三类问题，其三个判别函数为：

$$d_1(X) = -x_1 + x_2 + 1 \qquad d_2(\boldsymbol{X}) = x_1 + x_2 - 4 \qquad d_3(\boldsymbol{X}) = -x_2 + 1$$

现有一模式，***X***=[7,5]^T，试判定应属于哪类？并画出三类模式的分布区域。

解：将***X***=[7,5]^{T}代入上三式，有：

$$d_1(\boldsymbol{X}) = -7+5+1 = -1 < 0$$

$$d_2(\boldsymbol{X}) = 7+5-4 = 8 > 0$$

$$d_3(\boldsymbol{X}) = -5+1 = -4 < 0$$

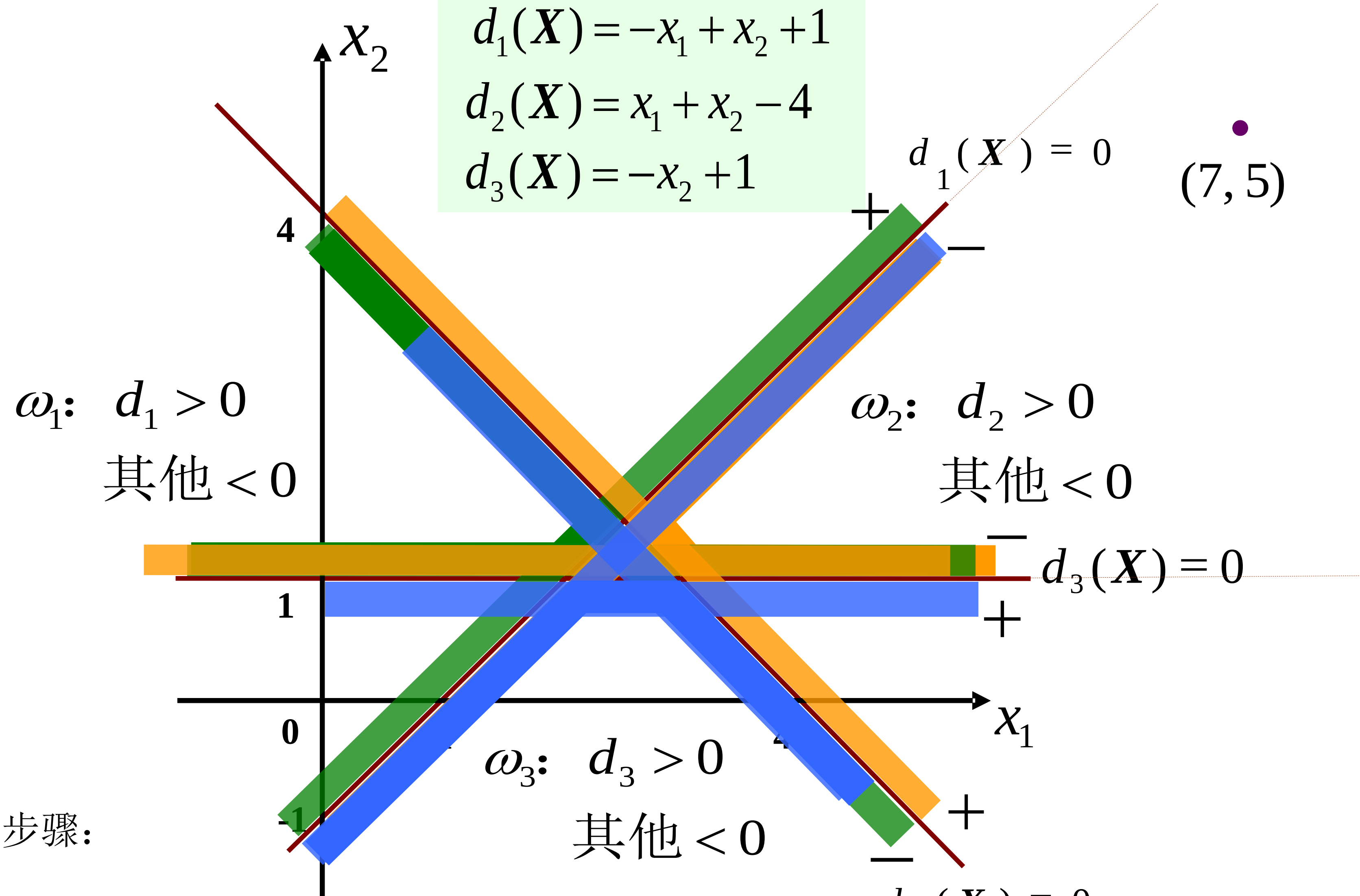
$$\because d_2(\boldsymbol{X}) > 0, \quad d_1(\boldsymbol{X}), d_3(\boldsymbol{X}) < 0 \quad \therefore \boldsymbol{X} \in \omega_2$$

三个判别界面分别为：

$$-x_1 + x_2 + 1 = 0 \qquad x_1 + x_2 - 4 = 0 \qquad -x_2 + 1 = 0$$

图示如下：

$$d_1(\boldsymbol{X}) = -x_1 + x_2 + 1$$
$$d_2(\boldsymbol{X}) = x_1 + x_2 - 4$$
$$d_3(\boldsymbol{X}) = -x_2 + 1$$



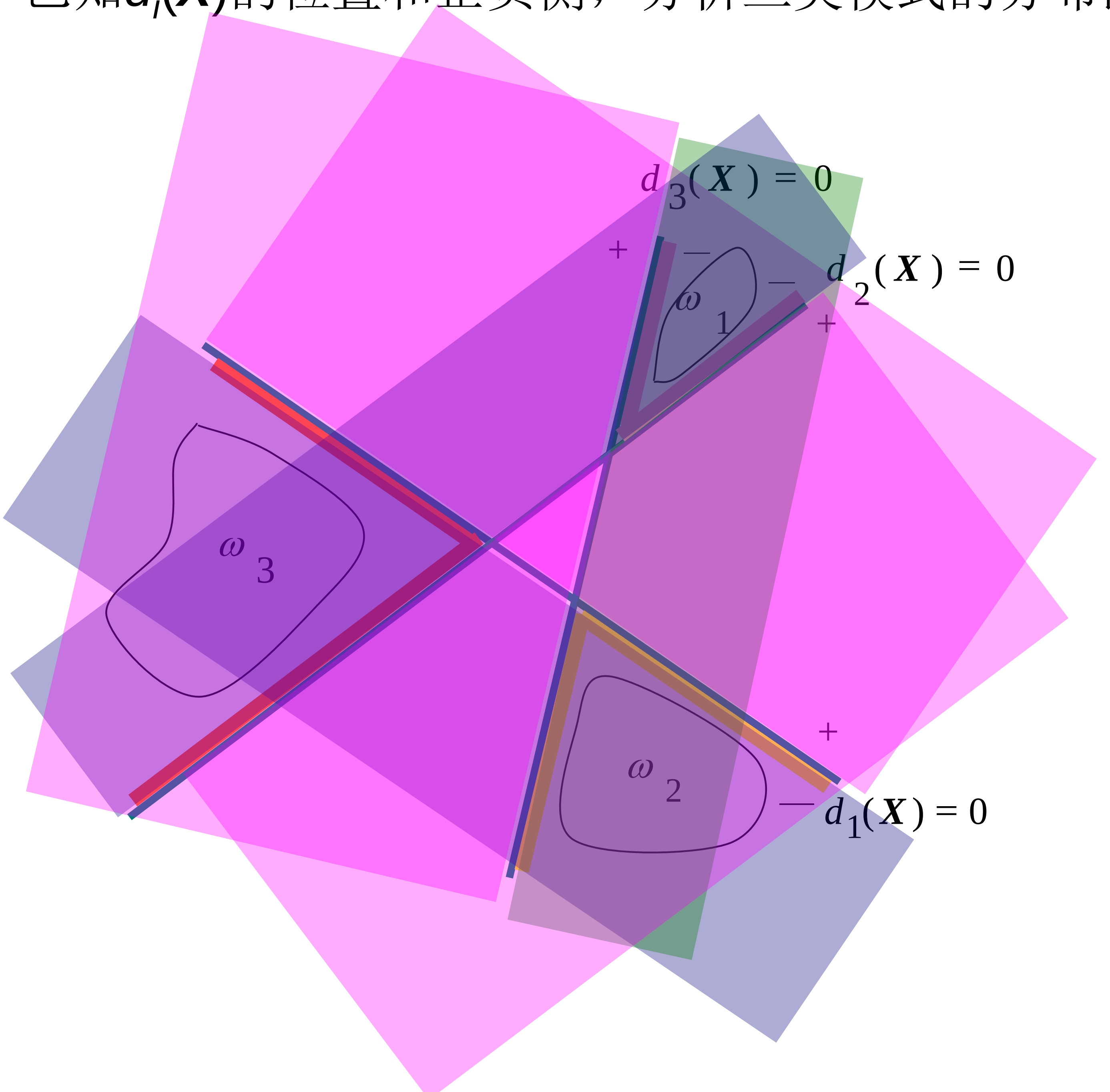
步骤：

a) 画出界面直线。

b) 判别界面正负侧：找特殊点代入。或者：画矢量W_0（以原点为起点），箭头指向的方向就是d(X)的正侧

c) 找交集（一正测、多负则侧）。

例3.2　已知$d_i(\boldsymbol{X})$的位置和正负侧，分析三类模式的分布区域。



（2）多类情况2：ω_i/ω_j 成对两分法

一个判别界面只能分开两个类别，不需要把其余所有的类别都分开。判决函数为：$d_{ij}(\boldsymbol{X})=\boldsymbol{W}_{ij}^{\mathrm{T}}\boldsymbol{X}$ 。这里 $d_{ji}=-d_{ij}$ 。

判别函数性质：

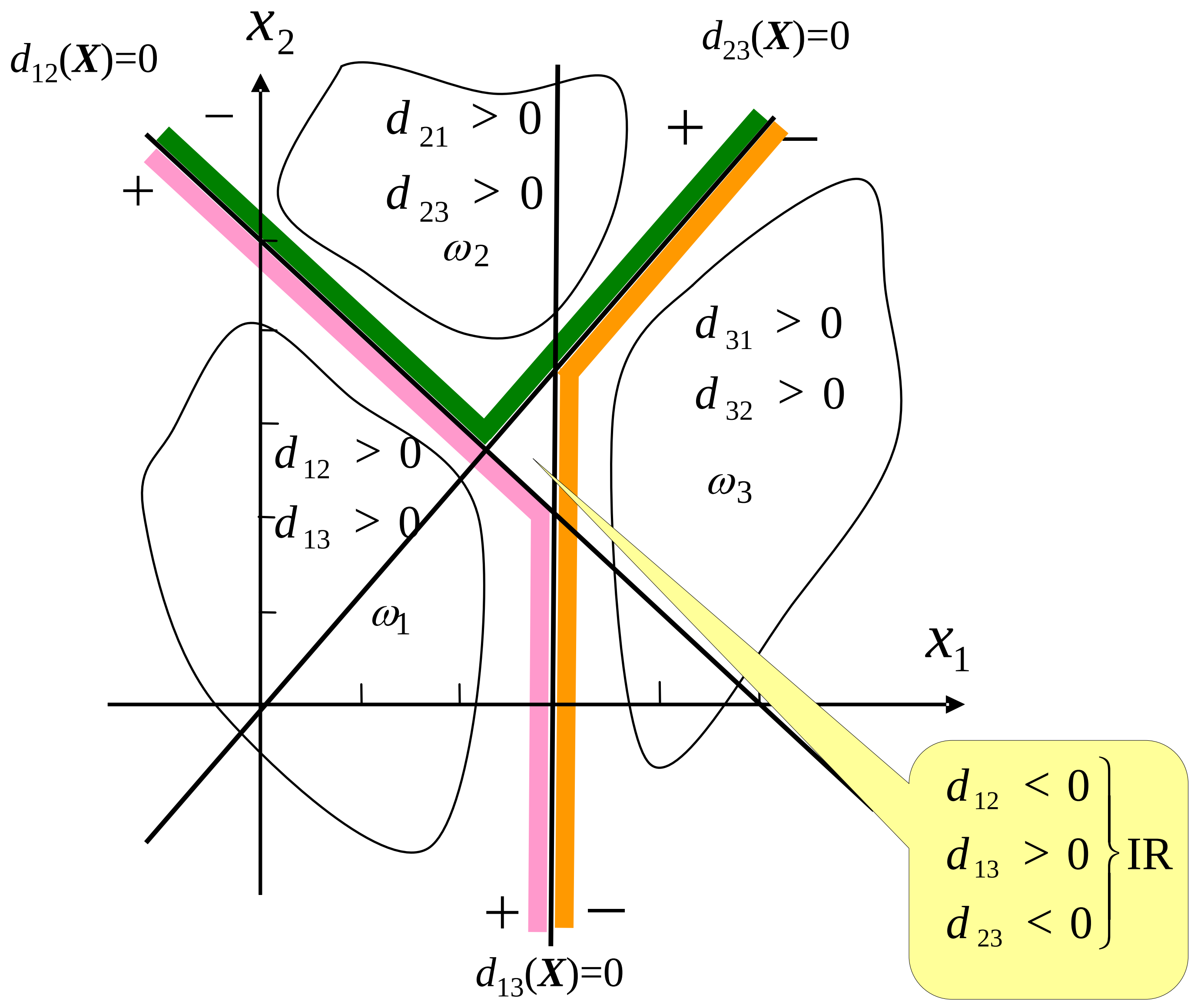
$$d_{ij}(\boldsymbol{X})>0,\ \forall j\neq i;\ i,j=1,2,\cdots,M,\quad \text{若}\ \boldsymbol{X}\in\omega_i$$

识别分类时：

在 *M* 类模式中，以*i*开始的*M*-1个判决函数全为正时，$\boldsymbol{X}\in\omega_i$。其中若有一个为负，则为IR区。
能用本方法分类的模式集称为：**成对线性可分的**。

如：对一个三类问题，如果$d_{12}(\boldsymbol{X})>0$ ，$d_{13}(\boldsymbol{X})>0$

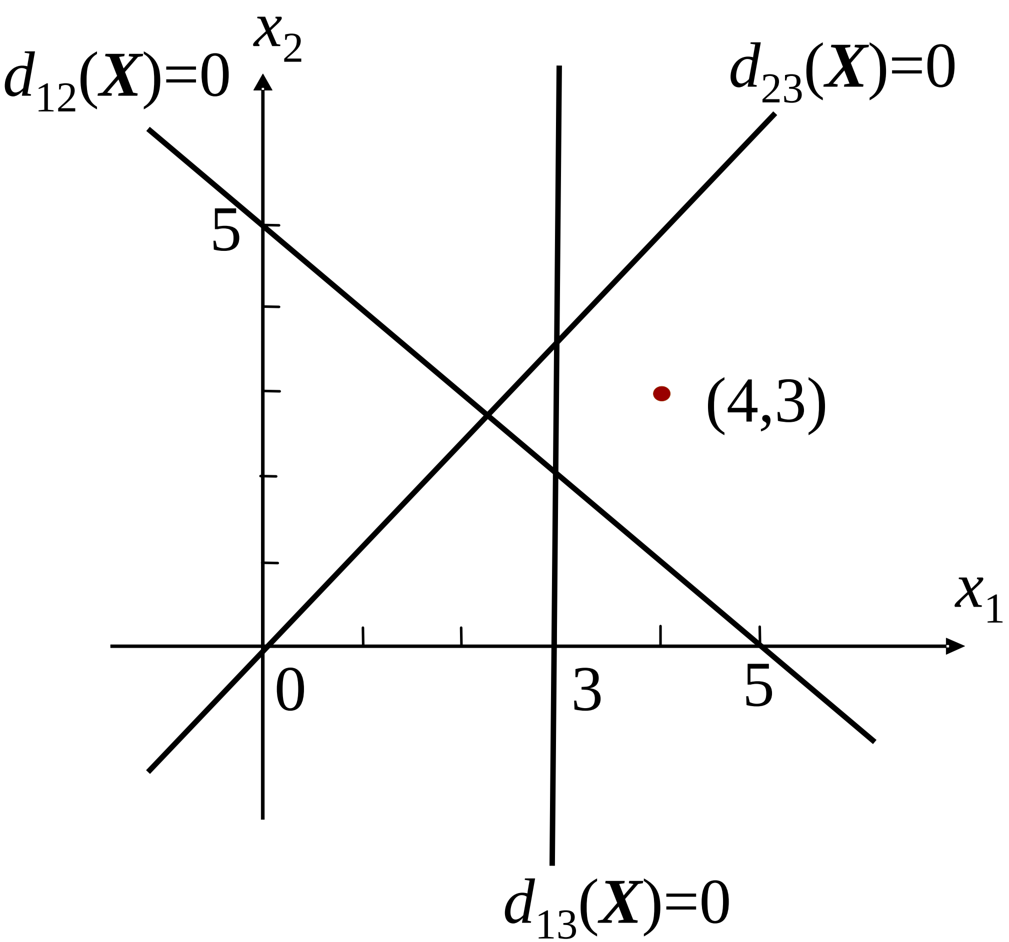
则$\boldsymbol{X}\in\omega_1$ 类，而 $d_{23}(\boldsymbol{X})$在判别 ω_1 类模式时不起作用。


$d_{12}(\boldsymbol{X})=0$
x_2
$d_{23}(\boldsymbol{X})=0$
$d_{21} > 0$
$d_{23} > 0$
ω_2
$d_{31} > 0$
$d_{32} > 0$
ω_3
$d_{12} > 0$
$d_{13} > 0$
ω_1
x_1
$d_{13}(\boldsymbol{X})=0$
$d_{12} < 0$
$d_{13} > 0$
$d_{23} < 0$
IR

例3.3 一个二维三类问题，已知三个判决函数为：

$$d_{12}(\boldsymbol{X})=-x_1-x_2+5 \quad d_{13}(\boldsymbol{X})=-x_1+3 \quad d_{23}(\boldsymbol{X})=-x_1+x_2$$

问模式 $\boldsymbol{X}=[4,3]^{\mathrm{T}}$ 属于哪类？



解1 先计算：$d_{12}(\boldsymbol{X})=-2,$

$d_{13}(\boldsymbol{X})=-1,$

$d_{23}(\boldsymbol{X})=-1$

然后求：$d_{21}(\boldsymbol{X})=2,$

$d_{31}(\boldsymbol{X})=1,$

$d_{32}(\boldsymbol{X})=1$

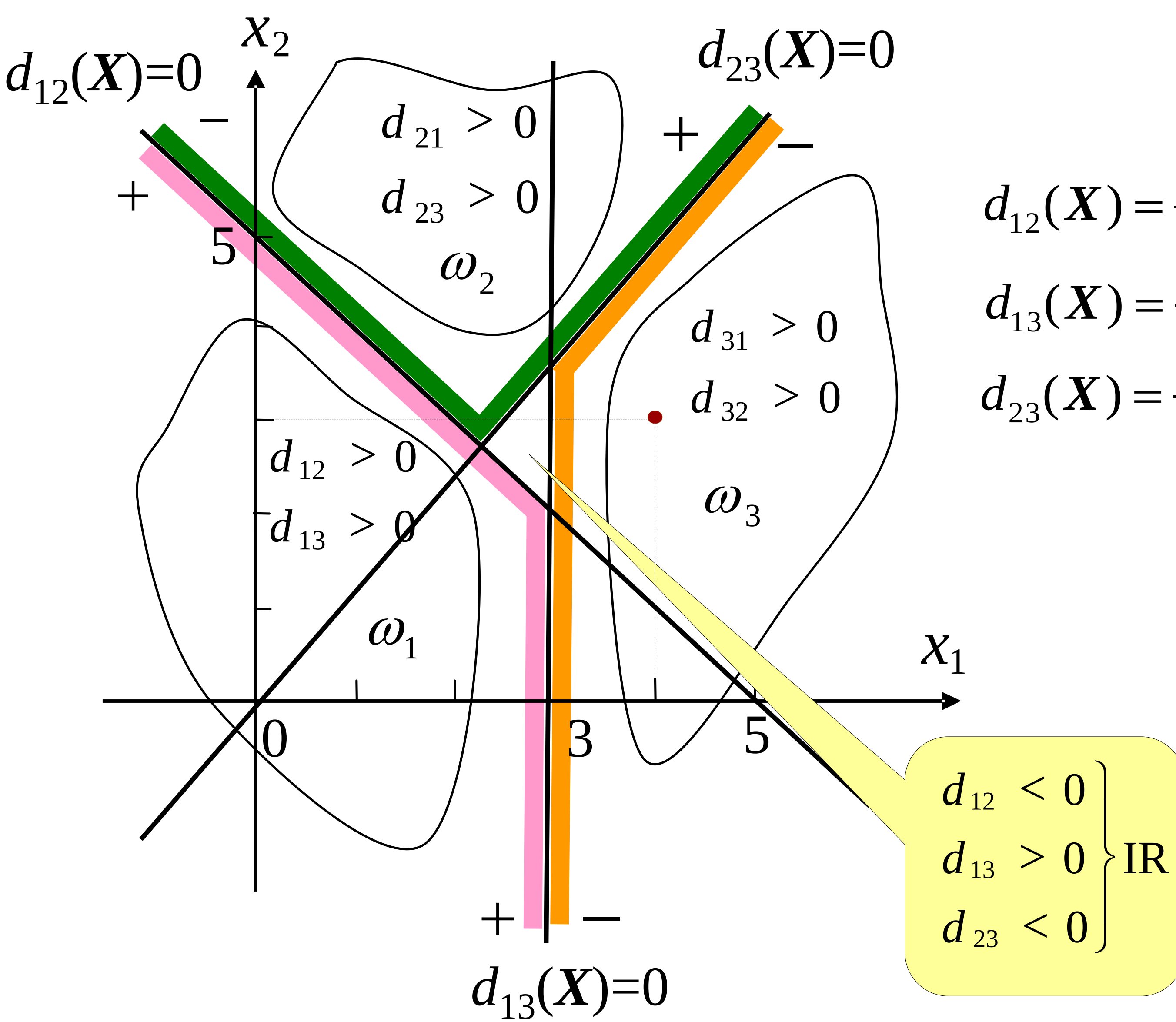
最后判：

$$\left.\begin{matrix} d_{31}(\boldsymbol{X})>0 \\ d_{32}(\boldsymbol{X})>0 \end{matrix}\right\} \Rightarrow$$

$$\boldsymbol{X}=[4,3]^{\mathrm{T}}\in\omega_3$$

与 $d_{12}(\boldsymbol{X})$值无关。

解2：先由三个判别函数，找出三个模式类的区域；
再由X所属区域判定其类别。



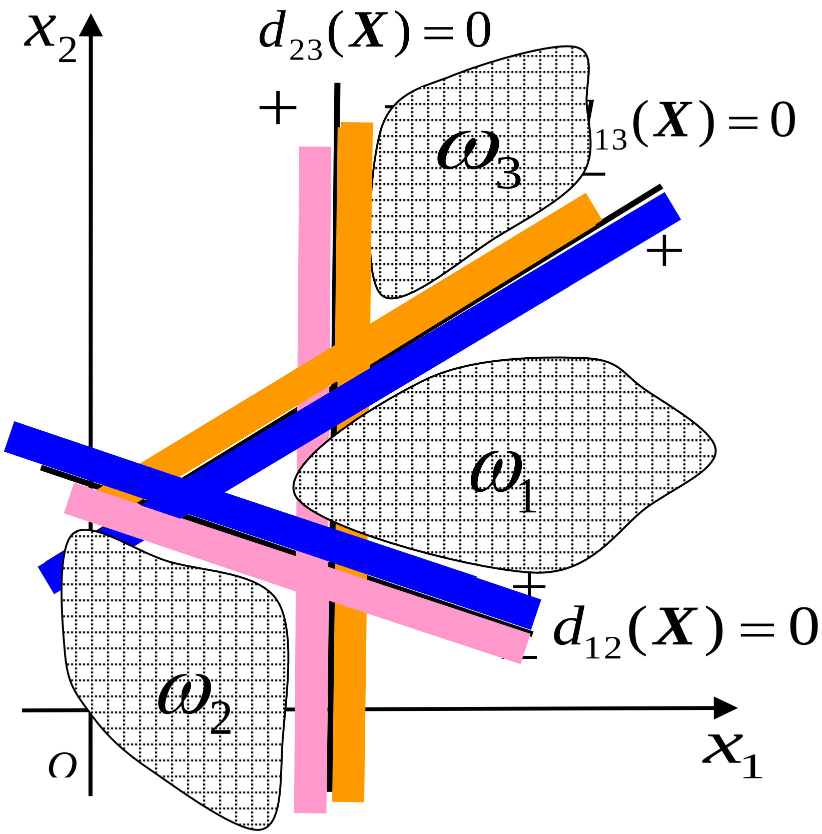
$$d_{12}(\boldsymbol{X}) = -x_1 - x_2 + 5$$

$$d_{13}(\boldsymbol{X}) = -x_1 + 3$$

$$d_{23}(\boldsymbol{X}) = -x_1 + x_2$$

分类时：每分离出一类，需要与模式有关的M-1个判决函数；要分开M类模式，共需$M(M$-1)/2个判决函数。对三类问题需要3(3-1)/2=3个判决函数。即：每次从M类中取出两类的组合：

$$C_M^2=\frac{M(M-1)}{2!}$$



例3.4 已知$d_{ij}(\boldsymbol{X})$的位置和正负侧，分析三类模式的分布区域 。

（3）多类情况3：ω_i/ω_j 成对两分法特例（没有不确定区）

当ω_i/ω_j成对两分法中的判别函数$d_{ij}(\boldsymbol{X})$，如果可以分解为

$$d_{ij}(\boldsymbol{X}) = d_i(\boldsymbol{X}) - d_j(\boldsymbol{X})$$

时，那么$d_i(\boldsymbol{X}) > d_j(\boldsymbol{X})$就相当于多类情况2中的$d_{ij}(\boldsymbol{X}) > 0$。

此时，对具有判别函数

$$d_i(\boldsymbol{X}) = \boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{X}, \quad i = 1, \cdots, M$$

的M类情况，等价的判别函数性质为：

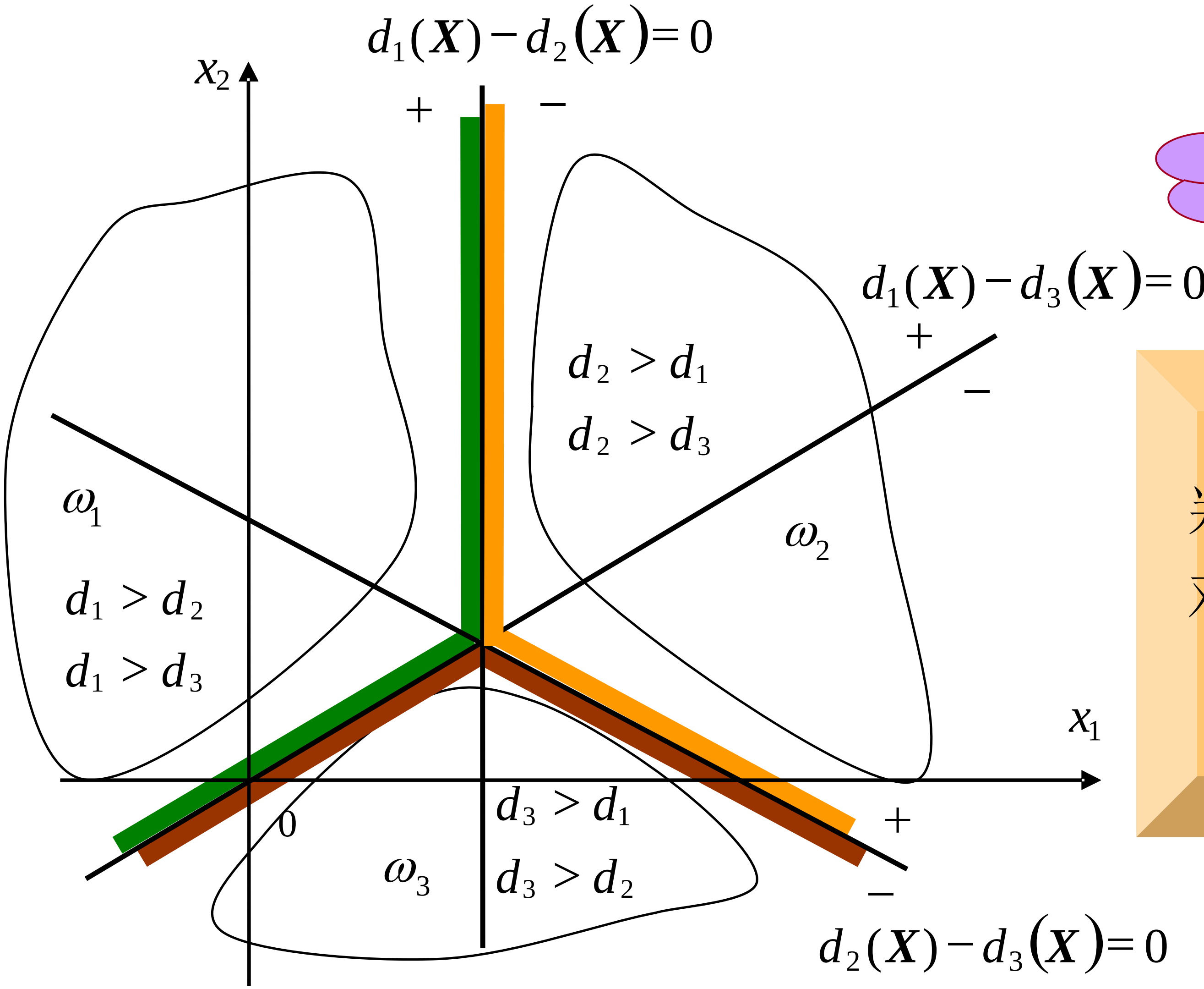
$$d_i(\boldsymbol{X}) > d_j(\boldsymbol{X}),\ \forall j \neq i;\ i, j = 1, 2, \cdots, M, \quad 若\ \boldsymbol{X} \in \omega_i$$

或：

$$d_i(\boldsymbol{X}) = \max\{d_k(\boldsymbol{X}),\quad k = 1, \cdots, M\}, \quad 若\boldsymbol{X} \in \omega_i$$

能用本方法分类的模式集称为：**线性可分的**。

注意：与前两种方法不同，这里是**以大小进行判定**的，不是仅仅以正负进行判定的。



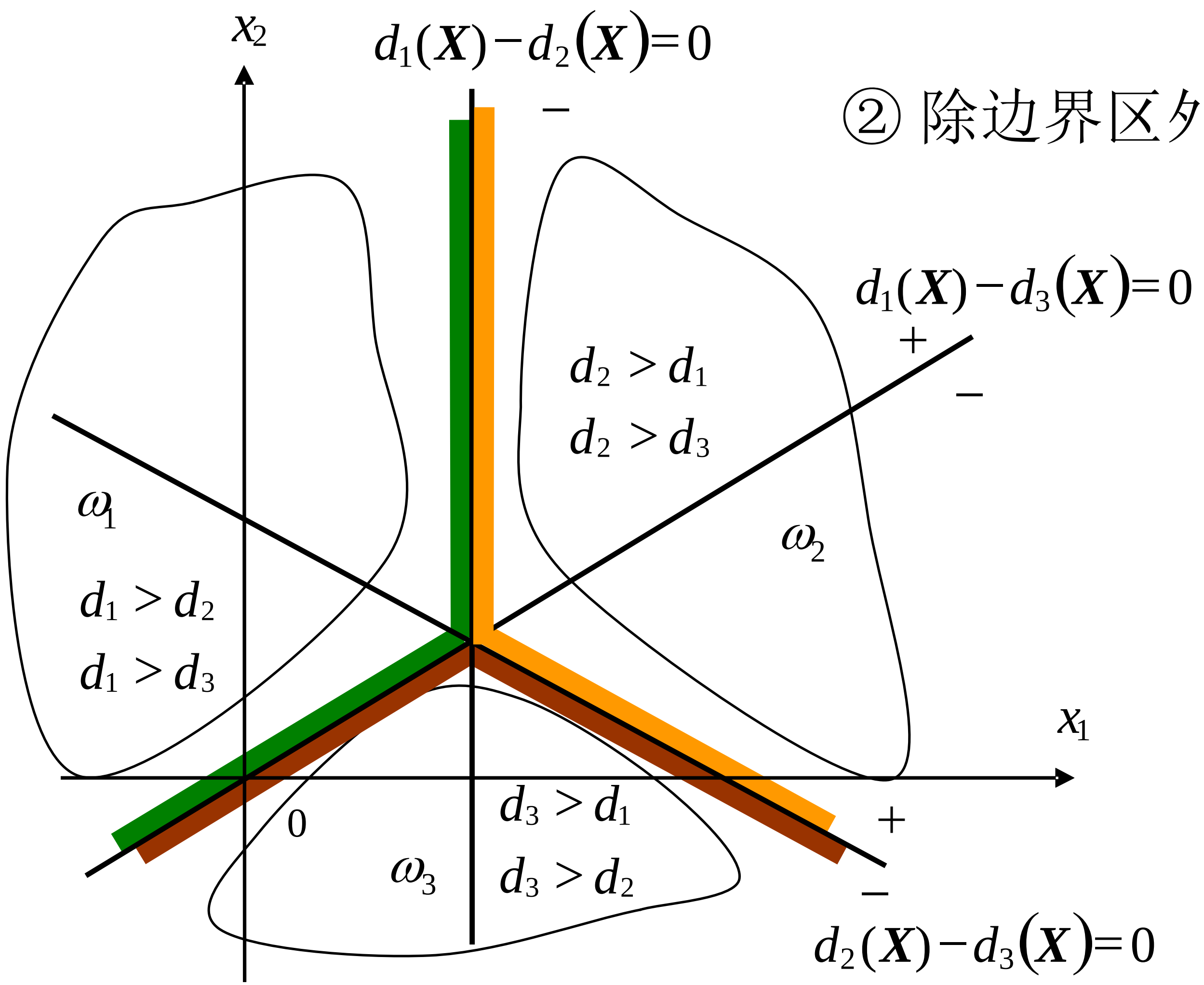
识别分类时：

判别界面做差值。

对ω_i类，应满足：

d_i>其他所有d_j

① 这是第二种情况的特例。因$d_{ij}(\boldsymbol{X}) = d_i(\boldsymbol{X}) - d_j(\boldsymbol{X})$，若在第三种情况下可分（线性可分），则在第二种情况下也可分（成对线性可分），**但反之不一定成立**。



② 除边界区外，**没有不确定区域**。

③ 把 M 类情况分成了$(M\text{-}1)$个两类问题。并且ω_i 类的判别界面全部与$\omega_j(\forall j \neq i)$ 类的判别界面相邻（向无穷远处延伸的区域除外）。

例3.5 一个二维三类模式分类器，满足多类情况3（成对两分法特例）要求的三个判决函数为：

$$d_1(\boldsymbol{X}) = -x_1 + x_2$$
$$d_2(\boldsymbol{X}) = x_1 + x_2 - 1$$
$$d_3(\boldsymbol{X}) = -x_2$$

试判断$\boldsymbol{X}_0=[1,1]^{\mathrm{T}}$属于哪一类，且分别给出三类的判决界面。

解：①

$$\left.\begin{aligned} d_1(\boldsymbol{X}_0) &= -1+1=0 \\ d_2(\boldsymbol{X}_0) &= 1+1-1=1 \\ d_3(\boldsymbol{X}_0) &= -1 \end{aligned}\right\} \Rightarrow \left.\begin{aligned} d_2(\boldsymbol{X}_0) &> d_1(\boldsymbol{X}_0) \\ d_2(\boldsymbol{X}_0) &> d_3(\boldsymbol{X}_0) \end{aligned}\right\} \Rightarrow \boldsymbol{X}_0 \in \omega_2$$

② ω_1 类的判决函数：

$$d_1(\boldsymbol{X}) - d_2(\boldsymbol{X}) = -2x_1 + 1$$

$$d_1(\boldsymbol{X}) - d_3(\boldsymbol{X}) = -x_1 + 2x_2$$

ω_2 类的判决函数：

$$d_2(\boldsymbol{X}) - d_1(\boldsymbol{X}) = 2x_1 - 1$$

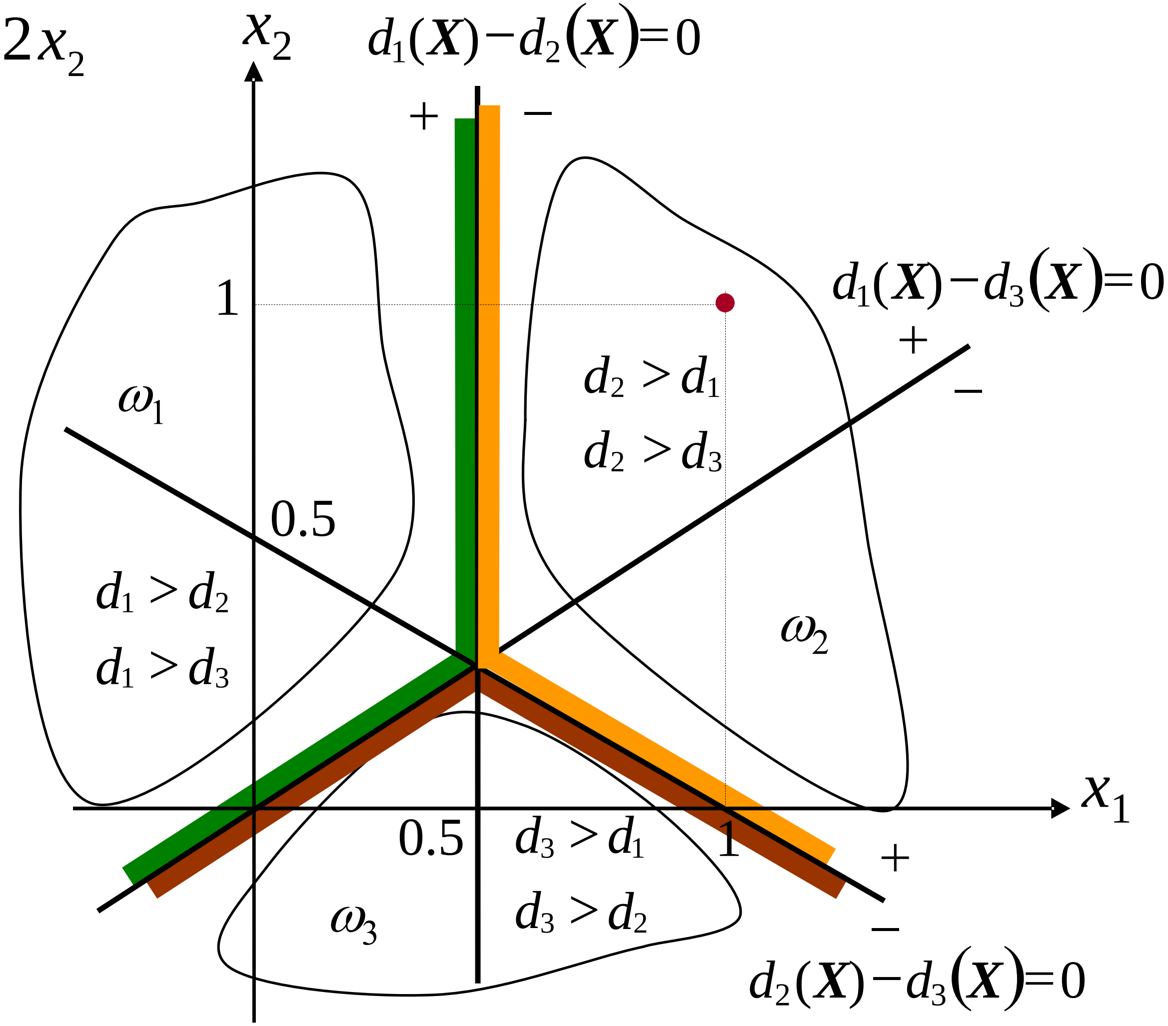
$$d_2(\boldsymbol{X}) - d_3(\boldsymbol{X}) = x_1 + 2x_2 - 1$$

ω_3 类的判决函数：

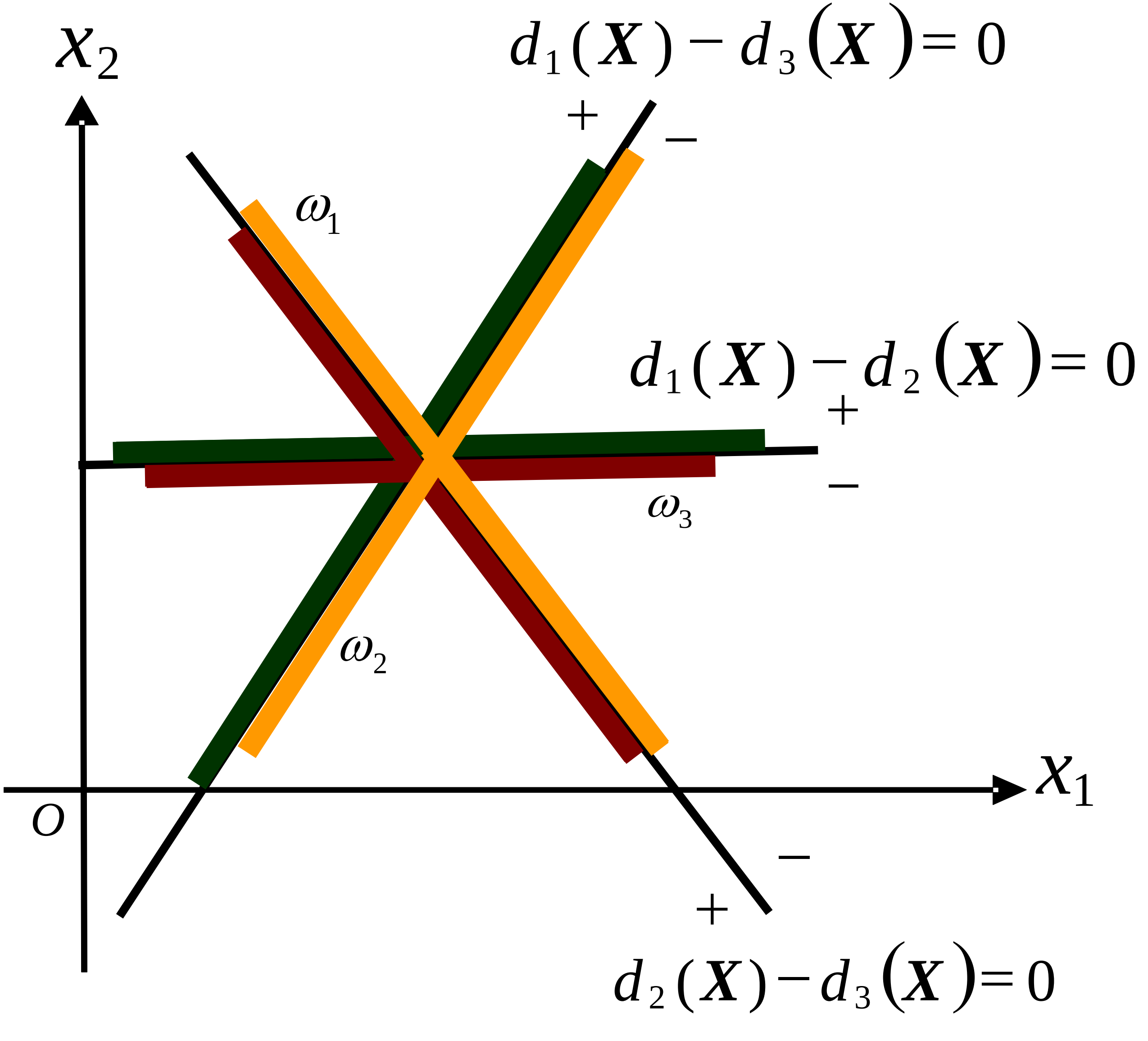
$$d_{31}(\boldsymbol{X}) = -d_{13}(\boldsymbol{X})$$

$$d_{32}(\boldsymbol{X}) = -d_{23}(\boldsymbol{X})$$

判决界面方程如图。



例3.6 已知判决界面的位置和正负侧，按多类情况3分析三类模式的分布区域。



3. 小结

（1）明确概念：线性可分。

一旦线性判别函数的系数$\boldsymbol{W}_k$被确定以后，这些函数就可以作为模式分类器，对未知模式进行分类。

（2） $\omega_i/\overline{\omega}_i$ 与ω_i/ω_j 两分法的比较：

对于M类模式的分类，$\omega_i/\overline{\omega}_i$ 是非两分法共需要M个判别函数，ω_i/ω_j 成对两分法需要$M(M-1)/2$个。当$M>3$时，后者需要更多的判别式（缺点），但对模式集进行线性可分的可能性要更大一些（优点）。

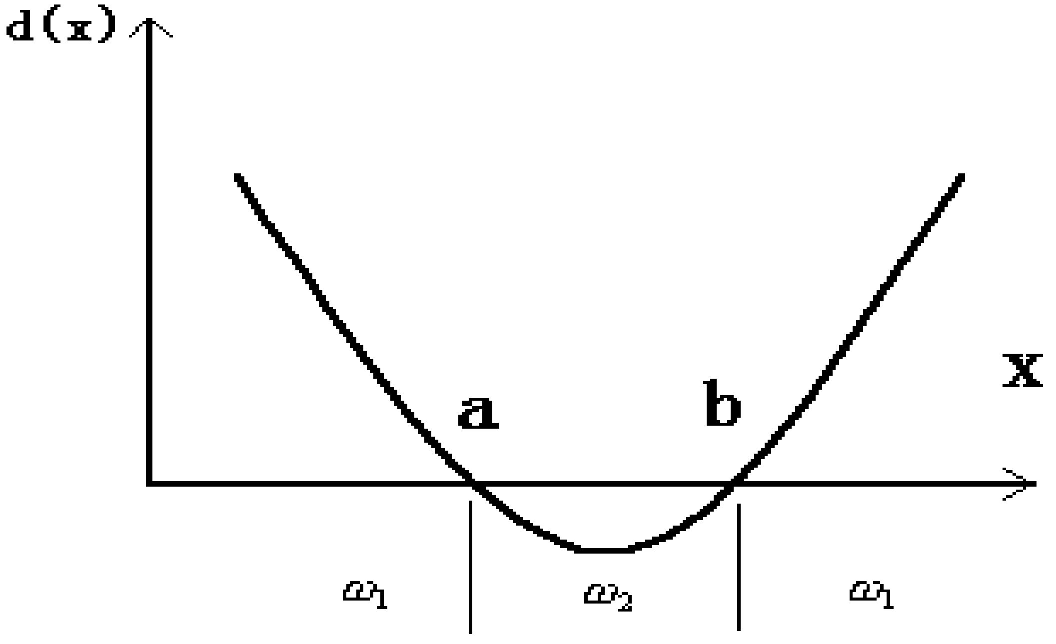
原因：

一种类别模式ω_j的分布要比M-1类模式$\overline{\omega}_i$的分布更为聚集，因此ω_i/ω_j两分法受到的限制比$\omega_i/\overline{\omega}_i$ 少，因此线性可分可能性大。

3.3 广义线性判别函数

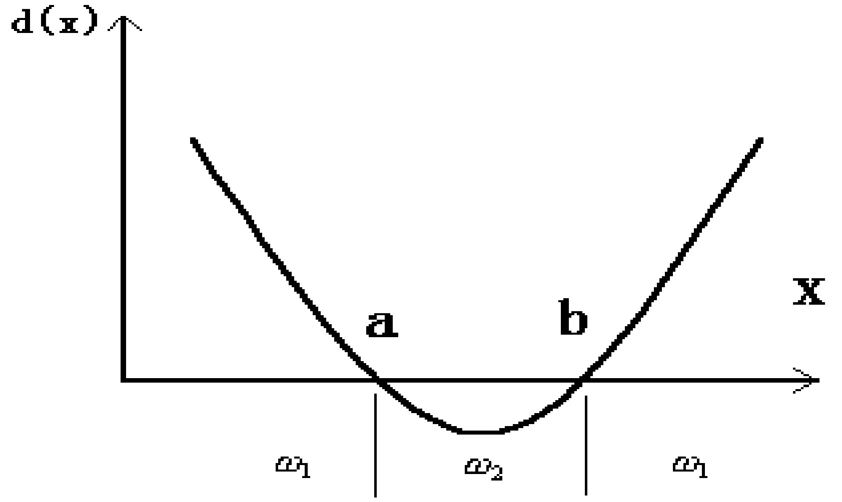
先看一个简单示例：

设一维两类模式 x 在一维空间——坐标轴上分布如图所示，两类的类域为 $\Omega_1:(-\infty,a)$ 或 (b,∞) 和 $\Omega_2:(a,b)$



一维特征空间中非线性可分的图示

显然，这两类不是线性可分的，因不能用一个(1-1)=0维超平面-分界点将两类分开。但是，对于一条穿过a、b两点的二次曲线，



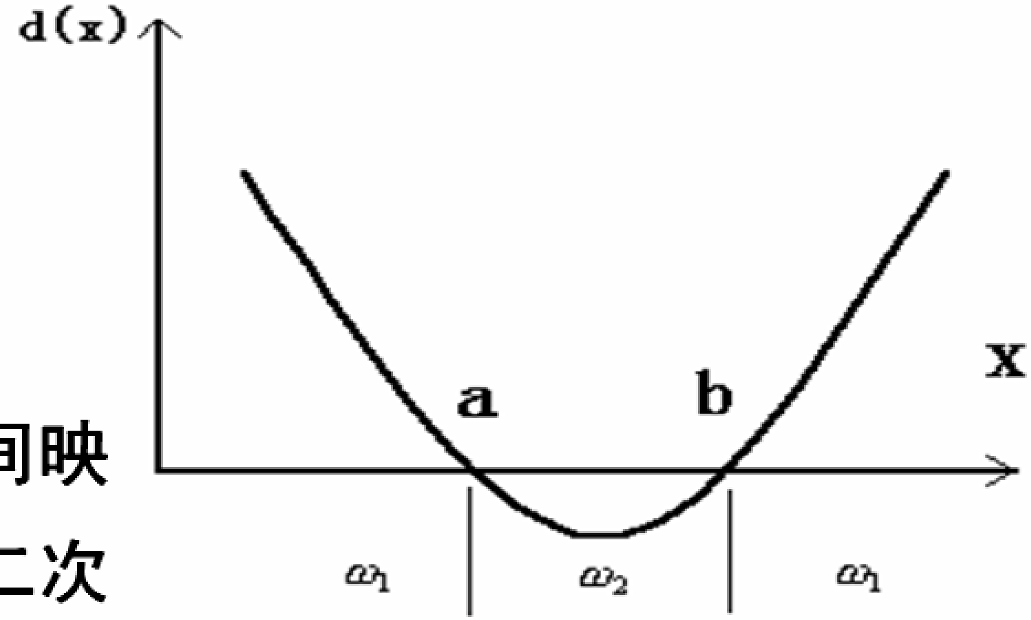
$$d(x)=(x-a)(x-b)=x^2-(a+b)x+ab$$

当 $x\in\omega_1$ 即 $x<a$ 或 $x>b$ 时， $d(x)>0$

当 $x\in\omega_2$ 即 $a<x<b$ 时， $d(x)<0$

如果作非线性变换：

$$\begin{cases} y_1 = x^2 \\ y_2 = x \end{cases}$$



则原来的一维特征空间映射为二维特征空间，二次判别函数 $d(x)$ 变换后在二维特征空间中映射成一次判别函数：$d(\vec{y}) = y_1 - (a+b)y_2 + ab$

原来的一维非线性可分的模式在所映射的二维特征空间中是线性可分的，即:

$$\begin{cases} \vec{y} \in \omega_1 & d(\vec{y}) > 0 \\ \vec{y} \in \omega_2 & d(\vec{y}) < 0 \end{cases}$$

广义线性判别的目的：

通过某映射，把模式空间X变成X^*，以便将X空间中非线性可分的模式集，变成在X^*空间中线性可分的模式集。

$$d(\boldsymbol{X})=w_1x_1+w_2x_2+\ldots+w_nx_n+w_{n+1}$$

非线性多项式判别函数——广义线性判别函数的一类

设一训练用模式集，$\{\boldsymbol{X}\}$在模式空间X中线性不可分，非线性多项式判别函数形式如下（关于系数是线性的）：

$$d(\boldsymbol{X})=w_1f_1(\boldsymbol{X})+w_2f_2(\boldsymbol{X})+\ldots+w_kf_k(\boldsymbol{X})+w_{k+1}=\sum_{i=1}^{k+1}w_if_i(\boldsymbol{X}) \qquad (3\text{-}9)$$

式中$\{f_i(\boldsymbol{X}),\quad i=1,2,\cdots,k\}$是模式$\boldsymbol{X}$的单值实函数，$f_{k+1}(\boldsymbol{X})=1$。

$f_i(\boldsymbol{X})$取什么形式及$d(\boldsymbol{X})$取多少项，取决于非线性边界的复杂程度。

广义线性函数的模式向量定义为：

$$\boldsymbol{X}^* = [x_1^*, x_2^*, \cdots, x_k^*, 1]^{\mathrm{T}} = \left[f_1(\boldsymbol{X}), f_2(\boldsymbol{X}), \cdots, f_k(\boldsymbol{X}), 1\right]^{\mathrm{T}} \quad (3\text{-}10)$$

这里X*空间的维数k高于X空间的维数n，(3-9)式可写为

$$d(\boldsymbol{X}) = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{X}^* = d(\boldsymbol{X}^*), \quad \boldsymbol{W} = \left[w_1, w_2, \cdots, w_k, w_{k+1}\right]^{\mathrm{T}} \quad (3\text{-}11)$$

上式不仅**关于系数是线性的，而且关于X*分量也是线性的**。

讨论广义线性判别函数不会失去一般性。

问题：

- 非线性变换可能非常复杂 。
- 维数大大增加，导致维数灾难。

随着**小样本学习理论**和**支持向量机**的迅速发展，广义线性判别函数的“**维数灾难**”问题在一定程度上找到了解决的办法。

3.4 线性判别函数的几何性质

3.4.1 模式空间与超平面

1. 概念

模式空间：以n维模式向量X的n个分量为坐标变量的欧氏空间。

模式向量：点、有向线段。

线性分类：用$d(\boldsymbol{X})$进行分类，相当于用超平面$d(\boldsymbol{X})=0$把模式空间分成不同的决策区域。

2. 讨论

设判别函数： $d(\boldsymbol{X})=\boldsymbol{W}_0^{\mathrm{T}}\boldsymbol{X}+w_{n+1}$

式中，$\boldsymbol{W}_0=[w_1, w_2, \cdots, w_n]^{\mathrm{T}}$，$\boldsymbol{X}=[x_1, x_2, \cdots, x_n]^{\mathrm{T}}$。

超平面：$d(\boldsymbol{X})=\boldsymbol{W}_0^{\mathrm{T}}\boldsymbol{X}+w_{n+1}=0$

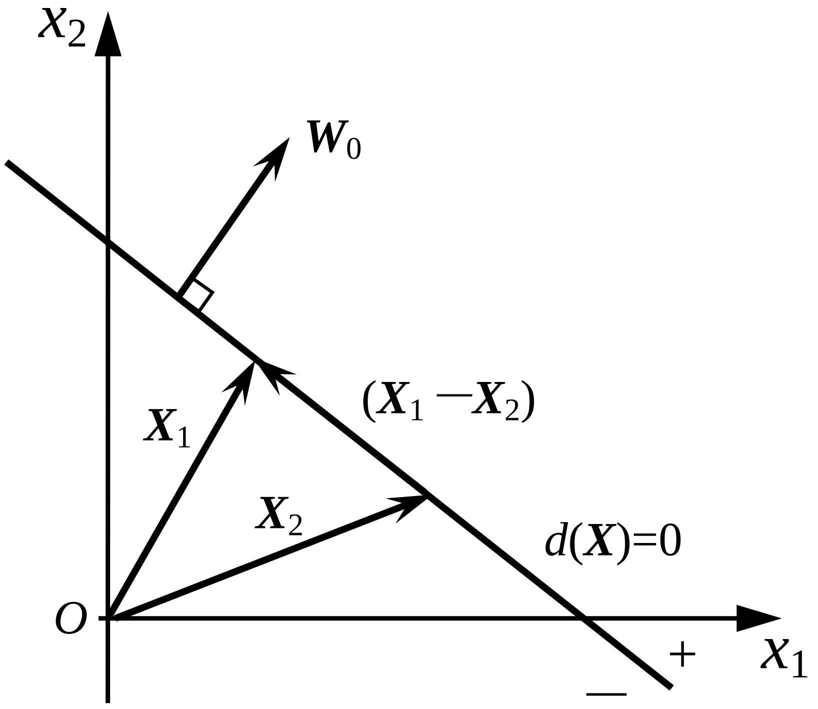
（1）模式向量X_1和X_2在超平面上

$$W_0^{\mathrm{T}} X_1 + w_{n+1} = W_0^{\mathrm{T}} X_2 + w_{n+1}$$

$$W_0^{\mathrm{T}} (X_1 - X_2) = 0$$

可见：**W_0是超平面的法向量，所指方向即为超平面的正侧。**

——Why？



记超平面的单位法线向量为U：

$$U = \frac{W_0}{\|W_0\|}$$

$$\|W_0\| = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$$

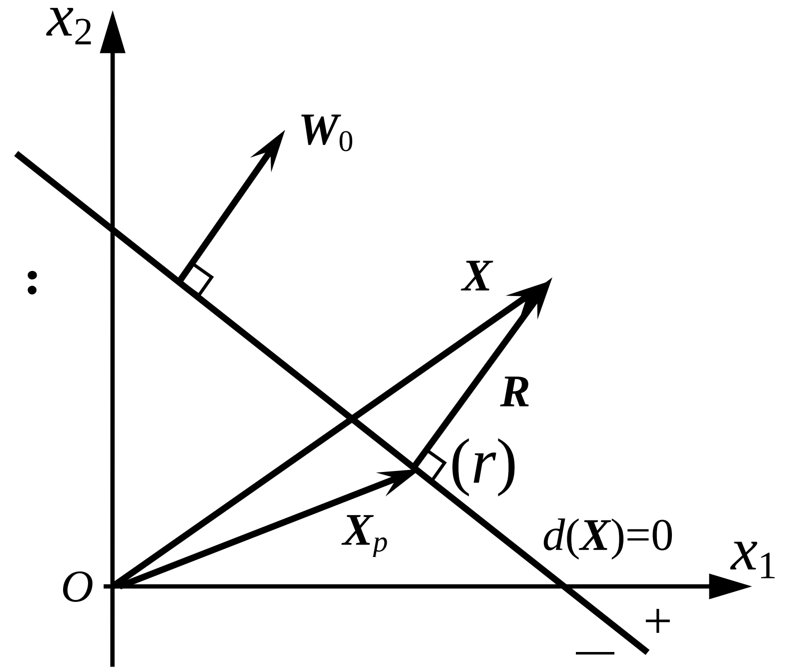
（2）$\boldsymbol{X}$不在超平面上

将X向超平面投影得向量X_p，

构造向量$\boldsymbol{R}$（投影点X_p到X的向量）：

$$\boldsymbol{R} = r \cdot \boldsymbol{U} = r \frac{W_0}{\|W_0\|}$$

r：$\boldsymbol{X}$到超平面的**代数距离**（待求）



$d(\boldsymbol{X}) = \boldsymbol{W}_0^{\mathrm{T}} \boldsymbol{X} + w_{n+1}$

$$\boldsymbol{X} = \boldsymbol{X}_p + \boldsymbol{R} = \boldsymbol{X}_p + r \frac{\boldsymbol{W}_0}{\|\boldsymbol{W}_0\|}$$

$$d(\boldsymbol{X}) = \boldsymbol{W}_0^{\mathrm{T}} (\boldsymbol{X}_p + r \frac{\boldsymbol{W}_0}{\|\boldsymbol{W}_0\|}) + w_{n+1} = (\boldsymbol{W}_0^{\mathrm{T}} \boldsymbol{X}_p + w_{n+1}) + \boldsymbol{W}_0^{\mathrm{T}} \cdot r \frac{\boldsymbol{W}_0}{\|\boldsymbol{W}_0\|}$$

$$= r \|\boldsymbol{W}_0\|$$

可见：判别函数$d(X)$ 正比于点X到超平面的代数距离。

求得$\boldsymbol{X}$到超平面的**代数距离**：$r \triangleq \dfrac{d(\boldsymbol{X})}{\|\boldsymbol{W}_0\|}$

换言之：点X到超平面的代数距离正比于$d(X)$函数值。

（3）X在原点

$$d(\boldsymbol{X}) = \boldsymbol{W}_0^{\mathrm{T}}\boldsymbol{X} + w_{n+1} = w_{n+1}$$

得：原点到超平面的代数距离

$$r_0 = \frac{w_{n+1}}{\|\boldsymbol{W}_0\|}$$

可见：原点在超平面的正负侧位置由阈值权w_{n+1}决定：

$w_{n+1} > 0$时，原点在超平面的正侧；

$w_{n+1} < 0$时，原点在超平面负侧；

$w_{n+1} = 0$ 时，超平面通过原点。

QUIZ：能否只用w_{n+1}的正负判别超平面的正负侧？ Why？

3.4.2 权空间与权向量解

（1）概念

权空间：以 $d(\boldsymbol{X}) = w_1x_1 + w_2x_2 + \ldots + w_nx_n + w_{n+1}$ 的权系数为坐标变量的（n+1）维欧氏空间，$\boldsymbol{X}$为已知。

增广权向量：$\boldsymbol{W} = [w_1, w_2, \cdots, w_n, w_{n+1}]^{\mathrm{T}}$

对应空间一，或者从原点引出的矢量。

（2）线性分类

判别函数形式已定，确定权向量==在权空间进行最优化搜索。

设增广样本向量：ω_1 类：$\boldsymbol{X}_{11}$，$\boldsymbol{X}_{12}$，…，$\boldsymbol{X}_{1p}$

ω_2 类：$\boldsymbol{X}_{21}$，$\boldsymbol{X}_{22}$，…，$\boldsymbol{X}_{2q}$

使$d(\boldsymbol{X})$将ω_1和 ω_2分开，需满足

$$d(\boldsymbol{X}_{1i}) > 0,\ i = 1,2,\cdots,p$$

$$d(\boldsymbol{X}_{2i}) < 0,\ i = 1,2,\cdots,q$$

对ω_2的q个增广模式乘以（—1），则要求增广权矢量满足：

$$d(\boldsymbol{X})>0 \text{ ，其中 } X=\begin{cases}\boldsymbol{X}_{1i}, & i=1,2,\cdots,p \\ -\boldsymbol{X}_{2i}, & i=1,2,\cdots,q\end{cases}$$

这就是 样本的符号规范化。

X：规范增广样本向量。

对每个已知的X，$d(\boldsymbol{X})=0$在权空间确定一个超平面；共$(p+q)$个。

在权空间中寻找权向量$\boldsymbol{W}$，使判别函数$d(\boldsymbol{X})$能把ω_1类和ω_2类分开，这样的权向量，必定在（$p+q$）个超平面的正侧的交迭区域里——称为**解区、解空间**。

现在，把$(p+q)$个已知类别模式作为约束条件，把权向量作为优化变量，进行优化搜索。

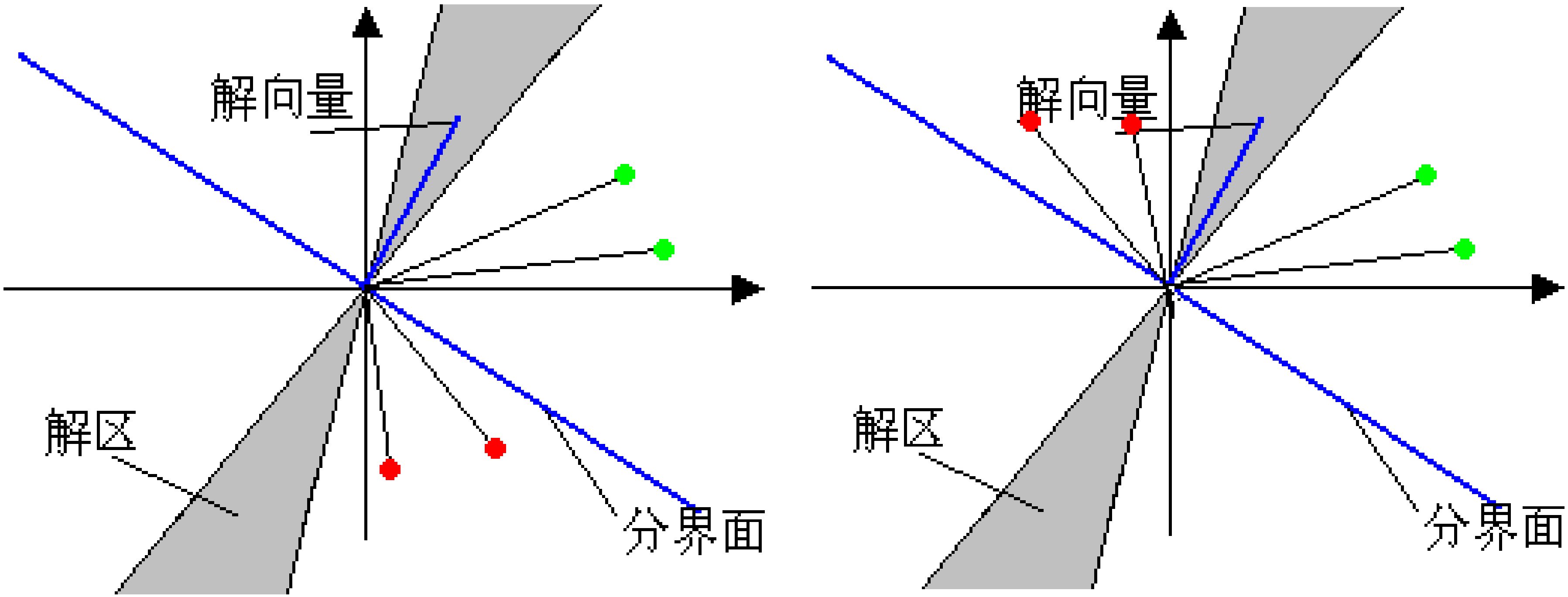
(3) 解空间

N 个训练模式将确定 N 个界面，每个界面都把权空间分为两个半空间，N 个正的半子空间的交空间是以权空间原点为顶点的凸多面锥。

满足上面各不等式的 $\vec{w}$ 必在该锥体中，即锥中每一点都是上面不等式组的解，解矢量不是唯一的，上述的凸多面锥包含了解的全体，称其为解区、解空间或解锥。

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。

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。



a: 第一类样本

(a) 未规范化

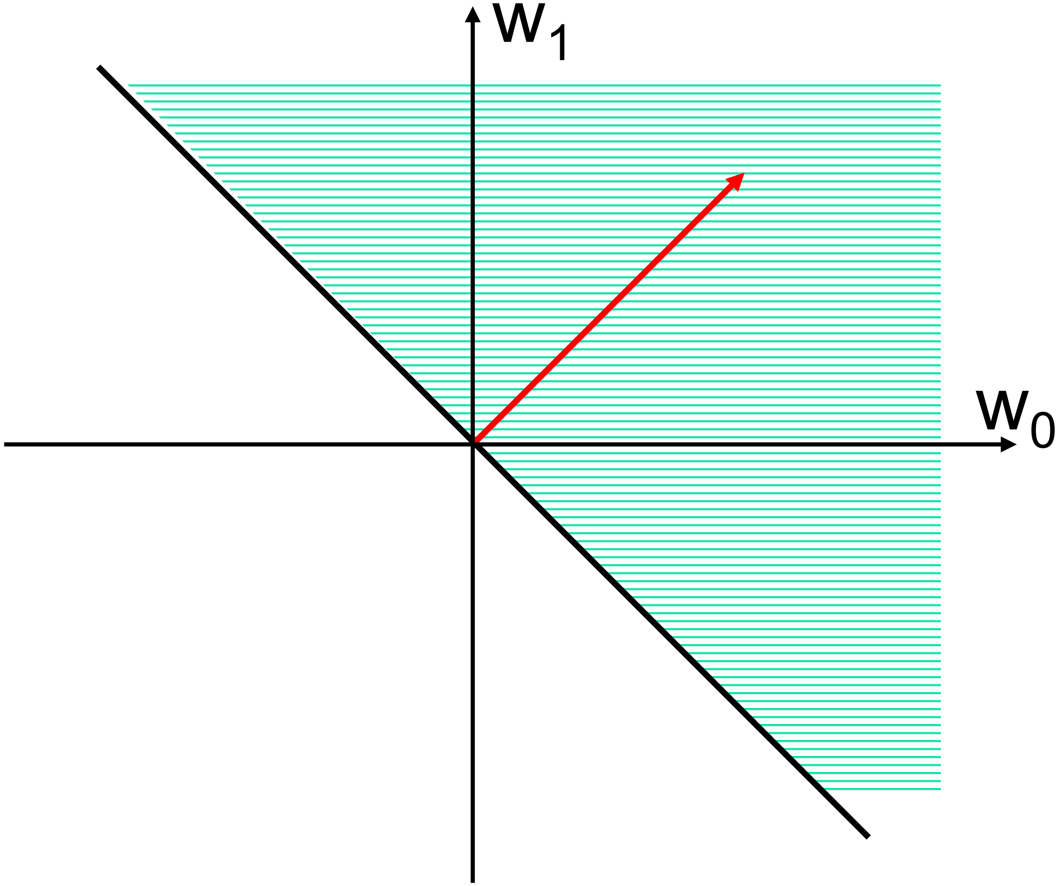
b: 第二类样本

(b) 规范化

样本的符号规范化示意图

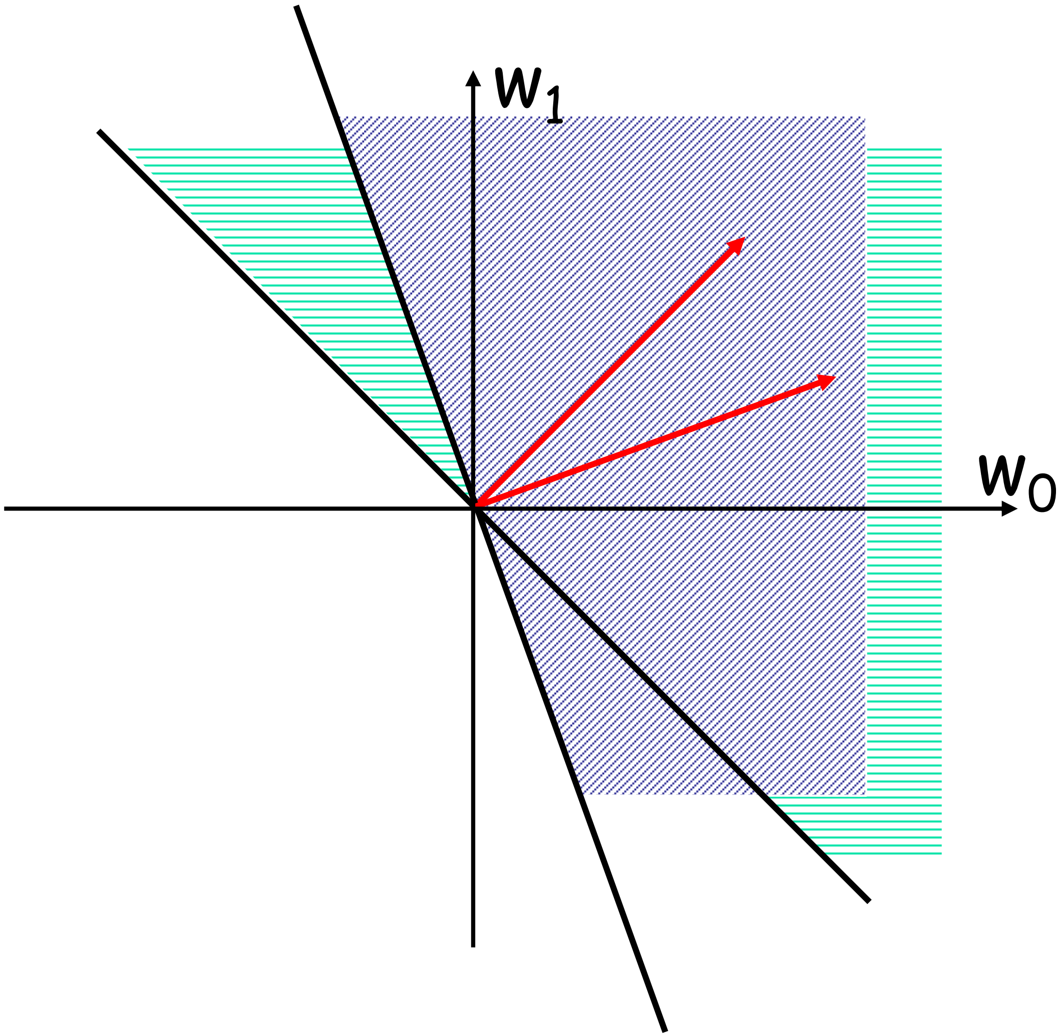
示例：权空间、解矢量与解空间

先看一个简单的情况。设4个一维数据$x_1=1$，$x_2=2$属于ω_1，$x_3=-1$，$x_4=-2$属于ω_2，求：将这两类区分开的线性判别函数的系数w_0，w_1。



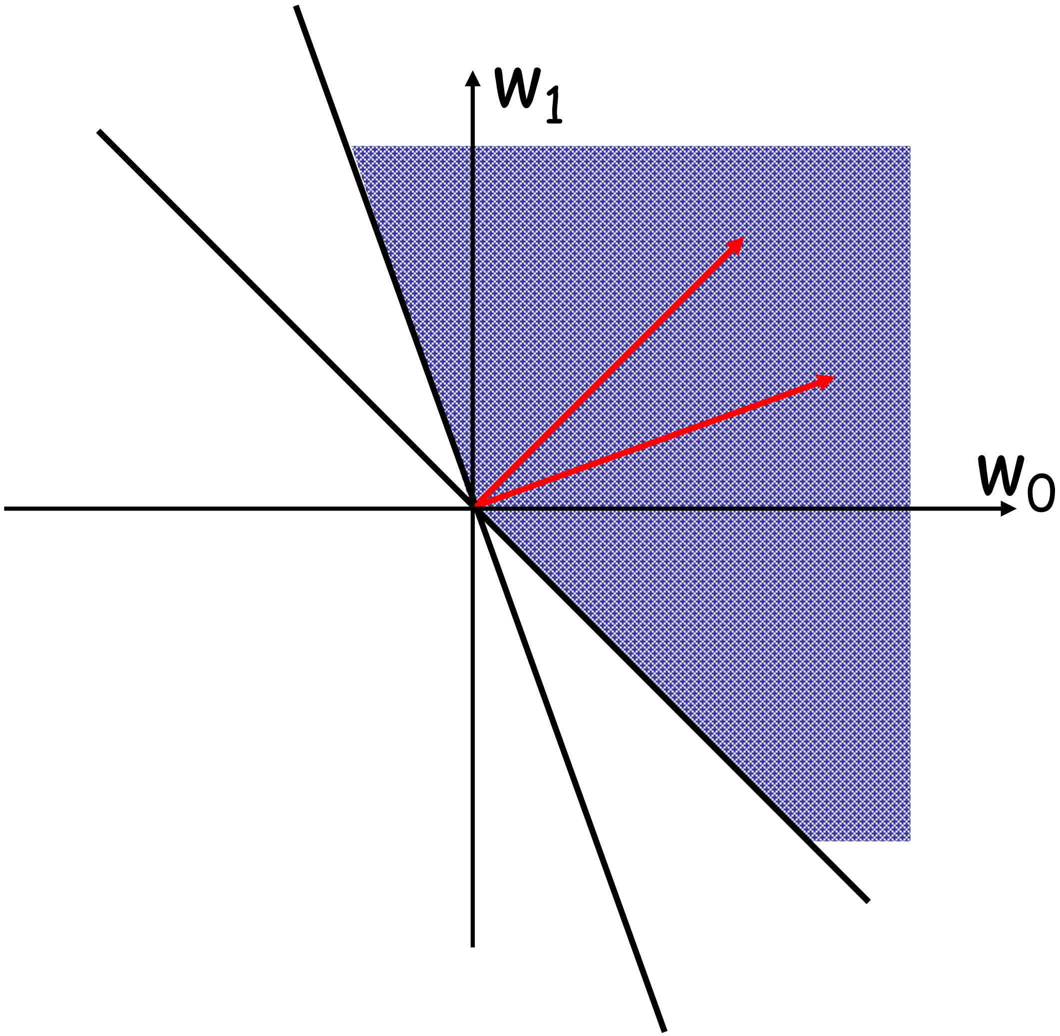
示例：权空间、解矢量与解空间

先看一个简单的情况。设4个一维数据 $x_1=1$，$x_2=2$属于 ω_1，$x_3=-1$，$x_4=-2$属于ω_2，求：将这两类区分开的线性判别函数的系数w_0，w_1。



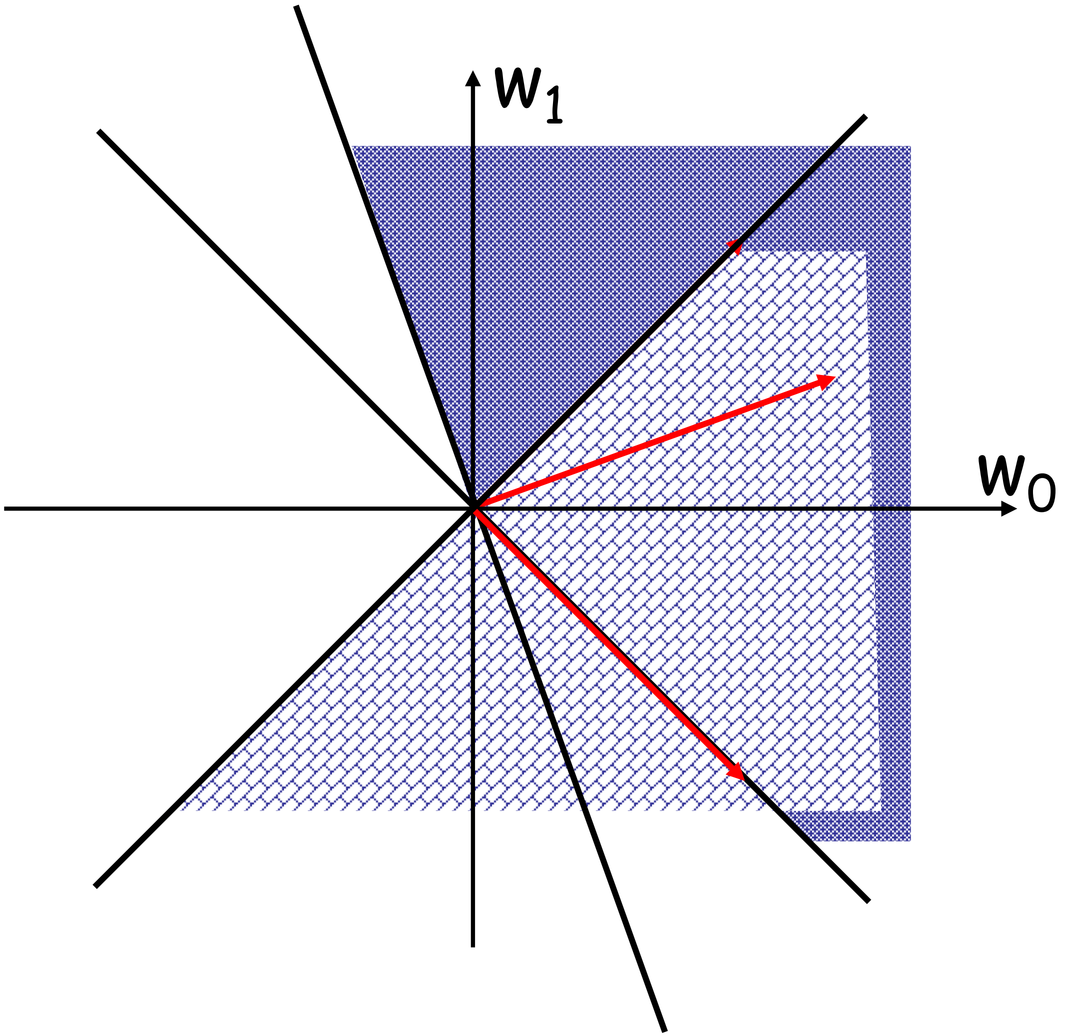
示例：权空间、解矢量与解空间

先看一个简单的情况。设4个一维数据 $x_1=1$，$x_2=2$属于 ω_1，$x_3=-1$，$x_4=-2$属于ω_2，求：将这两类区分开的线性判别函数的系数w_0，w_1。



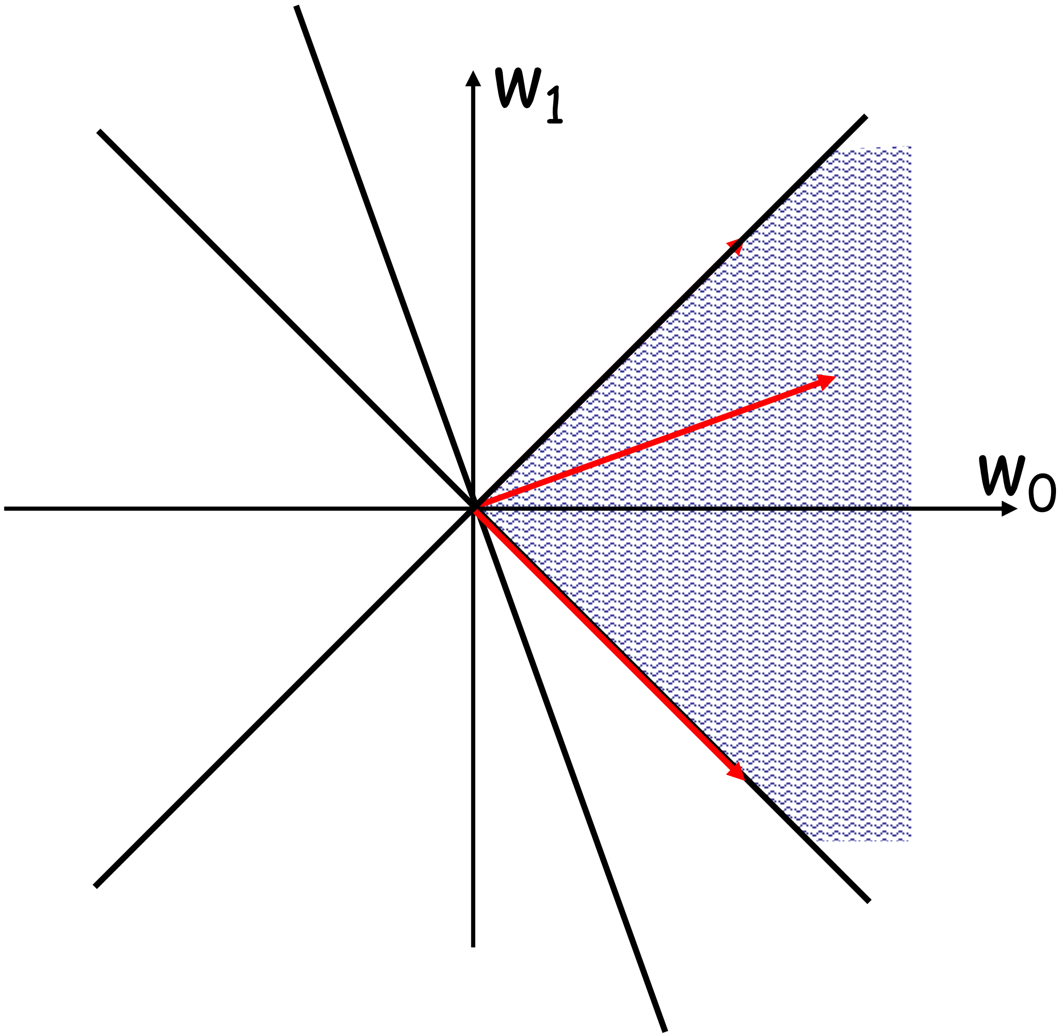
示例：权空间、解矢量与解空间

先看一个简单的情况。设4个一维数据 **x_1=1，x_2=2属于 ω_1, x_3=-1，x_4=-2属于ω_2，求：将这两类区分开的线性判别函数的系数**w_0，w_1。



示例：权空间、解矢量与解空间

先看一个简单的情况。设4个一维数据$x_1=1$，$x_2=2$属于ω_1，$x_3=-1$，$x_4=-2$属于ω_2，求：将这两类区分开的线性判别函数的系数w_0，w_1。



示例：权空间、解矢量与解空间

先看一个简单的情况。设4个一维数据 $x_1=1$，$x_2=2$属于 ω_1, $x_3=-1$，$x_4=-2$属于ω_2，求：将这两类区分开的线性判别函数的系数w_0，w_1。

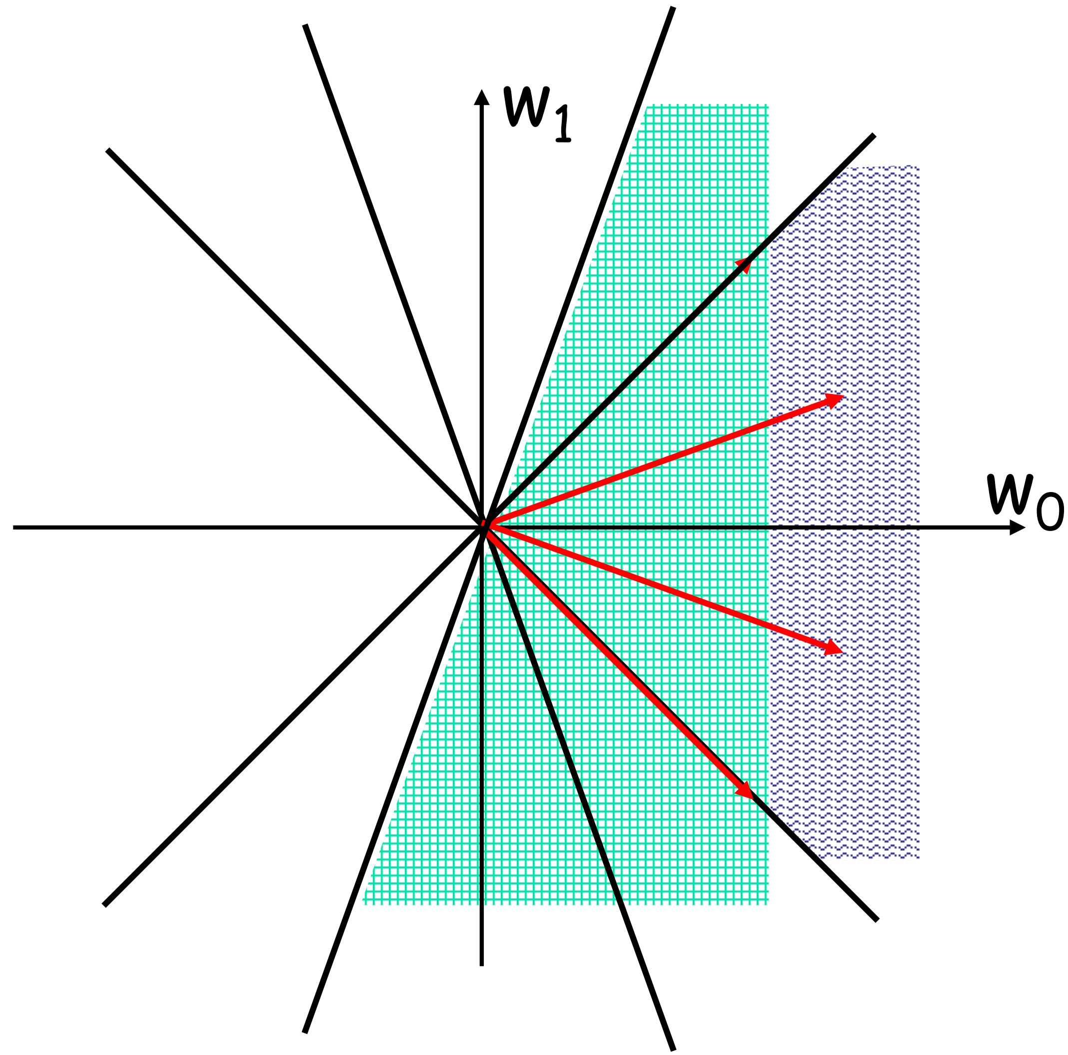
解：

$X^1=(1,1)$,

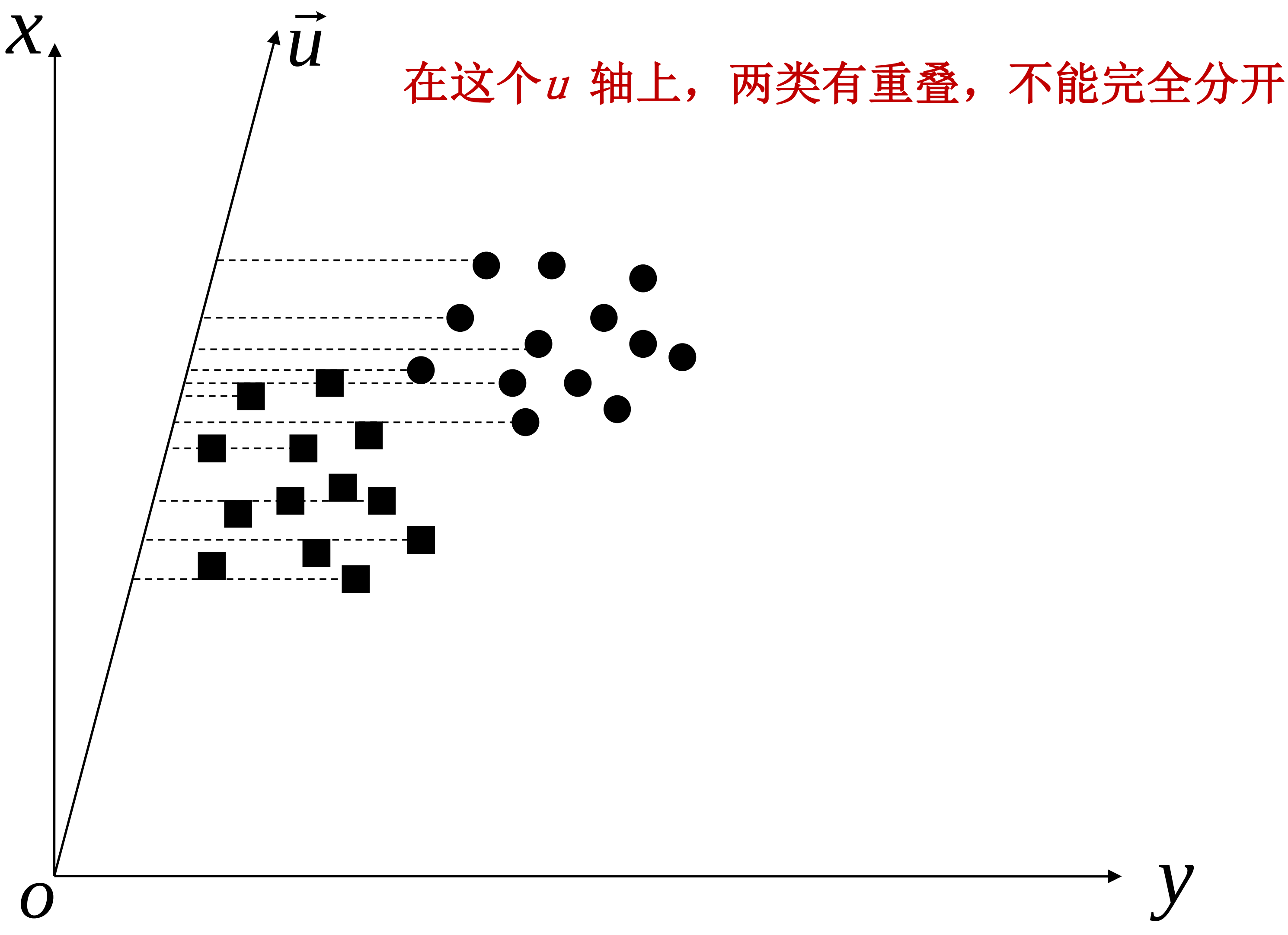
$X^1=(2,1)$,

$X^3=(-1,1)$取反$=(1,-1)$,

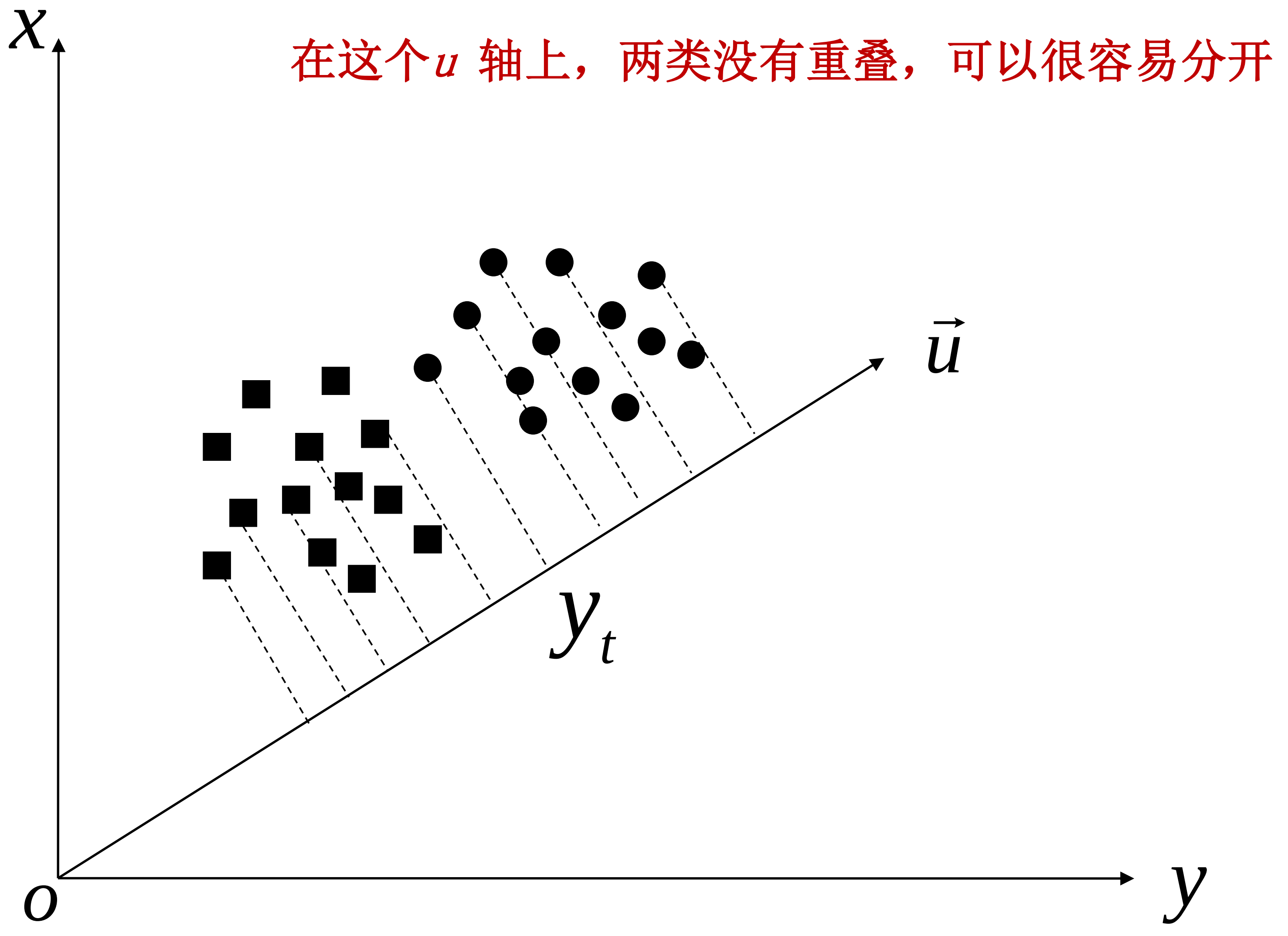
$X^4=(-2,1)$取反$=(2,-1)$,



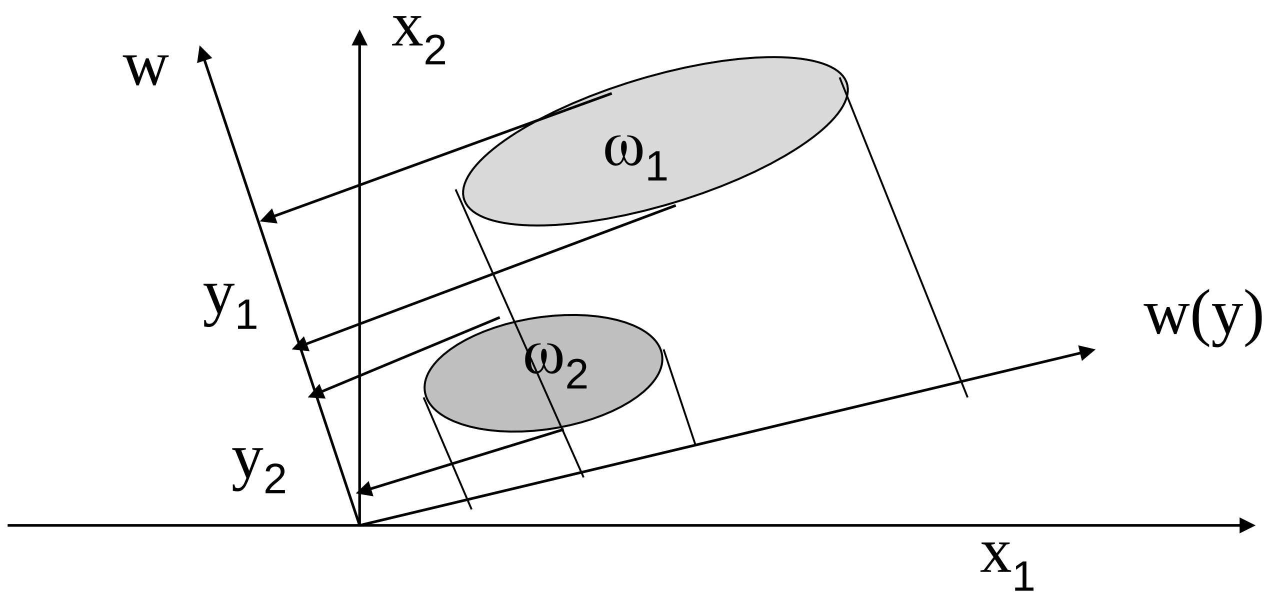
3.5 Fisher线性判决（孙即祥：§3.4）



二维模式向一维空间投影示意图



二维模式向一维空间投影示意图



w
x_2
ω_1
y_1
w(y)
ω_2
y_2
x_1

3.5 Fisher线性判决

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。【孙，§3.4】

思想： 多维 $\Rightarrow$ Fisher变换 $\Rightarrow$ 利于分类的一维

方法： 求权矢量 $\vec{w}$ $\Rightarrow$求满足上述目标的投影轴的方向 $\vec{w}_0$ 和在一维空间中确定判别规则。

Fisher线性判别(FDA)的基本思想

- 线性判别函数 $d(\mathbf{x})=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}$的几个等价含义:
  - 样本向量$\boldsymbol{x}$各分量的线性加权（$\boldsymbol{x}$与权向量u的向量内积）
  - 如果$\boldsymbol{u}$为单位长度，则可视作向量$\boldsymbol{x}$在$\boldsymbol{u}$上的投影长度
- 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准则函数：用投影后数据的统计特征（均值和离散度）的某个优化函数，作为判别分类效果的标准。具体而言：
  - 同类模式向量投影后的点，尽可能聚集
  - 异类模式向量投影后的点，尽可能分离
- 三个步骤
  - （1）确定Fisher判别问题的目标函数
  - （2）求解Fisher最佳投影矢量
  - （3）求解FIsher判别准则函数

(1) 求解Fish准则函数

- 用 $\vec{u}$ 表示待求的 $\vec{w}_0$。

设给定n维训练模式 $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N$，其中有$N_1$和$N_2=N-N_1$ 个模式分属w_1类和w_2类，分别记为 $\left\{\vec{x}_j^{(1)}\right\}$ 和 $\left\{\vec{x}_j^{(2)}\right\}$，各类模式均值矢量为

$$\vec{m}_i = \frac{1}{N_i}\sum_j \vec{x}_j^{(i)} \quad (i=1,2)$$

各类类内离差阵S_{W_i}和总的类内离差阵S_W分别为：

$$S_{W_i} = \sum_j (\vec{x}_j^{(i)} - \vec{m}_i)(\vec{x}_j^{(i)} - \vec{m}_i)' \qquad S_W = S_{W_1} + S_{W_2}$$

取类间离差阵为： $S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)'$

作变换，n维矢量 $\vec{x}$ 在以矢量 $\vec{u}$ 为方向的轴上进行投影：

$$y_j^{(i)} = \vec{u}'\vec{x}_j^{(i)} \qquad \text{i=1,2}$$

变换后在一维 y 空间中各类模式的均值为

$$\widetilde{m}_i = \frac{1}{N_i}\sum_j y_j^{(i)} = \frac{1}{N_i}\sum_j \vec{u}'x_j^{(i)} = \vec{u}'\vec{m}_i \quad \text{i=1,2}$$

类内离差度 $\tilde{s}_{W_i}^2$ 和总的类内离差度 $\tilde{s}_W^2$ 为：

$$\tilde{s}_{W_i}^2 = \sum_j (y_j^{(i)} - \tilde{m}_i)^2 = \sum_j (\vec{u}'\vec{x}_j^{(i)} - \vec{u}'\vec{m}_i)^2 = \vec{u}'S_{W_i}\vec{u}$$

$$\tilde{s}_W^2 = \tilde{s}_{W_1}^2 + \tilde{s}_{W_2}^2 = \vec{u}'(S_{W_1} + S_{W_2})\vec{u} = \vec{u}'S_W\vec{u}$$

类间离差度为：

$$\tilde{s}_B^2 = (\tilde{m}_1 - \tilde{m}_2)^2 = (\vec{u}'\vec{m}_1 - \vec{u}'\vec{m}_2)(\vec{u}'\vec{m}_1 - \vec{u}'\vec{m}_2)' = \vec{u}'S_B\vec{u}$$

希望经投影后，类内离差度 $\tilde{S}_W^2$ 越小越好，类间离差度 $\tilde{S}_B^2$ 越大越好，根据这个目标作准则函数

$$J_F(\vec{u}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{s}_{W_1}^2 + \tilde{s}_{W_2}^2} = \frac{\vec{u}' S_B \vec{u}}{\vec{u}' S_W \vec{u}}$$

并使其最大，上式称为Fisher准则函数。

（2）求解Fisher最佳鉴别矢量

利用二次型关于矢量求导的公式可得：

$$\frac{\partial J_F}{\partial \vec{u}} = \frac{\partial}{\partial \vec{u}}\left[\frac{\vec{u}'S_B\vec{u}}{\vec{u}'S_W\vec{u}}\right] = \frac{2(\vec{u}'S_W\vec{u})S_B\vec{u} - 2(\vec{u}'S_B\vec{u})S_W\vec{u}}{(\vec{u}'S_W\vec{u})^2} = \vec{\phi}$$

令 $\lambda = \dfrac{\vec{u}'S_B\vec{u}}{\vec{u}'S_W\vec{u}}$

可得： $S_B\vec{u} = \lambda S_W\vec{u}$

当N较大时，S_W通常是非奇异的, 于是有：

$$S_W^{-1} S_B \vec{u} = \lambda \vec{u}$$

上式表明，$\vec{u}$ 是矩阵 $S_W^{-1} S_B$ 相应于特征值λ的特征矢量。对于两类问题，S_B 的秩为1，因此，$S_W^{-1} S_B$只有一个非零特征值， 其所对应的特征矢量 $\vec{u}$ 称为Fisher最佳鉴别矢量，由上式可得：

$$\lambda \vec{u} = S_W^{-1} S_B \vec{u} = S_W^{-1} (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)' \vec{u}$$

$$\lambda\vec{u} = S_W^{-1}S_B\vec{u} = S_W^{-1}(\vec{m}_1-\vec{m}_2)(\vec{m}_1-\vec{m}_2)'\vec{u}$$

- 上式右边后两项因子的乘积为一标量，令其为 α ，于是可得

$$\vec{u} = \frac{\alpha}{\lambda}S_W^{-1}(\vec{m}_1-\vec{m}_2)$$

- 式中 $\frac{\alpha}{\lambda}$ 为一标量因子，其不改变轴的方向，可以取为**1**,于是有

$$\vec{u} = S_W^{-1}(\vec{m}_1-\vec{m}_2)$$

此时的 $\vec{u}$ 可使**Fisher**准则函数取最大值，即是n维空间到一维空间投影轴的最佳方向，由

$$\vec{u} = S_W^{-1}(\vec{m}_1 - \vec{m}_2) \text{ 和 } S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)'$$

J_F最大值为:

$$J_F(\vec{u}) = \frac{\vec{u}'S_B\vec{u}}{\vec{u}'S_W\vec{u}}$$

$$= \frac{(\vec{m}_1 - \vec{m}_2)'S_W^{-1}(\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)'S_W^{-1}(\vec{m}_1 - \vec{m}_2)}{(\vec{m}_1 - \vec{m}_2)'S_W^{-1}S_W S_W^{-1}(\vec{m}_1 - \vec{m}_2)}$$

$$= (\vec{m}_1 - \vec{m}_2)'S_W^{-1}(\vec{m}_1 - \vec{m}_2)$$

即

$$J_F = (\vec{m}_1 - \vec{m}_2)' S_W^{-1} (\vec{m}_1 - \vec{m}_2)$$

称 $$y = (\vec{m}_1 - \vec{m}_2)' S_W^{-1} \vec{x}$$

为**Fisher变换函数**

m_1-m_2是一向量，对与**(m_1-m_2)**平行的向量投影可使两均值点的距离最远。但是如从使类间分得较开，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**m_1-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

（3）求解Fisher判别函数

由于变换后的模式是一维的，因此判别界面实际上是各类模式所在轴上的一个点，所以可以根据训练模式确定一个阈值 y_t，于是**Fisher判别规则**为：

$$\vec{u}'\vec{x} = y \gtrless y_t \Rightarrow \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

判别阈值可取两个类心在u方向上轴的投影连线的中点作为阈值，即：

$$y_t = \frac{\tilde{m}_1 + \tilde{m}_2}{2}$$

Fisher方法实现步骤总结:

（1） 把来自两类ω_1/ω_2的训练样本集X分成与ω_1对应的子集X_1和与ω_2对应的子集X_2。

（2） 由 $\vec{m}_i = \frac{1}{N_i}\sum_j \vec{x}_j^{(i)} \quad (i=1,2)$ ，计算$\boldsymbol{m}_i$。

（3） 由 $S_{W_i} = \sum_j (\vec{x}_j^{(i)} - \vec{m}_i)(\vec{x}_j^{(i)} - \vec{m}_i)'$ ，计算各类的类内离差阵S_{W1}，S_{W2}。

（4） 计算类内总离差阵$S_W=S_{W1}+S_{W2}$。

（5） 计算S_W的逆矩阵S_W^{-1}。

（6） 按 $\vec{u} = S_W^{-1}(\vec{m}_1 - \vec{m}_2)$ 求解μ。

（7） 计算$\widetilde{m}_i$。

$$\tilde{m}_i = \frac{1}{N_i}\sum_j y_j^{(i)} = \frac{1}{N_i}\sum_j \vec{u}' x_j^{(i)} = \vec{u}'\vec{m}_i$$

（8） 计算y_t。$y_t = \dfrac{\tilde{m}_1 + \tilde{m}_2}{2}$

（9） 对未知模式x判定模式类。

$$\vec{u}'\vec{x} = y \gtrless y_t \Rightarrow \vec{x} \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

随堂练习

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。

课后作业

- 见另文。
- 下次上课前提交。
- 最好使用电子档。

End of This Part