第三章 判别域代数界面方程法

3.1 判别域界面方程分类的概念

3.2 线性判别函数

3.3 判别函数值的鉴别意义、权空间及解空间

3.4 fisher线性判别

3.5 线性可分条件下判别函数权矢量算法

3.6 一般情况下的判别函数权矢量算法

3.7 非线性判别函数

3.8 最近邻方法

3.3 判别函数值的鉴别意义、权空间及解空间

- 判别函数值的大小、正负的数学意义

n 维特征空间 X^n 中, 两类问题的线性判别界面方程为:

$$d(\vec{x}) = \vec{w}_0'\vec{x} + w_{n+1} = w_1x_1 + w_2x_2 + \cdots + w_nx_n + w_{n+1} = 0$$

3.3 判别函数值的鉴别意义、权空间及解空间

➢ 判别函数值的大小、正负的数学意义

$$d(\vec{x})=\vec{w}_0'\vec{x}+w_{n+1}=w_1x_1+w_2x_2+\cdots+w_nx_n+w_{n+1}=0$$

此方程表示一超平面 π。它有以下三个性质：

(1) 系数矢量 $\vec{w}_0=(w_1,w_2,....w_n)'$ 是该平面的法矢量。

(2) 判别函数 $d(\vec{x})$ 的绝对值正比于 $\vec{x}$ 到超平面 $d(\vec{x})=0$ 的距离。

(3) 判别函数值的正负表示出特征点位于哪个半空间中。

➢ 权空间、解矢量与解空间

（1）权空间

增广特征矢量与增广权矢量是对称的，判别函数可以写成 $d(\vec{x})=\vec{w}'\vec{x}=\vec{x}'\vec{w}=x_1w_1+x_2w_2+\cdots+x_nw_n+w_{n+1}$

这里 $x_1,x_2,\cdots,x_n,1$ 则应视为相应的 w_i 的"权"

$\vec{x}$ 指向平面 $\vec{x}'\vec{w}=0$ 的正侧，即该半空间中的任一点 $\vec{w}$ 都使 $\vec{x}'\vec{w}>0$

而 $\vec{x}$ 背向的半子空间中任一点 $\vec{w}$ 都有 $\vec{x}'\vec{w}<0$

(2) 解矢量

对于两类问题，在对待分类模式进行分类之前，应根据已知类别的增广训练模式 $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N$ 确定线性判别函数 $d(\vec{x}) = \vec{w}'\vec{x}$

当训练模式 $\vec{x}_j \in \omega_1$ 时有 $\vec{w}'\vec{x}_j > 0$

当训练模式 $\vec{x}_j \in \omega_2$ 时有 $\vec{w}'\vec{x}_j < 0$

这时的 $\vec{w}$ 称为解矢量，记为 $\vec{w}^*$

(2) 解矢量

当训练模式 $\vec{x}_j \in \omega_1$ 时有 $\vec{w}\vec{x}_j > 0$

当训练模式 $\vec{x}_j \in \omega_2$ 时有 $\vec{w}\vec{x}_j < 0$

为方便求 $\vec{w}^*$，常需要将已知类别的训练模式符号规范化：当 $\vec{x}$ 属于 ω_1 类时，不改变其符号；当属于 ω_2 类时，改变其符号。

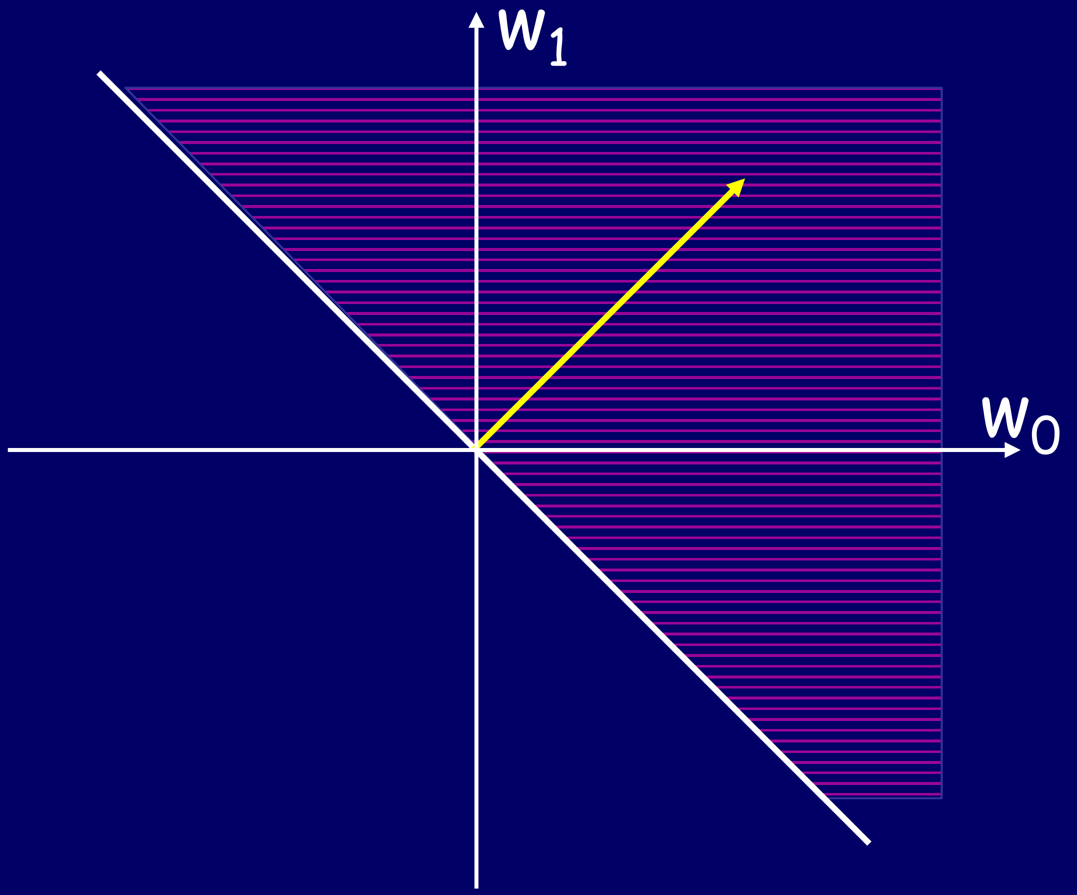
这样对所有的训练模式： $\vec{w}\vec{x}_j > 0$

(2) 解矢量

对于一个训练模式$\vec{x}_j$，界面 $d(\vec{x}_j)=\vec{x}_j'\vec{w}=0$ 过增广权空间原点且将其分成两个子空间，界面 $d(\vec{x}_j)=0$ 的法矢量$\vec{x}_j$指向正半子空间，所谓正半子空间是指该子空间中任一点 $\vec{w}$都使 $\vec{w}'\vec{x}_j>0$。显然，解矢量 $\vec{w}^*$必在正半子空间中。

(3) 解空间

先看一个简单的情况。设一维数据1，1/2属于ω_1, -1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。

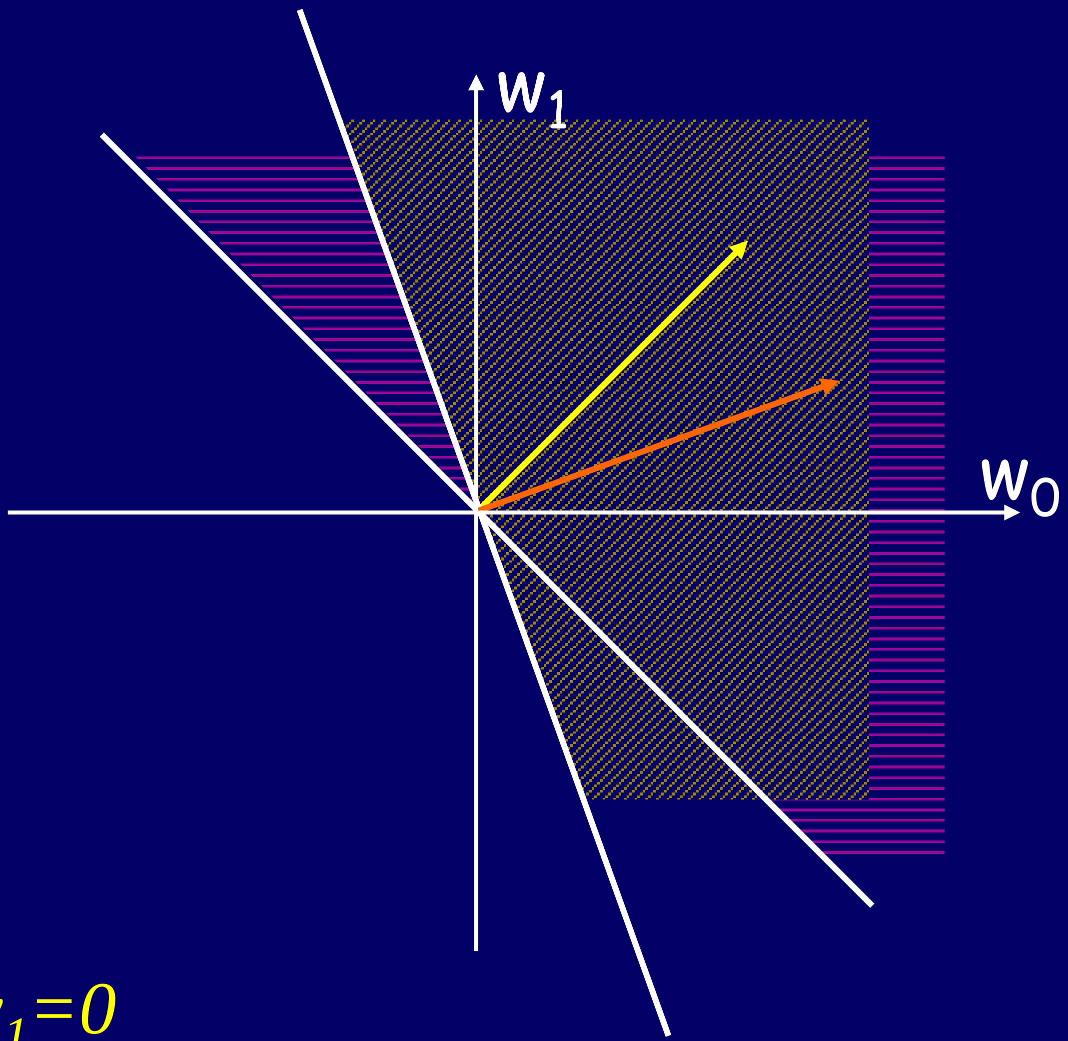


(1,1)

$w_0+w_1=0$

(3) 解空间

先看一个简单的情况。设一维数据1，1/2属于ω_1, -1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。

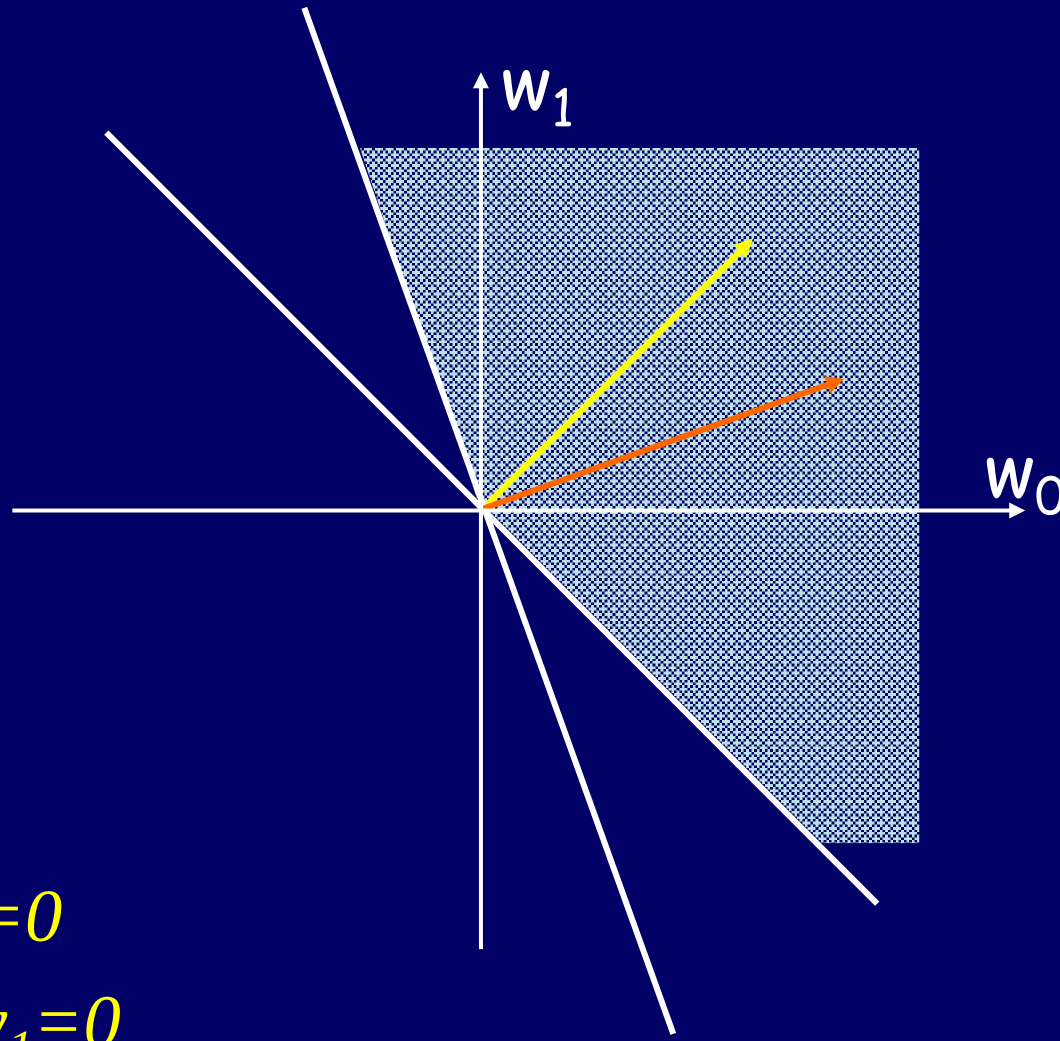


(1/2,1)　　　　$w_0/2+w_1=0$

3.3 判别函数值的鉴别意义、权空间及解空间

(3) 解空间

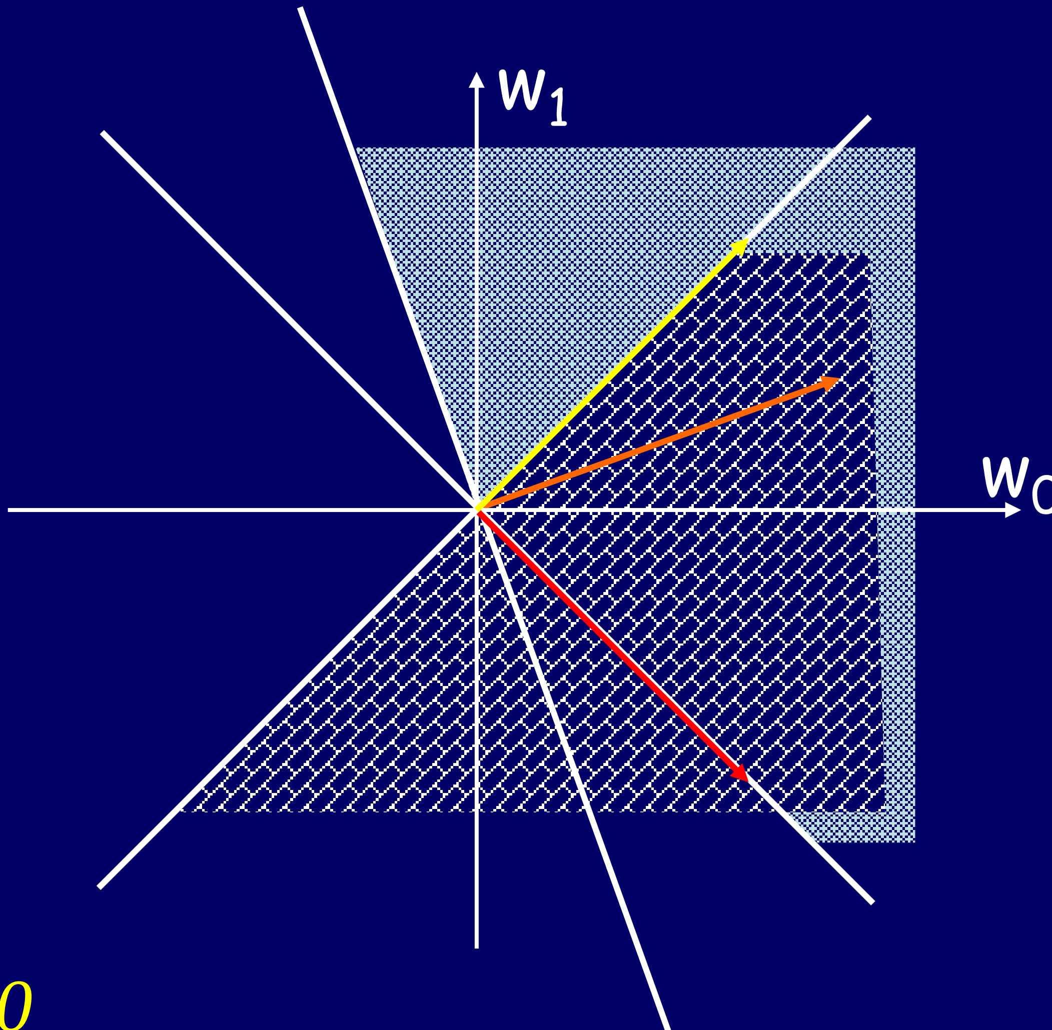
先看一个简单的情况。设一维数据1，1/2属于ω_1, -1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。



$$\begin{cases} w_0+w_1=0 \\ w_0/2+w_1=0 \end{cases}$$

(3) 解空间

先看一个简单的情况。设一维数据 1，1/2属于ω_1, -1，-1/2属于ω_2 求将ω_1和ω_2区分开的 w_0 ，w_1。

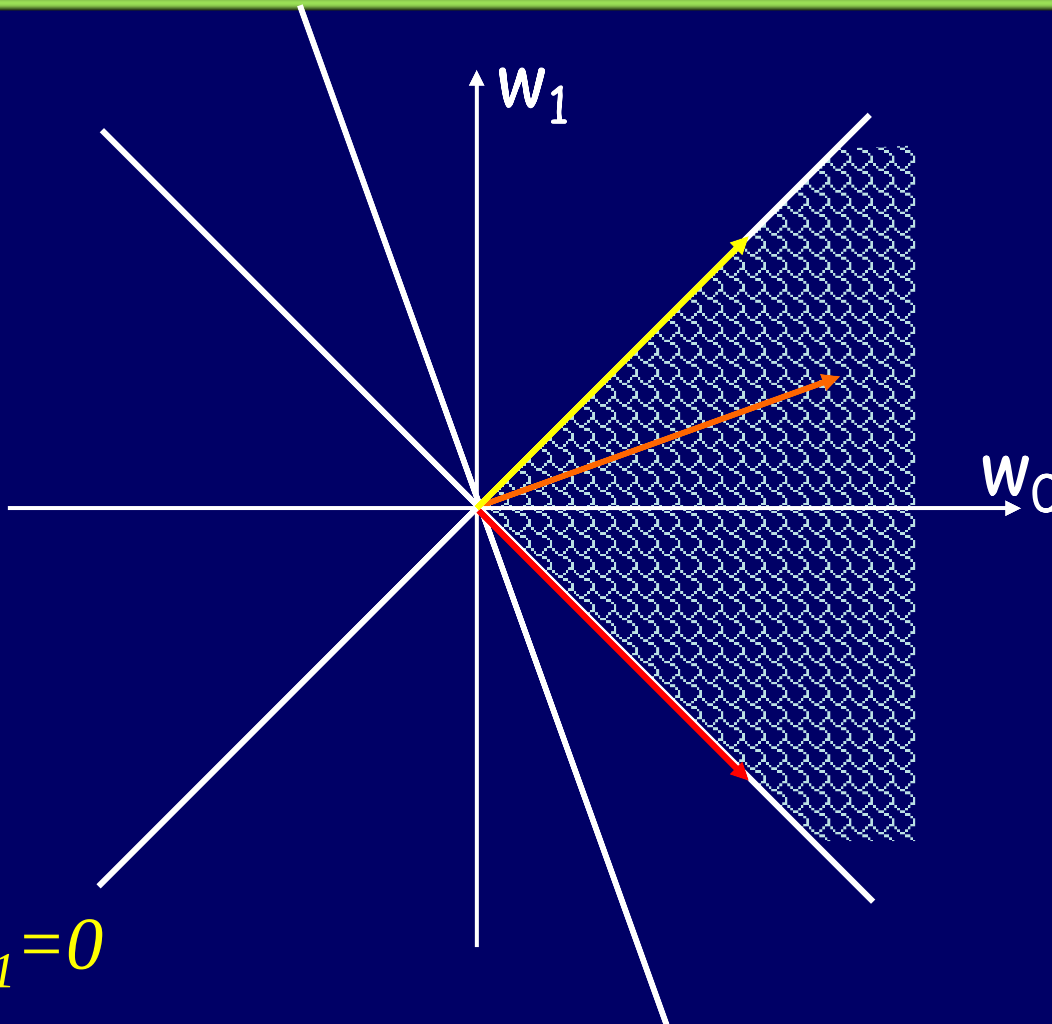


(1,-1)

$w_0 - w_1 = 0$

(3) 解空间

先看一个简单的情况。设一维数据1，1/2属于ω_1，-1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。

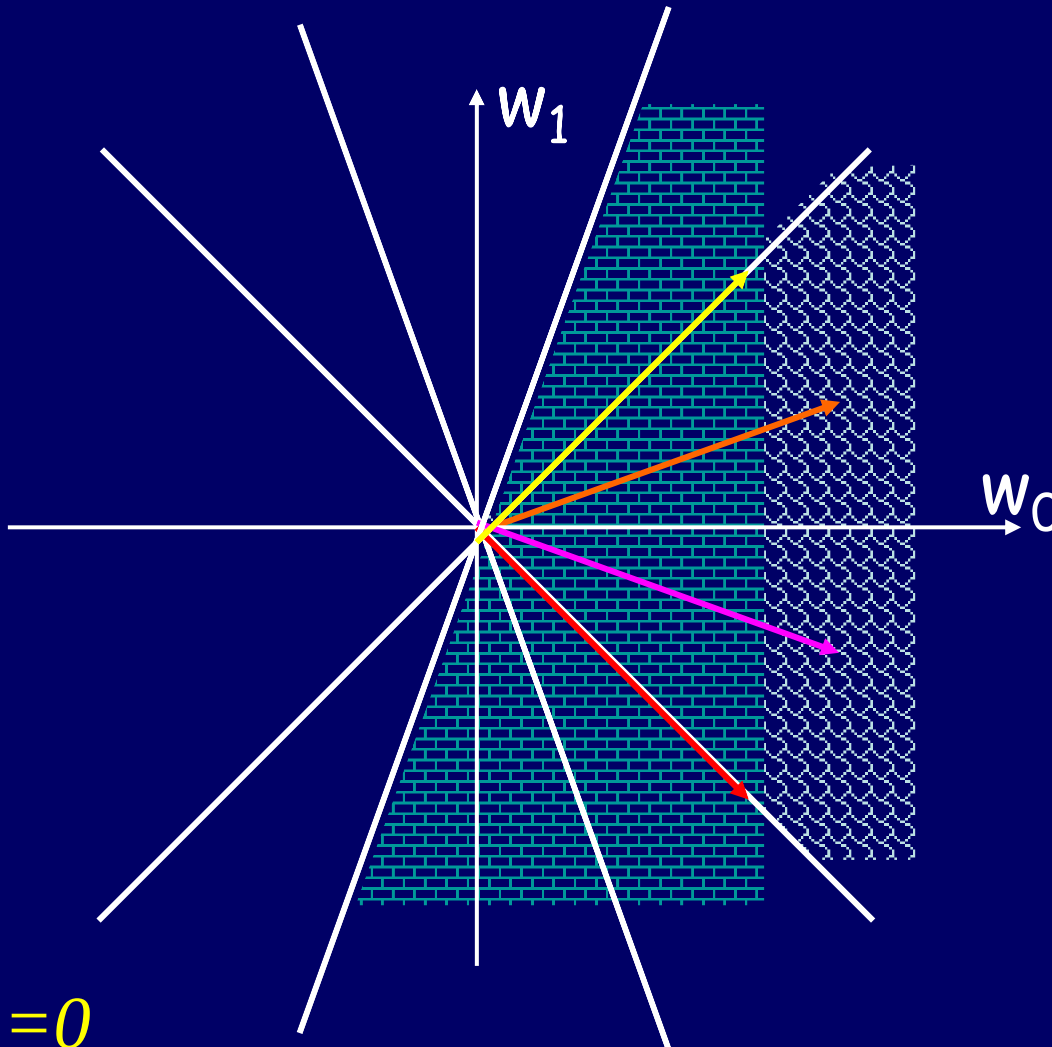


$$\begin{cases} w_0+w_1=0 \\ w_0/2+w_1=0 \\ w_0-w_1=0 \end{cases}$$

3.3 判别函数值的鉴别意义、权空间及解空间

(3) 解空间

先看一个简单的情况。设一维数据1，1/2属于ω_1，-1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。

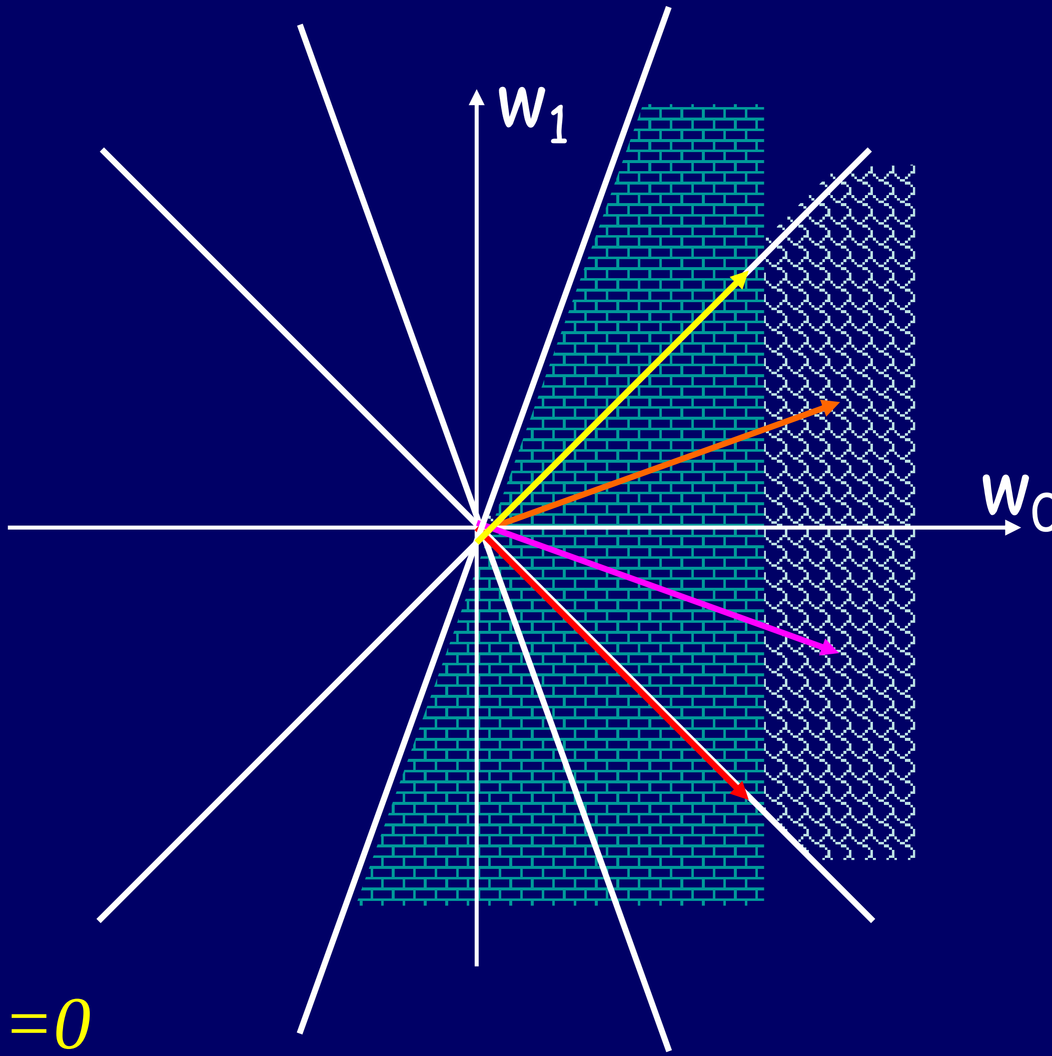


(1/2,-1)

$w_0/2-w_1=0$

(3) 解空间

先看一个简单的情况。设一维数据1，1/2属于ω_1，-1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。



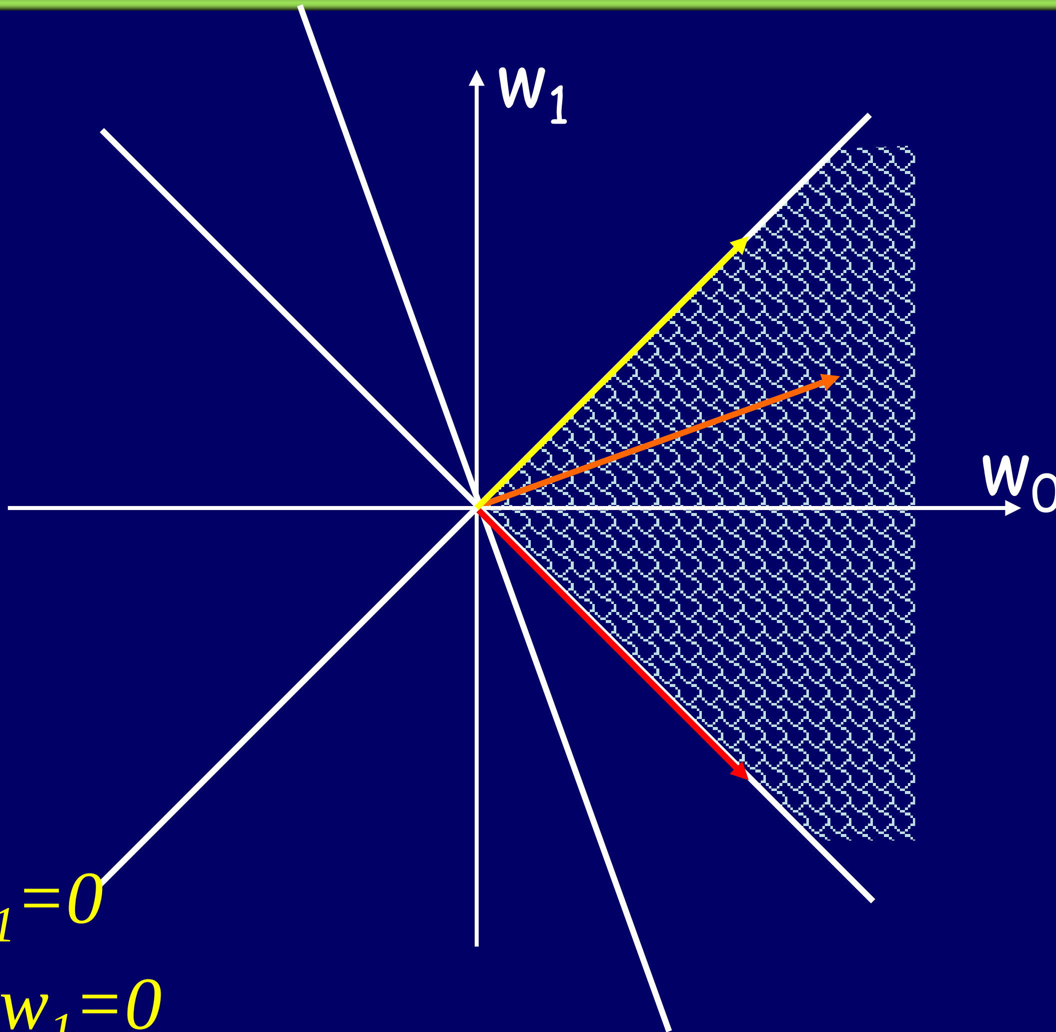
(1/2,-1)

$w_0/2-w_1=0$

3.3 判别函数值的鉴别意义、权空间及解空间

(3) 解空间

先看一个简单的情况。设一维数据1，1/2属于ω_1，-1，-1/2属于ω_2 求将ω_1和ω_2区分开的w_0，w_1。



$$\begin{cases} w_0+w_1=0 \\ w_0/2+w_1=0 \\ w_0-w_1=0 \\ w_0/2-w_1=0 \end{cases}$$

(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。

➢ 权空间、解矢量与解空间

(3) 解空间

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。

➢ 权空间、解矢量与解空间

(4) 余量

为使解矢量可靠，使解区更小，可以采取增加训练模式数以及引入余量 b，使 $\vec{w}'\vec{x} \geq b$ ，从而达到更好的效果。

即由 $\vec{w}\vec{x}_j \geq b > 0 \ (j = 1,2,\cdots,N)$ 所确定的凸多面锥在 $\vec{w}\vec{x}_j > 0 \ (j = 1,2,\cdots,N)$所确定的多面锥的内部，并且它的边界离开原解区边界的距离为 $b/\|\vec{x}_j\|$ 。

- 权空间、解矢量与解空间

(4) 余量

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。

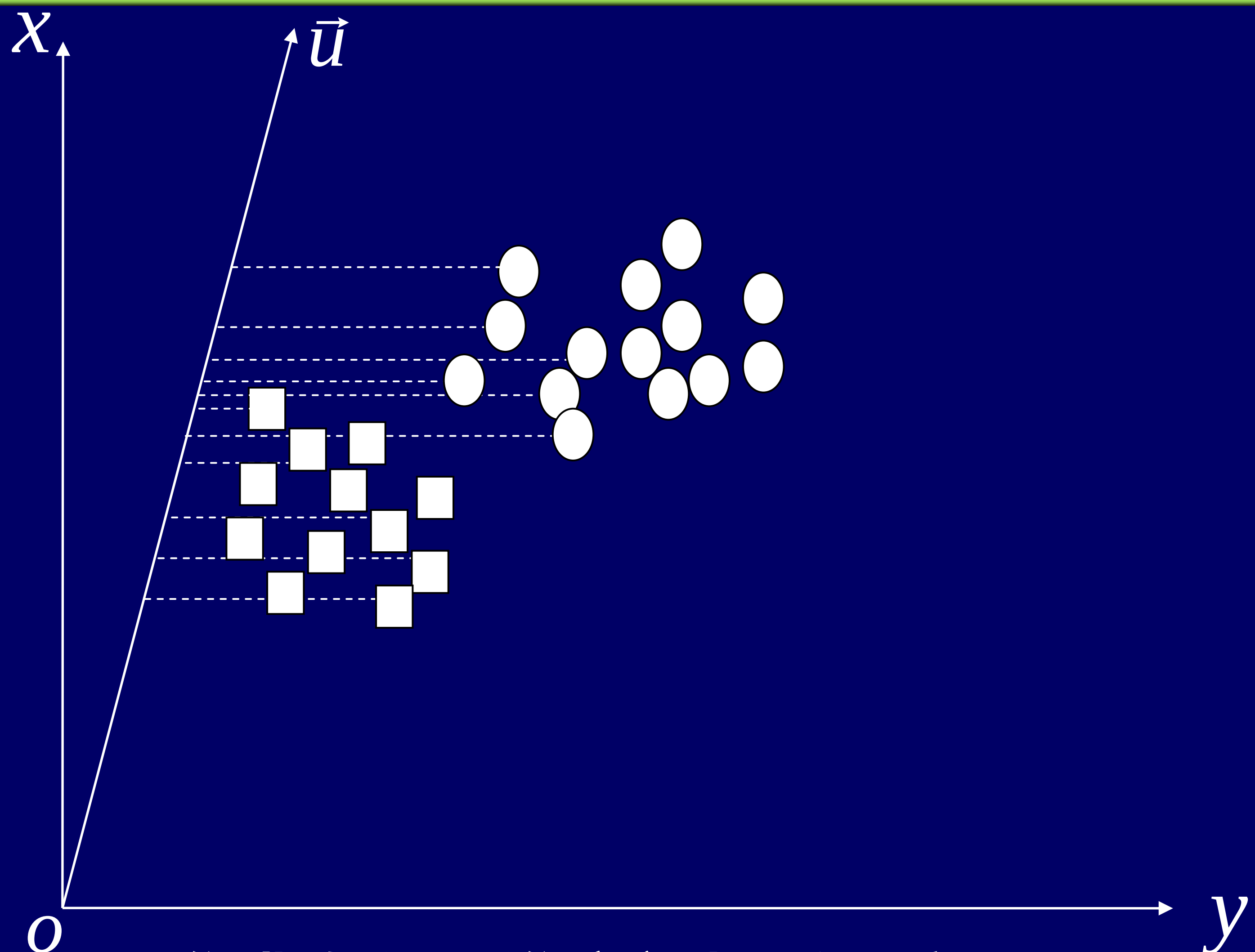
第三章 判别域代数界面方程法

3.4 Fisher线性判别

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

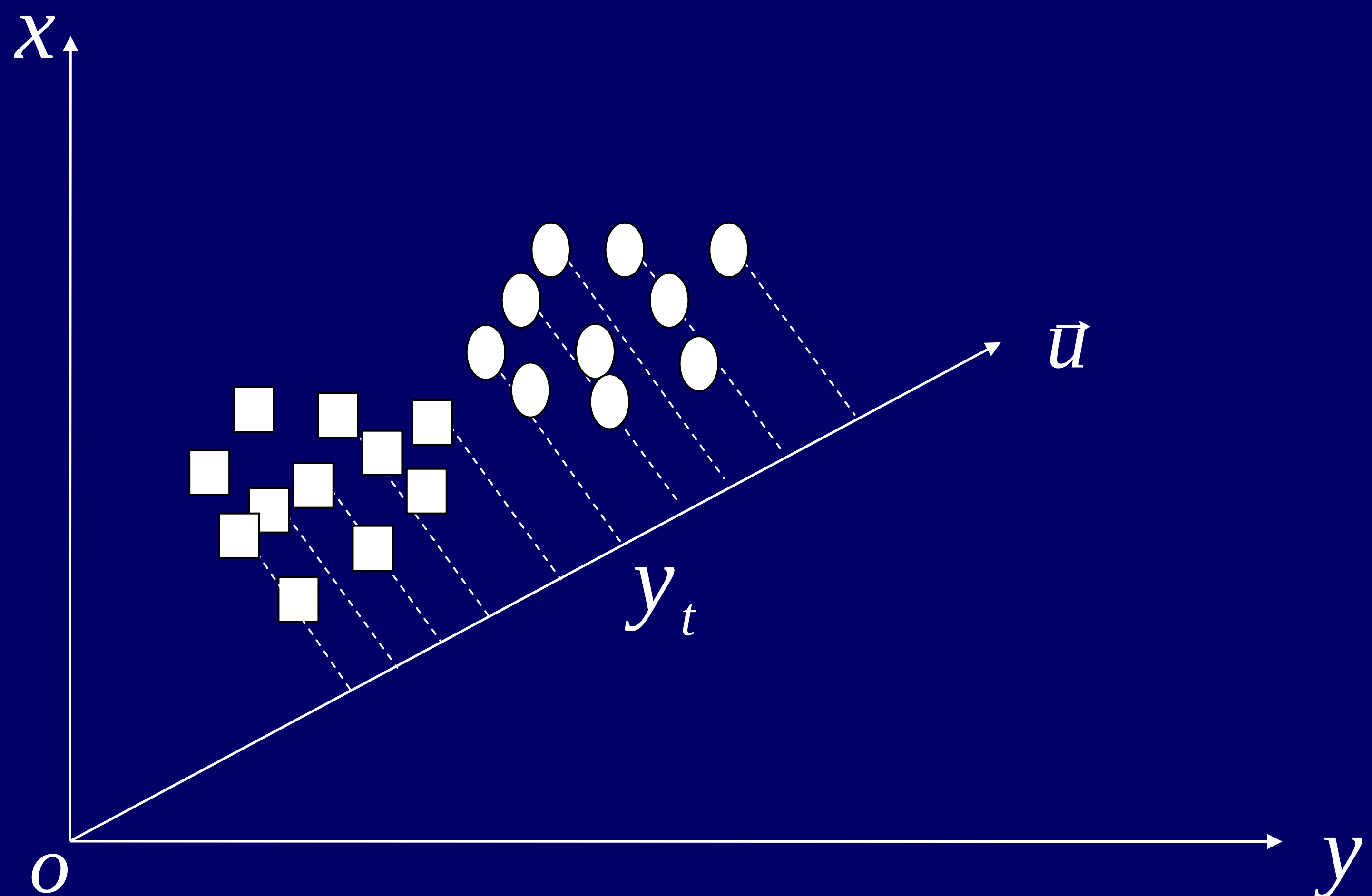
方法：求权矢量 $\vec{w}$ $\Rightarrow$求满足上述目标的投影轴的方向 $\vec{w}_0$ 和在一维空间中确定判别规则。

3.4 Fisher线性判别



二维模式向一维空间投影示意图

3.4 Fisher线性判别



二维模式向一维空间投影示意图

(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ω_1类和ω_2类，分别记为$\{\vec{x}_j^{(1)}\}$和 $\{\vec{x}_j^{(2)}\}$ ，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：

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

(1) 求解Fish准则函数

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：

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

(1) 求解Fish准则函数

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

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

(1)求解Fish准则函数

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

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

(1) 求解Fish准则函数

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

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

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

(1) 求解Fish准则函数

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：

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

(1) 求解Fish准则函数

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： $S_B\vec{u} = \lambda S_W\vec{u}$

（2）求解Fisher最佳鉴别矢量

当N 较大时，S_W 通常是非奇异的，于是有：

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

(2) 求解Fisher最佳鉴别矢量

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：

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

（2）求解Fisher最佳鉴别矢量

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

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

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

（2）求解Fisher最佳鉴别矢量

此时的 $\vec{u}$ 可使**Fisher**准则函数取最大值。J_F最大值为:

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

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

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

（2）求解Fisher最佳鉴别矢量

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

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

为Fisher变换函数

（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） 计算m_i

（3） 计算总得类内离差阵S_W

（4） 计算S_W的逆矩阵S_W^{-1}

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

（6） 计算 $\tilde{m}_i$ 和 y_t

（7） 对未知模式x判定模式类。

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

以100元A面数据和50元A面数据为例

100元A面:(64, 76, 99, 84, 98, 95, 88, 83), …

50元A面:(65, 67, 82, 80, 89, 94, 86, 92), …

$u=S_W^{-1}(m_1-m_2)$ =

(1.09, 0.0284, -0.184, 0.0928, 0.835, -0.832, -0.376, -0.33)

$y_t=5.67$

100元60点y值:(3.62 — 20.5)

3.62,	7.3,	11.9,	14.4,	16.9,	19.0,	20.3,	20.5
19.8,	17.9,	16.8,	15.6,	15.8,	15.1,	15.1,	15.5
13.7,	11.6,	9.08,	9.57,	9.26,	9.78,	8.74,	10.3
11.0,	13.5,	15.1,	14.6,	15.1,	14.6,	16.1,	16.2
17.7,	16.1,	15.2,	13.3,	12.3,	11.9,	11.9,	12.7
14.3,	15.7,	16.5,	14.0,	12.8,	11.5,	10.1,	11.0
10.8,	11.0,	12.6,	13.5,	15.4,	15.4,	13.9,	13.5
13.8,	13.3,	13.6,	13.0,				

y_t=5.67

50元60点y值: (-9.72 — 3.4)

-1.69,	-0.57,	1.89,	3.4,	1.99,	1.4,	-1.71,	-2.29
-3.43,	-3.5,	-3.56,	-3.84,	-2.34,	0.144,	1.67,	1.99
-1.12,	-3.5,	-3.83,	-4.24,	-3.72,	-2.99,	-3.82,	-3.45
-2.61,	-1.82,	-2.75,	-2.94,	-2.82,	-3.05,	-3.24,	-2.1
-1.88,	-1.92,	-1.55,	-1.55,	-1.47,	-2.82,	-2.3,	-2.65
-3.04,	-4.75,	-3.25,	-5.5,	-7.78,	-9.72,	-9.18,	-5.42
-2.55,	0.413,	1.18,	1.23,	-1.99,	-2.48,	-2.31,	-1.86
-0.38,	-1.21,	-2.55,	-5.08,				

谢谢！