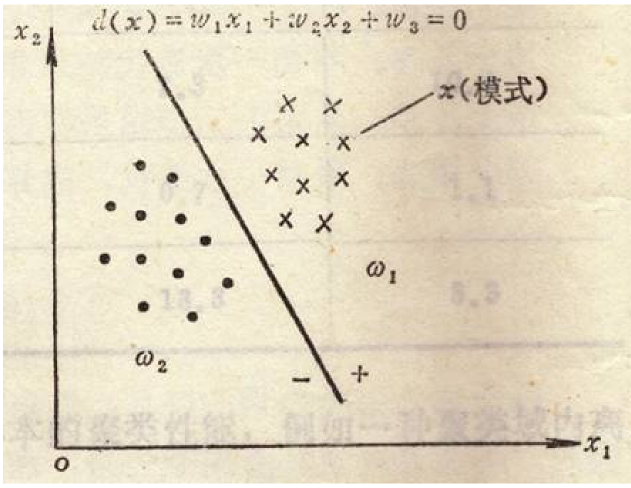
- 两类问题的判别函数（以二维模式样本为例）

若 x 是二维模式样本 $x = (x_1\ x_2)^T$，用 x_1 和 x_2 作为坐标分量，得到模式的平面图：



这时，若这些分属于 ω_1 和 ω_2 两类的模式可用一个直线方程 d(x)=0 来划分

$$d(x) = w_1x_1 + w_2x_2 + w_3 = 0$$

其中 x_1、x_2 为坐标变量，w_1、w_2、w_3 为参数方程，则将一个不知类别的模式代入 d(x)，有

- 若 $d(x) > 0$，则 $x \in \omega_1$
- 若 $d(x) < 0$，则 $x \in \omega_2$

此时，d(x)=0 称为判别函数。

- n 维线性判别函数的一般形式

一个 n 维线性判别函数的一般形式：

$$d(x) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_{n+1} = w_0^T x + w_{n+1}$$

其中 $\mathbf{w_0} = (w_1, w_2, \ldots, w_n)^T$ 称为权向量（或参数向量）, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$。

d(x)也可表示为：

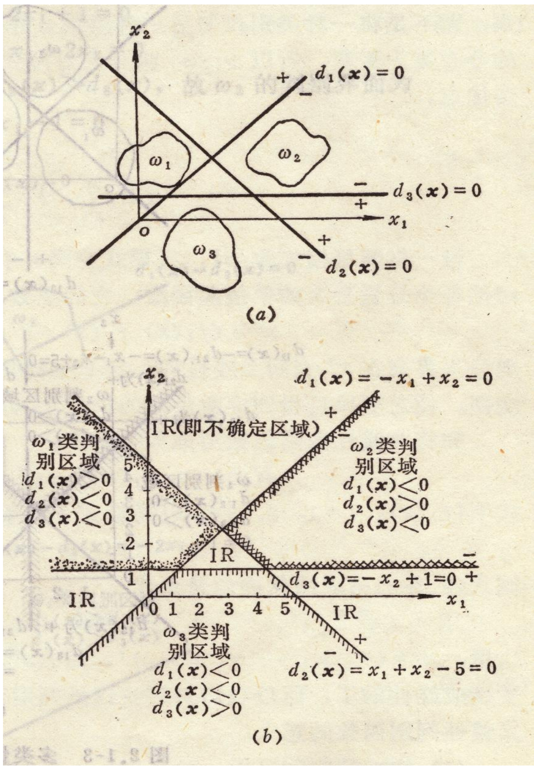
$$d(x) = w^T x$$

其中，$x = (x_1, x_2, \ldots, x_n, 1)^T$ 称为增广模式向量，$w = (w_1, w_2, \ldots, w_{n+1})^T$ 称为增广权向量。

- 两类情况：判别函数 d(x)

$$d(x) = w^T x = \begin{cases} > 0 & \text{if } x \in \omega_1 \\ \leq 0 & \text{if } x \in \omega_2 \end{cases}$$

● 多类情况 1



用线性判别函数将属于 ω_i 类的模式与不属于 ω_i 类的模式分开，其判别函数为：

$$d_i(x) = w_i^T x = \begin{cases} > 0 & \text{if } x \in \omega_i \\ \le 0 & \text{if } x \notin \omega_i \end{cases}$$

$$i = 1, 2, \ldots, M$$

这种情况称为 $\omega_i / \overline{\omega}_i$ 两分法，即把 M 类多类问题分成 M 个两类问题，因此共有 M 个判别函数，对应的判别函数的权向量为 w_i, $i = 1, 2, \ldots, M$。

图例：对一个三类情况，每一类模式可用一个简单的直线判别界面将它与其它类模式分开。

例如对 $x \in \omega_1$ 的模式，应同时满足：$d_1(x)>0$，$d_2(x)<0$，$d_3(x)<0$

不确定区域：若对某一模式区域，$d_i(x)>0$ 的条件超过一个，或全部 $d_i(x)<0$，$i = 1, 2, \ldots, M$，则分类失败，这种区域称为不确定区域(IR)。

例：设有一个三类问题，其判别式为：

$$d_1(x)= -x_1 + x_2,\quad d_2(x)= x_1 + x_2 - 5,\quad d_3(x)= -x_2 + 1$$

则对一个模式 $x=(6, 5)^T$，判断其属于哪一类。

将 $x=(6, 5)^T$ 代入上述判别函数，得：

$d_1(x) = -1$，故 $d_1(x)<0$

$d_2(x) = 6$，故 $d_2(x)>0$

$d_3(x) = -4$，故 $d_3(x)<0$

从而 $x \in \omega_2$

假若 $x=(3, 5)^T$，则

$d_1(x) = 2>0$

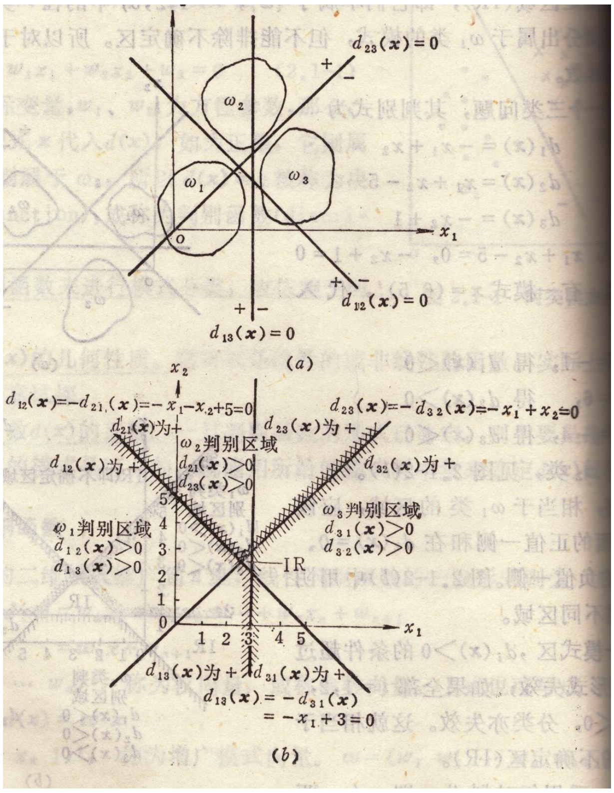
$d_2(x) = 3>0$

$d_3(x) = -4<0$

分类失败。

● 多类情况 2

采用每对划分，即 ω_i/ω_j 两分法，此时一个判别界面只能分开两种类别，但不能把它与其余所有的界面分开。



其判别函数为：

$$d_{ij}(x)=w_{ij}^{T}x$$

若 $d_{ij}(x)>0$，$\forall j\neq i$，则 $x\in\omega_i$

重要性质：$d_{ij}=-d_{ji}$

图例：对一个三类情况，$d_{12}(x)=0$ 仅能分开 ω_1 和 ω_2 类，不能分开 ω_1 和 ω_3 类。

要分开 M 类模式，共需 M(M-1)/2 个判别函数。

不确定区域：若所有 $d_{ij}(x)$，找不到 $\forall j\neq i$，$d_{ij}(x)>0$ 的情况。

例：设有一个三类问题，其判别函数为：

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

若 $x=(4,3)^T$，则：$d_{12}(x)=-2$，$d_{13}(x)=-1$，$d_{23}(x)=-1$

有：$\begin{cases}d_{12}(x)<0\\ d_{13}(x)<0\end{cases}$，$\begin{cases}d_{21}(x)=-d_{12}(x)>0\\ d_{23}(x)<0\end{cases}$，$\begin{cases}d_{31}(x)=-d_{13}(x)>0\\ d_{32}(x)=-d_{23}(x)>0\end{cases}$

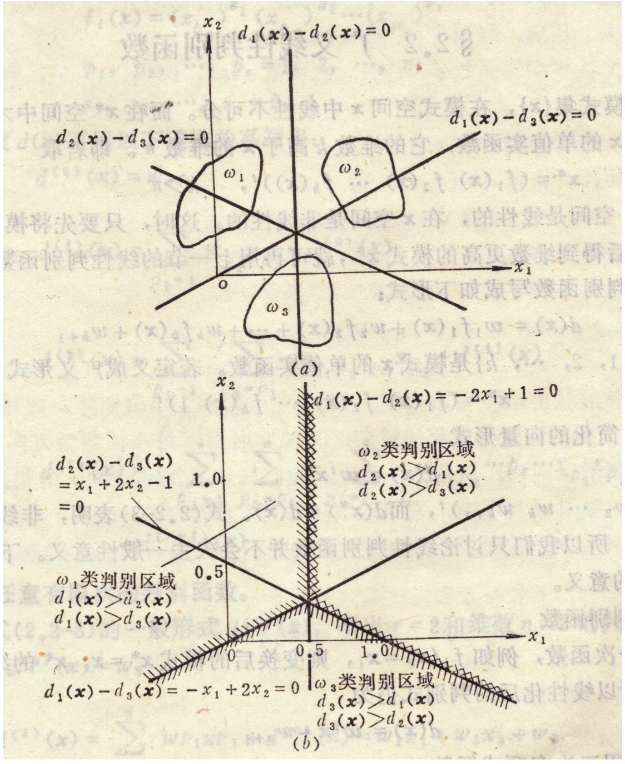
从而 $x\in\omega_3$

若 $x=(2.8,2.5)^T$，则：$d_{12}(x)=-0.3$，$d_{13}(x)=0.2$，$d_{23}(x)=-0.3$

有：$\begin{cases}d_{12}(x)<0\\ d_{13}(x)>0\end{cases}$，$\begin{cases}d_{21}(x)>0\\ d_{23}(x)<0\end{cases}$，$\begin{cases}d_{31}(x)<0\\ d_{32}(x)>0\end{cases}$

分类失败。

- 多类情况 3（多类情况 2 的特例）



这是没有不确定区域的 ω_i/ω_j 两分法。假若多类情况 2 中的 d_{ij} 可分解成：$d_{ij}(x) = d_i(x) - d_j(x) = (w_i - w_j)^T x$，则 $d_{ij}(x)>0$ 相当于 $d_i(x)>d_j(x)$，$\forall j \neq i$，这时不存在不确定区域。此时，对 M 类情况应有 M 个判别函数：

$$d_k(x) = w_k^T x,\ k = 1,2,\ldots,M$$

即 $d_i(x)>d_j(x)$，$\forall j \neq i$，$i, j = 1,2,\ldots,M$，则 $x \in \omega_i$，也可写成，若 $d_i(x)=\max\{d_k(x), k=1,2,\ldots,M\}$，则 $x \in \omega_i$。

该分类的特点是把 M 类情况分成 M−1 个两类问题。

例：设有一个三类问题的模式分类器，其判别函数为：

$$d_1(x)= -x_1 + x_2，\ d_2(x)= x_1 + x_2 - 1，\ d_3(x)= -x_2$$

属于 ω_1 类的区域应满足 $d_1(x)>d_2(x)$ 且 $d_1(x)>d_3(x)$，ω_1 类的判别界面为：

$$d_{12}(x)= d_1(x)-d_2(x) = -2x_1 + 1 = 0$$

$$d_{13}(x)= d_1(x)-d_3(x) = -x_1 + 2x_2 = 0$$

属于 ω_2 类的区域应满足 $d_2(x)>d_1(x)$ 且 $d_2(x)>d_3(x)$，ω_2 类的判别界面为：

$d_{21}(x)= d_2(x)-d_1(x) = 2x_1 - 1 = 0$，可看出 $d_{21}(x)=-d_{12}(x)$

$d_{23}(x)= d_2(x)-d_3(x) = x_1 + 2x_2 - 1= 0$

同理可得 ω_3 类的判别界面为：

$d_{31}(x) = -d_{13}(x) = x_1 - 2x_2 = 0$

$d_{32}(x) = -d_{23}(x) = -x_1 - 2x_2 + 1= 0$

若有模式样本 $x=(1, 1)^T$，则：$d_1(x) = 0$，$d_2(x) = 1$，$d_3(x) = -1$

从而：$d_2(x)>d_1(x)$且 $d_2(x)>d_3(x)$，故 $x \in \omega_2$

- 广义线性判别函数的描述

一个非线性判别函数可如下表示：

$$d(x) = w_1 f_1(x) + w_2 f_2(x) + \cdots + w_k f_k(x) + w_{k+1}$$

其中$\{f_i(x), i = 1,2,\ldots,k\}$是模式 x 的单值实函数。若定义成广义形式：

$$x^* = (f_1(x), f_2(x), \ldots, f_k(x), 1)^T$$

此时有：

$$d(x^*) = w^T x^*，\text{其中 } w = (w_1, w_2, \ldots, w_k, w_{k+1})^T$$

该式表明，非线性判别函数已被变换成广义线性，因此只讨论线性判别函数不会失去一般性意义。

- $f_i(x)$为 r 次多项式函数，x 为 n 维模式，则有

$$f_i(x) = x_{p_1}^{s_1} x_{p_2}^{s_2} \cdots x_{p_r}^{s_r}, \quad p_1, p_2, \ldots, p_r = 1,2,\ldots,n, \quad s_1, s_2, \ldots, s_r = 0,1$$

此时，判别函数 d(x)可用以下递推关系给出：

常数项：$d^{(0)}(x) = w_{n+1}$

一次项：$d^{(1)}(x) = \sum_{p_1=1}^{n} w_{p_1} x_{p_1} + d^{(0)}(x)$

二次项：$d^{(2)}(x) = \sum_{p_1=1}^{n} \sum_{p_2=p_1}^{n} w_{p_1 p_2} x_{p_1} x_{p_2} + d^{(1)}(x)$

r 次项：

$$d^{(r)}(x) = \sum_{p_1=1}^{n} \sum_{p_2=p_1}^{n} \cdots \sum_{p_r=p_{r-1}}^{n} w_{p_1 p_2 \cdots p_r} x_{p_1} x_{p_2} \cdots x_{p_r} + d^{(r-1)}(x)$$

d(x)总项数的讨论：对于 n 维 x 向量，若用 r 次多项式，d(x)的权系数的总项数为：$N_w = C_{n+r}^{r} = \dfrac{(n+r)!}{r!n!}$

当 r=2 时：$N_w = C_{n+2}^{2} = \dfrac{(n+2)!}{2!n!} = \dfrac{(n+2)(n+1)}{2}$

当 r=3 时：$N_w = C_{n+3}^{3} = \dfrac{(n+3)!}{3!n!} = \dfrac{(n+3)(n+2)(n+1)}{6}$

- 用 $d^{(r)}(x)$ 写出 r=2 和 n=2 时的判别函数

常数项：$d^{(0)}(x) = w_{n+1} = w_3$

一次项：$d^{(1)}(x) = \sum_{p_1=1}^{n} w_{p_1} x_{p_1} + d^{(0)}(x) = w_1 x_1 + w_2 x_2 + w_3$

二次项：
$$
\begin{aligned}
d^{(2)}(x) &= \sum_{p_1=1}^{n} \sum_{p_2=p_1}^{n} w_{p_1 p_2} x_{p_1} x_{p_2} + d^{(1)}(x) \\
&= w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + w_1 x_1 + w_2 x_2 + w_3
\end{aligned}
$$

- 线性判别函数

取 $f_i(x)$为一次函数，例如 x_i，则变换后的模式 $x^*=x$，x^*的维数 k 为 x 的维数 n，此时广义线性化后的判别式仍为：

$$d(x) = w^T x + w_{n+1}$$

- $f_i(x)$选用二次多项式函数

1. x 是二维的情况，即 $x =(x_1\ x_2)^T$。若原判别函数为：

$$d(x) = w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 + w_1x_1 + w_2x_2 + w_3$$

要线性化为 $d(x^*) = w^T x^*$，须定义：

$$x^* = (x_1^2\ \ x_1x_2\ \ x_2^2\ \ x_1\ \ x_2\ \ 1)^T$$

$$w = (w_{11}\ \ w_{12}\ \ w_{22}\ \ w_1\ \ w_2\ \ w_3)^T$$

此时，只要把模式空间 x^*中的分量定义成 x 的单值实函数，x^*即变成线性可分。此时 x^*的维数（这里为 5）大于 x 的维数（这里为 2）。

2. x 是 n 维的情况，此时原判别函数设为：

$$d(x) = \sum_{j=1}^{n} w_{jj}x_j^2 + \sum_{j=1}^{n-1}\sum_{k=j+1}^{n} w_{jk}x_jx_k + \sum_{j=1}^{n} w_jx_j + w_{n+1}$$

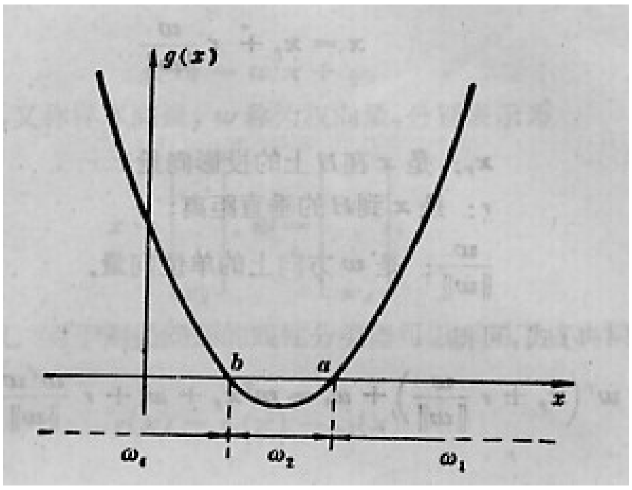
式中各项的组成应包含 x 的各个分量的二次项、一次项和常数项，其中平方项 n 个，二次项 n(n-1)/2 个，一次项 n 个，常数项一个，其总项数为：

$$n + n(n-1)/2 + n + 1 = (n+1)(n+2)/2 > n$$

显然，对于 $d(x^*) = w^T x^*$，x^*的维数大于 x 的维数，w 分量的数目也与 x^*的维数相应。x^*的各分量的一般化形式为：

$$f_i(x) = x_{p_1}^s x_{p_2}^t,\ p_1, p_2 = 1, 2, \ldots, n, \quad s, t = 0, 1$$

- 广义线性判别实例



如图所示，设有一维样本空间 X，所希望的分类是：

$$若 x \le b 或 x \ge a, x \in \omega_1;\quad 若 b < x < a, x \in \omega_2$$

显然没有一个线性判别函数能在一维空间中解决上述问题。

要在一维空间中分类，只有定义判别函数

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

将此分类问题转化到二维空间，令

$$x_1 = f_1(x) = x^2,\quad x_2 = f_2(x) = x$$

则可以定义线性判别函数

$$d(x) = x_1-(a+b)x_2+ab = w^T x$$

此时

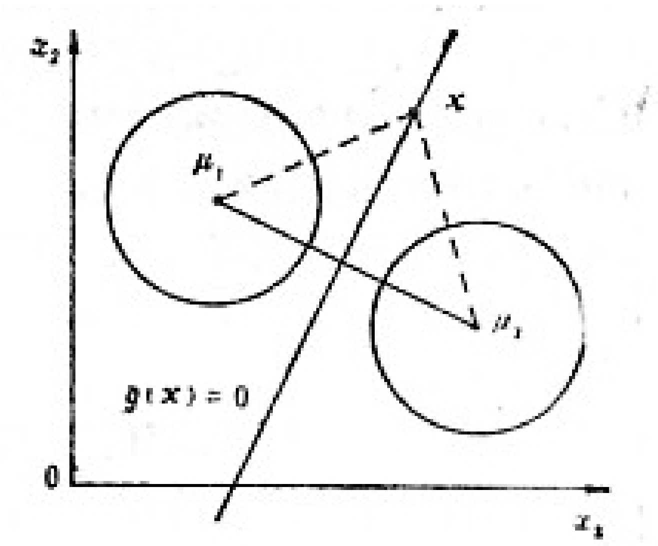
$$x = (x_1\ \ x_2\ \ 1)^T,\quad w = (1\ \ -(a+b)\ \ ab)^T$$

- 最小距离分类

设μ$_1$和μ$_2$为两个模式类ω$_1$和ω$_2$的聚类中心，定义决策规则：

$$\|x-\mu_1\|^2-\|x-\mu_2\|^2\begin{cases}<0 & x\in\omega_1\\ >0 & x\in\omega_2\end{cases}$$

这时的决策面是两类期望连线的垂直平分面，这样的分类器称为最小距离分类器。



- 分类描述

设有判别函数：$d(x)=w^Tx$，其中 $x=(x_1\ x_2 \dots x_n\ 1)^T$，$w=(w_1\ w_2 \dots w_n\ w_{n+1})^T$

判别界面为：$w^Tx=0$

对两类问题，ω_1类有模式$\{x_1\ x_2\}$，ω_2类有模式$\{x_3\ x_4\}$，则应满足如下条件：

$$\left.\begin{matrix} w^Tx_1>0 & w^Tx_2>0 \\ w^Tx_3<0 & w^Tx_4<0 \end{matrix}\right\}$$

若将属于ω_2类的模式都乘以(-1)，则上式可写成：

$$\left.\begin{matrix} w^Tx_1>0 & w^Tx_2>0 \\ w^T(-x_3)>0 & w^T(-x_4)>0 \end{matrix}\right\}$$

因此，若权向量能满足上述四个条件，则 $w^Tx=0$ 为所给模式集的判别界面。

- Fisher 准则函数

Fisher 准则函数定义为：

$$J_F(w)=\frac{(\tilde{m}_1-\tilde{m}_2)^2}{\tilde{S}_1^2+\tilde{S}_2^2}$$

其中，$(\tilde{m}_1-\tilde{m}_2)$是两类均值之差，$\tilde{S}_i^2$ 是样本类内离散度。显然，应该使 $J_F(w)$的分子尽可能大而分母尽可能小，即应寻找使 $J_F(w)$尽可能大的 w 作为投影方向。但上式中并不显含 w，因此须设法将 $J_F(w)$变成 w 的显函数。

由各类样本的均值可推出：

$$\tilde{m}_i=\frac{1}{N_i}\sum_{y\in\Gamma_i'}y=\frac{1}{N_i}\sum_{x\in\Gamma_i}w^Tx=w^T\left(\frac{1}{N_i}\sum_{x\in\Gamma_i}x\right)=w^Tm_i$$

这样，Fisher 准则函数 $J_F(w)$的分子可写成：

$$\begin{aligned}(\tilde{m}_1-\tilde{m}_2)^2&=(w^Tm_1-w^Tm_2)^2\\&=(w^Tm_1-w^Tm_2)(w^Tm_1-w^Tm_2)^T\\&=(w^Tm_1-w^Tm_2)(m_1^Tw-m_2^Tw)\\&=w^T(m_1-m_2)(m_1-m_2)^Tw=w^TS_bw\end{aligned}$$

现在再来考察 $J_F(w)$的分母与 w 的关系：

$$\begin{aligned}\tilde{S}_i^2&=\sum_{y\in\Gamma_i'}(y-\tilde{m}_i)^2=\sum_{x\in\Gamma_i}(w^Tx-w^Tm_i)^2\\&=w^T\left[\sum_{x\in\Gamma_i}(x-m_i)(x-m_i)^T\right]w=w^TS_iw\end{aligned}$$

因此，

$$\tilde{S}_1^2+\tilde{S}_2^2=w^T(S_1+S_2)w=w^TS_ww$$

将上述各式代入 $J_F(w)$，可得：

$$J_F(w)=\frac{w^TS_bw}{w^TS_ww}$$

其中 S_b 为样本类间离散度矩阵，S_w 为总样本类内离散度矩阵。

- Fisher 准则函数中的基本参量

1．在 d 维 X 空间

（1）各类样本的均值向量 $\mathbf{m_i}$

$$m_i = \frac{1}{N_i}\sum_{x\in\Gamma_i} x,\ i = 1,2$$

（2）样本类内离散度矩阵 S_i 和总样本类内离散度矩阵 S_w

$$S_i = \sum_{x\in\Gamma_i}(x - m_i)(x - m_i)^T,\ i = 1,2$$

$$S_w = S_1 + S_2$$

其中 S_w 是对称半正定矩阵，而且当 N>d 时通常是非奇异的。

（3）样本类间离散度矩阵 S_b

$$S_b = (m_1 - m_2)(m_1 - m_2)^T$$

S_b 是对称半正定矩阵。

2．在一维 Y 空间

（1）各类样本的均值 $\tilde{m}_i$

$$\tilde{m}_i = \frac{1}{N_i}\sum_{y\in\Gamma_i'} y,\ i = 1,2$$

（2）样本类内离散度 $\tilde{S}_i^2$ 和总样本类内离散度 $\tilde{S}_w$

$$\tilde{S}_i^2 = \sum_{y\in\Gamma_i'}(y - \tilde{m}_i)^2,\ i = 1,2$$

$$\tilde{S}_w = \tilde{S}_1^2 + \tilde{S}_2^2$$

● 最佳变换向量 w*的求取

为求使 $J_F(w)=w^T S_b w / w^T S_w w$ 取极大值时的 w*，可以采用 Lagrange 乘数法求解。令分母等于非零常数，即：

$$w^T S_w w = c \neq 0$$

定义 Lagrange 函数为：

$$L(w,\lambda)=w^T S_b w-\lambda(w^T S_w w-c)$$

其中 λ 为 Lagrange 乘子。将上式对 w 求偏导数，可得：

$$\frac{\partial L(w,\lambda)}{\partial w}=2(S_b w-\lambda S_w w)$$

令偏导数为零，有；

$$S_b w^*-\lambda S_w w^*=0$$

即

$$S_b w^*=\lambda S_w w^*$$

其中 w*就是 $J_F(w)$的极值解。因为 S_w 非奇异，将上式两边左乘 S_w^{-1}，可得：

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

上式为求一般矩阵 $S_w^{-1} S_b$ 的特征值问题。利用 $S_b=(m_1-m_2)(m_1-m_2)^T$ 的定义，将上式左边的 $S_b w^*$ 写成：

$$S_b w^*=(m_1-m_2)(m_1-m_2)^T w^*=(m_1-m_2)R$$

其中 $R=(m_1-m_2)^T w^*$ 为一标量，所以 $S_b w^*$ 总是在向量 (m_1-m_2) 的方向上。因此 λ w*可写成：

$$\lambda w^*=S_w^{-1}(S_b w^*)=S_w^{-1}(m_1-m_2)R$$

从而可得：

$$w^* = \frac{R}{\lambda} S_w^{-1}(m_1 - m_2)$$

由于我们的目的是寻找最佳的投影方向，w*的比例因子对此并无影响，因此可忽略比例因子 R/ λ , 有：

$$w^* = S_w^{-1}(m_1 - m_2)$$

- 感知器的训练算法

已知两个训练模式集分别属于ω_1类和ω_2类，权向量的初始值为w(1)，可任意取值。若$x_k \in \omega_1, w^T(k)x_k > 0$，若$x_k \in \omega_2, w^T(k)x_k \le 0$，则在用全部训练模式集进行迭代训练时，第k次的训练步骤为：

- 若$x_k \in \omega_1$且$w^T(k)x_k \le 0$，则分类器对第k个模式x_k做了错误分类，此时应校正权向量，使得$w(k+1) = w(k) + Cx_k$，其中C为一个校正增量。
- 若$x_k \in \omega_2$且$w^T(k)x_k > 0$，同样分类器分类错误，则权向量应校正如下：$w(k+1) = w(k) - Cx_k$
- 若以上情况不符合，则表明该模式样本在第k次中分类正确，因此权向量不变，即：$w(k+1) = w(k)$

若对$x \in \omega_2$的模式样本乘以(−1)，则有：

$$w^T(k)x_k \le 0 \text{ 时，} w(k+1) = w(k) + Cx_k$$

此时，感知器算法可统一写成：

$$w(k+1) = \begin{cases} w(k) & \text{if } w^T(k)x_k > 0 \\ w(k) + Cx_k & \text{if } w^T(k)x_k \le 0 \end{cases}$$

- 感知器的训练算法实例

将属于ω_2的训练样本乘以（-1），并写成增广向量的形式。

$x_{①}=(0\ \ 0\ \ 1)^T$，$x_{②}=(0\ \ 1\ \ 1)^T$，$x_{③}=(-1\ \ 0\ \ -1)^T$，$x_{④}=(-1\ \ -1\ \ -1)^T$

第一轮迭代：取 C=1，$w(1)=(0\ \ 0\ \ 0)^T$

因 $w^T(1)x_{①}=(0\ \ 0\ \ 0)(0\ \ 0\ \ 1)^T=0\not>0$，故 $w(2)=w(1)+x_{①}=(0\ \ 0\ \ 1)^T$

因 $w^T(2)x_{②}=(0\ \ 0\ \ 1)(0\ \ 1\ \ 1)^T=1>0$，故 $w(3)=w(2)=(0\ \ 0\ \ 1)^T$

因 $w^T(3)x_{③}=(0\ \ 0\ \ 1)(-1\ \ 0\ \ -1)^T=-1\not>0$，故 $w(4)=w(3)+x_{③}=(-1\ \ 0\ \ 0)^T$

因 $w^T(4)x_{④}=(-1\ \ 0\ \ 0)(-1\ \ -1\ \ -1)^T=1>0$，故 $w(5)=w(4)=(-1\ \ 0\ \ 0)^T$

这里，第 1 步和第 3 步为错误分类，应“罚”。

因为只有对全部模式都能正确判别的权向量才是正确的解，因此需进行第二轮迭代。

第二轮迭代：

因 $w^T(5)x_{①}=(-1\ \ 0\ \ 0)(0\ \ 0\ \ 1)^T=0\not>0$，故 $w(6)=w(5)+x_{①}=(-1\ \ 0\ \ 1)^T$

因 $w^T(6)x_{②}=(-1\ \ 0\ \ 1)(0\ \ 1\ \ 1)^T=1>0$，故 $w(7)=w(6)=(-1\ \ 0\ \ 1)^T$

因 $w^T(7)x_{③}=(-1\ \ 0\ \ 1)(-1\ \ 0\ \ -1)^T=0\not>0$，故 $w(8)=w(7)+x_{③}=(-2\ \ 0\ \ 0)^T$

因 $w^T(8)x_{④}=(-2\ \ 0\ \ 0)(-1\ \ -1\ \ -1)^T=2>0$，故 $w(9)=w(8)=(-2\ \ 0\ \ 0)^T$

需进行第三轮迭代。

第三轮迭代：

因 $w^T(9)x_{①}=(-2\ \ 0\ \ 0)(0\ \ 0\ \ 1)^T=0\not>0$，故 $w(10)=w(9)+x_{①}=(-2\ \ 0\ \ 1)^T$

因 $w^T(10)x_{②}=(-2\ \ 0\ \ 1)(0\ \ 1\ \ 1)^T=1>0$，故 $w(11)=w(10)=(-2\ \ 0\ \ 1)^T$

因 $w^T(11)x_{③}=(-2\ \ 0\ \ 1)(-1\ \ 0\ \ -1)^T=1>0$，故 $w(12)=w(11)=(-2\ \ 0\ \ 1)^T$

因 $w^T(12)x_{④}=(-2\ \ 0\ \ 1)(-1\ \ -1\ \ -1)^T=1>0$，故 $w(13)=w(12)=(-2\ \ 0\ \ 1)^T$

需进行第四轮迭代。

第四轮迭代：

因 $w^T(13)x_{①}=1>0$，故 $w(14)=w(13)=(-2\ \ 0\ \ 1)^T$

因 $w^T(14)x_{②}=1>0$，故 $w(15)=w(10)=(-2\ \ 0\ \ 1)^T$

因 $w^T(15)x_{③}=1>0$，故 $w(16)=w(11)=(-2\ \ 0\ \ 1)^T$

因 $w^T(16)x_{④}=1>0$，故 $w(17)=w(12)=(-2\ \ 0\ \ 1)^T$

该轮的迭代全部正确，因此解向量 $w=(-2\ \ 0\ \ 1)^T$，相应的判别函数为：

$$d(x)=-2x_1+1$$

- 感知器算法判别函数的推导

多类情况 3：对 M 类模式存在 M 个判别函数$\{d_i, i = 1,2,\ldots,M\}$，

若$x \in \omega_i$，则$d_i > d_j$，$\forall j \neq i$。

设有 M 种模式类别ω_1，ω_2，…，ω_M，若在训练过程的第 k 次迭代时，一个属于ω_i类的模式样本 x 送入分类器，则应先计算出 M 个判别函数：

$$d_j(k) = w_j(k)x, \quad j = 1,2,\ldots,M$$

若$d_i(k) > d_j(k)$，$j = 1,2,\ldots,M, \forall j \neq i$的条件成立，则权向量不变，即

$$w_j(k+1) = w_j(k), \quad j = 1,2,\ldots,M$$

若其中第 l 个权向量使得$d_i(k) \leq d_l(k)$，则相应的权向量应做调整，即

$$\begin{cases} w_i(k+1) = w_i(k) + Cx \\ w_l(k+1) = w_l(k) - Cx \\ w_j(k+1) = w_j(k),\ j = 1,2,\ldots,M,\ j \neq i,\ j \neq l \end{cases}$$

其中 C 是一个正常数。权向量的初始值 $w_i(1)$，$i = 1,2,\ldots,M$ 可视情况任意选择。

- 感知器算法判别函数的推导实例

给出三类模式的训练样本：

ω_1:{$(0\ \ 0)^T$}， ω_2:{$(1\ \ 1)^T$}， ω_3:{$(-1\ \ 1)^T$}

将模式样本写成增广形式：

$x_{①}=(0\ \ 0\ \ 1)^T$， $x_{②}=(1\ \ 1\ \ 1)^T$， $x_{③}=(-1\ \ 1\ \ 1)^T$

取初始值 $w_1(1)=w_2(1)=w_3(1)=(0\ \ 0\ \ 0)^T$，C=1。

第一轮迭代（k=1）：以 $x_{①}=(0\ \ 0\ \ 1)^T$ 作为训练样本

$d_1(1)=w_1^T(1)\,x_{①}=(0\ \ 0\ \ 0)(0\ \ 0\ \ 1)^T=0$

$d_2(1)=w_2^T(1)\,x_{①}=(0\ \ 0\ \ 0)(0\ \ 0\ \ 1)^T=0$

$d_3(1)=w_3^T(1)\,x_{①}=(0\ \ 0\ \ 0)(0\ \ 0\ \ 1)^T=0$

因 $d_1(1)\ngtr d_2(1)$，$d_1(1)\ngtr d_3(1)$，故

$w_1(2)=w_1(1)+x_{①}=(0\ \ 0\ \ 1)^T$

$w_2(2)=w_2(1)-x_{①}=(0\ \ 0\ \ -1)^T$

$w_3(2)=w_3(1)-x_{①}=(0\ \ 0\ \ -1)^T$

第二轮迭代（k=2）：以 $x_{②}=(1\ \ 1\ \ 1)^T$ 作为训练样本

$d_1(2)=w_1^T(2)\,x_{②}=(0\ \ 0\ \ 1)(1\ \ 1\ \ 1)^T=1$

$d_2(2)=w_2^T(2)\,x_{②}=(0\ \ 0\ \ -1)(1\ \ 1\ \ 1)^T=-1$

$d_3(2)=w_3^T(2)\,x_{②}=(0\ \ 0\ \ -1)(1\ \ 1\ \ 1)^T=-1$

因 $d_2(2)\ngtr d_1(2)$，$d_2(2)\ngtr d_3(2)$，故

$w_1(3)=w_1(2)-x_{②}=(-1\ \ -1\ \ 0)^T$

$w_2(3)=w_2(2)+x_{②}=(1\ \ 1\ \ 0)^T$

$w_3(3)=w_3(2)-x_{②}=(-1\ \ -1\ \ -2)^T$

第三轮迭代（k=3）：以 $x_{③}=(-1\ \ 1\ \ 1)^T$ 作为训练样本

$d_1(3)=w_1^T(3)x_{③}=(-1\ \ -1\ \ 0)(-1\ \ 1\ \ 1)^T=0$

$d_2(3)=w_2^T(3)x_{③}=(1\ \ 1\ \ 0)(-1\ \ 1\ \ 1)^T=0$

$d_3(3)=w_3^T(3)x_{③}=(-1\ \ -1\ \ -2)(-1\ \ 1\ \ 1)^T=-2$

因 $d_3(3)\ngtr d_1(3)$，$d_3(3)\ngtr d_2(3)$，故

$w_1(4)=w_1(3)-x_{③}=(0\ \ -2\ \ -1)^T$

$w_2(4)=w_2(3)-x_{③}=(2\ \ 0\ \ -1)^T$

$w_3(4)=w_3(3)+x_{③}=(-2\ \ 0\ \ -1)^T$

第四轮迭代（k=4）：以 $x_{①}=(0\ \ 0\ \ 1)^T$ 作为训练样本

$d_1(4)=w_1^T(4)x_{①}=(0\ \ -2\ \ -1)(0\ \ 0\ \ 1)^T=-1$

$d_2(4)=w_2^T(4)x_{①}=(2\ \ 0\ \ -1)(0\ \ 0\ \ 1)^T=-1$

$d_3(4)=w_3^T(4)x_{①}=(-2\ \ 0\ \ -1)(0\ \ 0\ \ 1)^T=-1$

因 $d_1(4)\ngtr d_2(4)$，$d_1(4)\ngtr d_3(4)$，故

$w_1(5)=w_1(4)+x_{①}=(0\ \ -2\ \ 0)^T$

$w_2(5)=w_2(4)-x_{①}=(2\ \ 0\ \ -2)^T$

$w_3(5)=w_3(4)-x_{①}=(-2\ \ 0\ \ -2)^T$

第五轮迭代（k=5）：以 $x_{②}=(1\ \ 1\ \ 1)^T$ 作为训练样本

$d_1(5)=w_1^T(5)x_{②}=(0\ \ -2\ \ 0)(1\ \ 1\ \ 1)^T=-2$

$d_2(5)=w_2^T(5)x_{②}=(2\ \ 0\ \ -2)(1\ \ 1\ \ 1)^T=0$

$d_3(5)=w_3^T(5)x_{②}=(-2\ \ 0\ \ -2)(1\ \ 1\ \ 1)^T=-4$

因 $d_2(5)>d_1(5)$，$d_2(5)>d_3(5)$，故

$w_1(6)=w_1(5)$

$w_2(6)=w_2(5)$

$w_3(6)=w_3(5)$

第六轮迭代（k=6）：以 $x_{③}=(-1\ \ 1\ \ 1)^T$ 作为训练样本

$d_1(6)=w_1^T(6)x_{③}=(0\ \ -2\ \ 0)(-1\ \ 1\ \ 1)^T=-2$

$d_2(6)=w_2^T(6)x_{③}=(2\ \ 0\ \ -2)(-1\ \ 1\ \ 1)^T=-4$

$d_3(6)=w_3^T(6)x_{③}=(-2\ \ 0\ \ -2)(-1\ \ 1\ \ 1)^T=0$

因 $d_3(6)>d_1(6)$，$d_3(6)>d_2(6)$，故

$w_1(7)=w_1(6)$

$w_2(7)=w_2(6)$

$w_3(7)=w_3(6)$

第七轮迭代（k=7）：以 $x_{①}=(0\ \ 0\ \ 1)^T$ 作为训练样本

$d_1(7)=w_1^T(7)x_{①}=(0\ \ -2\ \ 0)(0\ \ 0\ \ 1)^T=0$

$d_2(7)=w_2^T(7)x_{①}=(2\ \ 0\ \ -2)(0\ \ 0\ \ 1)^T=-2$

$d_3(7)=w_3^T(7)x_{①}=(-2\ \ 0\ \ -2)(0\ \ 0\ \ 1)^T=-2$

因 $d_1(7)>d_2(7)$，$d_1(7)>d_3(7)$，分类结果正确，故权向量不变。

由于第五、六、七次迭代中 $x_{①}$、$x_{②}$、$x_{③}$均已正确分类，所以权向量的解为：

$w_1=(0\ \ -2\ \ 0)^T$

$w_2=(2\ \ 0\ \ -2)^T$

$w_3=(-2\ \ 0\ \ -2)^T$

三个判别函数：

$d_1(x)=-2x_2$

$d_2(x)=2x_1-2$

$d_3(x)=-2x_1-2$

- 梯度法定义

设函数 f(y)是向量 $y = (y_1, y_2, \ldots, y_n)^T$ 的函数，则 f(y)的梯度定义为:

$$\nabla f(y) = \frac{d}{dy} f(y) = \left(\frac{\partial f}{\partial y_1} \ \frac{\partial f}{\partial y_2} \ \cdots \ \frac{\partial f}{\partial y_n} \right)^T$$

- 从 w(k)导出 w(k+1)的一般关系式

$$w(k+1) = w(k) - C\left\{\frac{\partial J(w,x)}{\partial w}\right\}_{w=w(k)} = w(k) - C \cdot \nabla J$$

C 是一个正的比例因子（步长）

- 梯度法实例

选准则函数 $J(w,x)=|w^Tx|-w^Tx$，考虑一维模式的情况。

设训练样本 x 为 1。

当错误分类时，$w^Tx<0$，J 为正值，$\bigtriangledown J=-2$，则 $w(k+1)=w(k)+2C$，对权向量进行校正。

当正确分类时，$w^Tx>0$，$J=0$，$\bigtriangledown J=0$，则 $w(k+1)=w(k)$，权向量不变。

- 固定增量的逐次调整算法

设取准则函数为：

$$J(w,x)=\frac{|w^Tx|-w^Tx}{2}$$

则 J 对 w 的微分式：

$$\frac{\partial J}{\partial w}=\frac{1}{2}[x\cdot \mathrm{sign}(w^Tx)-x]$$

定义：

$$\mathrm{sign}(w^Tx)=\begin{cases}+1 & \text{if } w^Tx>0\\ -1 & \text{if } w^Tx\le 0\end{cases}$$

则由梯度法中 w(k+1)和 w(k)的关系有：

$$w(k+1)=w(k)+\frac{C}{2}[x_k-x_k\cdot \mathrm{sign}(w^T(k)x_k)]$$

其中 x_k 是训练模式样本，k 是指第 k 次迭代。

$$w(k+1)=w(k)+C\cdot\begin{cases}0 & \text{if } w^Tx>0\\ x_k & \text{if } w^Tx\le 0\end{cases}$$

可以看出，当 $w^Tx>0$ 时，则 w(k+1) = w(k)，此时不对权向量进行修正；当 $w^Tx\le 0$ 时，则 w(k+1) = w(k) + Cx_k，需对权向量进行校正，初始权向量 w(1)的值可任选，显然这就是前面所说的感知器算法，因此感知器算法是梯度法的一个特例。在上式中 C 是预先选好的固定值，在迭代过程中，只要 $w^Tx\le 0$，就要对权向量修正 Cx_k 值，因此称为固定增量算法。

- H-K 算法

H-K 算法是求解 Xw=b，式中 $b=(b_1, b_2, \ldots, b_n)^T$，b 的所有分量都是正值。这里要同时计算 w 和 b，我们已知 X 不是 N*N 的方阵，通常是行多于列的 N*(n+1)阶的长方阵，属于超定方程，因此一般情况下，Xw=b 没有唯一确定解，但可求其线性最小二乘解。

设 Xw=b 的线性最小二乘解为 w*，即使||Xw*-b||=极小

采用梯度法，定义准则函数：

$$J(w,x,b)=\frac{1}{2}\sum_{i=1}^{n}(w^T x_i-b_i)^2=\frac{1}{2}\|Xw-b\|^2=\frac{1}{2}(Xw-b)^T(Xw-b)$$

当 Xw=b 的条件满足时，J 达到最小值。由于上式中包括的 $\sum_{i=1}^{n}(w^T x_i-b_i)^2$ 项为两个数量方差的和，且我们将使其最小化，因此也称之为最小均方误差算法。

使函数 J 同时对变量 w 和 b 求最小。对于 w 的梯度为：

$$\frac{\partial J}{\partial w}=X^T(Xw-b)$$

使 $\frac{\partial J}{\partial w}=0$，得 $X^T(Xw-b)=0$，从而 $X^TXw=X^Tb$。因为 X^TX 为(n+1)*(n+1)阶方阵，因此可求得解：

$$w=(X^TX)^{-1}X^Tb=X^{\#}b$$

这里 $X^{\#}=(X^TX)^{-1}X^T$ 称为 X 的伪逆，X 是 N*(n+1)阶的长方阵。

由上式可知，只要求出 b 即可求得 w。利用梯度法可求得 b 的迭代公式为：

$$b(k+1) = b(k) - C \cdot \left(\frac{\partial J}{\partial b} \right)_{b=b(k)}$$

根据上述约束条件，在每次迭代中，b(k)的全部分量只能是正值。由 J 的准则函数式，J 也是正值，因此，当取校正增量 C 为正值时，为保证每次迭代中的 b(k)都是正值，应使$\left(\frac{\partial J}{\partial b} \right)_{b=b(k)}$为非正值。在此条件下，准则函数 J 的微分为：

$$-2\left(\frac{\partial J}{\partial b} \right)_{b=b(k)} = (Xw - b) + |Xw - b|$$

该式满足以下条件：

$$若[Xw(k) - b(k)] > 0，则 -\left(\frac{\partial J}{\partial b} \right)_{b=b(k)} = Xw(k) - b(k)$$

$$若[Xw(k) - b(k)] < 0，则 -\left(\frac{\partial J}{\partial b} \right)_{b=b(k)} = 0$$

由 b 的迭代式和微分，有：

$$b(k+1) = b(k) + \delta b(k)$$

$$\delta b(k) = C[Xw(k) - b(k) + |Xw(k) - b(k)|]$$

将此式代入 $w=X^{\#}b$，有：

$$w(k+1) = X^{\#}b(k+1) = X^{\#}[b(k) + \delta b(k)] = w(k) + X^{\#} \delta b(k)$$

为简化起见，令 e(k) = Xw(k) – b(k)，可得 H-K 算法的迭代式。

设初值为 b(1)，其每一分量均为正值，则：

$$w(1) = X^{\#}b(1)$$

$$e(k) = Xw(k) - b(k)$$

$$\begin{aligned} w(k+1) &= w(k) + X^{\#}\{C[Xw(k) - b(k) + |Xw(k) - b(k)|]\} \\ &= w(k) + CX^{\#}[e(k) + |e(k)|] \end{aligned}$$

由于

$$X^{\#}e(k) = X^{\#}[Xw(k) - b(k)] = (X^TX)^{-1}X^T[Xw(k) - b(k)]$$

$$= w(k) - X^{\#}b(k) = 0$$

因此

$$w(k+1) = w(k) + CX^{\#}|e(k)|$$

$$b(k+1) = b(k) + C[Xw(k) - b(k) + |Xw(k) - b(k)|]$$

$$= b(k) + C[e(k) + |e(k)|]$$

- 分类器的不等式方程

求两类问题的解相当于求一组线性不等式的解，因此，若给出分别属于ω_1和ω_2的两个模式样本的训练样本集，即可求出其权向量 w 的解，其性质应满足：

$$\mathrm{w}^{\mathrm{T}}\mathrm{x} > 0, \forall \mathrm{x} \in \omega_1$$

$$\mathrm{w}^{\mathrm{T}}\mathrm{x} < 0, \forall \mathrm{x} \in \omega_2$$

将属于ω_2的模式乘以(−1)，可得对于全部模式都有 $\mathrm{w}^{\mathrm{T}}\mathrm{x} > 0$ 的条件。

设两类模式的训练样本总数为 N，写成增广形式，则有不等式组:

$$\mathrm{X}\mathbf{w} > \mathbf{0}$$

式中：

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_i^T \\ -x_{i+1}^T \\ \vdots \\ -x_N^T \end{pmatrix}, \quad \begin{pmatrix} x_1^T \\ \vdots \\ x_i^T \end{pmatrix} \in \omega_1, \quad \begin{pmatrix} -x_{i+1}^T \\ \vdots \\ -x_N^T \end{pmatrix} \in \omega_2$$

$$\mathbf{w} = (\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_n, \mathrm{w}_{n+1})^{\mathrm{T}}$$

其中，$\mathbf{0}$ 是零向量，$\mathrm{x}_{\mathrm{i}}^{\mathrm{T}}$ 是第 i 个 n 维模式样本的增广向量，即 $x_i = (x_{i_1}, x_{i_2}, \cdots, x_{i_n}, 1)^T$，$i = 1,2,\cdots,N$，它包括分属于$\omega_1$和$\omega_2$中全部供训练用的样本，但属于$\omega_2$类的模式应乘以(−1)，所以 X 是一个 N*(n+1)阶的矩阵。

- 模式类别可分性的判别

当不等式组 Xw>0 有解时，该算法对 $0 < C \le 1$ 收敛，可求得解 w。

(i) 若 e(k)=0，即 Xw(k)=b(k)>0，有解。

(ii) 若 e(k)>0，此时隐含 $Xw(k) \ge b(k) > 0$ 的条件，有解。若继续进行迭代，可使 e(k)−>0。

(iii)若 e(k)的全部分量停止变为正值（但不是全部为零），表明该模式类别线性不可分。因此，若 e(k)没有一个分量为正值，则 b(k)不会再变化，所以不能求得解。

- LMSE 算法实例

1. 有解情况

已知模式样本集：ω_1: $\{(0\ 0)^T, (0\ 1)^T\}$，ω_2: $\{(1\ 0)^T, (1\ 1)^T\}$

模式的增广矩阵 X 为：$X=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{pmatrix}$

其伪逆矩阵为：$X^{\#}=(X^TX)^{-1}X^T=\frac{1}{2}\begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$

取 $b(1)=(1\ 1\ 1\ 1)^T$ 和 C=1，由 H-K 算法的迭代式：

$$w(1)=X^{\#}b(1)=(-2\ 0\ 1)^T$$

因 $Xw(1)=(1\ 1\ 1\ 1)^T$，即 $e(1)= Xw(1)-b(1)= (0\ 0\ 0\ 0)^T$，故 w(1)是解。

2. 无解情况

已知模式样本集：ω_1: $\{(0\ 0)^T, (1\ 1)^T\}$，ω_2: $\{(0\ 1)^T, (1\ 0)^T\}$

模式的增广矩阵 X 为：$X=\begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & -1 \end{pmatrix}$

其伪逆矩阵为：$X^{\#}=(X^TX)^{-1}X^T=\frac{1}{2}\begin{pmatrix} -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix}$

取 $b(1)=(1\ 1\ 1\ 1)^T$ 和 C=1，由 H-K 算法的迭代式：

$$w(1)=X^{\#}b(1)=(0\ 0\ 0)^T$$

则：$e(1)= Xw(1)-b(1)= (-1\ -1\ -1\ -1)^T$，全部分量均为负，无解。

- 判别函数产生逐步分析

设初始势函数 $K_0(x) = 0$

第一步：加入第一个训练样本 x_1，则有

$$K_1(x)=\begin{cases} K(x,x_1) & \text{if } x_1 \in \omega_1 \\ -K(x,x_1) & \text{if } x_1 \in \omega_2 \end{cases}$$

这里第一步积累势函数 $K_1(x)$描述了加入第一个样本时的边界划分。当样本属于 ω_1 时，势函数为正；当样本属于 ω_2 时，势函数为负。

第二步：加入第二个训练样本 x_2，则有

（i） 若 $x_2 \in \omega_1$ 且 $K_1(x_2)>0$，或 $x_2 \in \omega_2$ 且 $K_1(x_2)<0$，则分类正确，此时 $K_2(x) = K_1(x)$，即积累势函数不变。

（ii） 若 $x_2 \in \omega_1$ 且 $K_1(x_2)<0$，则

$$K_2(x)=K_1(x)+K(x,x_2)=\pm K(x,x_1)+K(x,x_2)$$

（iii） 若 $x_2 \in \omega_2$ 且 $K_1(x_2)>0$，则

$$K_2(x)=K_1(x)-K(x,x_2)=\pm K(x,x_1)-K(x,x_2)$$

以上（ii）、（iii）两种情况属于错分。假如 x_2 处于 $K_1(x)$定义的边界的错误一侧，则当 $x \in \omega_1$ 时，积累位势 $K_2(x)$要加 $K(x, x_2)$，当 $x \in \omega_2$ 时，积累位势 $K_2(x)$要减 $K(x, x_2)$。

第 K 步：设 $K_k(x)$为加入训练样本 x_1，x_2，…，x_k 后的积累位势，则加入第(k+1)个样本时，$K_{k+1}(x)$决定如下：

（i） 若 $x_{k+1} \in \omega_1$ 且 $K_k(x_{k+1})>0$，或 $x_{k+1} \in \omega_2$ 且 $K_k(x_{k+1})<0$，则分类正确，此时 $K_{k+1}(x) = K_k(x)$，即积累位势不变。

（ii） 若 $x_{k+1} \in \omega_1$ 且 $K_k(x_{k+1})<0$，则 $K_{k+1}(x)=K_k(x)+K(x,x_{k+1})$

（iii） 若 $x_{k+1} \in \omega_2$ 且 $K_k(x_{k+1})>0$，则 $K_{k+1}(x)=K_k(x)-K(x,x_{k+1})$

因此，积累位势的迭代运算可写成：$K_{k+1}(x)=K_k(x)+r_{k+1}K(x,x_{k+1})$，$r_{k+1}$为校正系数：

$$r_{k+1}=\begin{cases}0 & \text{if } x_{k+1}\in\omega_1 \text{ and } K_k(x_{k+1})>0\\ 0 & \text{if } x_{k+1}\in\omega_2 \text{ and } K_k(x_{k+1})<0\\ 1 & \text{if } x_{k+1}\in\omega_1 \text{ and } K_k(x_{k+1})<0\\ -1 & \text{if } x_{k+1}\in\omega_2 \text{ and } K_k(x_{k+1})>0\end{cases}$$

若从给定的训练样本集$\{x_1, x_2, \ldots, x_k, \ldots\}$中去除不使积累位势发生变化的样本，即使$K_j(x_{j+1})>0$且$x_{j+1}\in\omega_1$，或$K_j(x_{j+1})<0$且$x_{j+1}\in\omega_2$的那些样本，则可得一简化的样本序列$\{\hat{x}_1,\hat{x}_2,\ldots,\hat{x}_j,\ldots\}$，它们完全是校正错误的样本。此时，上述迭代公式可归纳为：

$$K_{k+1}(x)=\sum_{\hat{x}_j}a_jK(x,\hat{x}_j)$$

其中

$$a_j=\begin{cases}+1 & for\ \hat{x}_j\in\omega_1\\ -1 & for\ \hat{x}_j\in\omega_2\end{cases}$$

也就是说，由 k+1 个训练样本产生的积累位势，等于ω_1类和ω_2类两者中的校正错误样本的总位势之差。

- 构成势函数的两种方式

第一类势函数：可用对称的有限多项式展开，即：

$$K(x, x_k)=\sum_{i=1}^{m}\varphi_i(x)\varphi_i(x_k)$$

其中$\{\varphi_i(x)\}$在模式定义域内为正交函数集。将这类势函数代入判别函数，有：

$$d_{k+1}(x)=d_k(x)+r_{k+1}\sum_{i=1}^{m}\varphi_i(x_{k+1})\varphi_i(x)=d_k(x)+\sum_{i=1}^{m}r_{k+1}\varphi_i(x_{k+1})\varphi_i(x)$$

得迭代关系：

$$d_{k+1}(x)=\sum_{i=1}^{m}C_i(k+1)\varphi_i(x)$$

其中

$$C_i(k+1)=C_i(k)+r_{k+1}\varphi_i(x_{k+1})$$

因此，积累位势可写成：

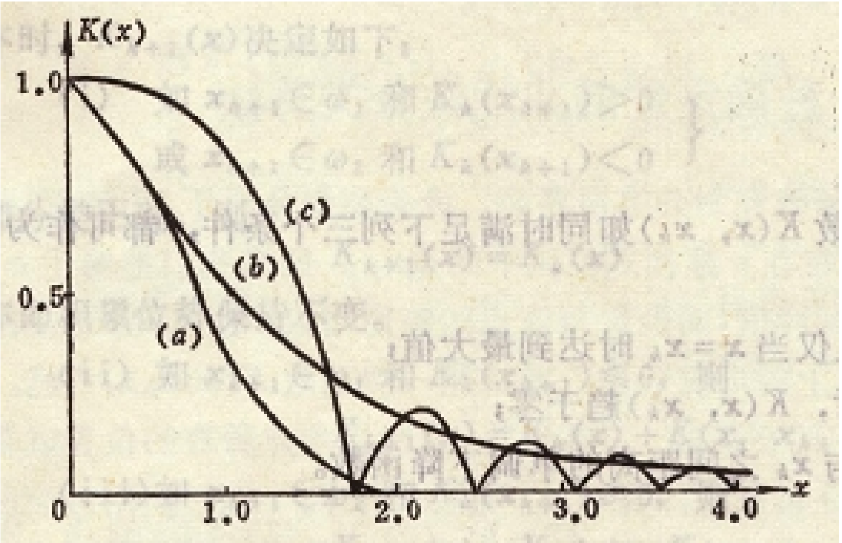
$$K_{k+1}(x)=\sum_{i=1}^{m}C_i(k+1)\varphi_i(x)$$，C_i可用迭代式求得。

第二类势函数：选择双变量 x 和 x_k 的对称函数作为势函数，即 $K(x, x_k) = K(x_k, x)$，并且它可展开成无穷级数，例如：

(a) $K(x, x_k)=e^{-\alpha\|x-x_k\|^2}$

(b) $K(x, x_k)=\dfrac{1}{1+\alpha\|x-x_k\|^2}$， α 是正常数

(c) $K(x, x_k)=\left|\dfrac{\sin\alpha\|x-x_k\|^2}{\alpha\|x-x_k\|^2}\right|$



K(x)
1.0
0.5
(a)
(b)
(c)
0
1.0
2.0
3.0
4.0
x

- 势函数法

实例 1：用第一类势函数的算法进行分类

（1）选择合适的正交函数集$\{\varphi_i(x)\}$

选择 Hermite 多项式，其正交域为(-∞，+∞)，其一维形式是

$$\phi_k = \frac{e^{-x^2/2}}{\sqrt{2^k \cdot k!\sqrt{\pi}}} H_k(x), \quad k = 0,1,2,\cdots$$

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}$$

其正交性：$\int_{-\infty}^{+\infty} H_m(x) H_n(x) e^{-x^2} dx = \begin{cases} 0 & m \neq n \\ 2^n n!\sqrt{\pi} & m = n \end{cases}$

其中，$H_k(x)$前面的乘式为正交归一化因子，为计算简便可省略。因此，Hermite 多项式前面几项的表达式为

$H_0(x)=1$，　$H_1(x)=2x$，　$H_2(x)=4x^2-2$，

$H_3(x)=8x^3-12x$，　$H_4(x)=16x^4-48x^2+12$

（2）建立二维的正交函数集

二维的正交函数集可由任意一对一维的正交函数组成，这里取四项最低阶的二维的正交函数

$$\varphi_1(x) = \varphi_1(x_1, x_2) = H_0(x_1) H_0(x_2) = 1$$

$$\varphi_2(x) = \varphi_2(x_1, x_2) = H_1(x_1) H_0(x_2) = 2x_1$$

$$\varphi_3(x) = \varphi_3(x_1, x_2) = H_0(x_1) H_1(x_2) = 2x_2$$

$$\varphi_4(x) = \varphi_4(x_1, x_2) = H_1(x_1) H_1(x_2) = 4x_1 x_2$$

（3）生成势函数

按第一类势函数定义，得到势函数

$$K(x,x_k)=\sum_{i=1}^{4}\varphi_i(x)\varphi_j(x)=1+4x_1x_{k_1}+4x_2x_{k_2}+16x_1x_2x_{k_1}x_{k_2}$$

其中 $x=(x_1,x_2)^T$， $x_k=(x_{k_1},x_{k_2})^T$

（4）通过训练样本逐步计算累积位势 K(x)

给定训练样本：ω_1类为 $x_{①}=(1\ \ 0)^T$，$x_{②}=(0\ \ -1)^T$

ω_2类为 $x_{③}=(-1\ \ 0)^T$，$x_{④}=(0\ \ 1)^T$

累积位势 K(x)的迭代算法如下

第一步：取 $x_{①}=(1\ \ 0)^T\in\omega_1$，故

$K_1(x)=K(x,x_{①})=1+4x_1\cdot1+4x_2\cdot0+16x_1x_2\cdot1\cdot0=1+4x_1$

第二步：取 $x_{②}=(0\ \ -1)^T\in\omega_1$，故 $K_1(x_{②})=1+4\cdot0=1$

因 $K_1(x_{②})>0$ 且 $x_{②}\in\omega_1$，故 $K_2(x)=K_1(x)=1+4x_1$

第三步：取 $x_{③}=(-1\ \ 0)^T\in\omega_2$，故 $K_2(x_{③})=1+4\cdot(-1)=-3$

因 $K_2(x_{③})<0$ 且 $x_{③}\in\omega_2$，故 $K_3(x)=K_2(x)=1+4x_1$

第四步：取 $x_{④}=(0\ \ 1)^T\in\omega_2$，故 $K_3(x_{④})=1+4\cdot0=1$

因 $K_3(x_{④})>0$ 且 $x_{④}\in\omega_2$，

故 $K_4(x)=K_3(x)-K(x,x_{④})=1+4x_1-(1+4x_2)=4x_1-4x_2$

将全部训练样本重复迭代一次，得

第五步：取 $x_{⑤}=x_{①}=(1\ \ 0)^T\in\omega_1$，$K_4(x_{⑤})=4$

故 $K_5(x)=K_4(x)=4x_1-4x_2$

第六步：取 $x_{⑥}=x_{②}=(0\ \ -1)^T\in\omega_1$，$K_5(x_{⑥})=4$

故 $K_6(x)=K_5(x)=4x_1-4x_2$

第七步：取 $x_{⑦}=x_{③}=(-1\ \ 0)^T\in\omega_2$，$K_6(x_{⑦})=-4$

故 $K_7(x)=K_6(x)=4x_1-4x_2$

第八步：取 $x_{⑧}=x_{④}=(0\ \ 1)^T\in\omega_2$，$K_7(x_{⑧})=-4$

故 $K_8(x)=K_7(x)=4x_1-4x_2$

以上对全部训练样本都能正确分类，因此算法收敛于判别函数

$d(x)=4x_1-4x_2$

- 势函数法

实例 2：用第二类势函数的算法进行分类

选择指数型势函数，取 $\alpha=1$，在二维情况下势函数为

$$K(x,x_k)=e^{-\|x-x_k\|^2}=e^{-[(x_1-x_{k_1})^2+(x_2-x_{k_2})^2]}$$

这里：ω_1 类为 $x_{①}=(0\ \ 0)^T$， $x_{②}=(2\ \ 0)^T$

ω_2 类为 $x_{③}=(1\ \ 1)^T$， $x_{④}=(1\ \ -1)^T$

可以看出，这两类模式是线性不可分的。算法步骤如下：

第一步：取 $x_{①}=(0\ \ 0)^T\in\omega_1$，则

$$K_1(x)=K(x,x_{①})=e^{-[(x_1-0)^2+(x_2-0)^2]}=e^{-(x_1^2+x_2^2)}$$

第二步：取 $x_{②}=(2\ \ 0)^T\in\omega_1$

因 $K_1(x_{②})=e^{-(4+0)}=e^{-4}>0$，

故 $K_2(x)=K_1(x)=e^{-(x_1^2+x_2^2)}$

第三步：取 $x_{③}=(1\ \ 1)^T\in\omega_2$

因 $K_2(x_{③})=e^{-(1+1)}=e^{-2}>0$，

故 $K_3(x)=K_2(x)-K(x,x_{③})=e^{-(x_1^2+x_2^2)}-e^{-[(x_1-1)^2+(x_2-1)^2]}$

第四步：取 $x_{④}=(1\ \ -1)^T\in\omega_2$

因 $K_3(x_{④})\ =e^{-(1+1)}-e^{-(0+4)}=e^{-2}-e^{-4}>0$，

故 $K_4(x)=K_3(x)-K(x,x_{④})$

$$=e^{-(x_1^2+x_2^2)}-e^{-[(x_1-1)^2+(x_2-1)^2]}-e^{-[(x_1-1)^2+(x_2+1)^2]}$$

需对全部训练样本重复迭代一次

第五步：取 $x_{⑤}=x_{①}=(0\ \ 0)^T\in\omega_1$，$K_4(x_{⑤})=e^0-e^{-2}-e^{-2}=1-2e^{-2}>0$

故 $K_5(x)=K_4(x)$

第六步：取 $x_{⑥}=x_{②}=(2\ \ 0)^T\in\omega_1$，$K_5(x_{⑥})=e^{-4}-e^{-2}-e^{-2}=e^{-4}-2e^{-2}<0$

故 $K_6(x)=K_5(x)+K(x,x_{⑥})$

$$=e^{-(x_1^2+x_2^2)}-e^{-[(x_1-1)^2+(x_2-1)^2]}-e^{-[(x_1-1)^2+(x_2+1)^2]}+e^{-[(x_1-2)^2+x_2^2]}$$

第七步：取 $x_{⑦}=x_{③}=(1\ \ 1)^T\in\omega_2$，$K_6(x_{⑦})=e^{-2}-e^{0}-e^{-4}+e^{-2}=2e^{-2}-e^{-4}-1<0$

故 $K_7(x)=K_6(x)$

第八步：取 $x_{⑧}=x_{④}=(1\ \ -1)^T\in\omega_2$，$K_7(x_{⑧})=e^{-2}-e^{-4}-e^{0}+e^{-2}=2e^{-2}-e^{-4}-1<0$

故 $K_8(x)=K_7(x)$

第九步：取 $x_{⑨}=x_{①}=(0\ \ 0)^T\in\omega_1$，$K_8(x_{⑨})=e^{0}-e^{-2}-e^{-2}+e^{-4}=1+e^{-4}-2e^{-2}>0$

故 $K_9(x)=K_8(x)$

第十步：取 $x_{⑩}=x_{②}=(2\ \ 0)^T\in\omega_1$，$K_9(x_{⑩})=e^{-4}-e^{-2}-e^{-2}+e^{0}=1+e^{-4}-2e^{-2}>0$

故 $K_{10}(x)=K_9(x)$

经过上述迭代，全部模式都已正确分类，因此算法收敛于判别函数

$$d(x)=e^{-(x_1^2+x_2^2)}-e^{-[(x_1-1)^2+(x_2-1)^2]}-e^{-[(x_1-1)^2+(x_2+1)^2]}+e^{-[(x_1-2)^2+x_2^2]}$$

- Lagrange 乘数法（详见相关数学文献）

Lagrange 乘数法是一种在等式约束条件下的优化算法，其基本思想是将等式约束条件下的最优化问题转化为无约束条件下的最优化问题。

问题：设目标函数为

$$y=f(x)，x=(x_1, x_2, \ldots, x_n)$$

求其在 m(m<n)个约束条件

$$g_k(x)=0，k=1,2,\ldots,m$$

下的极值。

描述：引进函数

$$L(x,\lambda_1,\lambda_2,\cdots,\lambda_m)=f(x)\ +\sum_{k=1}^{m}\lambda_k g_k(x)$$

其中 λ_k，k=1,2,…,m 为待定常数。将 L 当作 n+m 个变量 $x_1, x_2, \ldots, x_n$ 和 λ_1，λ_2, …，λ_m 的无约束的函数，对这些变量求一阶偏导数可得稳定点所要满足的方程：

$$\frac{\partial L}{\partial x_i}=0,\quad i=1,2,\cdots,n$$
$$g_k=0,\quad k=1,2,\cdots,m$$