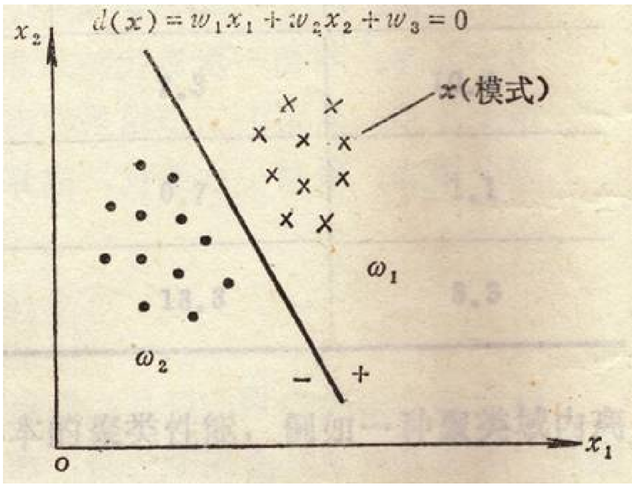
- 两类问题的判别函数（以二维模式样本为例）

若 $\boldsymbol{x}$ 是二维模式样本 $\boldsymbol{x} = (x_1\ x_2)^T$，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 $d(\boldsymbol{x})=0$ 来划分

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

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 $d(\boldsymbol{x})$，有

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

此时，$d(\boldsymbol{x})=0$ 称为决策面/判别界面方程。

- 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：

$$d(\boldsymbol{x}) = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_{n+1} = \boldsymbol{w}_0^T \boldsymbol{x} + w_{n+1}$$

其中 $\boldsymbol{w}_0 = (w_1, w_2, \ldots, w_n)^T$ 称为权向量（或参数向量），$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$。

$d(\boldsymbol{x})$也可表示为：

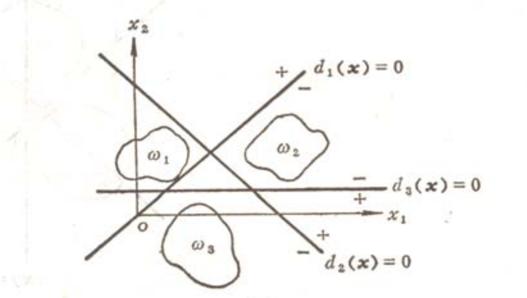
$$d(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$$

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

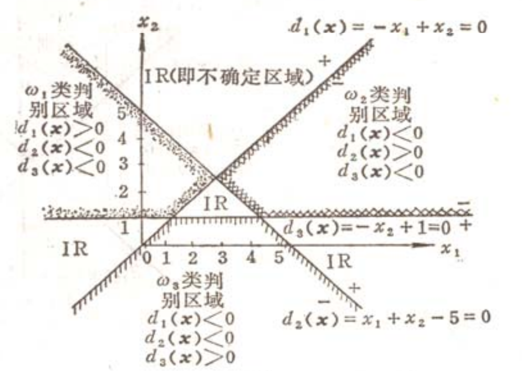
- 两类情况：判别函数 $d(\boldsymbol{x})$

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

● 多类情况 1



(a)



(b)

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：

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

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

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 $\boldsymbol{w}_i$, $i = 1, 2, \ldots, M$。

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

不确定区域：若对某一模式区域，$d_i(\boldsymbol{x})>0$ 的条件超过一个，或全部 $d_i(\boldsymbol{x})<0$，$i = 1, 2, \ldots, M$，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：

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

则对一个模式 $\boldsymbol{x}=(6, 5)^T$，判断其属于哪一类。

将 $\boldsymbol{x}=(6, 5)^T$ 代入上述判别函数，得：

$d_1(\boldsymbol{x}) = -1$，故 $d_1(\boldsymbol{x})<0$

$d_2(\boldsymbol{x}) = 6$ ，故 $d_2(\boldsymbol{x})>0$

$d_3(\boldsymbol{x}) = -4$，故 $d_3(\boldsymbol{x})<0$

从而 $\boldsymbol{x} \in \omega_2$

假若 $\boldsymbol{x}=(3, 5)^{\mathrm{T}}$，则

$d_1(\boldsymbol{x}) = 2>0$

$d_2(\boldsymbol{x}) = 3>0$

$d_3(\boldsymbol{x}) = -2<0$

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。

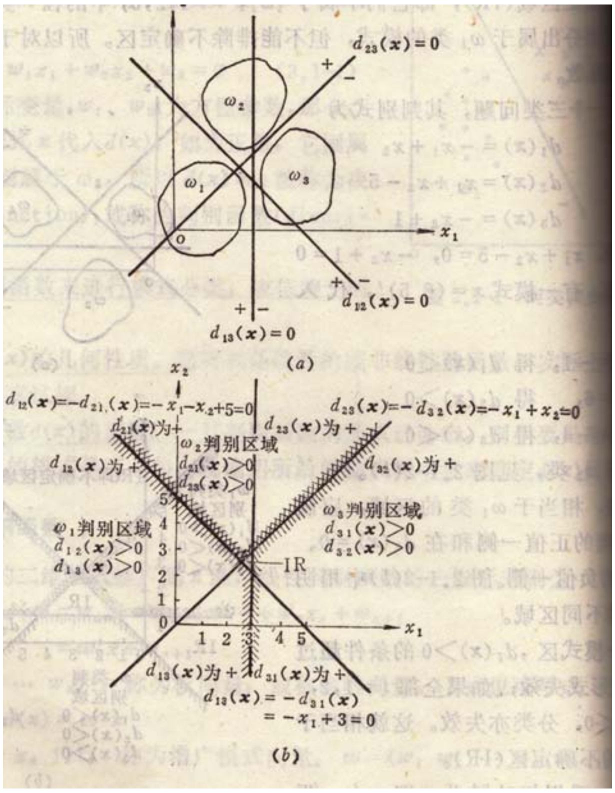
其判别函数为：

$$d_{ij}(\boldsymbol{x}) = \boldsymbol{w}_{ij}^T \boldsymbol{x}$$

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

重要性质：$d_{ij} = -d_{ji}$

图例：对一个三类情况，$d_{12}(\boldsymbol{x})=0$ 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 $d_{ij}(\boldsymbol{x})$，找不到$\forall j \neq i$，$d_{ij}(\boldsymbol{x})>0$ 的情况。

例：设有一个三类问题，其判别函数为：

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

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

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

从而 $\boldsymbol{x} \in \omega_3$

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

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

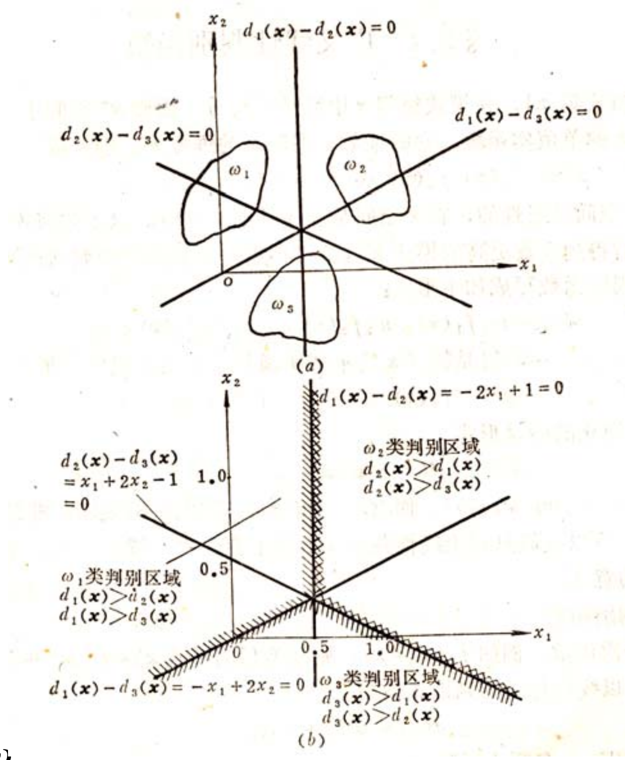
分类失败。

- 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：$d_{ij}(\boldsymbol{x}) = d_i(\boldsymbol{x}) - d_j(\boldsymbol{x}) = (\boldsymbol{w}_i - \boldsymbol{w}_j)^T \boldsymbol{x}$，则 $d_{ij}(\boldsymbol{x})>0$ 相当于 $d_i(\boldsymbol{x})> d_j(\boldsymbol{x})$，$\forall j \neq i$，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：

$$d_k(\boldsymbol{x}) = \boldsymbol{w}_k^T \boldsymbol{x}, k = 1, 2, \cdots, M$$

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



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：

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

属于 ω_1 类的区域应满足 $d_1(\boldsymbol{x})>d_2(\boldsymbol{x})$且 $d_1(\boldsymbol{x})>d_3(\boldsymbol{x})$，$\omega_1$ 类的判别界面为：

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

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

属于 ω_2 类的区域应满足 $d_2(\boldsymbol{x})> d_1(\boldsymbol{x})$且 $d_2(\boldsymbol{x})> d_3(\boldsymbol{x})$，$\omega_2$ 类的判别界面为：

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

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

同理可得 ω_3 类的判别界面为：

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

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

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

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

- 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：

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

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

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

此时有：

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

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。

- 线性判别函数

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

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

- $f_i(\boldsymbol{x})$选用二次多项式函数

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

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

要线性化为 $d(\boldsymbol{x}^*) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}^*$，须定义：

$$\boldsymbol{x}^* = (x_1^2 \quad x_1x_2 \quad x_2^2 \quad x_1 \quad x_2 \quad 1)^{\mathrm{T}}$$

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

此时，只要把模式空间 $\boldsymbol{x}^*$中的分量定义成 $\boldsymbol{x}$ 的单值实函数，$\boldsymbol{x}^*$即变成线性可分。此时 $\boldsymbol{x}^*$的维数（这里为 6）大于 $\boldsymbol{x}$ 的维数（这里为 2）。

2. $\boldsymbol{x}$ 是 n 维的情况，此时原判别函数设为：

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

式中各项的组成应包含 $\boldsymbol{x}$ 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 $d(\boldsymbol{x}^*) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}^*$，$\boldsymbol{x}^*$的维数大于 $\boldsymbol{x}$ 的维数，$\boldsymbol{w}$ 分量的数目也与 $\boldsymbol{x}^*$的维数相应。$\boldsymbol{x}^*$的各分量的一般化形式为：

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

- $f_i(\boldsymbol{x})$为 r 次多项式函数，$\boldsymbol{x}$ 为 n 维模式，则有

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

此时，判别函数 $d(\boldsymbol{x})$可用以下递推关系给出：

常数项：$d^{(0)}(\boldsymbol{x}) = w_{n+1}$

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

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

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

- $d(\boldsymbol{x})$总项数的讨论：对于 n 维 $\boldsymbol{x}$ 向量，若用 r 次多项式，$d(\boldsymbol{x})$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：$N_w = C_{n+3}^{3} = \dfrac{(n+3)!}{3!n!} = \dfrac{(n+3)(n+2)(n+1)}{6}$

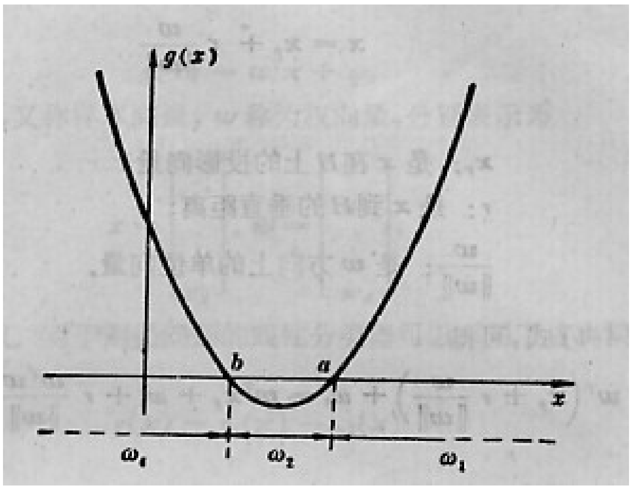
- 用 $d^{(r)}(x)$ 写出 r=2 和 n=2 时的判别函数

常数项：$d^{(0)}(\boldsymbol{x}) = w_{n+1} = w_3$

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

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

- 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

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

则可以定义线性判别函数

$$d(\boldsymbol{x}) = x_1-(a+b)x_2+\mathrm{ab} = w^T\boldsymbol{x}$$

此时

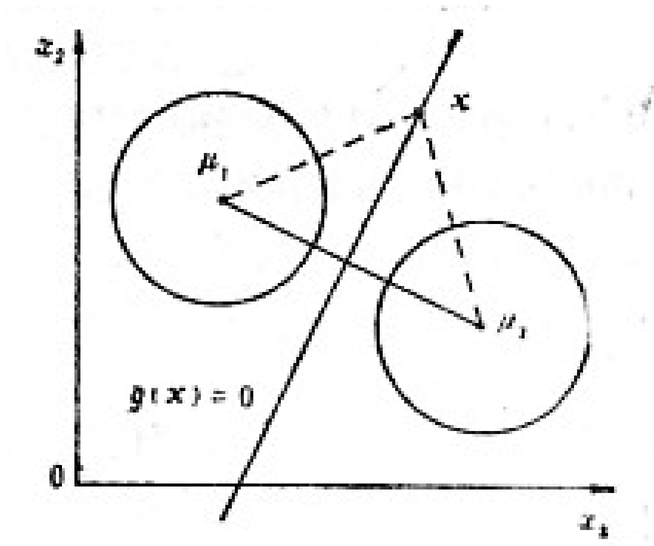
$$\boldsymbol{x} = (x_1\ x_2\ \ 1)^T, w = (1\ -(a+b)\ ab)^T$$

- 最小距离分类

设 $\boldsymbol{\mu}_1$ 和 $\boldsymbol{\mu}_2$ 为两个模式类 ω_1 和 ω_2 的聚类中心，定义决策规则：

$$\|\boldsymbol{x}-\boldsymbol{\mu}_1\|^2-\|\boldsymbol{x}-\boldsymbol{\mu}_2\|^2\begin{cases}<0 & \boldsymbol{x}\in\omega_1\\ >0 & \boldsymbol{x}\in\omega_2\end{cases}$$

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：$d(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$，其中 $\boldsymbol{x}=(x_1\ x_2\ldots x_{\mathrm{n}}\ 1)^{\mathrm{T}}$，$\boldsymbol{w}=(w_1\ w_2\ldots w_{\mathrm{n}}\ w_{\mathrm{n+1}})^{\mathrm{T}}$

判别界面为：$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=0$

对两类问题，ω_1类有模式$\{\boldsymbol{x}_1\ \boldsymbol{x}_2\}$，$\omega_2$类有模式$\{\boldsymbol{x}_3\ \boldsymbol{x}_4\}$，则应满足如下条件：

$$\left.\begin{array}{ll} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_1>0 & \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_2>0 \\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_3<0 & \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_4<0 \end{array}\right\}$$

若将属于ω_2类的模式都乘以(-1)，则上式可写成：

$$\left.\begin{array}{ll} \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_1>0 & \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_2>0 \\ \boldsymbol{w}^{\mathrm{T}}(-\boldsymbol{x}_3)>0 & \boldsymbol{w}^{\mathrm{T}}(-\boldsymbol{x}_4)>0 \end{array}\right\}$$

因此，若权向量能满足上述四个条件，则 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=0$ 为所给模式集的判别界面。

- Fisher 准则函数中的基本参量

1．在 d 维 $\boldsymbol{X}$ 空间

（1）各类样本的均值向量 $\boldsymbol{m_i}$

$$\boldsymbol{m}_i = \frac{1}{N_i}\sum_{x\in\Gamma_i}\boldsymbol{x}, i = 1,2,$$

（2）样本类内离散度矩阵 $\boldsymbol{S}_{\mathrm{i}}$ 和总样本类内离散度矩阵 $\boldsymbol{S}_{\mathrm{w}}$

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

其中 $\boldsymbol{S}_{\mathrm{w}}$ 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 $\boldsymbol{S}_{\mathrm{b}}$

$$\boldsymbol{S}_b = (\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T$$

$\boldsymbol{S}_{\mathrm{b}}$ 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 $\tilde{S}_w$

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

- Fisher 准则函数

Fisher 准则函数定义为：

$$J_{\mathrm{F}}(\boldsymbol{w})=\frac{(\tilde{m}_1-\tilde{m}_1)^2}{\tilde{S}_1^2+\tilde{S}_2^2}$$

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 $J_{\mathrm{F}}(\boldsymbol{w})$的分子尽可能大而分母尽可能小，即应寻找使 $J_{\mathrm{F}}(\boldsymbol{w})$尽可能大的 $\boldsymbol{w}$ 作为投影方向。但上式中并不显含 $\boldsymbol{w}$，因此须设法将 $J_{\mathrm{F}}(\boldsymbol{w})$变成 $\boldsymbol{w}$ 的显函数。

由各类样本的均值可推出：

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

这样，Fisher 准则函数 $J_{\mathrm{F}}(\boldsymbol{w})$的分子可写成：

$$\begin{aligned}\left(\tilde{m}_1-\tilde{m}_2\right)^2&=\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_1-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_2\right)^2\\&=\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_1-\boldsymbol{w}^T\boldsymbol{m}_2\right)\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_1-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_2\right)^{\mathrm{T}}\\&=\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{m}_1-\boldsymbol{w}^T\boldsymbol{m}_2\right)\left(\boldsymbol{m}_1^{\mathrm{T}}\boldsymbol{w}-\boldsymbol{m}_2^{\mathrm{T}}\boldsymbol{w}\right)\\&=\boldsymbol{w}^{\mathrm{T}}\left(\boldsymbol{m}_1-\boldsymbol{m}_2\right)\left(\boldsymbol{m}_1-\boldsymbol{m}_2\right)^{\mathrm{T}}\boldsymbol{w}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}\end{aligned}$$

现在再来考察 $J_{\mathrm{F}}(\boldsymbol{w})$的分母与 $\boldsymbol{w}$ 的关系：

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

因此，

$$\tilde{S}_1^2+\tilde{S}_2^2=\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{S}_1+\boldsymbol{S}_2)\boldsymbol{w}=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}$$

将上述各式代入 $J_F(\boldsymbol{w})$，可得：

$$J_F(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \boldsymbol{S}_b \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_w \boldsymbol{w}}$$

其中 $\boldsymbol{S}_b$ 为样本类间离散度矩阵，$\boldsymbol{S}_w$ 为总样本类内离散度矩阵。

- 最佳变换向量 $\boldsymbol{w}$*的求取

为求使 $J_{\mathrm{F}}(\boldsymbol{w})=\dfrac{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}}{\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}}$ 取极大值时的 $\boldsymbol{w}$*，可以采用 Lagrange 乘数法求解。令分母等于非零常数，即：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}=c\neq 0$$

定义 Lagrange 函数为：

$$L(\boldsymbol{w},\lambda)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}-\lambda(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}-c)$$

其中 λ 为 Lagrange 乘子。将上式对 $\boldsymbol{w}$ 求偏导数，可得：

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial \boldsymbol{w}}=\boldsymbol{S}_b\boldsymbol{w}+\boldsymbol{S}_b^T\boldsymbol{w}-\lambda(\boldsymbol{S}_w\boldsymbol{w}+\boldsymbol{S}_w^T\boldsymbol{w})=2\boldsymbol{S}_b\boldsymbol{w}-\lambda 2\boldsymbol{S}_w\boldsymbol{w}$$

令偏导数为零，有:

$$\boldsymbol{S}_b\boldsymbol{w}^*-\lambda\boldsymbol{S}_w\boldsymbol{w}^*=0$$

注：

y是标量，$\boldsymbol{x}=\begin{bmatrix}x_1 & x_2 & \cdots & x_d\end{bmatrix}^T$ 是d维列向量,则 $\dfrac{\partial y}{\partial \boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2} & \cdots & \dfrac{\partial y}{\partial x_d}\end{bmatrix}^T$

常用求导公式：

$$\frac{\partial \boldsymbol{Ax}}{\partial \boldsymbol{x}}=\boldsymbol{A}^T$$

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{x}}{\partial \boldsymbol{x}}=2\boldsymbol{x}$$

$$\frac{\partial \boldsymbol{x}^T\boldsymbol{Ax}}{\partial \boldsymbol{x}}=\boldsymbol{Ax}+\boldsymbol{A}^T\boldsymbol{x}$$

即

$$\boldsymbol{S}_b\boldsymbol{w}^*=\lambda\boldsymbol{S}_w\boldsymbol{w}^*$$

其中 $\boldsymbol{w}$*就是 $J_{\mathrm{F}}(\boldsymbol{w})$的极值解。因为 $\boldsymbol{S}_{\mathrm{w}}$ 非奇异，将上式两边左乘 $\boldsymbol{S}_w^{-1}$，可得：

$$\boldsymbol{S}_w^{-1}\boldsymbol{S}_b\boldsymbol{w}^*=\lambda\boldsymbol{w}^*$$

上式为求一般矩阵$\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$的特征值问题。利用$\boldsymbol{S}_b=(\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T$的定义，将上式左边的$\boldsymbol{S}_b\boldsymbol{w}^*$写成：

$$\boldsymbol{S}_b\boldsymbol{w}^*=(\boldsymbol{m}_1-\boldsymbol{m}_2)(\boldsymbol{m}_1-\boldsymbol{m}_2)^T\boldsymbol{w}^*=(\boldsymbol{m}_1-\boldsymbol{m}_2)R$$

其中$R=(\boldsymbol{m}_1-\boldsymbol{m}_2)^T\boldsymbol{w}^*$为一标量，所以$\boldsymbol{S}_b\boldsymbol{w}^*$总是在向量$(\boldsymbol{m}_1-\boldsymbol{m}_2)$的方向上。因此$\lambda\boldsymbol{w}^*$可写成：

$$\lambda\boldsymbol{w}^*=\boldsymbol{S}_w^{-1}\boldsymbol{S}_b\boldsymbol{w}^*=\boldsymbol{S}_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)R$$

从而可得：

$$\boldsymbol{w}^*=\frac{R}{\lambda}\boldsymbol{S}_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)$$

由于我们的目的是寻找最佳的投影方向，$\boldsymbol{w}^*$的比例因子对此并无影响，因此可忽略比例因子$\frac{R}{\lambda}$, 有：

$$\boldsymbol{w}^*=\boldsymbol{S}_w^{-1}(\boldsymbol{m}_1-\boldsymbol{m}_2)$$

- 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 $\boldsymbol{w}(1)$，可任意取值。若 $\boldsymbol{x}^k \in \omega_1, \boldsymbol{w}^T(k)\boldsymbol{x}^k > 0$，若 $\boldsymbol{x}^k \in \omega_2, \boldsymbol{w}^T(k)\boldsymbol{x}^k \le 0$，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 $\boldsymbol{x}^k \in \omega_1$ 且 $\boldsymbol{w}^T(k)\boldsymbol{x}^k \le 0$，则分类器对第 k 个模式 $\boldsymbol{x}^k$ 做了错误分类，此时应校正权向量，使得 $\boldsymbol{w}(\mathrm{k}+1) = \boldsymbol{w}(\mathrm{k}) + C\boldsymbol{x}^k$，其中 C 为一个校正增量。
- 若 $\boldsymbol{x}^k \in \omega_2$ 且 $\boldsymbol{w}^T(k)\boldsymbol{x}^k > 0$，同样分类器分类错误，则权向量应校正如下：$\boldsymbol{w}(\mathrm{k}+1) = \boldsymbol{w}(\mathrm{k}) - C\boldsymbol{x}^k$
- 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：$\boldsymbol{w}(\mathrm{k}+1) = \boldsymbol{w}(\mathrm{k})$

若对 $\boldsymbol{x}^k \in \omega_2$ 的模式样本乘以(-1)，则有：

$\boldsymbol{w}^T(k)\boldsymbol{x}^k \le 0$ 时，$\boldsymbol{w}(\mathrm{k}+1) = \boldsymbol{w}(\mathrm{k}) + C\boldsymbol{x}^k$

此时，感知器算法可统一写成：

$$\boldsymbol{w}(k+1) = \begin{cases} \boldsymbol{w}(k) & , \text{ if } \boldsymbol{w}^T(k)\boldsymbol{x}^k > 0 \\ \boldsymbol{w}(k) + \mathrm{C}\boldsymbol{x}^k & \text{ if } \boldsymbol{w}^T(k)\boldsymbol{x}^k \le 0 \end{cases}$$

- 感知器的训练算法实例

将属于ω₂的训练样本乘以（−1），并写成增广向量的形式。

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

第一轮迭代：取 $C=1$，$\boldsymbol{w}(1)=(0\ 0\ 0)^T$

因 $\boldsymbol{w}^T(1)\,\boldsymbol{x}^1=(0\ 0\ 0)(0\ 0\ 1)^T=0 \ngtr 0$，故 $\boldsymbol{w}(2)=\boldsymbol{w}(1)+\boldsymbol{x}^1=(0\ 0\ 1)^T$

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

因 $\boldsymbol{w}^T(3)\,\boldsymbol{x}^3=(0\ 0\ 1)(-1\ 0\ -1)^T=-1 \ngtr 0$，故 $\boldsymbol{w}(4)=\boldsymbol{w}(3)+\boldsymbol{x}^3=(-1\ 0\ 0)^T$

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

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 $\boldsymbol{w}^T(5)\boldsymbol{x}_{①}=(-1\ 0\ 0)(0\ 0\ 1)^T=0 \ngtr 0$，故 $\boldsymbol{w}(6)=\boldsymbol{w}(5)+\boldsymbol{x}_{①}=(-1\ 0\ 1)^T$

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

因 $\boldsymbol{w}^T(7)\boldsymbol{x}_{③}=(-1\ 0\ 1)(-1\ 0\ -1)^T=0 \ngtr 0$，故 $\boldsymbol{w}(8)=\boldsymbol{w}(7)+\boldsymbol{x}_{③}=(-2\ 0\ 0)^T$

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

需进行第三轮迭代。

第三轮迭代：

因 $\boldsymbol{w}^T(9)\,\boldsymbol{x}^1=(-2\ 0\ 0)(0\ 0\ 1)^T=0 \ngtr 0$，故 $\boldsymbol{w}(10)=\boldsymbol{w}(9)+\boldsymbol{x}^1=(-2\ 0\ 1)^T$

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

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

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

需进行第四轮迭代。

第四轮迭代：

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

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

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

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

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

$$d(\boldsymbol{x})=-2x_1+1$$

- 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 $\boldsymbol{x}_k \in \omega_i$，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 $\boldsymbol{x}$ 送入分类器，则应先计算出 M 个判别函数：

$$d_j(k) = \boldsymbol{w}_j(k)\boldsymbol{x}, j = 1,2,\cdots,\mathrm{M}$$

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

$$\boldsymbol{w}_j(k+1) = \boldsymbol{w}_j(k), j = 1,2,\cdots,M$$

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

$$\begin{aligned}
&\boldsymbol{w}_i(k+1) = \boldsymbol{w}_i(k) + Cx \\
&\boldsymbol{w}_l(k+1) = \boldsymbol{w}_l(k) - Cx \\
&\boldsymbol{w}_j(k+1) = \boldsymbol{w}_j(k), j = 1,2,\cdots,M \ \ ,j \neq i, j \neq l
\end{aligned}$$

其中 C 是一个正常数。权向量的初始值 $\boldsymbol{w}_i(l)$，$i = 1,2,\ldots,M$ 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：

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

取初始值 $\boldsymbol{w}_1(1)=\boldsymbol{w}_2(1)=\boldsymbol{w}_3(1)=(0\ 0\ 0)^T$，$C$=1。

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

$d_1(1)=\boldsymbol{w}_1^T(1)\boldsymbol{x}^1=(0\ 0\ 0)(0\ 0\ 1)^T=0$

$d_2(1)=\boldsymbol{w}_2^T(1)\boldsymbol{x}^1=(0\ 0\ 0)(0\ 0\ 1)^T=0$

$d_3(1)=\boldsymbol{w}_3^T(1)\boldsymbol{x}^1=(0\ 0\ 0)(0\ 0\ 1)^T=0$

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

$\boldsymbol{w}_1(2)=\boldsymbol{w}_1(1)+\boldsymbol{x}^1=(0\ 0\ 1)^T$

$\boldsymbol{w}_2(2)=\boldsymbol{w}_2(1)-\boldsymbol{x}^1=(0\ 0\ -1)^T$

$\boldsymbol{w}_3(2)=\boldsymbol{w}_3(1)-\boldsymbol{x}^1=(0\ 0\ -1)^T$

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

$d_1(2)=\boldsymbol{w}_1^T(2)\boldsymbol{x}^2=(0\ 0\ 1)(1\ 1\ 1)^T=1$

$d_2(2)=\boldsymbol{w}_2^T(2)\boldsymbol{x}^2=(0\ 0\ -1)(1\ 1\ 1)^T=-1$

$d_3(2)=\boldsymbol{w}_3^T(2)\boldsymbol{x}^2=(0\ 0\ -1)(1\ 1\ 1)^T=-1$

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

$\boldsymbol{w}_1(3)=\boldsymbol{w}_1(2)-\boldsymbol{x}^2=(-1\ -1\ 0)^T$

$\boldsymbol{w}_2(3)=\boldsymbol{w}_2(2)+\boldsymbol{x}^2=(1\ 1\ 0)^{\mathrm{T}}$

$\boldsymbol{w}_3(3)=\boldsymbol{w}_3(2)-\boldsymbol{x}^2=(-1\ -1\ -2)^{\mathrm{T}}$

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

$\mathrm{d}_1(3)=\boldsymbol{w}_1^T(3)\boldsymbol{x}^3=(-1\ -1\ 0)(-1\ 1\ 1)^{\mathrm{T}}=0$

$\mathrm{d}_2(3)=\boldsymbol{w}_2^T(3)\boldsymbol{x}^3=(1\ 1\ 0)(-1\ 1\ 1)^{\mathrm{T}}=0$

$\mathrm{d}_3(3)=\boldsymbol{w}_2^T(3)\boldsymbol{x}^3=(-1\ -1\ -2)(-1\ 1\ 1)^{\mathrm{T}}=-2$

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

$\boldsymbol{w}_1(4)=\boldsymbol{w}_1(3)-\boldsymbol{x}^3=(0\ -2\ -1)^{\mathrm{T}}$

$\boldsymbol{w}_2(4)=\boldsymbol{w}_2(3)-\boldsymbol{x}^3=(2\ 0\ -1)^{\mathrm{T}}$

$\boldsymbol{w}_3(4)=\boldsymbol{w}_3(3)+\boldsymbol{x}^3=(-2\ 0\ -1)^{\mathrm{T}}$

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

$d_1(4)=\boldsymbol{w}_1^T(4)\boldsymbol{x}^1=(0\ -2\ -1)(0\ 0\ 1)^{\mathrm{T}}=-1$

$d_2(4)=\boldsymbol{w}_2^T(4)\boldsymbol{x}^1=(2\ 0\ -1)(0\ 0\ 1)^{\mathrm{T}}=-1$

$d_3(4)=\boldsymbol{w}_3^T(4)\boldsymbol{x}^1=(-2\ 0\ -1)(0\ 0\ 1)^{\mathrm{T}}=-1$

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

$\boldsymbol{w}_1(5)=\boldsymbol{w}_1(4)+\boldsymbol{x}^1=(0\ -2\ 0)^{\mathrm{T}}$

$\boldsymbol{w}_2(5)=\boldsymbol{w}_2(4)-\boldsymbol{x}^1=(2\ 0\ -2)^{\mathrm{T}}$

$\boldsymbol{w}_3(5)=\boldsymbol{w}_3(4)-\boldsymbol{x}^1=(-2\ 0\ -2)^{\mathrm{T}}$

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

$d_1(5)=\boldsymbol{w}_1^T(5)\boldsymbol{x}^2=(0\ -2\ 0)(1\ 1\ 1)^{\mathrm{T}}=-2$

$d_2(5)=\boldsymbol{w}_2^T(5)\boldsymbol{x}^2=(2\ 0\ -2)(1\ 1\ 1)^{\mathrm{T}}=0$

$d_3(5)=\boldsymbol{w}_3^T(5)\boldsymbol{x}^2=-(-2\ 0\ -2)(1\ 1\ 1)^{\mathrm{T}}=-4$

因 $d_2(5)>d_1(5)$，$d_2(5)>d_3(5)$，故

$\boldsymbol{w}_1(6)=\boldsymbol{w}_1(5)$

$\boldsymbol{w}_2(6)=\boldsymbol{w}_2(5)$

$\boldsymbol{w}_3(6)=\boldsymbol{w}_3(5)$

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

$d_1(6)=\boldsymbol{w}_1^T(6)\boldsymbol{x}^3=(0\ -2\ 0)(-1\ 1\ 1)^{\mathrm{T}}=-2$

$d_2(6)=\boldsymbol{w}_2^T(6)\boldsymbol{x}^3=(2\ 0\ -2)(-1\ 1\ 1)^{\mathrm{T}}=-4$

$d_3(6)=\boldsymbol{w}_3^T(6)\boldsymbol{x}^3=(-2\ 0\ -2)(-1\ 1\ 1)^{\mathrm{T}}=0$

因 $d_3(6)>d_1(6)$，$d_3(6)>d_2(6)$，故

$\boldsymbol{w}_1(7)=\boldsymbol{w}_1(6)$

$\boldsymbol{w}_2(7)=\boldsymbol{w}_2(6)$

$\boldsymbol{w}_3(7)=\boldsymbol{w}_3(6)$

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

$d_1(7)=\boldsymbol{w}_1^T(7)\boldsymbol{x}^1=(0\ -2\ 0)(0\ 0\ 1)^{\mathrm{T}}=0$

$d_2(7)=\boldsymbol{w}_2^T(7)\boldsymbol{x}^1=(2\ 0\ -2)(0\ 0\ 1)^{\mathrm{T}}=-2$

$d_3(7)=\boldsymbol{w}_3^T(7)\boldsymbol{x}^1=(-2\ 0\ -2)(0\ 0\ 1)^{\mathrm{T}}=-2$

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 $\boldsymbol{x}^1$、$\boldsymbol{x}^2$、$\boldsymbol{x}^3$ 均已正确分类，所以权向量的解为：

$\boldsymbol{w}_1=(0\ -2\ 0)^T$

$\boldsymbol{w}_2=(2\ 0\ -2)^T$

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

三个判别函数：

$d_1(\boldsymbol{x})=-2x_2$

$d_2(\boldsymbol{x})=2x_1-2$

$d_3(\boldsymbol{x})=-2x_1-2$

- 梯度法定义

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

$$\nabla f(\boldsymbol{y}) = \frac{d}{d\boldsymbol{y}} f(\boldsymbol{y}) = \begin{pmatrix} \frac{\partial f}{\partial y_1} & \frac{\partial f}{\partial y_2} & \cdots & \frac{\partial f}{\partial y_n} \end{pmatrix}^{\mathrm{T}}$$

- 从 $\boldsymbol{w}(k)$导出 $\boldsymbol{w}(k+1)$的一般关系式

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

C 是一个正的比例因子（步长）

- 梯度法实例

选准则函数 $J(\boldsymbol{w},\boldsymbol{x})=|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}|-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$，考虑一维模式的情况。

设训练样本 $\boldsymbol{x}$ 为 1。

当错误分类时，$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}<0$，$J$ 为正值，$\nabla J=-2$，则 $\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+2\boldsymbol{C}$，对权向量进行校正。

当正确分类时，$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}>0$，$J=0$，$\nabla J=0$，则 $\boldsymbol{w}(k+1)=\boldsymbol{w}(k)$，权向量不变。

- 固定增量的逐次调整算法

设取准则函数为：

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

则 J 对 $\boldsymbol{w}$ 的微分式：

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

定义：

$$sign(\boldsymbol{w}^T \boldsymbol{x}) = \begin{cases} +1 & if\ \boldsymbol{w}^T \boldsymbol{x} > 0 \\ -1 & if\ \boldsymbol{w}^T \boldsymbol{x} \le 0 \end{cases}$$

则由梯度法中 $\boldsymbol{w}(k+1)$和 $\boldsymbol{w}(k)$的关系有：

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

其中 $\boldsymbol{x}^k$ 是训练模式样本，k 是指第 k 次迭代。

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + C \cdot \begin{cases} 0 & if\ \boldsymbol{w}^T \boldsymbol{x}^k > 0 \\ \boldsymbol{x}^k & if\ \boldsymbol{w}^T \boldsymbol{x}^k \le 0 \end{cases}$$

可以看出，当 $\boldsymbol{w}^T \boldsymbol{x}^k > 0$ 时，则 $\boldsymbol{w}(k+1) = \boldsymbol{w}(k)$，此时不对权向量进行修正；当 $\boldsymbol{w}^T \boldsymbol{x}^k \le 0$ 时，则 $\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + C\boldsymbol{x}^k$，需对权向量进行校正，初始权向量 $\boldsymbol{w}(1)$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 $\boldsymbol{w}^T \boldsymbol{x}^k \le 0$，就要对权向量修正 $C\boldsymbol{x}^k$ 值，因此称为固定增量算法。

- 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$\boldsymbol{w}$的解，其性质应满足：

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} > 0, \forall \boldsymbol{x} \in \omega_1$$
$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} < 0, \forall \boldsymbol{x} \in \omega_2$$

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

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

式中：

$$\boldsymbol{X} = \begin{pmatrix} \left(\boldsymbol{x}^1\right)^{\mathrm{T}} \\ \vdots \\ \left(\boldsymbol{x}^i\right)^{\mathrm{T}} \\ -\left(\boldsymbol{x}^{i+1}\right)^{\mathrm{T}} \\ \vdots \\ -\left(\boldsymbol{x}^N\right)^{\mathrm{T}} \end{pmatrix}, \begin{pmatrix} \left(\boldsymbol{x}^1\right)^{\mathrm{T}} \\ \vdots \\ \left(\boldsymbol{x}^i\right)^{\mathrm{T}} \end{pmatrix} \in \omega_1, \begin{pmatrix} \left(\boldsymbol{x}^{i+1}\right)^{T} \\ \vdots \\ \left(\boldsymbol{x}^N\right)^{\mathrm{T}} \end{pmatrix} \in \omega_2$$

$$\boldsymbol{w} = (w_1, w_2, \ldots, w_n, w_{n+1})^{\mathrm{T}}$$

其中，$\mathbf{0}$ 是零向量，$\left(\boldsymbol{x}^i\right)^{\mathrm{T}}$是第 i 个 n 维模式样本的增广向量，即$\left(\boldsymbol{x}^i\right)^{\mathrm{T}} = (x_{i_1}, x_{i_2}, \cdots, x_{i_n}, 1)^T, i = 1, 2, \cdots, N$，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$\boldsymbol{X}$是一个$N$*($n$+1)阶的矩阵。

- H-K 算法

H-K 算法是求解 $\boldsymbol{Xw}=\boldsymbol{b}$，式中 $\boldsymbol{b}=(b_1, b_2, \ldots, b_N)^T$，$\boldsymbol{b}$ 的所有分量都是正值。这里要同时计算 $\boldsymbol{w}$ 和 $\boldsymbol{b}$，我们已知 $\boldsymbol{X}$ 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，$\boldsymbol{Xw}=\boldsymbol{b}$ 没有唯一确定解，但可求其线性最小二乘解。

设 $\boldsymbol{Xw}=\boldsymbol{b}$ 的线性最小二乘解为 $\boldsymbol{w}^*$，即使$\|\boldsymbol{Xw}^*-\boldsymbol{b}\|$=极小

采用梯度法，定义准则函数：

$$J(\boldsymbol{w},\boldsymbol{x},\boldsymbol{b})=\frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{w}^T\boldsymbol{x}_i-\boldsymbol{b}_i\right)^2=\frac{1}{2}\left\|\boldsymbol{w}^T\boldsymbol{x}_i-\boldsymbol{b}_i\right\|^2=\frac{1}{2}\left(\boldsymbol{Xw}-\boldsymbol{b}\right)^T\left(\boldsymbol{Xw}-\boldsymbol{b}\right)$$

当 $\boldsymbol{Xw}=\boldsymbol{b}$ 的条件满足时，J 达到最小值。由于上式中包括的 $\sum_{i=1}^{N}\left(\boldsymbol{w}^T\boldsymbol{x}_i-\boldsymbol{b}_i\right)^2$ 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 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 求最小。对于 $\boldsymbol{w}$ 的梯度为：

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

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

$$\boldsymbol{w}=(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{X}^{\#}\boldsymbol{b}$$

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

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

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

根据上述约束条件，在每次迭代中，$\boldsymbol{b}(k)$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 $\boldsymbol{b}(k)$都是正值，应使$\left(\frac{\partial J}{\partial \boldsymbol{b}}\right)_{\boldsymbol{b}=\boldsymbol{b}(k)}$为非正值。在此条件下，准则函数 J 的微分为：

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

该式满足以下条件：

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

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

由 $\boldsymbol{b}$ 的迭代式和微分，有：

$$\boldsymbol{b}(k+1)=\boldsymbol{b}(k)+\delta\,\boldsymbol{b}(k)$$

$$\delta\,\boldsymbol{b}(k)=C[\boldsymbol{Xw}(k)-\boldsymbol{b}(k)+|\boldsymbol{Xw}(k)-\boldsymbol{b}(k)|]$$

将此式代入 $\boldsymbol{w}=\boldsymbol{X}^{\#}\boldsymbol{b}$，有：

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

为简化起见，令 $\boldsymbol{e}(k)=\boldsymbol{Xw}(k)-\boldsymbol{b}(k)$，可得 H-K 算法的迭代式。

设初值为 $\boldsymbol{b}(1)$，其每一分量均为正值，则：

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

$$\boldsymbol{e}(k)=\boldsymbol{Xw}(k)-\boldsymbol{b}(k)$$

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

由于

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

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

因此

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + C\boldsymbol{X}^{\#}|\boldsymbol{e}(k)|$$

$$\boldsymbol{b}(k+1) = \boldsymbol{b}(k) + C[\boldsymbol{X}\boldsymbol{w}(k) - \boldsymbol{b}(k) + |\boldsymbol{X}\boldsymbol{w}(k) - \boldsymbol{b}(k)|]$$

$$= \boldsymbol{b}(k) + \boldsymbol{C}[\boldsymbol{e}(k) + |\boldsymbol{e}(k)|]$$

- 模式类别可分性的判别

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

(i) 若 $\boldsymbol{e}(k)=0$，即 $\boldsymbol{Xw}(k)=\boldsymbol{b}(k)>0$，有解。

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

(iii)若 $\boldsymbol{e}(k)$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 $\boldsymbol{e}(k)$没有一个分量为正值，则 $\boldsymbol{b}(k)$不会再变化，所以不能求得解。

- LMSE 算法实例：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：$\boldsymbol{X}=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{pmatrix}$

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

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

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

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

- LMSE 算法实例：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：$\boldsymbol{X}=\begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & -1 \end{pmatrix}$

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

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

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

则：$\boldsymbol{e}(1)=\boldsymbol{Xw}(1)-\boldsymbol{b}(1)=(-1\ -1\ -1\ -1)^T$，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 $\boldsymbol{x}^1$，则有

$$K_1(\boldsymbol{x}) = \begin{cases} K(\boldsymbol{x}, \boldsymbol{x}^1) & if\ \ \boldsymbol{x}^1 \in \omega_1 \\ -K(\boldsymbol{x}, \boldsymbol{x}^1) & if\ \ \boldsymbol{x}^1 \in \omega_2 \end{cases}$$

这里第一步积累势函数 $K_1(\boldsymbol{x})$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 $\boldsymbol{x}^2$，则有

（i） 若 $\boldsymbol{x}^2 \in \omega_1$ 且 $K_1(\boldsymbol{x}^2)>0$，或 $\boldsymbol{x}^2 \in \omega_2$ 且 $K_1(\boldsymbol{x}^2)<0$，则分类正确，此时 $K_2(\boldsymbol{x}) = K_1(\boldsymbol{x})$，即积累势函数不变。

（ii） 若 $\boldsymbol{x}^2 \in \omega_1$ 且 $K_1(\boldsymbol{x}^2)<0$，则

$$K_2(\boldsymbol{x}) = K_1(\boldsymbol{x}) + K(\boldsymbol{x}, \boldsymbol{x}^2) = \pm K(\boldsymbol{x}, \boldsymbol{x}^1) + K(\boldsymbol{x}, \boldsymbol{x}^2)$$

（iii） 若 $\boldsymbol{x}^2 \in \omega_2$ 且 $K_1(\boldsymbol{x}^2)>0$，则

$$K_2(\boldsymbol{x}) = K_1(\boldsymbol{x}) - K(\boldsymbol{x}, \boldsymbol{x}^2) = \pm K(\boldsymbol{x}, \boldsymbol{x}^1) - K(\boldsymbol{x}, \boldsymbol{x}^2)$$

以上（ii）、（iii）两种情况属于错分。假如 $\boldsymbol{x}^2$ 处于 $K_1(\boldsymbol{x})$定义的边界的错误一侧，则当 $\boldsymbol{x}^2 \in \omega_1$ 时，积累位势 $K_2(\boldsymbol{x})$要加 $K(\boldsymbol{x}, \boldsymbol{x}^2)$，当 $\boldsymbol{x}^2 \in \omega_2$ 时，积累位势 $K_2(\boldsymbol{x})$要减 $K(\boldsymbol{x}, \boldsymbol{x}^2)$。

第 K 步：设 $K_k(\boldsymbol{x})$为加入训练样本 $\boldsymbol{x}^1$，$\boldsymbol{x}^2$，…，$\boldsymbol{x}^k$ 后的积累位势，则加入第(k+1)个样本时，$K_{k+1}(\boldsymbol{x})$决定如下：

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

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

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

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

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

若从给定的训练样本集$\{\boldsymbol{x}^1,\ \boldsymbol{x}^2,\ \ldots,\ \boldsymbol{x}^k\ ,\ldots\}$中去除不使积累位势发生变化的样本，即使$K_j(\boldsymbol{x}^{j+1})>0$且$\boldsymbol{x}^{j+1}\in\omega_1$，或$K_j(\boldsymbol{x}^{j+1})<0$且$\boldsymbol{x}^{j+1}\in\omega_2$的那些样本，则可得一简化的样本序列$\{\widehat{\boldsymbol{x}}^1,\widehat{\boldsymbol{x}}^2,\ldots,\widehat{\boldsymbol{x}}^j,\ldots\}$，它们完全是校正错误的样本。此时，上述迭代公式可归纳为：

$$K_{k+1}(\boldsymbol{x})=\sum_{\widehat{x}_j}a_jK(\boldsymbol{x},\widehat{\boldsymbol{x}}^j)$$

其中

$$a_j=\begin{cases}+1 & for\ \widehat{\boldsymbol{x}}^j\in\omega_1\\ -1 & for\ \widehat{\boldsymbol{x}}^j\in\omega_2\end{cases}$$

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：

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

其中$\{\varphi_i(\boldsymbol{x})\}$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：

$$\begin{aligned}d_{k+1}(\boldsymbol{x})&=d_k(\boldsymbol{x})+r_{k+1}\sum_{i=1}^{m}\varphi_i(\boldsymbol{x}_{k+1})\varphi_i(\boldsymbol{x})\\&=d_k(\boldsymbol{x})+\sum_{i=1}^{m}r_{k+1}\varphi_i(\boldsymbol{x}_{k+1})\varphi_i(\boldsymbol{x})\end{aligned}$$

得迭代关系：

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

其中

$$\boldsymbol{C}_i(k+1)=\boldsymbol{C}_i(k)+\mathrm{r}_{k+1}\,\varphi_i(\boldsymbol{x}_{k+1})$$

因此，积累位势可写成：

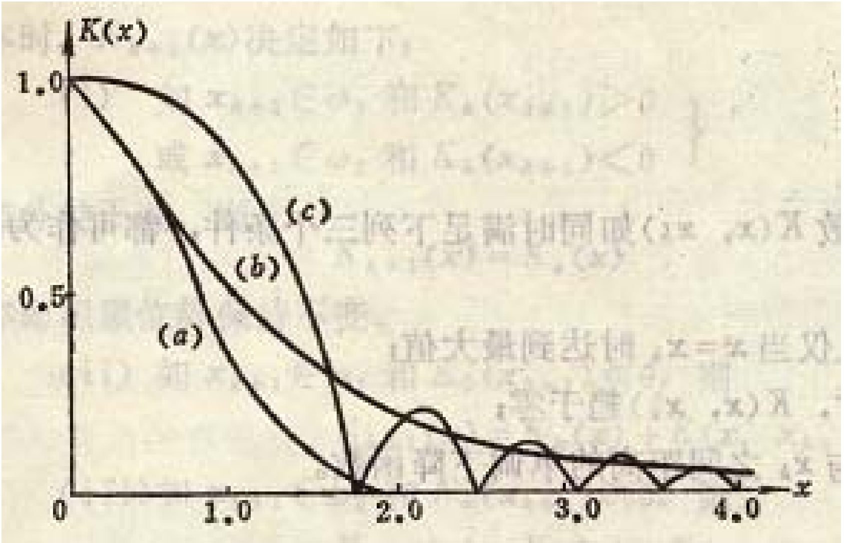
$K_{k+1}(\boldsymbol{x})=\sum_{i=1}^{m}\boldsymbol{C}_i(k+1)\varphi_i(\boldsymbol{x})$， $\boldsymbol{C}_i$ 可用迭代式求得。

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

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

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

(c) $K(\boldsymbol{x},\boldsymbol{x}_k)=\dfrac{\sin\alpha\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}{\alpha\|\boldsymbol{x}-\boldsymbol{x}_k\|^2}$


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


K(x)
1.0
(c)

- 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$\{\varphi_i(\boldsymbol{x})\}$

选择 Hermite 多项式，其正交域为(-∞，+∞)，其一维形式是

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

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

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，$H_k(\boldsymbol{x})$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

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

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

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

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

（3）生成势函数

按第一类势函数定义，得到势函数

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

其中 $\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}},\boldsymbol{x}_k=(x_{k_1},x_{k_2})^{\mathrm{T}}$

（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 $\boldsymbol{x}^1=(1\ 0)^{\mathrm{T}}$, $\boldsymbol{x}^2=(0\ \text{-}1)^{\mathrm{T}}$

ω_2 类为 $\boldsymbol{x}^3=(\text{-}1\ 0)^{\mathrm{T}}$, $\boldsymbol{x}^4=(0\ 1)^{\mathrm{T}}$

累积位势 $K(x)$的迭代算法如下

第一步：取 $\boldsymbol{x}^1=(1\ 0)^{\mathrm{T}}\in\omega_1$，故

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

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

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

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

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

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

因 $K_3(\boldsymbol{x}^4)>0$ 且 $\boldsymbol{x}^4\in\omega_2$，

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

将全部训练样本重复迭代一次，得

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

故 $K_5(\boldsymbol{x})=K_4(\boldsymbol{x})=4x_1-4x_2$

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

故 $K_6(\boldsymbol{x})=K_5(\boldsymbol{x})=4x_1-4x_2$

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

故 $K_7(\boldsymbol{x})=K_6(\boldsymbol{x})=4x_1-4x_2$

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

故 $K_8(\boldsymbol{x})=K_7(\boldsymbol{x})=4x_1-4x_2$

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

$d(\boldsymbol{x})=4x_1-4x_2$

- 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 α=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 $\boldsymbol{x}^1=(0\ 0)^T$, $\boldsymbol{x}^2=(2\ 0)^T$

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

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 $\boldsymbol{x}^1=(0\ 0)^T\in\omega_1$，则

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

第二步：取 $\boldsymbol{x}^2=(2\ 0)^T\in\omega_1$

因 $K_1(\boldsymbol{x}^2)=e^{-(4+0)}=e^{-4}>0$，

故 $K_2(\boldsymbol{x})=K_1(\boldsymbol{x})=e^{-(x_1^2+x_2^2)}$

第三步：取 $\boldsymbol{x}^3=(1\ 1)^T\in\omega_2$

因 $K_2(\boldsymbol{x}^3)=e^{-(1+1)}=e^{-2}>0$，

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

第四步：取 $\boldsymbol{x}^4=(1\ -1)^T\in\omega_2$

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

故 $K_4(\boldsymbol{x})=K_3(\boldsymbol{x})-K(\boldsymbol{x},\boldsymbol{x}^4)$

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

需对全部训练样本重复迭代一次

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

故 $K_5(\boldsymbol{x})=K_4(\boldsymbol{x})$

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

故 $K_6(\boldsymbol{x})=K_5(\boldsymbol{x})+K(\boldsymbol{x},\boldsymbol{x}^6)$

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

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

故 $K_7(\boldsymbol{x})=K_6(\boldsymbol{x})$

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

故 $K_8(\boldsymbol{x})=K_7(\boldsymbol{x})$

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

故 $K_9(\boldsymbol{x})=K_8(\boldsymbol{x})$

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

故 $K_{10}(\boldsymbol{x})=K_9(\boldsymbol{x})$

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

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