第三章 判别域代数界面方程法

第三章 判别域代数界面方程法

3.5 线性可分条件下判别函数权矢量算法

3.5.1 感知器算法

- 基本思想

用训练模式检验初始的或迭代中的增广权矢量w的合理性，当不合理时，对其进行校正。

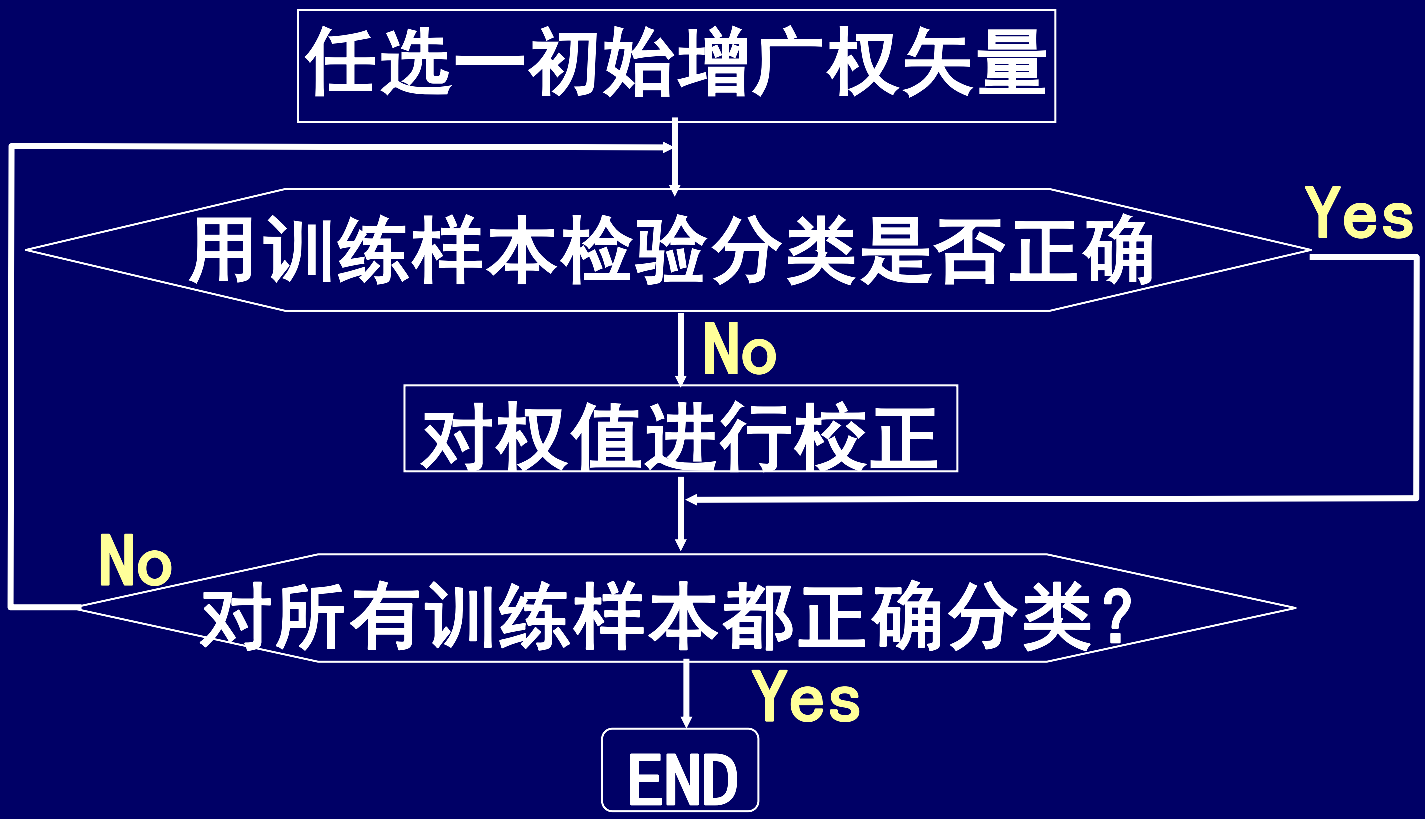
- 算法实质

校正方法实际上是最优化技术中的梯度下降法。本算法也是人工神经网络理论中的线性阈值神经元的学习算法。

3.5　线性可分条件下判别函数权矢量算法

3.5.1 感知器算法

算法流程



3.5.1 感知器算法

一、算法原理步骤

设给定一个增广的训练模式集 $\{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N\}$，其中每个模式类别已知，它们分属 ω_1 类和 ω_2 类。

(1) 置步数 κ =1，令增量 ρ =某正的常数，分别赋给初始增广权矢量 $\vec{w}(1)$ 的各分量较小的任意值；

(2) 输入训练模式 $\vec{x}_k$ ，计算判别函数值 $\vec{w}'(k)\vec{x}_k$

(3)调整增广权矢量，规则是:

-- 如果 $\vec{x}_k \in \omega_1$ 和 $\vec{w}'(k)\vec{x}_k \le 0$，则 $\vec{w}(k+1) = \vec{w}(k) + \rho\vec{x}_k$

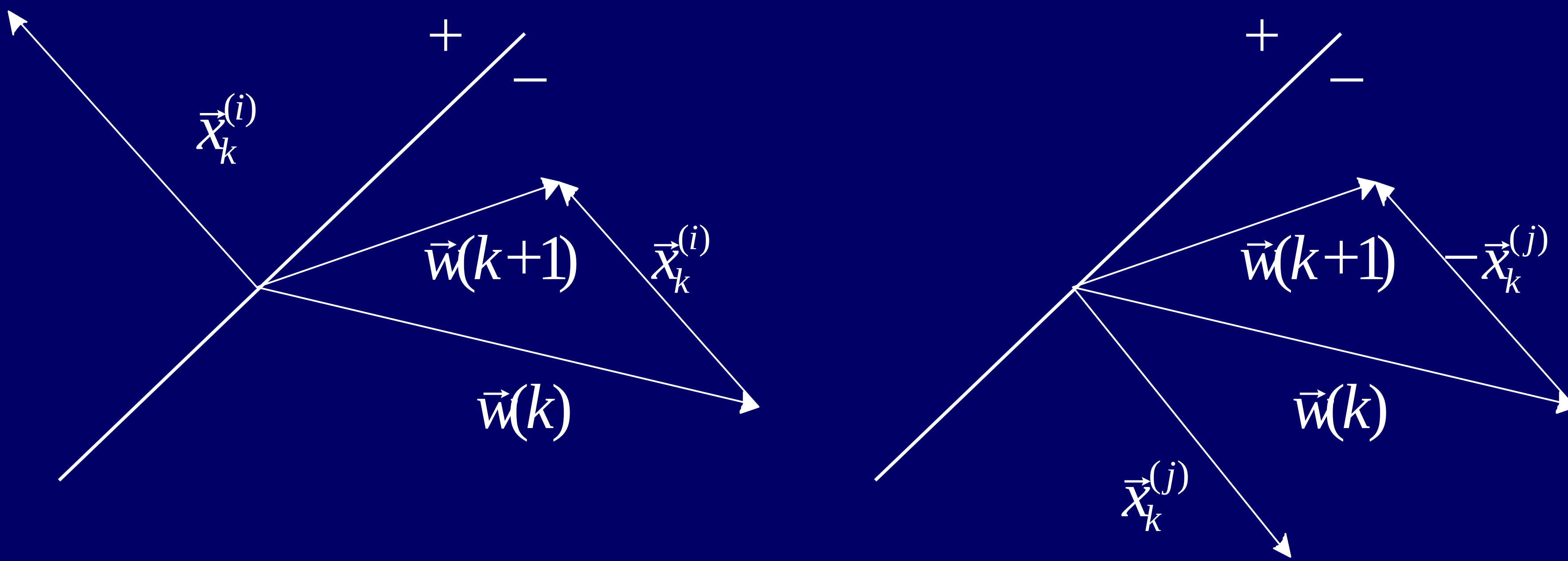
-- 如果 $\vec{x}_k \in \omega_2$ 和 $\vec{w}'(k)\vec{x}_k \ge 0$，则 $\vec{w}(k+1) = \vec{w}(k) - \rho\vec{x}_k$

-- 如果 $\vec{x}_k \in \omega_1$ 和 $\vec{w}'(k)\vec{x}_k > 0$，

或 $\vec{x}_k \in \omega_2$ 和 $\vec{w}'(k)\vec{x}_k \le 0$，则 $\vec{w}(k+1) = \vec{w}(k)$

(4)如果 $k < N$，令 $k = k+1$，返至(2)。如果 $k = N$，检验判别函数 $\vec{w}'\vec{x}$ 对 $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N$ 是否都能正确分类。若是，结束；若不是,令 $k=1$，返至(2)。

3.5　线性可分条件下判别函数权矢量算法



权空间中感知器算法权矢量校正过程示意图

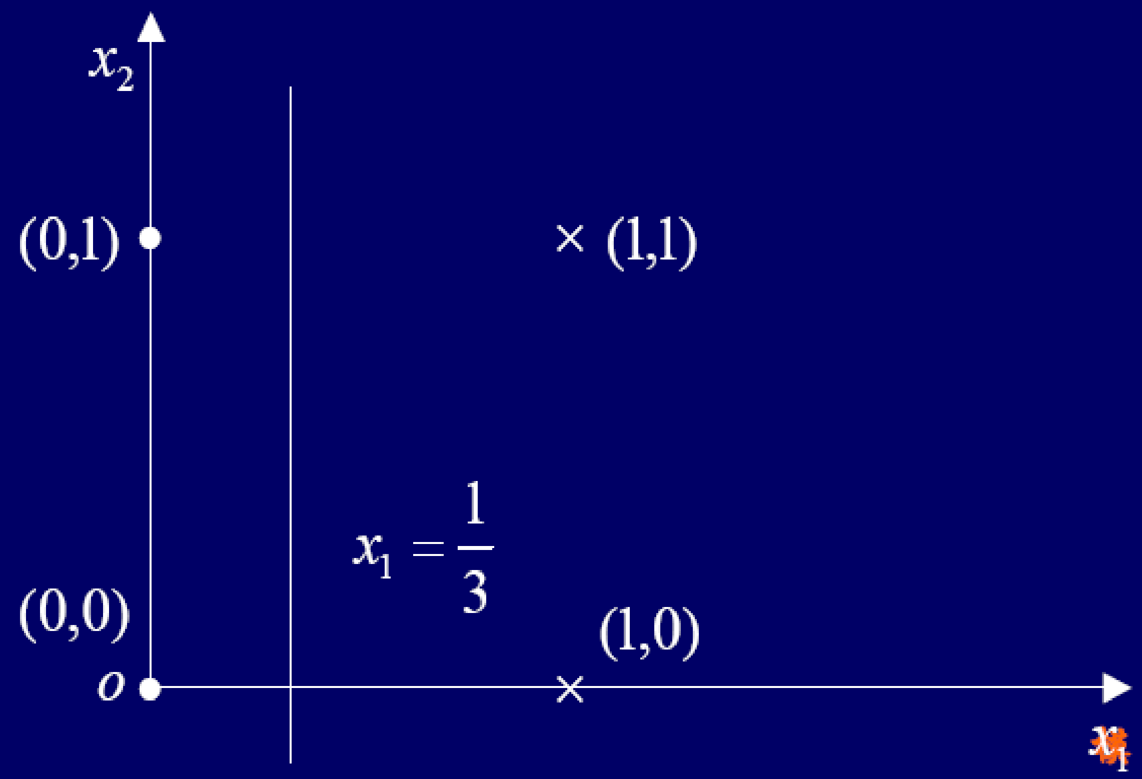
3.5.1 感知器算法

二、收敛定理

如果训练模式是线性可分的，感知器训练算法在有限次迭代后便可以收敛到正确的解矢量 $\vec{w}^*$。

3.5 线性可分条件下判别函数权矢量算法

例题：已知两类模式训练样本如下图所示，其中 $(0,0)'$ 、$(0,1)'$ **属于** ω_1**类，**$(1,0)'$、$(1,1)'$**属于** ω_2 **类，试用感知器算法求** $\vec{w}^*$。



3.5 线性可分条件下判别函数权矢量算法

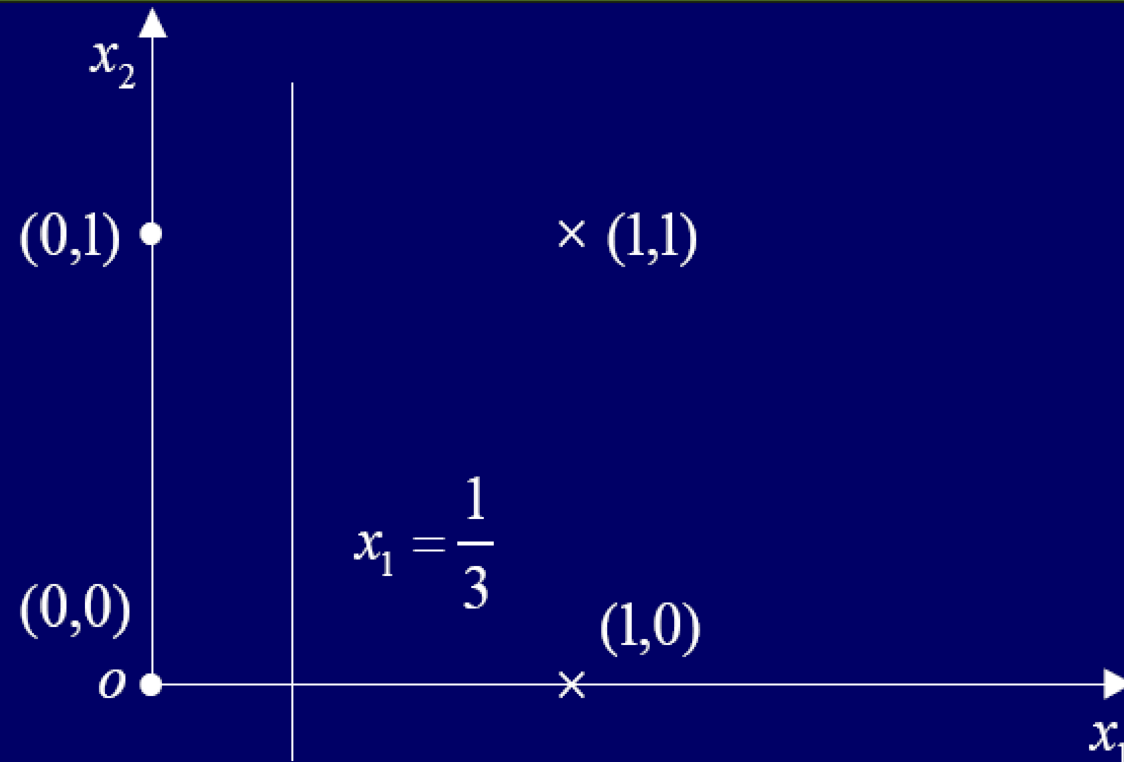
(1) 训练样本分量增广化及符号规范化。

$$\vec{x}_1 = (0,0,1)'$$

$$\vec{x}_2 = (0,1,1)'$$

$$\vec{x}_3 = (-1,0,-1)'$$

$$\vec{x}_4 = (-1,-1,-1)'$$



(2) 运用感知器训练算法。任意给增广权矢量赋初值

$\vec{w}(1) = (1,1,1)'$ **取增量** $\rho = 1$ **，则：**

$$\vec{x}_1 = (0,0,1)' \quad \vec{x}_2 = (0,1,1)' \quad \vec{w}(1) = (1,1,1)'$$

$$\vec{x}_3 = (-1,0,-1)' \quad \vec{x}_4 = (-1,-1,-1)' \quad \rho = 1$$

$$k=1,\ \vec{x}_k = \vec{x}_1,\ d(\vec{x}_k) = w'(k)\vec{x}_k = 1 > 0,\quad \vec{w}(2) = \vec{w}(1)$$

$$k=2,\ \vec{x}_k = \vec{x}_2,\ d(\vec{x}_k) = w'(k)\vec{x}_k = 2 > 0,\quad \vec{w}(3) = \vec{w}(2)$$

$$k=3,\ \vec{x}_k = \vec{x}_3,\ d(\vec{x}_k) = w'(k)\vec{x}_k = -2 < 0,\ \vec{w}(4) = \vec{w}(3) + \vec{x}_3 = (0,1,0)'$$

$$k=4,\ \vec{x}_k = \vec{x}_4,\ d(\vec{x}_k) = w'(k)\vec{x}_k = -1 < 0,\ \vec{w}(5) = \vec{w}(4) + \vec{x}_4 = (-1,0,-1)'$$

$$k=5,\ \vec{x}_k = \vec{x}_1,\ d(\vec{x}_k) = w'(k)\vec{x}_k = -1 < 0,\ \vec{w}(6) = \vec{w}(5) + \vec{x}_1 = (-1,0,0)'$$

$$k=6,\ \vec{x}_k = \vec{x}_2,\ d(\vec{x}_k) = w'(k)\vec{x}_k = 0,\quad \vec{w}(7) = \vec{w}(6) + \vec{x}_2 = (-1,1,1)'$$

$$k=7,\ \vec{x}_k = \vec{x}_3,\ d(\vec{x}_k) = w'(k)\vec{x}_k = 0,\quad \vec{w}(8) = \vec{w}(7) + \vec{x}_3 = (-2,1,0)'$$

$$k=8,\ \vec{x}_k = \vec{x}_4,\ d(\vec{x}_k) = w'(k)\vec{x}_k = 1 > 0,\quad \vec{w}(9) = \vec{w}(8)$$

$$k=9,\ \vec{x}_k = \vec{x}_1,\ d(\vec{x}_k) = w'(k)\vec{x}_k = 0,\quad \vec{w}(10) = \vec{w}(9) + \vec{x}_1 = (-2,1,1)'$$

$\vec{x}_1 = (0,0,1)' \quad \vec{x}_2 = (0,1,1)' \quad \vec{w}(1) = (1,1,1)'$

$\vec{x}_3 = (-1,0,-1)' \quad \vec{x}_4 = (-1,-1,-1)' \quad \rho = 1$

$k=9\ ,\ \vec{x}_k = \vec{x}_1\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 0\ ,\quad \vec{w}(10) = \vec{w}(9) + \vec{x}_1 = (-2,1,1)'$

$k=10\ ,\ \vec{x}_k = \vec{x}_2\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 2 > 0\ ,\ \vec{w}(11) = \vec{w}(10)$

$k=11\ ,\ \vec{x}_k = \vec{x}_3\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 1 > 0\ ,\ \vec{w}(12) = \vec{w}(11)$

$k=12\ ,\ \vec{x}_k = \vec{x}_4\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 0\ ,\quad \vec{w}(13) = \vec{w}(12) + \vec{x}_4 = (-3,0,0)'$

$k=13\ ,\ \vec{x}_k = \vec{x}_1\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 0\ ,\quad \vec{w}(14) = \vec{w}(13) + \vec{x}_1 = (-3,0,1)'$

$k=14\ ,\ \vec{x}_k = \vec{x}_2\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 1 > 0\ ,\ \vec{w}(15) = \vec{w}(14)$

$k=15\ ,\ \vec{x}_k = \vec{x}_3\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 2 > 0\ ,\ \vec{w}(16) = \vec{w}(15)$

$k=16\ ,\ \vec{x}_k = \vec{x}_4\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 2 > 0\ ,\ \vec{w}(17) = \vec{w}(16)$

$k=17\ ,\ \vec{x}_k = \vec{x}_1\ ,\ d(\vec{x}_k) = \vec{w}'(k)\vec{x}_k = 1 > 0\ ,\ \vec{w}(18) = \vec{w}(17)$

由上面的结果可以看出，经过对 $\vec{x}_1$、$\vec{x}_2$、$\vec{x}_3$、$\vec{x}_4$一轮迭代后表明，使用$\vec{w}(14)$已能对所有训练样本正确分类，增广权矢量的值不再变化，所以算法收敛于 $\vec{w}(14)$，$\vec{w}(14)$就是所求的解向量，即 $\vec{w}^* = (-3,0,1)'$

由此还可以得到区分界面为$-3x_1+1=0$，即 $x_1 = \dfrac{1}{3}$

3.5.2 一次准则函数及梯度下降法

一、梯度下降法

采用最优化技术求解线性判别函数中的增广权矢量时，首先需构造准则函数，负梯度方向是函数下降最快的方向。在解矢量对应于一次准则函数极小值的情况中，可采用梯度下降法沿负梯度方向选择适当的步长进行搜索，求解函数的极小值点w^*。

3.5.2 一次准则函数及梯度下降法

二、一次准则函数

构造准则函数： $J(\vec{w}) = k(|\vec{w}'\vec{x}| - \vec{w}'\vec{x}) \quad (k > 0)$

当 $\vec{w}'\vec{x} < 0$ 时 $|\vec{w}'\vec{x}| - \vec{w}'\vec{x} = -2\vec{w}'\vec{x} > 0$

当 $\vec{w}'\vec{x} \geq 0$ 时 $|\vec{w}'\vec{x}| - \vec{w}'\vec{x} = 0$

故有极小值： $J_{\min}(\vec{w}) = 0$

对于已符号规范化的训练模式，寻求使 $J(\vec{w})$

取极小的 $\vec{w}$， 此时的 $\vec{W}$ 满足 $\vec{w}'\vec{x} \geq 0$

3.5 线性可分条件下判别函数权矢量算法

令k = 1/2，求得准则函数的梯度

$$\nabla J(\vec{w}) = \frac{\partial J}{\partial \vec{w}} = \frac{1}{2}\left[\vec{x}\,\mathrm{sgn}(\vec{w}'\vec{x}) - \vec{x}\right]$$

其中符号函数 $\mathrm{sgn}(\vec{w}'\vec{x}) = \begin{cases} 1 & \vec{w}'\vec{x} > 0 \\ -1 & \vec{w}'\vec{x} \le 0 \end{cases}$

增广权矢量的修正迭代公式为：

$$\vec{w}(k+1) = \vec{w}(k) - \rho_k \nabla J(\vec{w}(k))$$

$$= \vec{w}(k) - \frac{\rho_k}{2}\left[\vec{x}_k\,\mathrm{sgn}\left[\vec{w}'(k)\vec{x}_k\right] - \vec{x}_k\right]$$

$$= \begin{cases} \vec{w}(k) & \text{若}\ \vec{w}'(k)\vec{x}_k > 0 \\ \vec{w}(k) + \rho_k \vec{x}_k & \text{若}\ \vec{w}'(k)\vec{x}_k \le 0\text{，}\ \rho_k > 0 \end{cases}$$

增广权矢量修正迭代公式：

$$\vec{w}(k+1)=\begin{cases}\vec{w}(k) & \text{若}\vec{w}'(k)\vec{x}_k>0\\ \vec{w}(k)+\rho_k\vec{x}_k & \text{若}\vec{w}'(k)\vec{x}_k\le 0\text{，}\rho_k>0\end{cases}$$

称为单样本修正，实际也可构造批量修正：

$$\vec{w}(k+1)=\vec{w}(k)+\rho_k\sum_{\vec{x}_j\in X_m}\vec{x}_j$$

式中的 X_m 是被 $\vec{w}(k)$ 错分类的模式集。

3.5 线性可分条件下判别函数权矢量算法

注 意

★当 ρ_k 为常数时，梯度下降法的迭代公式与感知器训练算法是一致的。

★ ρ_k 取常数时，这种梯度下降法也称为固定增量法。若 ρ_k 取得较小，收敛速度则较慢，若 ρ_k 取得较大，收敛速度加快，但当搜索接近极值点时，可能产生过调引起振荡。

★改进的办法是使 ρ_k 随 k 而变化，我们称之为可变增量法。

3.5.3　感知器训练算法在多类问题中的应用

判别规则：

对于c类问题，应建立c个判别函数：

$$d_i(x) = w_i' x_i \qquad (i=1, 2, \ldots, c)$$

如果$x \in \omega_i$，则有$w_i' x > w_j' x \quad (\forall\ j \neq i)$

因此判别规则是：

若$d_i(x) > d_j(x) \quad \forall\ j \neq i$　　则判$x \in \omega_i$

3.5　线性可分条件下判别函数权矢量算法

算法步骤：

(1) 赋初值，分别给c个权矢量w_i $(i=1,2,\ldots,c)$赋任意的初值，选择正常数ρ，置步数k=1。

(2) 输入已知类别的增广训练模式x_k，计算c个判别函数

$$d_i(x_k) = w_i' x_k \qquad (i=1,2,\ldots,c)$$

(3) 修正权矢量，修正规则是：

if $(x_k \in \omega_i)$ and $(d_i(x_k) > d_j(x_k))$ $(\forall j \neq i)$ then

$$w_i(k+1) = w_i(k) \qquad (i=1,2,\ldots,c)$$（正确分类

if $(x_k \in \omega_i)$ and $(d_i(x_k) \leq d_l(x_k))$ $(l \neq i)$ then

$$w_i(k+1) = w_i(k) + \rho x_k$$

$$w_l(k+1) = w_l(k) - \rho x_k$$

$$w_j(k+1) = w_j(k) \quad (\forall\ j \neq i, l)$$

例题：已知训练样本 $(0,0)^T \in \omega_1$，$(1,1)^T \in \omega_2$，$(-1,1)^T \in \omega_3$，试求解向量 w_1、w_2 和 w_3。

解:（1）训练样本分量增广化。将训练样本变成增广训练模式：$x_1=(0,0,1)^T$，$x_2=(1,1,1)^T$，$x_3=(-1,1,1)^T$，这里的下标恰是所属类别，各类样本不需符号规范化。

（2）运用感知器训练算法。置k=1，增量ρ=1，赋初值：$w_1=(0,0,0)^T$，$w_2=(0,0,0)^T$，$w_3=(0,0,0)^T$，进行迭代运算：

3.5 线性可分条件下判别函数权矢量算法

$(0, 0, 1)^T \in \omega_1$， $(1, 1, 1)^T \in \omega_2$ ， $(-1, 1, 1)^T \in \omega_3$

$k=1$, $x_k=x_1 \in \omega_1$, 因为 $d_1(x_1)=d_2(x_1)=0$， $d_1(x_1)=d_3(x_1)=0$,

错分，所以： $w_1(2)=w_1(1)+ x_1=(0, 0, 1)^T$

$w_2(2)=w_2(1)- x_1=(0, 0, -1)^T$

$w_3(2)=w_3(1)- x_1=(0, 0, -1)^T$

$k=2$, $x_k=x_2 \in \omega_2$,

因为 $d_2(x_2)=-1<d_1(x_2)=1$， $d_2(x_2)=d_3(x_2)=-1$， 错分,

所以 $w_1(3)=w_1(2)- x_2=(-1, -1, 0)^T$

$w_2(3)=w_2(2)+ x_2=(1, 1, 0)^T$

$w_3(3)=w_3(2)- x_2=(-1, -1, -2)^T$

$k=3,\ x_k=x_3\in\omega_3,$

因为$d_3(x_3)=-2<d_1(x_3)=0$，$d_3(x_3)=d_2(x_3)=0$，错分，

所以 $w_1(4)=w_1(3)-\ x_3=(\ 0,-2,-1)^T$

$w_2(4)=w_2(3)-\ x_3=(\ 2,\ \ 0,-1)^T$

$w_3(4)=w_3(3)+\ x_3=(-2,\ \ 0,-1)^T$

$k=4,\ x_k=x_1\in\omega_1,$

因为$d_1(x_1)=d_2(x_1)=-1$，$d_1(x_1)=d_3(x_1)=-1$，错分，

所以 $w_1(5)=w_1(4)+\ x_1=(\ 0,-2,\ \ 0)^T$

$w_2(5)=w_2(4)-\ x_1=(\ 2,\ \ 0,-2)^T$

$w_3(5)=w_3(4)-\ x_1=(-2,\ \ 0,-2)^T$

3.5 线性可分条件下判别函数权矢量算法

$k=5,\ x_k=x_2\in\omega_2$,

因为$d_2(x_2)=0>d_1(x_2)=-2$，$d_2(x_2)=0>d_3(x_2)=-4$，正确，

所以 $w_1(6)=w_1(5)=(\ 0, -2,\ \ 0)^T$

$w_2(6)=w_2(5)=(\ 2,\ \ 0, -2)^T$

$w_3(6)=w_3(5)=(-2,\ \ 0, -2)^T$

$k=6,\ x_k=x_3\in\omega_3$,

因为$d_3(x_3)=0>d_1(x_3)=-2$，$d_3(x_3)=0>d_2(x_3)=-4$，正确，

所以 $w_1(7)=w_1(6)=(\ 0, -2,\ \ 0)^T$

$w_2(7)=w_2(6)=(\ 2,\ \ 0, -2)^T$

$w_3(7)=w_3(6)=(-2,\ \ 0, -2)^T$

$k=7, x_k=x_1 \in \omega_1$,

因为$d_1(x_1)=0>d_2(x_1)=-2$，$d_1(x_1)=0>d_3(x_1)=-2$，正确，

三个权矢量不再变化，因此可以确定所有训练样本均已被正确分类，由此得到三个解矢量：$w_1^*=w_1(5)$，$w_2^*=w_2(5)$，$w_3^*=w_3(5)$，同时可得三个判别函数：

$$d_1(x) = -2x_2$$

$$d_2(x) = 2x_1-2$$

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

谢谢！