图像分析与识别

第4章 模式识别—
4.2线性和非线性判别分析

信控学院 蔡利梅

4.2.1概述

（1）贝叶斯决策的局限性

- 前提：对先验概率和类概率密度函数有充分的先验知识；或有足够多的样本，可以较好地进行概率密度估计。
- 局限：
  - 若前提条件不满足，采用最优方法设计出的分类器往往不具有最优的性质
  - 估计：实际问题中，得到的只是样本集，样本的分布形式很难确定，进行估计需要大量样本；当样本数有限时，概率密度函数估计问题往往是一个比分类更难的一般性问题

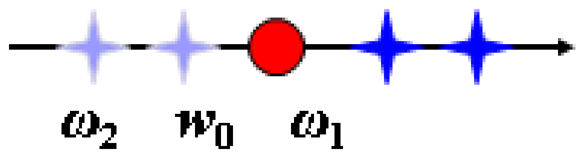
实际问题中，不去估计类条件概率，直接利用样本集设计分类器。

（2）利用样本集直接设计分类器的思路

- 首先给定某个判别函数，利用样本集去确定判别函数中的未知参数。
- 判别函数分类
  - 线性判别函数
  - 非线性判别函数

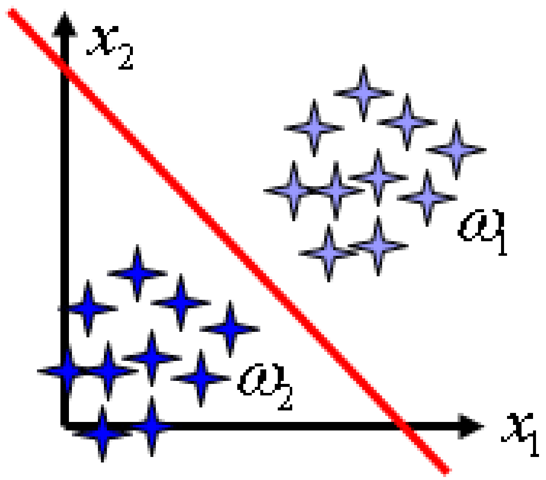
（3）线性判别函数

- 实例分析
  - 一维数据
  - 二维数据



两类的分界点为w_0，

判别函数表示为：$g(x)=w_1x-w_0$



两类的判别函数表示为：

$$g(X)=w_1x_1+w_2x_2+w_0$$
$$=W^TX+w_0$$

W、X均为二维列向量

- 一般表达式 $\boxed{g(X)=W^T X+w_0}$

W、X均为n维列向量，W称为权向量（系数）

- 决策规则

$$\boxed{\begin{array}{l}\text{若}g(X)>0\text{，则决策}X\in\omega_1\text{；}\\ \text{若}g(X)<0\text{，决策}X\in\omega_2\text{；}\\ \text{若}g(X)=0\text{，任意分类或拒绝决策。}\end{array}}$$

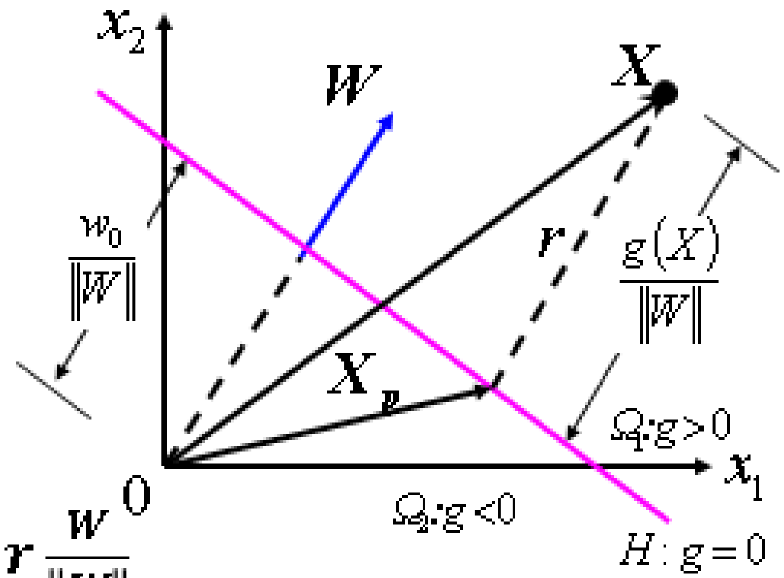
- 决策面方程 $g(X)=0$

几何解释

设X_1和X_2都在决策面H上，有：

$W^T X_1 + w_0 = W^T X_2 + w_0$,

$W^T (X_1 - X_2) = 0$



W和超平面H上任一向量正交，

即 **W是H的法向量**

特征空间某点X，表示成：$X = X_p + r\frac{W}{\|W\|}$

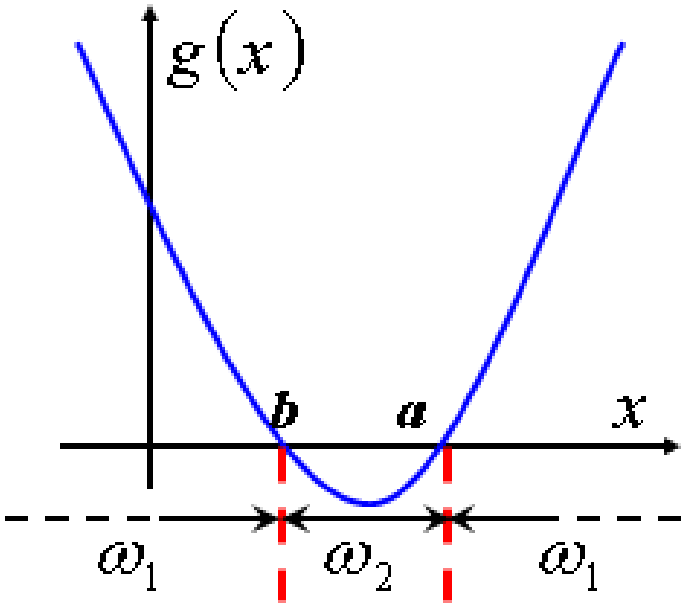
$$g(X) = W^T\left(X_p + r\frac{W}{\|W\|}\right) + w_0 = W^T X_p + w_0 + r\frac{W^T W}{\|W\|} = r\|W\|$$

$$\boxed{r = \frac{g(X)}{\|W\|}}$$

若X为原点，$g(X) = w_0$，到超平面的距离：$r = \frac{w_0}{\|W\|}$

（4）广义线性判别函数

把一些高次判别函数作适当变换，变换成一次的线性判别函数，称为广义线性判别函数。



如图所示，一维样本空间x，
如果$x<b$或$x>a$，则$x\in\omega_1$，
如果$b<x<a$，则$x\in\omega_2$。

采用线性判别函数无法分类

但二次判别函数适用

$$g(x)=(x-a)(x-b)$$

$$g(x)>0 \quad \text{即} x<b \text{或} x>a \quad \text{则} x\in\omega_1$$

$$g(x)<0 \quad \text{即} \ b<x<a \quad \text{则} x\in\omega_2$$

二次判别函数一般表达式：$\boxed{g(X) = C_0 + C_1 X + C_2 X^2}$

选择$X \to Y$的映射，变换二次函数为Y的线性函数

$$g(X) = A^T Y + a_0 \qquad a_0 = C_0$$

$$Y = [y_1 \quad y_2]^T = [X \quad X^2]^T \quad A = [a_1 \quad a_2]^T = [C_1 \quad C_2]^T$$

意义：经过以上变换，可以用简单的线性判别函数来解决复杂问题，但增加了维数。

（5）广义齐次线性判别函数

对于线性判别函数：$g(X) = W^T X + w_0$，可以进行类似的变换，变成广义齐次线性判别函数：

$$g(X) = W^T X + w_0 = \sum_{i=1}^{n} w_i x_i + w_0 = \sum_{i=0}^{n} a_i y_i = A^T Y$$

$$Y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n & 1 \end{bmatrix}^T = \begin{bmatrix} X & 1 \end{bmatrix}^T$$ ：增广样本向量

$$A = \begin{bmatrix} w_1 & w_2 & \cdots & w_n & w_0 \end{bmatrix}^T = \begin{bmatrix} W & w_0 \end{bmatrix}^T$$ ：增广权向量

经过变换，**维数增加一维**，但分界面变成了通过原点的**超平面**，给解决问题带来了方便。

（7）线性判别函数的设计

- 核心思想
  - 根据样本集去确定权向量W和w_0，或A
- 确定的方法
  - 首先要有一个准则函数，根据这个准则函数去找出满足要求的尽可能好的结果
  - 分类器的设计转化为求准则函数的极值
- 两个关键问题
  - 寻找合适的准则函数
  - 如何对准则函数求最优

- 设计步骤
  - 生成样本集：一般通过抽样生成，个别情况下要转化成增广样本集。
  - 确定准则函数
    - 极值对应最好的决策
    - 是W、w_0或A等参数的函数
  - 求最优值W^*、w_0^*或A^*

4.2.2最小二乘法

（1）基本概念

- 线性可分

- 对于样本集：$Z_1, Z_2, \cdots, Z$，Z_N为n维增广样本向量，分别来自ω_1、ω_2类，如果存在一个权向量A，使得对于任何$Z \in \omega_1$，$A^T Z > 0$，对于任何$Z \in \omega_2$，$A^T Z < 0$，称这组样本集是线性可分的。
- 利用线性判别函数对样本集进行分类，首要的前提条件是样本集线性可分。

■ 样本的规范化

- □ 若样本集$Y_1, Y_2, \cdots, Y_N$线性可分，则必存在权向量A，使得$A^T Y_i > 0$时，$Y_i \in \omega_1$；$A^T Y_j < 0$时，$Y_j \in \omega_2$。
- □ 令$Y_j' = -Y_j$，线性可分表示为：$A^T Y_k' > 0$时，Y_k分类正确。
- □ **构造新的样本集**：$Y_k' = \begin{cases} Y_i & Y_i \in \omega_1 \\ -Y_j & Y_j \in \omega_2 \end{cases}$，称为样本的规范化，$Y_k'$叫做规范化增广样本向量，以后仍用Y表示。

解向量和解区

- 满足$A^TY_k > 0$的权向量A^*称为解向量。
- 解向量往往不唯一，而是由无穷多个解向量组成的区域，称这样的区域为解区。
- 解区之中的向量都能满足要求，为避免求解向量时的算法不至收敛到解区边界的点上，引入余量b>0，$A^TY_k > b > 0$的解作为$A^TY_k > 0$的解。

（2）算法原理

设样本集为$\mathscr{X}=\{X_1,X_2,\cdots,X_N\}$，$X_i, i=1,2,\cdots,N$为$n$维样本向量，规范化增广样本向量$Z_i$为$\hat{n}=n+1$维，设余量$b_i>0$，线性判别函数权向量A满足$A^T Z_i=b_i>0$

矩阵形式 $\mathbf{Z}A=b$

$$\mathbf{Z}=\begin{pmatrix} Z_1^T \\ \vdots \\ Z_N^T \end{pmatrix}=\begin{pmatrix} z_{11} & \cdots & z_{1\hat{n}} \\ \vdots & \ddots & \vdots \\ z_{N1} & \cdots & z_{N\hat{n}} \end{pmatrix}$$

$$b=\begin{pmatrix} b_1 & b_2 & \cdots & b_N \end{pmatrix}^T$$

平方误差函数

$$J_S(A)=\|\mathbf{Z}A-b\|^2=\sum_{i=1}^{N}\left(A^T Z_i-b_i\right)^2$$

求最优 $A^*=\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T b=\mathbf{Z}^+ b$

判别函数 $g(X)=A^T Z$

4.2.3多元线性回归

- 如果仅对初始样本进行增广化，$g(X) = W^T X + w_0 = A^T Z$，$Z = [X \quad 1]^T$，$g(X) > 0$，$X \in \omega_1$，$g(X) < 0$，$X \in \omega_2$。
- 若$A^T Z_i = b_i$，b_i中对应第二类的元素应为负值；若取$b = \left(\underbrace{1 \quad \cdots \quad 1}_{N_1} \quad \underbrace{-1 \quad \cdots \quad -1}_{N_2}\right)^T$，实际是样本的类别标号向量$Y = (y_1 \ y_2 \ \cdots \ y_N)^T$，$y_i \in \{+1, -1\}$。
- 令$y_i = W^T X_i + w_0 = A^T Z_i$，去求解$W$、$w_0$或$A$，称为**多元线性回归**，可以采用最小二乘法方法，$A^* = \left(Z^T Z\right)^{-1} Z^T Y$。

例，对两类数据$\omega_1:\{(0\quad 0)^T,(0\quad 1)^T\}$，$\omega_2:\{(1\quad 0)^T,(1\quad 1)^T\}$利用多元线性回归求解线性判别函数的权向量。

$$A^* = \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T Y = \boldsymbol{Z}^+ Y$$

$$Y = (1\quad 1\quad -1\quad -1)^T$$

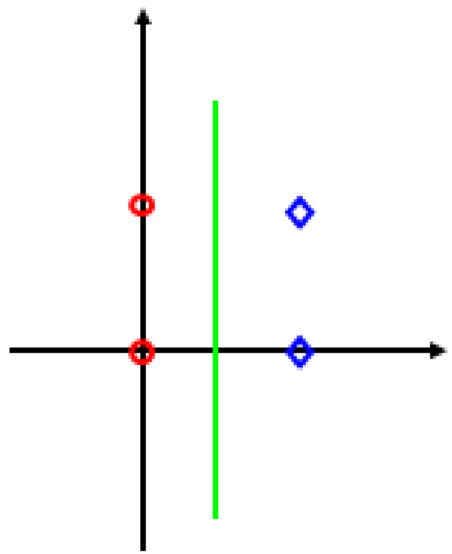
$$\boldsymbol{Z} = \begin{pmatrix} Z_1^T \\ \vdots \\ Z_N^T \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{Z}^T\boldsymbol{Z} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 2 & 4 \end{pmatrix}$$

$$\left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1} = \frac{1}{4}\begin{pmatrix} 4 & 0 & -2 \\ 0 & 4 & -2 \\ -2 & -2 & 3 \end{pmatrix}$$

$$Z^{+}=\left(Z^{T}Z\right)^{-1}Z^{T}=\frac{1}{4}\begin{pmatrix}4 & 0 & -2\\ 0 & 4 & -2\\ -2 & -2 & 3\end{pmatrix}\begin{pmatrix}0 & 0 & 1 & 1\\ 0 & 1 & 0 & 1\\ 1 & 1 & 1 & 1\end{pmatrix}=\frac{1}{4}\begin{pmatrix}-2 & -2 & 2 & 2\\ -2 & 2 & -2 & 2\\ 3 & 1 & 1 & -1\end{pmatrix}$$

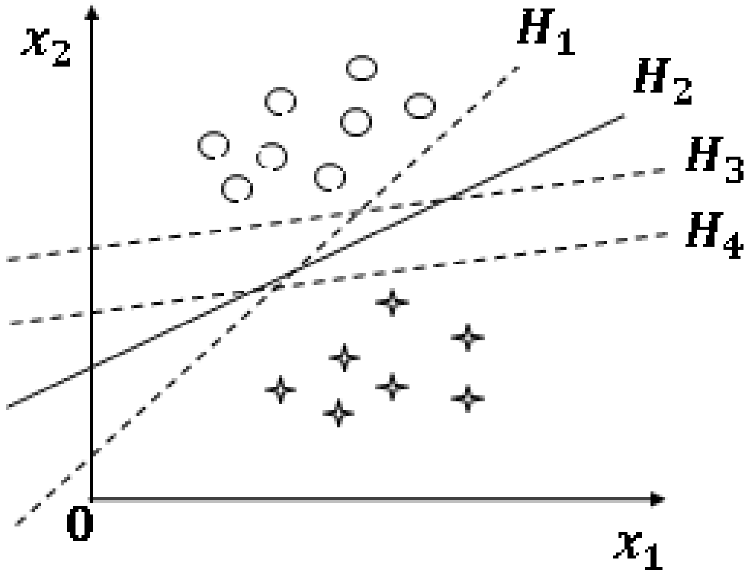
$$A^{*}=Z^{+}Y=\frac{1}{4}\begin{pmatrix}-2 & -2 & 2 & 2\\ -2 & 2 & -2 & 2\\ 3 & 1 & 1 & -1\end{pmatrix}\begin{pmatrix}1\\ 1\\ -1\\ -1\end{pmatrix}=\begin{pmatrix}-2\\ 0\\ 1\end{pmatrix}$$



判别函数 $$g(X)=A^{T}Z=\begin{pmatrix}-2 & 0 & 1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ 1\end{pmatrix}=-2x_1+1$$

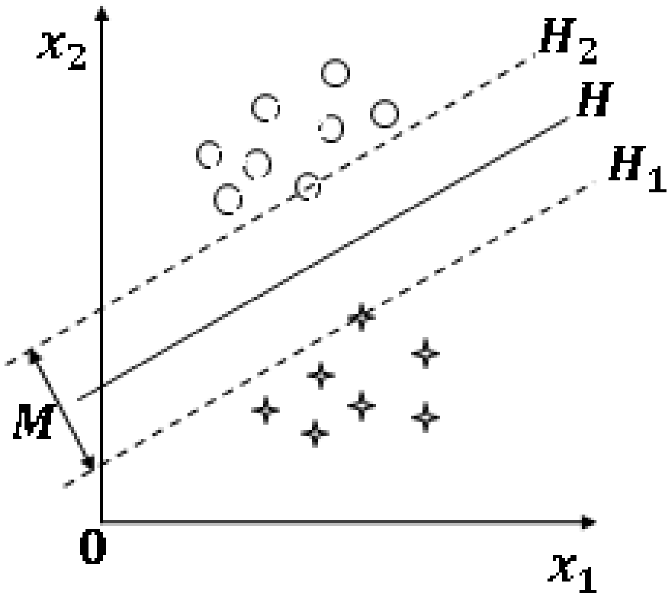
4.2.4支持向量机

（1）最优分类超平面与线性支持向量机



样本集线性可分，H_1、H_2、H_3、H_4四条分类线都能够把两类训练样本没有错误地分开，很明显，H_2的分类性能优于其他三条。

定义**最优分类超平面**



二维两类线性可分情况：

- 实心点和空心点：两类的训练样本
- H：把两类没有错误地分开的分类线
- H_1、H_2：过两类样本中离分类线最近的点且平行于分类线的直线

需要超平面与离它最近的点之间有最大**确信度**，即**间隔最大**，两类样本中离分类面最近的样本到分类面的距离

最优分类面

要求分类面不但能将两类无错误地分开，而且要使两类的分类间隔最大。

对线性可分判别函数引入余量，对于所有的X_i，满足：

$$\begin{cases} W^T X_i + w_0 \geq 1, y_i = +1 \\ W^T X_i + w_0 \leq -1, y_i = -1 \end{cases}$$，合并为一个式子：

$$y_i[W^T X_i + w_0] \geq 1, i = 1,2,\cdots,N$$

判别函数为：$f(X) = \mathrm{sgn}[g(X)] = \mathrm{sgn}[W^T X + w_0]$

因为$g(X_i)=y_i[W^TX_i+w_0]\geq 1$，距离分类面最近样本的$|g(X_i)|=1$，分类间隔等于$\frac{2}{\|W\|}$。

要使分类间隔最大，即使$\|W\|$最小；要使分类面对所有样本正确分类，即$y_i[W^TX_i+w_0]\geq 1$；满足这两点就是最优分类面。

$$\max_{W,w_0} \frac{2}{\|W\|}$$
$$s.t.\quad y_i\left(W^TX_i+w_0\right)\geq 1, i=1,2,\cdots,N$$

$$\min_{W,w_0} \frac{1}{2}\|W\|^2$$
$$s.t.\quad y_i\left(W^TX_i+w_0\right)\geq 1, i=1,2,\cdots,N$$

在$y_i[W^TX_i+w_0]\geq 1$的约束下，求$\|W\|^2$的最小值

■ 求解最优分类面

□ 构建拉格朗日函数

$$L(W,w_0,\Lambda)=\frac{1}{2}\|W\|^2-\sum_{i=1}^{N}\lambda_i\left[y_i\left(W^TX_i+w_0\right)-1\right]$$

$$\begin{cases}\lambda_i\geq 0\\ y_i\left(W^TX_i+w_0\right)-1\geq 0\\ \lambda_i\left[y_i\left(W^TX_i+w_0\right)-1\right]=0\end{cases}$$

$$\min_{W,w_0}\max_{\lambda\geq 0}L(W,w_0,\Lambda)$$

$$\max_{\lambda\geq 0}\min_{W,w_0}L(W,w_0,\Lambda)$$

□ 求解

令$L(W,w_0,\lambda)$对W、w_0的偏导为零

$$W=\sum_{i=1}^{N}\lambda_iy_iX_i \qquad \sum_{i=1}^{N}\lambda_iy_i=0$$

代入$L(W,w_0,\lambda)$中，消去W和w_0：

$$L(\lambda)=\sum_{i=1}^{N}\lambda_i-\frac{1}{2}\sum_{i,j}^{N}\lambda_i\lambda_jy_iy_jX_i^TX_j$$

最终得到最优的权向量是训练向量的线性组合

$$W^* = \sum_{k=1}^{N} \lambda_k y_k X_k$$

$\sum_{k=1}^{N} \lambda_k y_k = 0$，且$\lambda_k$是$L(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j X_i^T X_j$最大的最优解，且$\lambda_k \geq 0$，$y_k\left[W^T X_k + w_0\right] = 1$时，$\lambda_k > 0$，否则$\lambda_k = 0$。

参与权向量运算的只有$y_k[W^T X_k + w_0] = 1$的样本，就是过两类样本中离分类面最近的点且平行于最优分类面的超平面H_1、H_2上的训练样本—**支持向量**

□ w_0的求解

用所有支持向量X_k对$y_k[W^T X_k + w_0] = 1$求解w_{0k}，再取平均

□ 判别函数

$$f(X) = \mathrm{sgn}[g(X)] = \mathrm{sgn}\left[W^{*T} X + w_0^*\right] = \mathrm{sgn}\left\{\sum_{k=1}^{N} \lambda_k y_k X_k^T X + w_0^*\right\}$$

由于最优分类面的解最后完全由支持向量决定，因此这种方法后来被称作**支持向量机(support vector machines-SVM)**

以上讨论仅是线性可分情况下的线性支持向量机。

（3）SMO算法

支持向量机的求解都依赖于下列式子的最优解

$$\begin{cases} \max\limits_{\lambda} L(\lambda) = \sum\limits_{i=1}^{N} \lambda_i - \dfrac{1}{2} \sum\limits_{i,j}^{N} \lambda_i \lambda_j y_i y_j X_i^T X_j \\ s.t. \quad 0 \le \lambda_k \le 1, k = 1,2,\cdots,N \\ \sum\limits_{k=1}^{N} \lambda_k y_k = 0 \end{cases}$$

$$\begin{cases} \min\limits_{\lambda} L(\lambda) = \dfrac{1}{2} \sum\limits_{i,j}^{N} \lambda_i \lambda_j y_i y_j X_i^T X_j - \sum\limits_{i=1}^{N} \lambda_i \\ s.t. \quad 0 \le \lambda_k \le 1, k = 1,2,\cdots,N \\ \sum\limits_{k=1}^{N} \lambda_k y_k = 0 \end{cases}$$

例：两类样本ω_1：$\{(0,0)^T\}$，ω_2：$\{(1,1)^T\}$，求支持向量机的判别函数。

解：$y_1 = 1$，$y_2 = -1$，$N = 2$；

$\because \sum_{k=1}^{N} \lambda_k y_k = 0$，$\therefore \lambda_1 = \lambda_2$，$0 \le \lambda_{1,2} \le 1$

$$L(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j X_i^T X_j$$

$$= 2\lambda_2 - \frac{1}{2}\begin{pmatrix} \lambda_1 \lambda_1 y_1 y_1 X_1^T X_1 + \lambda_1 \lambda_2 y_1 y_2 X_1^T X_2 \\ + \lambda_2 \lambda_1 y_2 y_1 X_2^T X_1 + \lambda_2 \lambda_2 y_2 y_2 X_2^T X_2 \end{pmatrix}$$

$$= 2\lambda_2 - \lambda_2^2 = -(\lambda_2 - 1)^2 + 1$$

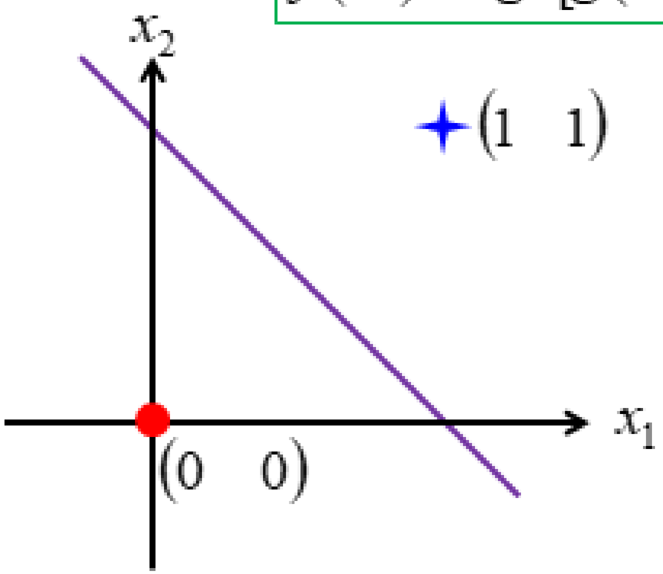
最大值为1，$\lambda_1 = \lambda_2 = 1$

$$W^* = \sum_{k=1}^{N} \lambda_k y_k X_k = (-1 \quad -1)^T$$

用所有支持向量X_k对$y_k[W^T X_k + w_0] = 1$求解w_{0k}，再取平均

$$\begin{cases} y_1\left[W^T X_1 + w_{01}\right] = 1 \\ y_2\left[W^T X_2 + w_{02}\right] = 1 \end{cases} \Rightarrow \begin{cases} w_{01} = 1 \\ w_{02} = 1 \end{cases} \Rightarrow w_0 = 1$$

判别函数：$f(X) = \mathbf{sgn}\left[g(X)\right] = \mathbf{sgn}\left[W^{*T} X + w_0^*\right] = \mathbf{sgn}\{-x_1 - x_2 + 1\}$



样本数增多，$L(\lambda)$求解会变得复杂，需要优化算法

SMO：Sequential Minimal Optimization，序列最小优化算法，一种高效求解支持向量机的算法

基本思路：在一次迭代中，只优化两个变量，固定其他变量，将一个大的优化问题分解为若干个小的优化问题求解。

（3）非线性可分情况

- **问题分析**

样本集$\{X_i\}, i=1,2,\cdots,N$，类别标号$y_i \in \{+1,-1\}$；
非线性可分：即有样本X_i**不满足**$y_i[W^TX_i+w_0]-1 \geq 0$

给每一个样本引入一个非负的松弛变量ξ_i，使得
$y_i[W^TX_i+w_0]-1+\xi_i \geq 0, i=1,2,\cdots,N$；
若样本X_i被正确分类，$y_i[W^TX_i+w_0]-1 \geq 0$， $\xi_i=0$；
若样本X_j被错误分类，$y_j[W^TX_j+w_0]-1<0$， $\xi_j>0$。

所有样本的松弛因子之和$\sum_{i=1}^{N}\xi_i$反应了在整个训练样本集上的**错分程度**：错分样本越多，$\sum_{i=1}^{N}\xi_i$越大；同时，样本错误的程度越大（错误的方向上远离分类面），$\sum_{i=1}^{N}\xi_i$也越大。**希望$\sum_{i=1}^{N}\xi_i$尽可能小。**

要求分类面不但要使两类的**分类间隔最大**，而且要**错分样本尽可能少且错误程度尽可能低。**

目标函数：

$$\min_{w,b}\frac{1}{2}\|W\|^2+C\sum_{i=1}^{N}\xi_i$$

C为常数，人为选择，反映两个目标的折中：C较小，对错误比较容忍，强调正确分类的样本的分类间隔；C较大，强调对分类错误的惩罚。

问题表述

训练样本集$\{X_i\}, i = 1, 2, \cdots, N$，类别标号$y_i \in \{+1, -1\}$；求解：

$$\min_{w,b} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$s.t. \quad y_i\left[W^T X_i + w_0\right] - 1 + \xi_i \geq 0, i = 1,2,\cdots,N$$

$$\text{且：} \quad \xi_i \geq 0, i = 1,2,\cdots,N$$

目标函数：

$$\min_{W,b,\xi_i} \max_{\lambda} L(W, w_0, \lambda) = \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\lambda_i\left\{y_i\left[W^T X_i + w_0\right] - 1 + \xi_i\right\} - \sum_{i=1}^{N}\gamma_i\xi_i$$

$$\lambda_i \geq 0, \quad \gamma_i \geq 0$$

- 方程求解

$\lambda_k > 0$的训练向量（**软间隔的支持向量**）的线性组合

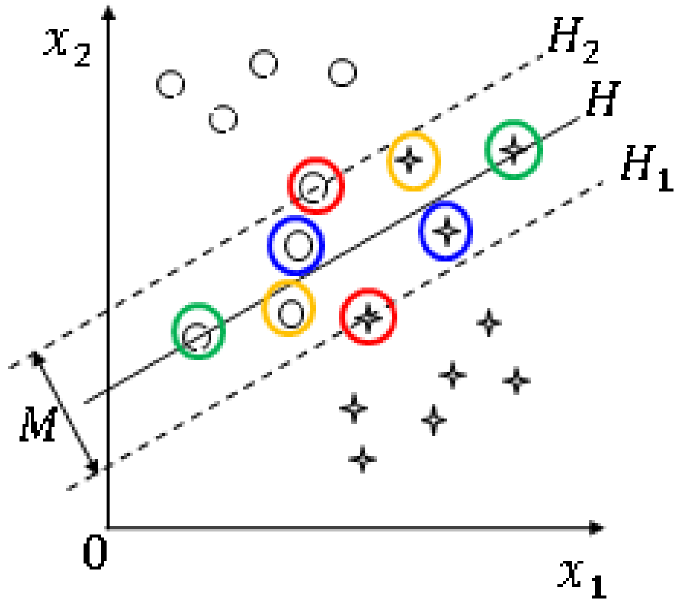
$$W^* = \sum_{k=1}^{N} \lambda_k y_k X_k,$$

$\sum_{k=1}^{N} \lambda_k y_k = 0$，且$\lambda_k$是$L(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j X_i^T X_j$最大的最优解，且$0 \le \lambda_k \le C$，$\lambda_k = C$时，$\xi_k > 0$，其余$\xi_k = 0$。

KKT条件：

$$\begin{cases}\lambda_k \geq 0, \gamma_k \geq 0 \\ \gamma_k \xi_k = 0 \\ \lambda_k\left[y_k\left(W^T X_k + w_0\right) - 1 + \xi_k\right] = 0\end{cases}$$

当$y_k(W^T X_k + w_0) - 1 + \xi_k = 0$时，$\lambda_k > 0$，否则$\lambda_k = 0$。参与权向量运算的只有$\lambda_k > 0$的样本，称为软间隔支持向量



1）分类正确但处于分类间隔边界面上，$0 < \lambda_k < C$，$\xi_k = 0$；

2）分类错误的样本，$\lambda_k = C$，$\xi_k > 0$。

$0 < \xi_k < 1$，支持向量X_k在间隔边界和分类超平面之间；

$\xi_k = 1$，支持向量X_k在分类超平面上；

$\xi_k > 1$，支持向量X_k位于分类超平面错分一侧。

w_0的求解：

用$0 < \lambda_k < C$的样本求解$y_k[W^T X_k + w_0] - 1 + \xi_k = 0$，得$w_0$

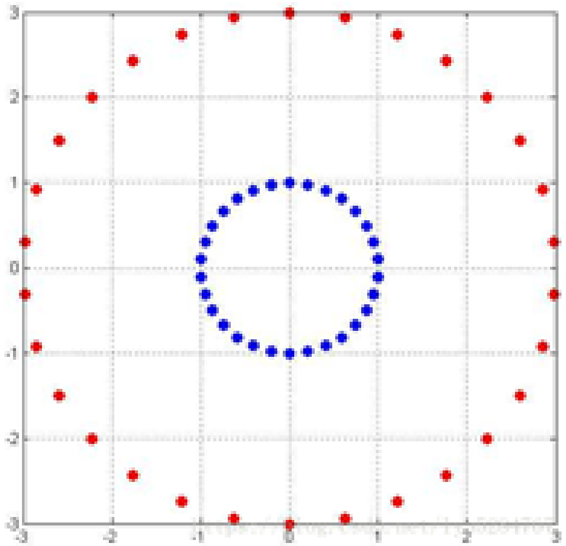
判别函数：

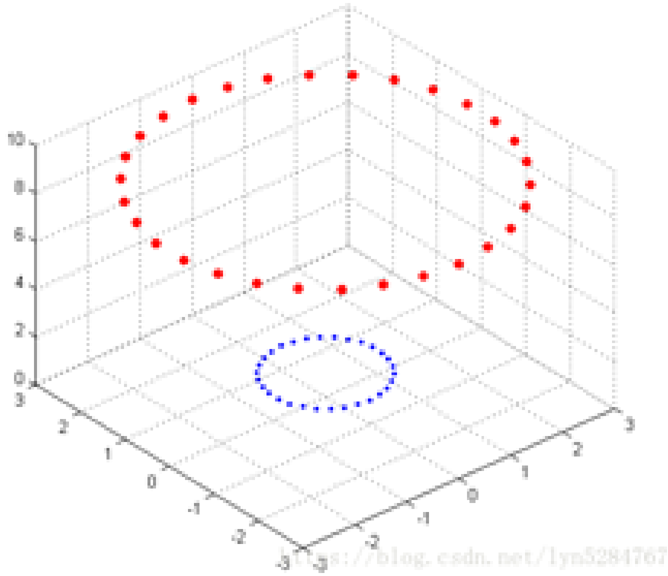
$$f(X) = \mathbf{sgn}[g(X)] = \mathbf{sgn}\left[W^{*T}X + w_0^*\right] = \mathbf{sgn}\left\{\sum_{k=1}^{N} \lambda_k y_k X_k^{\ T} X + w_0^*\right\}$$

（4）核函数变换与非线性支持向量机

- **问题分析**

对于原空间中的非线性问题，通过特征变换将到新空间，在这个新空间中求取最优线性分类面。





问题求解

对特征X进行非线性变换，记新特征为$Z = \varphi(X)$；

新空间中求解的最优分类面为：

$$W^* = \sum_{k=1}^{N} \lambda_k y_k \varphi(X_k) \text{且} \sum_{k=1}^{N} \lambda_k y_k = 0$$

λ_k是$L(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{k,j}^{N} \lambda_k \lambda_j y_k y_j \varphi(X_j)^T \varphi(X_k)$最大的最优解，且$\lambda_k \geq 0$，$y_k\left[W^T \varphi(X_k)\right] = 1$时，$\lambda_k > 0$，否则$\lambda_k = 0$。

w_0的求解：

用所有支持向量X_k对$y_k[W^T\varphi(X_k)+w_0]=1$求解$b_k$，再取平均，即对 $y_k\left[\sum_{k=1}^{N}\lambda_k y_k\varphi^T(X_k)\varphi(X_k)+w_0\right]=1$ 求解。

判别函数：

$$f(Z)=\mathbf{sgn}\left[g(Z)\right]=\mathbf{sgn}\left[W^{*T}\varphi(X_k)+w_0\right]=\mathbf{sgn}\left\{\sum_{k=1}^{N}\lambda_k y_k\varphi^T(X_k)\varphi(X_k)+w_0^*\right\}$$

结论分析

- 无论变换的具体形式如何，变换对支持向量机的影响是把两个样本在原特征空间中的内积变成了新空间中的内积：

$$<X_j, X_k> \Rightarrow <\varphi(X_j), \varphi(X_k)> = K(X_j, X_k)$$

- 因此，变换空间的线性支持向量机求解可以在原空间通过核函数$K(X_j, X_k)$进行，避免高维空间的计算。
- 进一步，可以采用直接设计核函数而不设计变换函数φ的方法求解非线性的支持向量机

■ 核函数

□ 定义

设χ是输入空间（欧氏空间R^n的子集或离散集合），H是特征空间，如果存在一个χ到H的映射

$$\varphi(X):\ \chi \to H$$

使得对所有的$X, X' \in \chi$，函数$K(X, X') = <\varphi(X), \varphi(X')>$，那么称$K(X, X')$为核函数，$\varphi(X)$为映射函数。

Mercer条件

对于任意的对称函数$K(X,X')$，是某个特征空间中的内积运算的充要条件是：对于任意的$\varphi \neq 0$且$\int \varphi^2(x)dx < \infty$，有

$$\iint K(x,x')\varphi(x)\varphi(x')dxdx' > 0$$

选择一个满足Mercer条件的核函数，可以构建非线性支持向量机。进一步证明，该条件可放松为满足如下条件的正定核：

$K(X,X')$是定义在空间χ上的对称函数，且对任意训练数据$X_i \in X, i = 1,2,\cdots,N$和任意的实系数$\alpha_i \in R, i = 1,2,\cdots,N$，有$\sum_{i,j}\alpha_i\alpha_j K\left(X_i,X_j\right) \geq 0$

□ 常用的核函数

多项式核函数 $K(X,X')=\left[<X,X'>+1\right]^q$

径向基(RBF)核函数 $K(X,X')=\exp\left(-\frac{\|X-X'\|^2}{\sigma^2}\right)$

Sigmoid函数 $K(X,X')=\tanh\left[v<X,X'>+c\right]$

支持向量机通过选择不同的核函数实现不同形式的非线性分类器；核函数需要针对具体问题来具体选择，很难有一个一般性的准则。

（5）支持向量机概括

- 线性支持向量机：即利用支持向量设计最优分类面；
- 非线性数据集设计线性支持向量机：引入余量；
- 非线性支持向量机：
  - 通过非线性变换将输入空间变换到高维空间，然后在这个新空间中求取最优线性分类面。
  - 非线性变换通过定义适当的内积核函数实现

$$\omega_1:\{(0\quad 0\quad 0)^T,(1\quad 0\quad 0)^T,(1\quad 0\quad 1)^T,(1\quad 1\quad 0)^T\}$$
$$\omega_2:\{(0\quad 0\quad 1)^T,(0\quad 1\quad 1)^T,(0\quad 1\quad 0)^T,(1\quad 1\quad 1)^T\}$$

$$\omega_1:\{(1\quad 1)^T,(-1\quad -1)^T\}$$
$$\omega_2:\{(1\quad -1)^T,(-1\quad 1)^T\}$$