河北师范大学软件学院
Software College of Hebei Normal University

分类模型

第一部分 线性分类模型

张朝晖

2018-2019学年 20180918-0919

河北师范大学软件学院
Software College of Hebei Normal University

1.分类模块

产生式分类模型 A.贝叶斯分类模型

判别式分类模型

线性分类模型
- B. Fisher判别分类
- C. 感知器分类模型
- D. 大间隔分类模型（线性*SVM*）

非线性分类模型
- E. 核*SVM*（非线性*SVM*）
- F. 核Fisher判别分类
- G. 神经网络

其它分类模型
- *H*.KNN分类模型
- *I*.决策树分类模型
- *J*.*Logistic*回归
- *K*.*Softmax*回归

2.聚类模块
- *L*.K-均值聚类
- *M*.高斯混合聚类
- *N*.DBSCAN聚类
- *O*.层次聚类

3.回归模块
- *P*.KNN回归
- *Q*.回归树
- *R*.最小二乘线性回归
- *S*.岭回归
- *T*.LASSO回归

4.集成学习
- *U*.*Bagging*
- *V*.随机森林
- *W*.*Boosting*

5.特征工程
- *X*.主成分分析*(PCA)*
- ...

6.评价模块 混淆矩阵（及其相关指标）、ROC曲线、交叉验证

主要内容

A.引言

B. Fisher线性判别分析

C.感知器

<u>D. 线性支持向量机(线性SVM)</u>

Support Vector Machine: SVM

学习要求：

- 理解SVM分类模型构建的基本思想
- 掌握SVM 目标函数的构造形式、意义
- 掌握SVM分类模型的学习步骤、使用流程
- 掌握SVM进行超参数寻优的典型方法
- 能够熟练应用SVM进行分类

关键词：

- 分类间隔、分类边界、判别函数
- 训练样本集的错分程度
- 分类超平面、超平面法向量、超参数
- 支持向量、非支持向量
- 判别函数、分类边界

- SVM(Support Vector Machine)是支持向量机的简称。“机”,对应机器学习中的“机器”，是“算法”的意思。

  所谓支持向量机就是一种与“支持向量”有关的机器学习算法。

- SVM 的求解最后转化成凸二次规划问题的求解，因此SVM 的解是全局唯一的最优解.

- SVM在解决小样本(样本数<特征维数)、非线性、高维模式分类问题中表现出许多特有的优势，并能够推广应用到函数拟合(回归)等其它机器学习问题中.

- SVM 用于分类，就是*SVC (Support Vector Classification)*；用于回归就是*SVR (Support Vector Regression)*。

第1部分

两类别分类情况下的线性SVC

主要内容

1 两类别分类问题：线性可分情况

2 两类别分类问题：近似线性可分情况

3 多类别的线性SVM

4 应用举例

SVM始于两类问题的解决需求

两类别分类问题的描述：

给定训练集 $T=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_N,y_N)\}$

其中 $\begin{cases}\boldsymbol{x}_i\in R^d \\ y_i\in\mathcal{Y}=\{1,-1\}\end{cases}\quad i=1,2,\ldots,N$

要求：寻找 R^d 空间上的一个实值函数 $g(\boldsymbol{x})$，从而可用决策函数 $f(\boldsymbol{x})=\operatorname{sgn}(g(\boldsymbol{x})):R^d\to\mathcal{Y}$；推断任意输入 $\boldsymbol{x}$ 对应的输出值 y.

线性可分问题-定义：

给定**训练集** $\boldsymbol{T}=\left\{\left(\boldsymbol{x}_1, y_1\right),\cdots,\left(\boldsymbol{x}_N, y_N\right)\right\}$

其中 $\boldsymbol{x}_i \in R^d, i=1,2,\ldots,N$

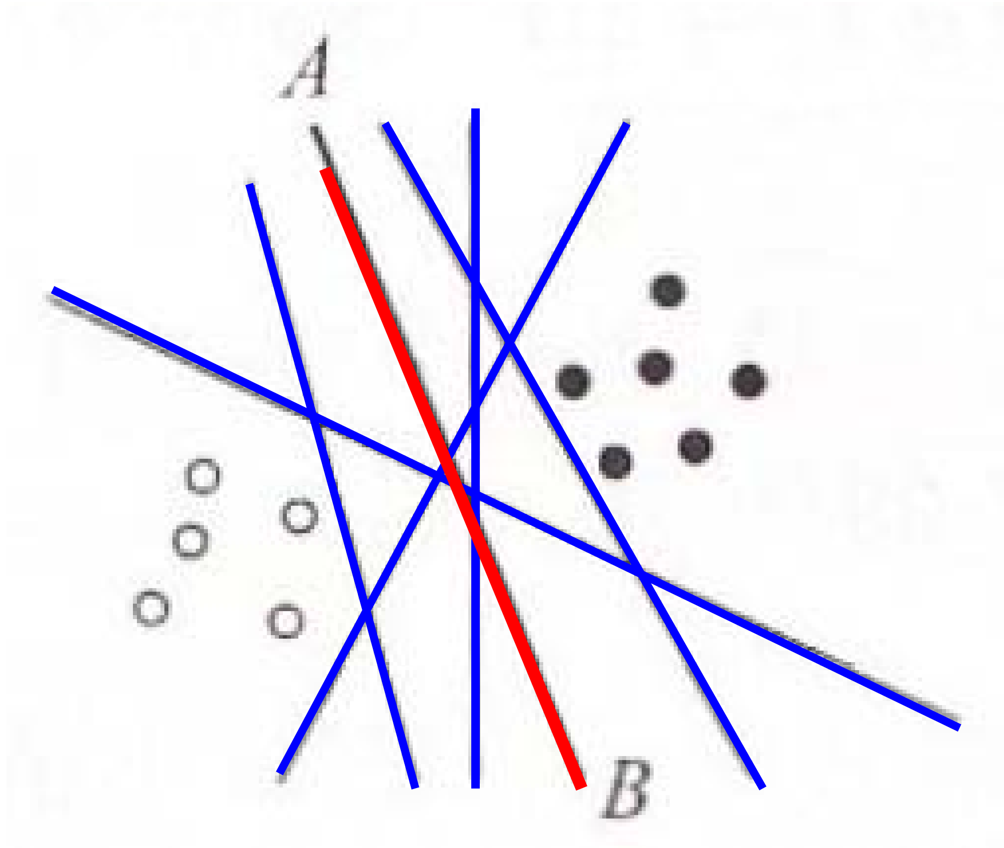
$y_i \in \mathcal{Y}=\{1,-1\}$

若存在$\omega \in R^d$，$b \in R$,以及正数ε，使得

$$\begin{cases}\text{对于所有}y_i=+1\text{的样本}\boldsymbol{x}_i\text{，有}g\left(\boldsymbol{x}_i\right)=\omega\bullet\boldsymbol{x}_i+b\geq\varepsilon \\ \text{对于所有}y_i=-1\text{的样本}\boldsymbol{x}_i\text{，有}g\left(\boldsymbol{x}_i\right)=\omega\bullet\boldsymbol{x}_i+b\leq-\varepsilon.\end{cases}$$

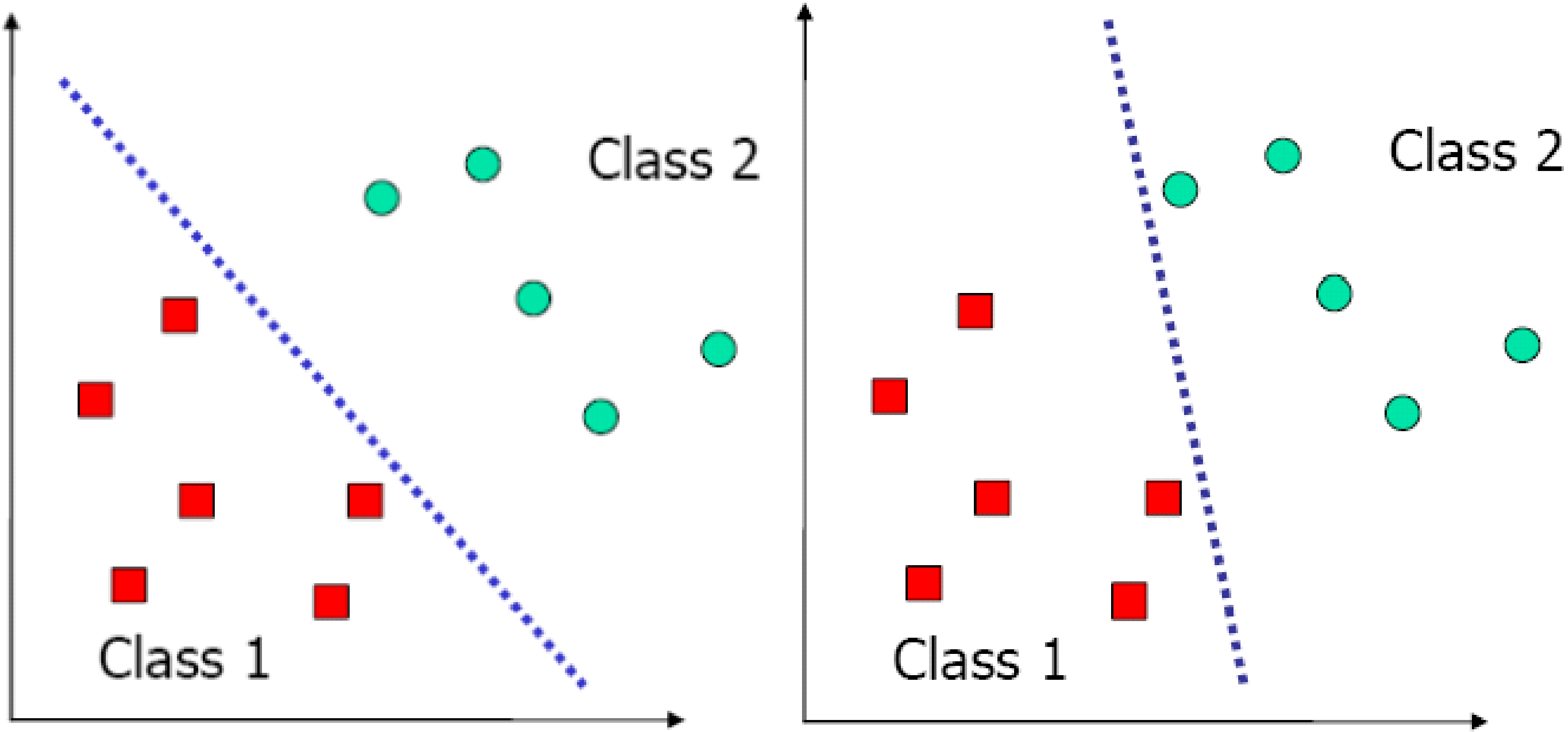
则称**训练集** $\boldsymbol{T}$ 线性可分。

1. 例： 二维、两类、线性可分情况



问题：可能存在许多分类边界（决策边界），若将两类完全分开，应当选择哪一个？

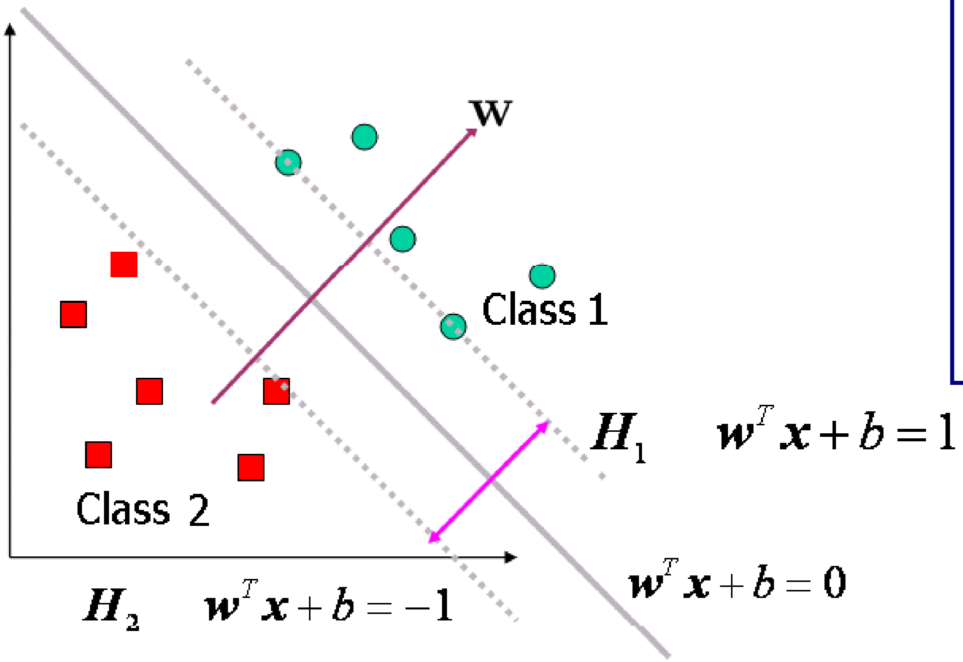
➤ 较差的分类边界(如图)



➤ 较好的分类边界：

分类间隔（margin）尽可能大；

分类边界尽可能远离两类样本数据



- **最优分类边界：**

  能将两类无错误分开；

  分类间隔(margin)最大

- H -----最优分类线；
- H_1, H_2----平行于H、且经过各类训练样本集中离H最近的训练样本
- 分类间隔(margin)---- H_1, H_2之间的距离
- 推广到高维空间：最优边界就变为最优超平面 (*optimal hyperplane*)

如何找到最优分类边界?

分类边界 $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$

判别函数 $g(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b$

2. 线性SVM的分类间隔

线性可分训练样本集 $\{\boldsymbol{x}_i, y_i\} \quad i=1,\dots,N \quad \boldsymbol{x}_i \in R^d$

类别标号 $y_i \in \{-1,1\}$

d维空间的线性判别函数 $g(\boldsymbol{x}) = \boldsymbol{w}\cdot\boldsymbol{x} + b$

判别函数归一化 $\begin{cases} \text{两类所有样本满足} & |g(\boldsymbol{x}_i)| \geq 1 \\ \text{离分类面最近样本} & |g(\boldsymbol{x}_i)| = 1 \end{cases}$

训练样本线性可分 $\begin{cases} g(\boldsymbol{x}_i) \geq 1 & \text{对于} y_i = 1 \\ g(\boldsymbol{x}_i) \leq -1 & \text{对于} y_i = -1 \end{cases}$

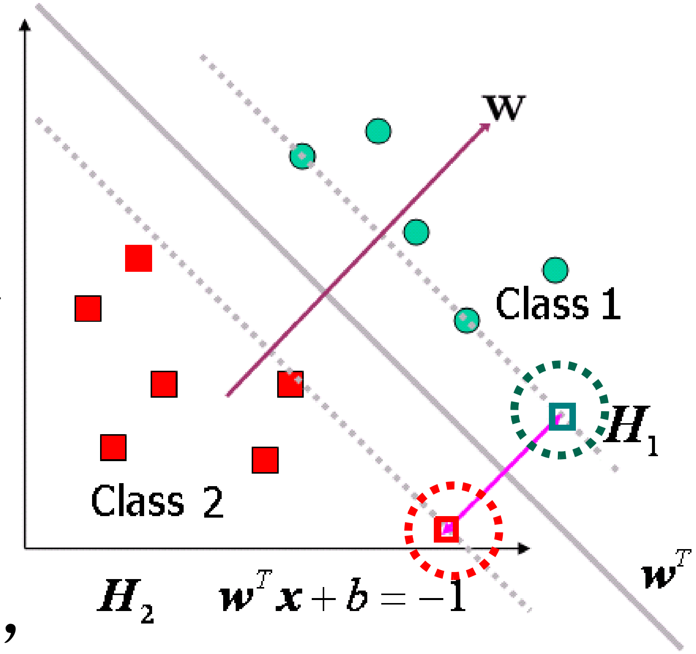
$$\Rightarrow \quad y_i \cdot g(\boldsymbol{x}_i) \geq 1$$

$$\Rightarrow y_i \cdot [\boldsymbol{w}\cdot\boldsymbol{x}_i + b] - 1 \geq 0$$

最优分类面H：$\boldsymbol{w}\cdot\boldsymbol{x}+b=0$

正平面H_1:　　$\boldsymbol{w}\cdot\boldsymbol{x}+b=1$

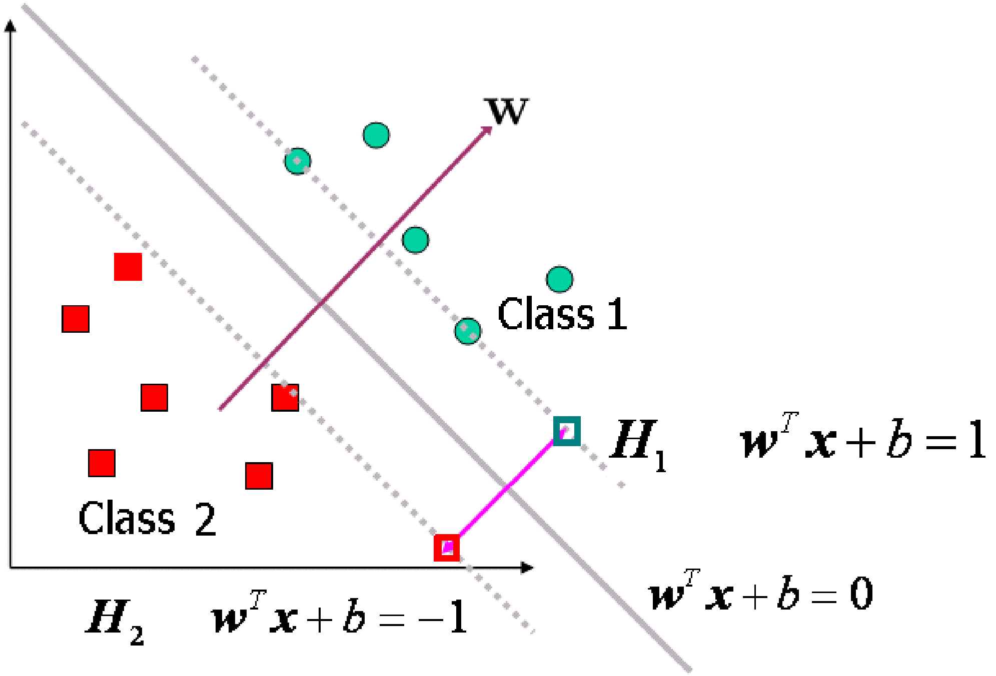
负平面H_2:　　$\boldsymbol{w}\cdot\boldsymbol{x}+b=-1$



设$\begin{cases}\boldsymbol{x}^+\text{为正平面}H_1\text{上任意点，}\\ \boldsymbol{x}^-\text{为负平面}H_2\text{上最接近}\boldsymbol{x}^+\text{的点，}\boldsymbol{x}^+=\boldsymbol{x}^-+\lambda\boldsymbol{w}\end{cases}$

则$\begin{cases}\text{正平面上}\quad \boldsymbol{w}\cdot\boldsymbol{x}^++b=1\\ \text{负平面上}\quad \boldsymbol{w}\cdot\boldsymbol{x}^-+b=-1\\ \text{两平面间隔}\quad margin=\left\|\boldsymbol{x}^+-\boldsymbol{x}^-\right\|=\left\|\lambda\boldsymbol{w}\right\|=\lambda\left\|\boldsymbol{w}\right\|\end{cases}$

$$\begin{cases} \text{正平面上：} \ \mathbf{w}\cdot\mathbf{x}^{+}+b=1 \\ \text{负平面上：} \ \mathbf{w}\cdot\mathbf{x}^{-}+b=-1 \\ \mathbf{x}^{+}=\mathbf{x}^{-}+\lambda\mathbf{w} \end{cases}$$



$$\rightarrow \mathbf{w}\cdot\left(\mathbf{x}^{-}+\lambda\mathbf{w}\right)+b=1$$

$$\rightarrow \mathbf{w}\cdot\mathbf{x}^{-}+b+\lambda\mathbf{w}\cdot\mathbf{w}=1$$

$$\rightarrow -1+\lambda\mathbf{w}\cdot\mathbf{w}=1$$

$$\rightarrow \lambda=\frac{2}{\mathbf{w}\cdot\mathbf{w}}=\frac{2}{\|\mathbf{w}\|^{2}}$$

$$\boxed{\text{分类间隔} \ margin=\left\|\mathbf{x}^{+}-\mathbf{x}^{-}\right\|=\|\lambda\mathbf{w}\|=\frac{2}{\|\mathbf{w}\|}}$$

3. 线性SVM—目标函数的构造

最优分类边界

$$\begin{cases} \text{分类间隔最大} \quad margin = \dfrac{2}{\|\mathbf{w}\|} & \Leftrightarrow \quad \text{使}\|\mathbf{w}\|\left(\text{或}\|\mathbf{w}\|^2\right)\text{最小} \\ \text{训练样本线性可分(均正确分类)} & \Leftrightarrow \quad y_i \cdot \left(\mathbf{w}\cdot\mathbf{x}_i + b\right) - 1 \geq 0 \\ & \quad\; i = 1, 2, \ldots, N \end{cases}$$

$$\phi\left(\mathbf{w}\right) = \frac{1}{2}\|\mathbf{w}\|^2 = \frac{1}{2}\left(\mathbf{w}\cdot\mathbf{w}\right) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

原始问题的目标函数(线性可分问题的最大间隔法)

约束优化问题(凸二次规划)

如何求解?

$$\begin{cases} \min\limits_{\mathbf{w},b} \quad \phi\left(\mathbf{w}\right) = \dfrac{1}{2}\left(\mathbf{w}\cdot\mathbf{w}\right) \\ s.t. \qquad y_i \cdot \left(\mathbf{w}\cdot\mathbf{x}_i + b\right) - 1 \geq 0 \quad i = 1, 2, \ldots, n \end{cases}$$

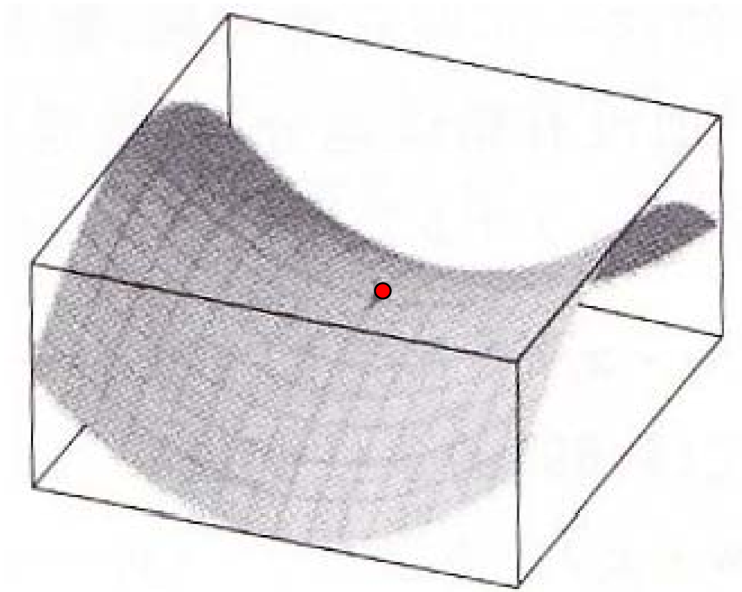
4. 线性SVM—目标函数的求解

构造***Lagrange***辅助目标函数，将**原始问题**转化为**对偶问题**.

$$L(\mathbf{w},b,\alpha)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})-\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]-1\right\}$$

其中非负*Lagrange*乘子向量$\boldsymbol{\alpha}=\left[\alpha_1 \quad \cdots \quad \alpha_N\right]^T$

显然：$L(\mathbf{w},b,\alpha)\le\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})$



原始问题等价为：

$$\min_{\mathbf{w},b}\max_{\alpha}L(\mathbf{w},b,\alpha)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})-\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]-1\right\}$$

$$L(\mathbf{w},b,\boldsymbol{\alpha})=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})-\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot\left[(\mathbf{w}\cdot\mathbf{x}_i)+b\right]-1\right\}$$

$$=\frac{1}{2}\left(\mathbf{w}^T\mathbf{w}\right)-\sum_{i=1}^{N}\alpha_i y_i\mathbf{w}^T\mathbf{x}_i-b\sum_{i=1}^{N}\alpha_i y_i+\sum_{i=1}^{N}\alpha_i$$

$L(\mathbf{w},b,\alpha)$在最优解处满足:

$$\begin{cases}\text{条件 1} \quad \dfrac{\partial L(\mathbf{w},b,\alpha)}{\partial \mathbf{w}}=0 \quad \rightarrow \quad \boxed{\mathbf{w}=\sum_{i=1}^{N}\alpha_i y_i\mathbf{x}_i} \\ \text{条件2} \quad \dfrac{\partial L(\mathbf{w},b,\alpha)}{\partial b}=0 \quad \rightarrow \quad \boxed{\sum_{i=1}^{N}\alpha_i y_i=0}\end{cases}$$

该条件下，$L(\boldsymbol{w},b,\alpha)$为

$$Q(\alpha)=\sum_{i=1}^{N}\alpha_i+\frac{1}{2}\left(\boldsymbol{w}^T\boldsymbol{w}\right)-\boldsymbol{w}^T\sum_{i=1}^{N}\alpha_i y_i\boldsymbol{x}_i$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\boldsymbol{w}^T\sum_{i=1}^{N}\alpha_i y_i\boldsymbol{x}_i$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\left(\sum_{j=1}^{N}\alpha_j y_j\boldsymbol{x}_j\right)^T\sum_{i=1}^{N}\alpha_i y_i\boldsymbol{x}_i$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\left(\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\boldsymbol{x}_j{}^T\boldsymbol{x}_i\right)$$

$$=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_j\cdot\boldsymbol{x}_i\right)$$

对偶问题

$$\max_{\boldsymbol{\alpha}}. \quad Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_j\cdot\boldsymbol{x}_i\right)$$

$$s.t. \quad \begin{cases}(1) \quad \sum_{i=1}^{N}\alpha_i y_i=0 \\ (2) \quad \alpha_i\geq 0 \qquad \text{对于} i=1,\ldots,N\end{cases}$$

等价形式

对偶问题

$$\min_{\boldsymbol{\alpha}}. \quad W(\boldsymbol{\alpha})=\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_j\cdot\boldsymbol{x}_i\right)-\sum_{i=1}^{N}\alpha_i$$

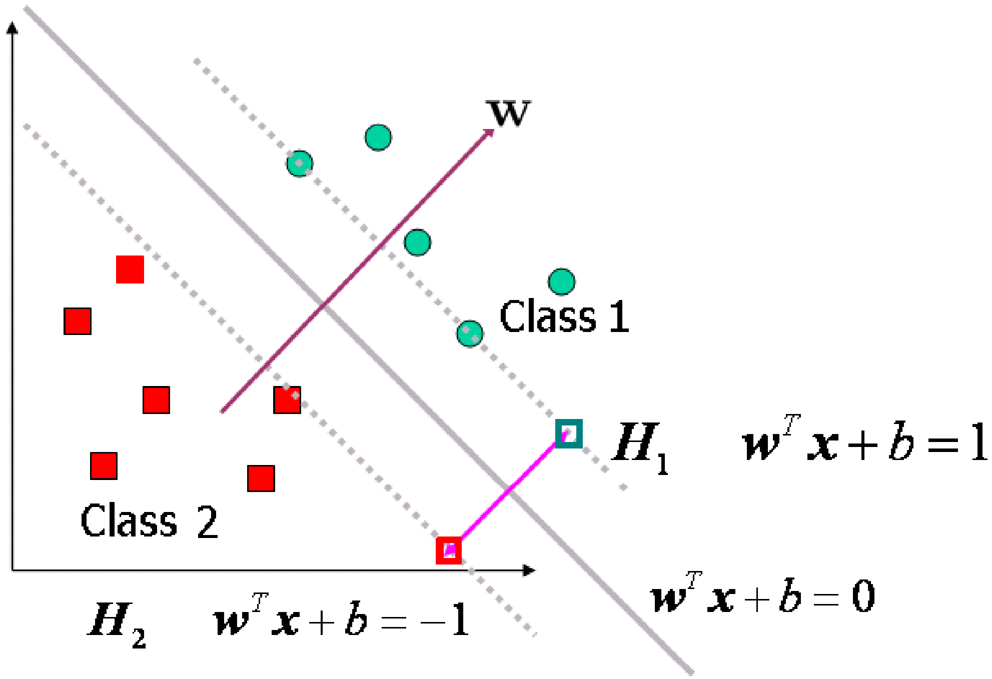
$$s.t. \quad \begin{cases}(1) \quad \sum_{i=1}^{N}\alpha_i y_i=0 \\ (2) \quad -\alpha_i\leq 0 \qquad \text{对于} i=1,\ldots,N\end{cases}$$

解得**非负**乘子向量的最优解：$\boldsymbol{\alpha}^*=\left[\alpha_1^* \quad \cdots \quad \alpha_N^*\right]^T$

线性SVM中，判别函数的最优参数w^*、b^*的确定：

(1) $\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$

$\boldsymbol{\alpha}^* = \begin{bmatrix} \alpha_1^* & \cdots & \alpha_N^* \end{bmatrix}^T$ 中多数分量为0



对$\boldsymbol{w}^*$有贡献的只有占少数的非零α_i^*

相应地，对应非零α_i^*的训练样本就称为**支持向量**（*Suport Vector*）

作为支持向量的训练样本就位于H_1, H_2超平面。

线性SVM中，判别函数的最优参数w^*、b^*的确定:

(2) $\boldsymbol{b}^*$的确定

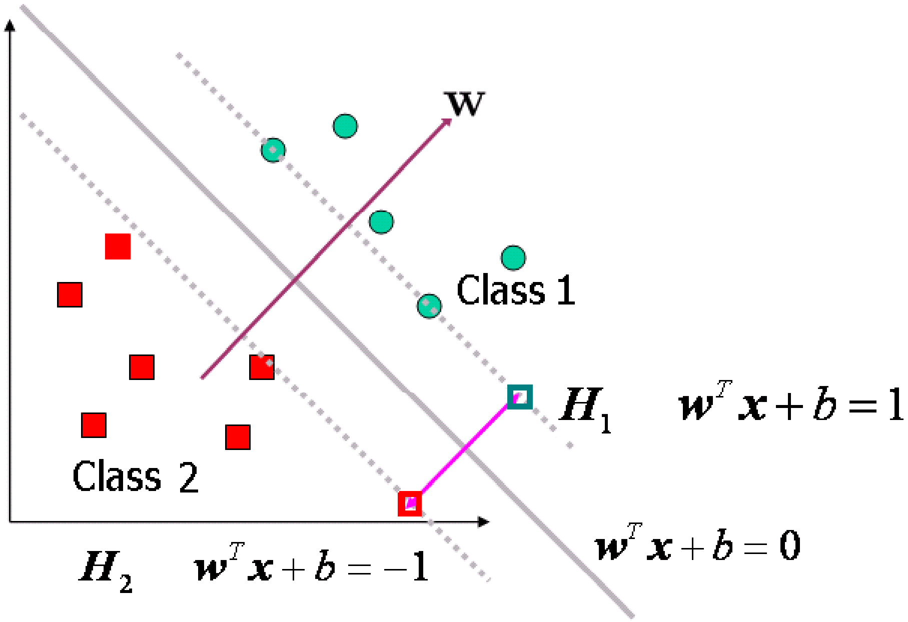
找到所有非零的α_i^*对应的支持向量，构成集合：

$$U=\left\{\left(\boldsymbol{x}_i,y_i\right)\mid\left(\boldsymbol{x}_i,y_i\right)\in \textbf{训练集}\boldsymbol{T}\text{，并且}\alpha_i^*>0\right\}$$

对于$\forall\left(\boldsymbol{x}_i,y_i\right)\in U$，必有：$y_i=\boldsymbol{w}^*\bullet\boldsymbol{x}_i+\boldsymbol{b}^*$

为了确保稳定性，取平均:

$$b^*=\frac{1}{\#U}\sum_{(\boldsymbol{x}_i,y_i)\in U}\left[y_i-\boldsymbol{w}^*\bullet\boldsymbol{x}_i\right]$$



$\#U$----集合U中元素数目(或表示为$|U|$)

原始问题的目标函数(线性可分问题的最大间隔法)

$$\begin{cases} \min\limits_{\boldsymbol{w},b} \quad \phi(\boldsymbol{w}) = \frac{1}{2}(\boldsymbol{w}\cdot\boldsymbol{w}) \\ s.t. \quad y_i \cdot (\boldsymbol{w}\cdot\boldsymbol{x}_i + b) - 1 \geq 0 \quad i = 1,2,\ldots,n \end{cases}$$

非负*Lagrange*乘子向量的最优解：$\boldsymbol{\alpha}^* = \left[\alpha_1^* \quad \cdots \quad \alpha_N^*\right]^T$

最优分类边界 $(\boldsymbol{w}^*\cdot\boldsymbol{x}) + b^* = 0$

$$\begin{cases} \boldsymbol{w}^* = \sum\limits_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i \\ b^* = \frac{1}{\#U} \sum\limits_{(\boldsymbol{x}_i, y_i)\in U} \left[y_i - \boldsymbol{w}^* \bullet \boldsymbol{x}_i\right] \\ U = \left\{(\boldsymbol{x}_i, y_i) \mid (\boldsymbol{x}_i, y_i) \in \textbf{训练集}\boldsymbol{T}，\text{并且}\alpha_i^* > 0\right\} \end{cases}$$

判别函数 $g(\boldsymbol{x}) = (\boldsymbol{w}^*\cdot\boldsymbol{x}) + b^* = \sum\limits_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b$

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[(\boldsymbol{w}^*\cdot\boldsymbol{x}) + b^*\right] = \operatorname{sgn}\left\{\sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^*\right\}$$

5. 利用线性SVM进行分类

对于未知类别的测试样本$\boldsymbol{x}$,

若 $f(\boldsymbol{x})=\operatorname{sgn}\left[\left(\boldsymbol{w}^{*}\cdot\boldsymbol{x}\right)+b^{*}\right]=1$

则 $\boldsymbol{x}$决策为第一类；否则$\boldsymbol{x}$决策为第二类。

6.线性可分SVC算法

(1)给定训练样本集$T=\{(\boldsymbol{x}_1,y_1),\cdots(\boldsymbol{x}_N,y_N)\}$

其中$\boldsymbol{x}_i\in R^d, y_i\in\mathcal{Y}=\{1,-1\}, i=1,\ldots,N$

训练样本的预处理

(2)构造并求解凸二次规划问题

$$\begin{cases}\min\limits_{\boldsymbol{\alpha}}. & W(\boldsymbol{\alpha})=\dfrac{1}{2}\sum\limits_{j=1}^{N}\sum\limits_{i=1}^{N}\alpha_i\alpha_j y_i y_j(\boldsymbol{x}_j\cdot\boldsymbol{x}_i)-\sum\limits_{i=1}^{N}\alpha_i \\ s.t. & \begin{cases}(1) & \sum\limits_{i=1}^{N}\alpha_i y_i=0 \\ (2) & -\alpha_i\le 0 \qquad \text{对于} i=1,\ldots,N\end{cases}\end{cases}$$

得解$\boldsymbol{\alpha}^*=[\alpha_1^*, \quad \cdots, \quad \alpha_N^*]^T$

(3) 计算 $\boldsymbol{w}^* = \sum_{j=1}^{N} \alpha_j^* y_j \boldsymbol{x}_j$

(4) 计算 $b^* = \frac{1}{\#U} \sum_{(\boldsymbol{x}_i, y_i) \in U} \left[y_i - \boldsymbol{w}^* \bullet \boldsymbol{x}_i \right]$

其中 $U = \left\{ \left(\boldsymbol{x}_i, y_i \right) \middle| \left(\boldsymbol{x}_i, y_i \right) \in \text{训练集} \boldsymbol{T}，\text{并且} \alpha_i^* > 0 \right\}$

(5) 分类模型的使用：

$$g(\boldsymbol{x}) = \left(\boldsymbol{w}^* \cdot \boldsymbol{x} \right) + b^* = \sum_{i=1}^{N} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^*$$

对于未知类别的测试样本 $\boldsymbol{x}$，

待决策样本的预处理

$$\begin{cases} \text{若} f(\boldsymbol{x}) = \operatorname{sgn}\left[g(\boldsymbol{x}) \right] = +1, \ \text{则决策} \boldsymbol{x} \text{为第1类} \\ \text{若} f(\boldsymbol{x}) = \operatorname{sgn}\left[g(\boldsymbol{x}) \right] = -1, \ \text{则决策} \boldsymbol{x} \text{为第2类} \end{cases}$$

主要内容

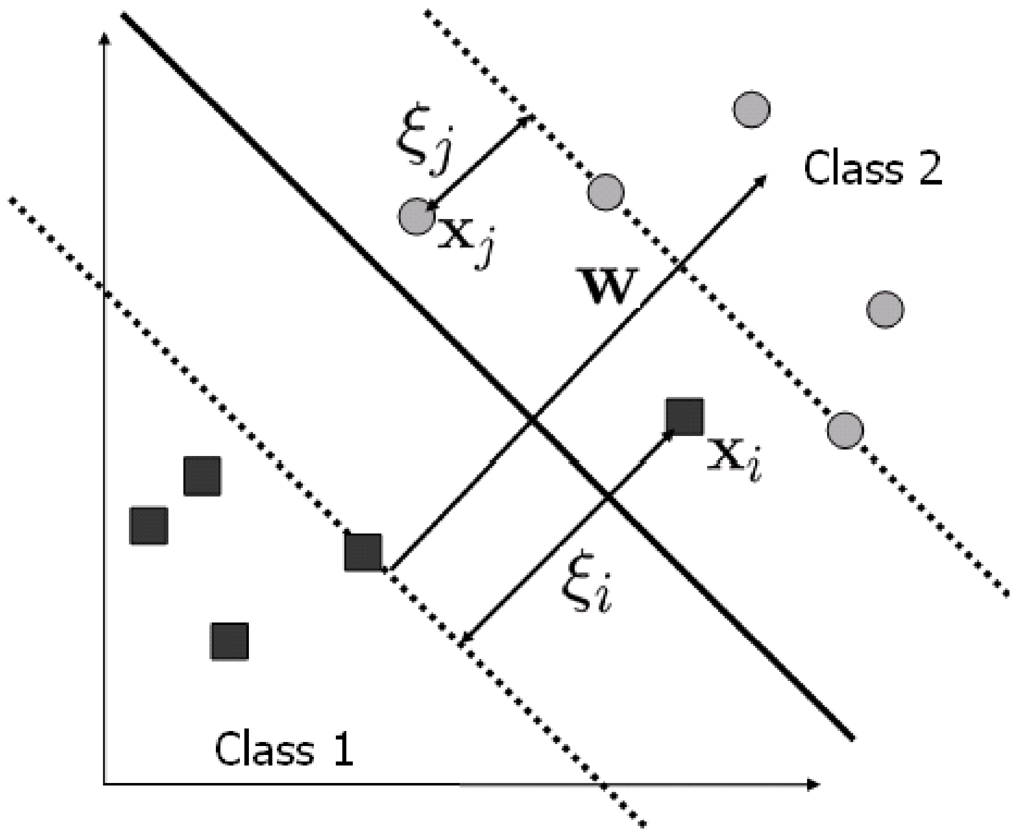
1　两类别分类问题：线性可分情况

2　两类别分类问题：近似线性可分情况

3　多类别的线性SVM

4　应用举例

1.例 二维特征空间、两类别、线性不可分情况



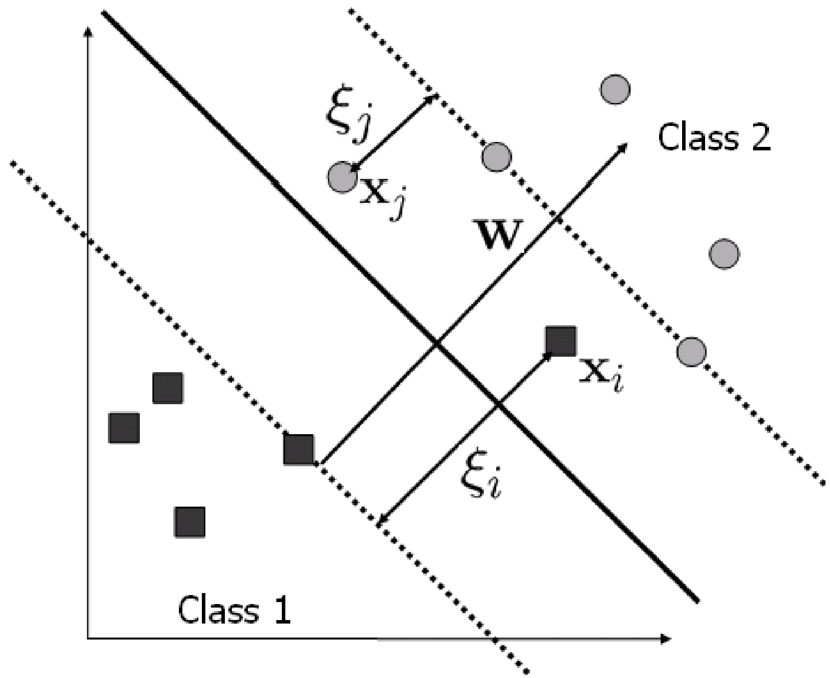
适用：含噪声的训练样本；

容许对训练样本进行线性错分

2.广义线性SVM—最优分类面的求取

(1)问题描述

$$\begin{cases} \text{已知：训练样本集} \{\boldsymbol{x}_i, y_i\}, i = 1, \dots, N \qquad \boldsymbol{x}_i \in R^d \\ \qquad\qquad \text{类别标号} \ y_i \in \{-1, 1\} \\ \text{估计：广义最优线性判别函数} \ g(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b \end{cases}$$



(2)目标函数的构造

折中：″最大分类间隔 + 训练样本集的最小错分程度″

归一化判别函数 $g(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$

对所有训练样本，引入松弛变量 $\xi_i \geq 0,\ i = 1,\dots,N$

$$for \quad i = 1,\dots,N \quad \begin{cases} g(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b \geq 1 - \xi_i & 若 y_i = +1 \\ g(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b \leq -(1 - \xi_i) & 若 y_i = -1 \end{cases}$$

$$\rightarrow y_i \cdot g(\mathbf{x}_i) + \xi_i \geq 1$$

$$\rightarrow y_i \cdot [\mathbf{w} \cdot \mathbf{x}_i + b] \geq 1 - \xi_i$$

目标函数构造的基本思想

$$
\begin{cases}
\text{分类间隔} margin = \dfrac{2}{\|\mathbf{w}\|} \text{尽可能大} & \Leftrightarrow \quad \dfrac{1}{2}(\mathbf{w}\cdot\mathbf{w})\downarrow \\
\text{训练样本集的错分程度尽可能低} & \Leftrightarrow \quad \sum_{i=1}^{N}\xi_i \downarrow
\end{cases}
$$

原始目标函数

松弛因子向量 $\boldsymbol{\xi}=[\xi_1 \quad \cdots \quad \xi_N]^T$

C - SVC (标准SVC)

$$
\min_{\mathbf{w},\xi,b} \quad \phi(\mathbf{w},\xi,b)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)
$$

$$
s.t. \quad \begin{cases} -\left[y_i\cdot(\mathbf{w}\cdot\mathbf{x}_i+b)-1+\xi_i\right]\le 0 \\ -\xi_i\le 0 \end{cases} \qquad i=1,2,\ldots,\mathrm{N}
$$

$C>0$——模型的超参数，控制错分样本的惩罚程度

(3)目标函数的求解

构造*Lagrange*辅助目标函数，将**原始问题**转化为**对偶问题**.

$$\begin{cases} \min \quad L(\mathbf{w},b,\xi;\boldsymbol{\alpha},\boldsymbol{\mu})=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right) \\ \qquad\qquad -\sum_{i=1}^{N}\alpha_i\left\{y_i\cdot[\mathbf{w}\cdot\mathbf{x}_i+b]-1+\xi_i\right\}-\sum_{i=1}^{N}\mu_i\xi_i, \\ \alpha_i\geq 0,\ \ i=1,2,\ldots,N \\ \mu_i\geq 0,\ \ i=1,2,\ldots,N \end{cases}$$

其中非负*Lagrange*乘子向量 $\begin{cases}\boldsymbol{\alpha}=[\alpha_1 \quad \cdots \quad \alpha_N]^T \\ \boldsymbol{\mu}=[\mu_1 \quad \cdots \quad \mu_N]^T\end{cases}$

显然：$L(\mathbf{w},b,\xi;\boldsymbol{\alpha},\boldsymbol{\mu})\leq\phi(\mathbf{w},\xi,b)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$

原始问题等价为： $\min\limits_{\mathbf{w},b,\xi}\max\limits_{\boldsymbol{\alpha},\boldsymbol{\mu}} L(\mathbf{w},b,\xi;\boldsymbol{\alpha},\boldsymbol{\mu})$

原始目标函数

C - $\boldsymbol{SVC}$（标准SVC）

$$\min_{\mathbf{w},\xi,b} \quad \phi(\mathbf{w},\xi,b)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$$

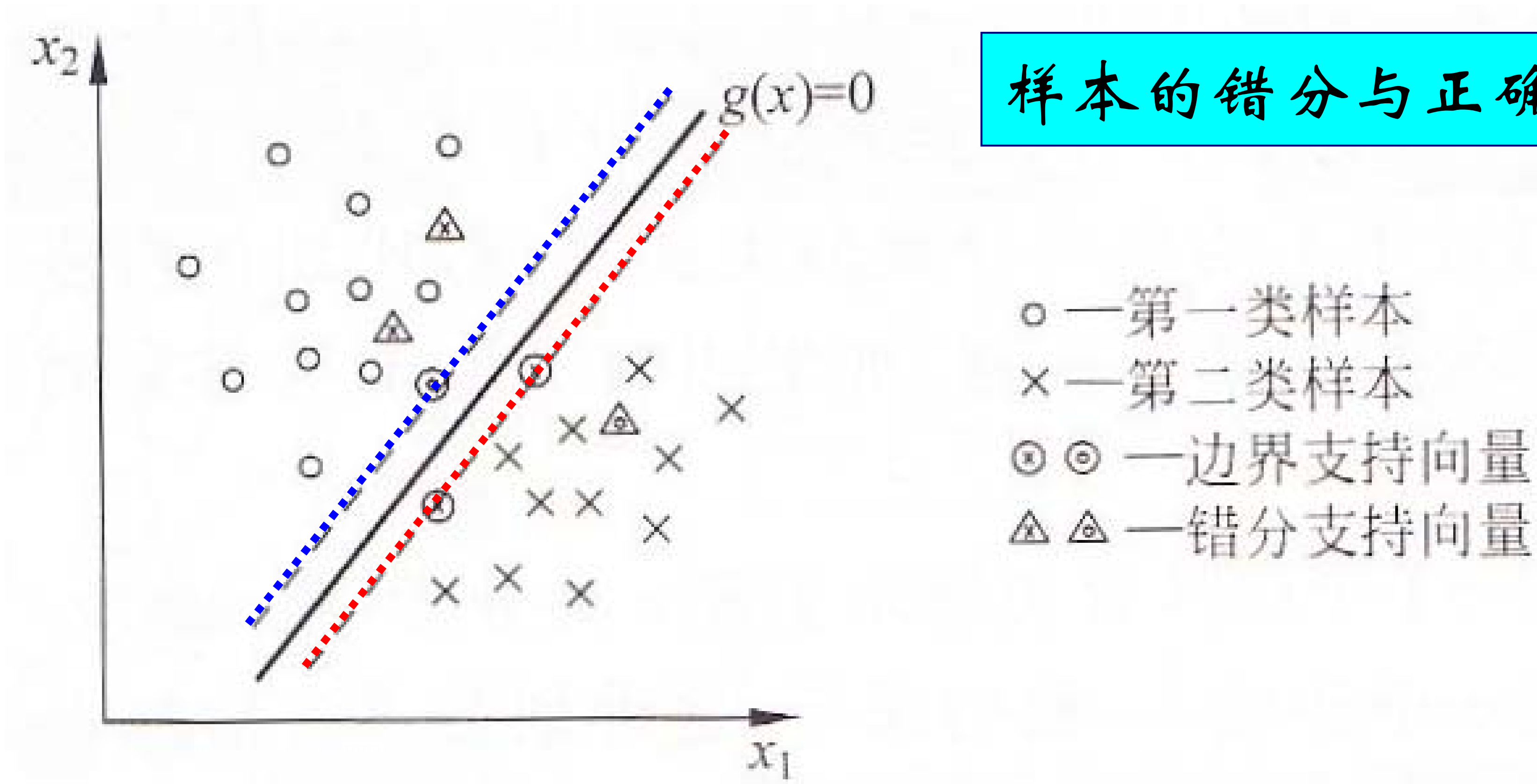
$$s.t. \quad \begin{cases} -\left[y_i\cdot(\mathbf{w}\cdot\mathbf{x}_i+b)-1+\xi_i\right]\le 0 \\ -\xi_i\le 0 \end{cases} \quad i=1,2,\ldots,\mathrm{N}$$

对偶目标函数

$$\max_{\boldsymbol{\alpha}}. \quad \sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j\left(\mathbf{x}_j\cdot\mathbf{x}_i\right)$$

$$s.t.\begin{cases} (1)\quad \sum_{i=1}^{N}\alpha_i y_i=0 \\ (2)\quad 0\le\alpha_i\le C,\quad i=1,\ldots,N \end{cases}$$

解得$\boldsymbol{\alpha}^*=\left[\alpha_1^*,\ldots,\alpha_N^*\right]^T$



样本的错分与正确分类

训练样本集T的划分 基于$\boldsymbol{\alpha}^*=\left[\alpha_1^* \quad \cdots \quad \alpha_N^*\right]^T$中各分量取值的三种可能划分

$$\begin{cases} \alpha_i^*=0 & \text{训练样本}\boldsymbol{x}_i\text{为非支持向量} \\ 0<\alpha_i^*<C & \text{训练样本}\boldsymbol{x}_i\text{为边界支持向量}(\text{位于}H_1,H_2\text{超平面}) \\ \alpha_i^*=C & \text{训练样本}\boldsymbol{x}_i\text{为错分支持向量} \end{cases}$$

训练样本集=$\{$**非支持向量**$\}\cup\{$**边界支持向量**$\}\cup\{$**错分支持向量**$\}$

广义线性SVM中，判别函数的最优参数w^*、b^*的确定：

A. $\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$

对$\boldsymbol{w}^*$有贡献的只有占少数的非零α_i^*

对应非零α_i^*的训练样本就称为**支持向量**(***Suport Vector***)

B. $\boldsymbol{b}^*$的确定

找到所有**边界支持向量**，构成集合：

$U=\left\{\left(\boldsymbol{x}_i, y_i\right) \mid \left(\boldsymbol{x}_i, y_i\right) \in \textbf{训练集}\boldsymbol{T}，\text{并且} 0<\alpha_i^*<C\right\}$

对于$\forall\left(\boldsymbol{x}_i, y_i\right) \in U$，必有：$y_i=\boldsymbol{w}^* \bullet \boldsymbol{x}_i+\boldsymbol{b}^*$

为了确保稳定性，取平均：

$$b^* = \frac{1}{\# U} \sum_{\left(\boldsymbol{x}_i, y_i\right) \in U}\left[y_i-\boldsymbol{w}^* \bullet \boldsymbol{x}_i\right]$$

$\# U$----集合U中元素数目(或表示为$|U|$)

(4)广义线性SVM的判别函数

广义最优分类决策函数：

$$f(\boldsymbol{x})=\operatorname{sgn}\left[\left(\boldsymbol{w}^{*}\bullet\boldsymbol{x}\right)+b^{*}\right]=\operatorname{sgn}\left\{\sum_{i=1}^{N}\alpha_i^{*}y_i\left(\boldsymbol{x}_i\bullet\boldsymbol{x}\right)+b^{*}\right\}$$

广义线性SVM（C - SVC）原始问题的目标函数

$$\min_{\mathbf{w},\xi,b} \ \phi(\mathbf{w},\xi,b)=\frac{1}{2}(\mathbf{w}\cdot\mathbf{w})+C\left(\sum_{i=1}^{N}\xi_i\right)$$

$$s.t.\quad \begin{cases}-\left[y_i\cdot(\mathbf{w}\cdot\mathbf{x}_i+b)-1+\xi_i\right]\le 0\\ -\xi_i\le 0\end{cases}\qquad i=1,2,\ldots,\mathrm{N}$$

非负Lagrange乘子向量的最优解：$\boldsymbol{\alpha}^*=\left[\alpha_1^* \ \cdots \ \alpha_N^*\right]^T$

最优分类边界 $(\mathbf{w}^*\cdot\mathbf{x})+b^*=0$

$$\begin{cases}\mathbf{w}^*=\sum_{i=1}^{N}\alpha_i^* y_i\mathbf{x}_i\\ b^*=\frac{1}{\#U}\sum_{(\mathbf{x}_i,y_i)\in U}\left[y_i-\mathbf{w}^*\bullet\mathbf{x}_i\right]\\ U=\left\{(\mathbf{x}_i,y_i)\,|\,(\mathbf{x}_i,y_i)\in\text{训练集}T\text{，并且}0<\alpha_i^*<C\right\}\end{cases}$$

判别函数 $g(\mathbf{x})=(\mathbf{w}^*\cdot\mathbf{x})+b^*=\sum_{i=1}^{N}\alpha_i^* y_i(\mathbf{x}_i\cdot\mathbf{x})+b$

$$f(\mathbf{x})=\mathrm{sgn}\left[(\mathbf{w}^*\cdot\mathbf{x})+b^*\right]=\mathrm{sgn}\left\{\sum_{i=1}^{N}\alpha_i^* y_i(\mathbf{x}_i\cdot\mathbf{x})+b^*\right\}$$

[线性C-SVC算法]

(1) 给定训练样本集 $T=\{(\mathbf{x}_1,y_1),\cdots(\mathbf{x}_N,y_N)\}$,

训练样本的预处理

其中 $\mathbf{x}_i\in R^d, y_i\in\mathcal{Y}=\{1,-1\}, i=1,\ldots,N$

(2) 选择适当的惩罚参数 C [问题：如何选择参数 C]

(3) 构造并求解凸二次规划问题

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}}. \quad \dfrac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_i\alpha_j y_i y_j(\mathbf{x}_j\cdot\mathbf{x}_i)-\sum_{i=1}^{N}\alpha_i \\ s.t. \quad \begin{cases} (1) \quad \sum_{i=1}^{N}\alpha_i y_i=0 \\ (2) \quad 0\le\alpha_i\le C \qquad 对于 i=1,\ldots,N \end{cases} \end{cases}$$

得解 $\boldsymbol{\alpha}^*=[\alpha_1^*, \quad \cdots, \quad \alpha_N^*]^T$

(4) 估计 $\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y_i \boldsymbol{x}_i$

(5) 估计 $\boldsymbol{b}^*$

找到所有**边界支持向量**，构成集合：

$$U = \left\{ \left(\boldsymbol{x}_i, y_i \right) \mid \left(\boldsymbol{x}_i, y_i \right) \in \textbf{训练集}\boldsymbol{T}，并且 0 < \alpha_i^* < C \right\}$$

对于 $\forall \left(\boldsymbol{x}_i, y_i \right) \in U$，必有：$y_i = \boldsymbol{w}^* \bullet \boldsymbol{x}_i + \boldsymbol{b}^*$

$$b^* = \frac{1}{\#U} \sum_{i \in U} \left[y_i - \sum_{j=1}^{N} \alpha_j^* y_j \left(\boldsymbol{x}_j \cdot \boldsymbol{x}_i \right) \right]$$

(6) 分类模型的使用：$g\left(\boldsymbol{x}\right) = \left(\boldsymbol{w}^* \cdot \boldsymbol{x} \right) + b^* = \sum_{i=1}^{N} \alpha_i^* y_i \left(\boldsymbol{x}_i \cdot \boldsymbol{x} \right) + b^*$

待决策样本的预处理

对于未知类别的测试样本 $\boldsymbol{x}$,

$$\begin{cases} 若 f\left(\boldsymbol{x}\right) = \operatorname{sgn}\left[g\left(\boldsymbol{x}\right) \right] = +1, & 则决策 \boldsymbol{x} 为第1类 \\ 若 f\left(\boldsymbol{x}\right) = \operatorname{sgn}\left[g\left(\boldsymbol{x}\right) \right] = -1, & 则决策 \boldsymbol{x} 为第2类 \end{cases}$$