

第三十一章 支持向量机

支持向量机是数据挖掘中的一项新技术，是借助于最优化方法来解决机器学习问题的新工具，最初由 V.Vapnik 等人提出，近几年来在其理论研究和算法实现等方面都取得了很大的进展，开始成为克服“维数灾难”和过学习等困难的强有力的手段，它的理论基础和实现途径的基本框架都已形成。

§1 支持向量分类机的基本原理

根据给定的训练集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (X \times Y)^l,$$

其中 $x_i \in X = R^n$ ， X 称为输入空间，输入空间中的每一个点 x_i 由 n 个属性特征组成， $y_i \in Y = \{-1, 1\}, i = 1, \dots, l$ 。寻找 R^n 上的一个实值函数 $g(x)$ ，以便用分类函数

$$f(x) = \text{sgn}(g(x)),$$

推断任意一个模式 x 相对应的 y 值的问题为分类问题。

1.1 线性可支持向量分类机

考虑训练集 T ，若 $\exists \omega \in R^n$ ， $b \in R$ 和正数 ε ，使得对所有使 $y_i = 1$ 的下标 i 有 $(\omega \cdot x_i) + b \geq \varepsilon$ （这里 $(\omega \cdot x_i)$ 表示向量 ω 和 x_i 的内积），而对所有使 $y_i = -1$ 的下标 i 有 $(\omega \cdot x_i) + b \leq -\varepsilon$ ，则称训练集 T 线性可分，称相应的分类问题是线性可分的。

记两类样本集分别为 $M^+ = \{x_i \mid y_i = 1, x_i \in T\}$ ， $M^- = \{x_i \mid y_i = -1, x_i \in T\}$ 。定义 M^+ 的凸包 $\text{conv}(M^+)$ 为

$$\text{conv}(M^+) = \left\{ x = \sum_{j=1}^{N_+} \lambda_j x_j \mid \sum_{j=1}^{N_+} \lambda_j = 1, \lambda_j \geq 0, j = 1, \dots, N^+; x_j \in M^+ \right\},$$

M^- 的凸包 $\text{conv}(M^-)$ 为

$$\text{conv}(M^-) = \left\{ x = \sum_{j=1}^{N_-} \lambda_j x_j \mid \sum_{j=1}^{N_-} \lambda_j = 1, \lambda_j \geq 0, j = 1, \dots, N^-; x_j \in M^- \right\}.$$

其中 N_+ 表示 +1 类样本集中样本点的个数， N_- 表示 -1 类样本集中样本点的个数，定理 1 给出了训练集 T 线性可分与两类样本集凸包之间的关系。

定理 1 训练集 T 线性可分的充要条件是， T 的两类样本集 M^+ 和 M^- 的凸包相离。如下图所示

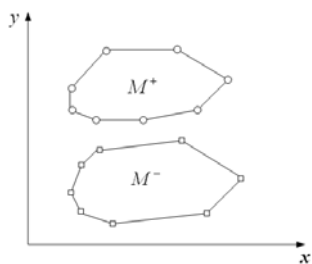


图 1 训练集 T 线性可分时两类样本点集的凸包

证明：①必要性

若 T 是线性可分的, $M^+ = \{x_i | y_i = 1, x_i \in T\}$, $M^- = \{x_i | y_i = -1, x_i \in T\}$, 由线性可分的定义可知, 存在超平面 $H = \{x \in R^n : (\omega \cdot x) + b = 0\}$ 和 $\varepsilon > 0$, 使得

$$(\omega \cdot x_i) + b \geq \varepsilon, \quad \forall x_i \in M^+ \text{ 且 } (\omega \cdot x_i) + b \leq -\varepsilon, \quad \forall x_i \in M^-.$$

而正类点集的凸包中的任意一点 x 和负类点集的凸包中的任意一点 x' 可分别表示为

$$x = \sum_{i=1}^{N_+} \alpha_i x_i \text{ 和 } x' = \sum_{j=1}^{N_-} \beta_j x'_j$$

其中 $\alpha_i \geq 0, \beta_j \geq 0$ 且 $\sum_{i=1}^{N_+} \alpha_i = 1, \sum_{j=1}^{N_-} \beta_j = 1$ 。

于是可以得到

$$(\omega \cdot x) + b = \left(\omega \cdot \sum_{i=1}^{N_+} \alpha_i x_i \right) + b = \sum_{i=1}^{N_+} \alpha_i ((\omega \cdot x_i) + b) \geq \varepsilon \sum_{i=1}^{N_+} \alpha_i = \varepsilon > 0$$

$$(\omega \cdot x') + b = \left(\omega \cdot \sum_{j=1}^{N_-} \beta_j x'_j \right) + b = \sum_{j=1}^{N_-} \beta_j ((\omega \cdot x'_j) + b) \leq -\varepsilon \sum_{j=1}^{N_-} \beta_j = -\varepsilon < 0$$

由此可见, 正负两类点集的凸包位于超平面 $(\omega \cdot x) + b = 0$ 的两侧, 故两个凸包相离。

②充分性

设两类点集 M^+ , M^- 的凸包相离。因为两个凸包都是闭凸集, 且有界, 根据凸集强分离定理, 可知存在一个超平面 $H = \{x \in R^n : (\omega \cdot x) + b = 0\}$ 强分离这两个凸包, 即存在正数 $\varepsilon > 0$, 使得对 M^+ , M^- 的凸包中的任意点 x 和 x' 分别有

$$(\omega \cdot x) + b \geq \varepsilon$$

$$(\omega \cdot x') + b \leq -\varepsilon$$

显然特别的, 对于任意的 $x_i \in M^+$, 有 $(\omega \cdot x_i) + b \geq \varepsilon$, 对于任意的 $x_i \in M^-$, 有 $(\omega \cdot x_i) + b \leq -\varepsilon$, 由训练集线性可分的定义可知 T 是线性可分的。

由于空间 R^n 中超平面都可以写为 $(\omega \cdot x) + b = 0$ 的形式, 参数 (ω, b) 乘以任意一个非零常数后得到的是同一个超平面, 定义满足条件

$$\begin{cases} y_i ((\omega \cdot x_i) + b) \geq 0, & i = 1, \dots, l \\ \min_{i=1, \dots, l} |(\omega \cdot x_i) + b| = 1 \end{cases}$$

的超平面为规范超平面。

定理 2 当训练集样本为线性可分时, 存在唯一的规范超平面 $(\omega \cdot x) + b = 0$, 使得

$$\begin{cases} (\omega \cdot x_i) + b \geq 1 & y_i = 1; \\ (\omega \cdot x_i) + b \leq -1 & y_i = -1. \end{cases} \quad (1)$$

证明: 规范超平面的存在性是显然的, 下证其唯一性。

假设其规范超平面有两个: $(\omega' \cdot x) + b' = 0$ 和 $(\omega'' \cdot x) + b'' = 0$ 。由于规范超平面满足条件

$$\begin{cases} y_i((\omega \cdot x_i) + b) \geq 0, & i=1, \dots, l, \\ \min_{i=1, \dots, l} |(\omega \cdot x_i) + b| = 1. \end{cases}$$

由第二个条件可知

$$\omega' = \omega'', \quad b' = b'',$$

或者

$$\omega' = -\omega'', \quad b' = -b''.$$

第一个条件说明 $\omega' = -\omega''$, $b' = -b''$ 不可能成立, 故唯一性得证。

式(1)中满足 $(\omega \cdot x_i) + b = \pm 1$ 成立的 x_i 称为普通支持向量, 对于线性可分的情况来说, 只有它们在建立分类超平面的时候起到了作用, 普通支持向量通常只占样本集很小的一部分, 故而也说明 SVM 具有稀疏性。对于 $y_i = 1$ 类的样本点, 其与规范超平面的间隔为

$$\min_{y_i=1} \frac{|(\omega \cdot x_i) + b|}{\|\omega\|} = \frac{1}{\|\omega\|},$$

对于 $y_i = -1$ 类的样本点, 其与规范超平面的间隔为

$$\min_{y_i=-1} \frac{|(\omega \cdot x_i) + b|}{\|\omega\|} = \frac{1}{\|\omega\|},$$

则普通支持向量间的间隔为 $\frac{2}{\|\omega\|}$ 。最优超平面即意味着最大化 $\frac{2}{\|\omega\|}$ 。如图 2 所示

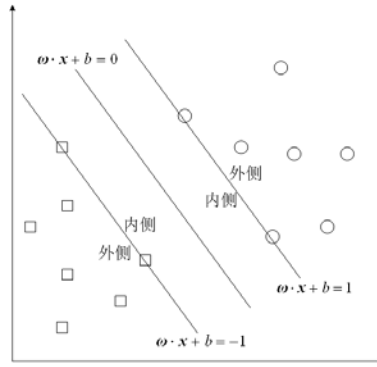


图 2 线性可分支持向量分类机

$(\omega \cdot x) + b = \pm 1$ 称为分类边界, 于是寻找最优超平面的问题可以转化为如下的二次规划问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i((\omega \cdot x_i) + b) \geq 1 \quad i=1, \dots, l \end{aligned} \quad (2)$$

该问题的特点是目标函数 $\frac{1}{2} \|\omega\|^2$ 是 ω 的凸函数, 并且约束条件都是线性的。

引入 Lagrange 函数

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^l \alpha_i (1 - y_i ((\omega \cdot x_i) + b))$$

其中 $\alpha = (\alpha_1, \dots, \alpha_l)^T \in R^{l+}$ 为 Lagrange 乘子。根据 wolfe 对偶的定义，通过对原问题中各变量的偏导置零可得

$$\begin{aligned} \frac{\partial L}{\partial \omega} = 0 &\Rightarrow \omega = \sum_{i=1}^l \alpha_i y_i x_i \\ \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned}$$

带入 Lagrange 函数化为原问题的 Lagrange 对偶问题

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{i=1}^l \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & \alpha_i \geq 0, i = 1, \dots, l \end{aligned} \quad (3)$$

求解上述最优化问题，得到最优解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ ，计算

$$\omega^* = \sum_{i=1}^l \alpha_i^* y_i x_i$$

由 KKT 互补条件知

$$\alpha_i^* (1 - y_i ((\omega^* \cdot x_i) + b^*)) = 0$$

可得，只有当 x_i 为支持向量的时候，对应的 α_i^* 才为正，否则皆为零。选择 α^* 的一个正分量 α_j^* ，并以此计算

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* (x_i \cdot x_j),$$

于是构造分类超平面 $(\omega^* \cdot x) + b^* = 0$ ，并由此求得决策函数

$$g(x) = \sum_{i=1}^l \alpha_i^* y_i (x_i \cdot x) + b^*,$$

得到分类函数

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i^* y_i (x_i \cdot x) + b^*) \quad (4)$$

从而对未知样本进行分类。

1.2 线性支持向量分类机

当训练集 T 的两类样本线性可分时，除了普通支持向量分布在两个分类边界 $(\omega \cdot x_i) + b = \pm 1$ 上外，其余的所有样本点都分布在分类边界以外。此时构造的超平面是硬间隔超平面。当训练集 T 的两类样本近似线性可分时，即允许存在不满足约束条件

$$y_i ((\omega \cdot x_i) + b) \geq 1$$

的样本点后，仍然能继续使用超平面进行划分。只是这时要对间隔进行“软化”，构造软间隔超平面。简言之就是在两个分类边界 $(\omega \cdot x_i) + b = \pm 1$ 之间允许出现样本点，这类样本点被称为边界支持向量。显然两类样本点集的凸包是相交的，只是相交的部分较小。线性支持向量分类机如图3所示。

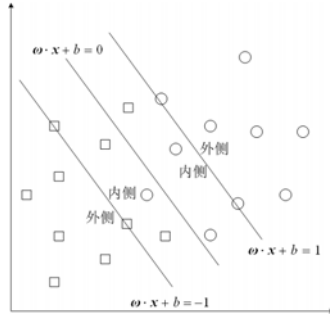


图3 线性支持向量分类机

软化的方法是通过引入松弛变量

$$\xi_i \geq 0, \quad i = 1, \dots, l$$

来得到“软化”的约束条件

$$y_i((\omega \cdot x_i) + b) \geq 1 - \xi_i \quad i = 1, \dots, l,$$

当 ξ_i 充分大时，样本点总是满足上述的约束条件，但是也要设法避免 ξ_i 取太大的值，为此要在目标函数中对它进行惩罚，得到如下的二次规划问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i, \\ \text{s.t.} \quad & y_i((\omega \cdot x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \dots, l. \end{aligned} \quad (5)$$

其中 $C > 0$ 是一个惩罚参数。其Lagrange函数如下

$$L(\omega, b, \xi, \alpha, \gamma) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i (y_i((\omega \cdot x_i) + b) - 1 + \xi_i) - \sum_{i=1}^l \gamma_i \xi_i,$$

其中 $\gamma_i \geq 0$, $\xi_i > 0$ 。原问题的对偶问题如下

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{i=1}^l \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, l \end{aligned} \quad (6)$$

求解上述最优化问题，得到最优解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ ，计算

$$\omega^* = \sum_{i=1}^l \alpha_i^* y_i x_i,$$

选择 α^* 的一个正分量 $0 < \alpha_j^* < C$ ，并以此计算

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* (x_i \cdot x_j)$$

于是构造分类超平面 $(\omega^* \cdot x) + b^* = 0$ ，并由此求得分类函数

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i^* y_i (x_i \cdot x) + b^*)$$

从而对未知样本进行分类，可见当 $C = \infty$ 时，就等价于线性可分的情形。

1.3 可分支支持向量分类机

当训练集 T 的两类样本点集重合的区域很大时，上述用来处理线性不可分问题的线性支持向量分类机就不可用了，可分支支持向量分类机给出了解决这种问题的一种有效途径。通过引进从输入空间 X 到另一个高维的 Hilbert 空间 H 的变换 $x \mapsto \varphi(x)$ 将原输入空间 X 的训练集

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (X \times Y)^l$$

转化为 Hilbert 空间 H 中的新的训练集：

$$\bar{T} = \{(\bar{x}_1, y_1), (\bar{x}_2, y_2), \dots, (\bar{x}_l, y_l)\} = \{(\varphi(x_1), y_1), (\varphi(x_2), y_2), (\varphi(x_l), y_l)\}$$

使其在 Hilbert 空间 H 中线性可分，Hilbert 空间 H 也称为特征空间。然后在空间 H 中求得超平面 $(\omega \cdot \varphi(x)) + b = 0$ ，这个超平面可以硬性划分训练集 \bar{T} ，于是原问题转化为如下的二次规划问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i ((\omega \cdot \varphi(x_i)) + b) \geq 1 \quad i = 1, \dots, l \end{aligned}$$

采用核函数 K 满足

$$K(x_i, x_j) = (\varphi(x_i) \cdot \varphi(x_j))$$

将避免在高维特征空间进行复杂的运算，不同的核函数形成不同的算法，主要的核函数有如下几类：

$$\text{线性内核函数} \quad K(x_i, x_j) = (x_i \cdot x_j)$$

$$\text{多项式核函数} \quad K(x_i, x_j) = [(x_i \cdot x_j) + 1]^q$$

$$\text{径向基核函数} \quad K(x_i, x_j) = \exp\left\{-\frac{\|x_i - x_j\|^2}{\sigma^2}\right\}$$

$$\text{S形内核函数} \quad K(x_i, x_j) = \tanh(v(x_i \cdot x_j) + c)$$

$$\text{傅立叶核函数} \quad K(x_i, x_j) = \sum_{k=1}^n \frac{1 - q^2}{2(1 - 2q \cos(x_{ik} - x_{jk}) + q^2)}$$

同样可以得到其 Lagrange 对偶问题如下

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i \cdot x_j) + \sum_{i=1}^l \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & \alpha_i \geq 0, i = 1, \dots, l \end{aligned}$$

若 K 是正定核, 则对偶问题是一个凸二次规划问题, 必定有解。求解上述最优化问题, 得到最优解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$, 选择 α^* 的一个正分量 α_j^* , 并以此计算

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i \cdot x_j)$$

构造分类函数

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \alpha_i^* K(x \cdot x_j) + b^*)$$

从而对未知样本进行分类。

1.4 C-支持向量分类机

当映射到高维 H 空间的训练集不能被硬性分划时, 需要对约束条件进行软化。结合 1.2 和 1.3 中所述, 得到如下的模型

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i \cdot x_j) + \sum_{i=1}^l \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, l \end{aligned} \quad (7)$$

得到最优解 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$, 选择 α^* 的一个正分量 $0 < \alpha_j^* < C$, 并以此计算

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i \cdot x_j)$$

构造决策函数

$$g(x) = \sum_{i=1}^l y_i \alpha_i^* K(x \cdot x_j) + b^*$$

构造分类函数

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \alpha_i^* K(x \cdot x_j) + b^*)$$

从而对未知样本进行分类。

当输入空间中两类样本点的分布区域严重重合时, 选择合适的核函数及其参数, 可以使映射到特征空间的每一类样本点的分布区域更为集中, 降低两类样本点分布区域的混合程度, 从而加强特征空间中两类样本集“线性可分”的程度, 来达到提高分类的精度和泛化性能的目的。但是就核函数及其参数的选取问题, 目前尚无理论依据, 同样的实验数据, 采用不同的核函数, 其精度往往相差很大, 即便是对于相同的核函数, 选取的参数不同, 分类的精度也会有较大的差别。在实际应用过程中, 往往针对具体的问题多次仿真试验, 找到适合该问题的核函数, 并决定其最佳参数。

下述定理给出了支持向量与 Lagrange 乘子之间的关系。

定理 3 对偶问题(7)的最优解为 $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$, 使得每个样本点 x_i 满足优化问题的 KKT 条件为

$$\begin{aligned} \alpha_i^* = 0 & \Rightarrow y_i g(x_i) > 1 \\ 0 < \alpha_i^* < C & \Rightarrow y_i g(x_i) = 1 \end{aligned}$$

$$\alpha_i^* = C \Rightarrow y_i g(x_i) < 1$$

其中 $0 < \alpha_i^* < C$ ，所对应的 x_i 就是普通支持向量(NSV)，位于分类间隔的边界 $g(x) = \pm 1$ 上，有 $|g(x_i)| = 1$ ； $\alpha_i^* = C$ 所对应的 x_i 就是边界支持向量(BSV)，代表了所有的错分样本点，位于分类间隔内部，有 $|g(x_i)| < 1$ 。BSV \cup NSV 就是支持向量集，是位于分类间隔两个边界上及其内部的所有训练样本集的子集。

§ 2 支持向量回归机

支持向量回归机在回归和函数逼近问题上有广泛的应用。考虑 l 个独立同分布的样本数据点

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (X \times Y)^l$$

其中 $x_i \in X = R^n, y_i \in Y = R, i = 1, \dots, l$ 。寻找 x 与 y 的函数 $y = f(x)$ 以便推断任意 x 所对应的 y 值，并且使得经验风险最小。

2.1 线性 ε -支持向量回归机

给定一个 $\varepsilon > 0$ ，线性超平面 $y = f(x) = (\omega \cdot x) + b$ 沿 y 方向依次上下平移 ε 所扫过的区域构成硬 ε -带超平面，并且超平面满足

$$-\varepsilon \leq (\omega \cdot x_i) + b - y_i \leq \varepsilon, \quad i = 1, \dots, l$$

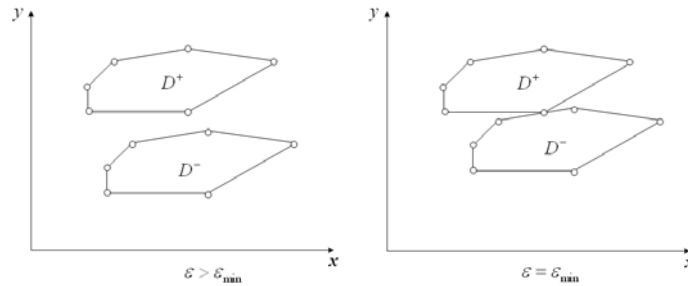
当 ε 足够大时，硬 ε -带超平面总是存在的，而最小能使硬 ε -带超平面存在的 ε 为 ε_{\min} ，是下列优化问题的最优值

$$\begin{aligned} \min_{\omega, b} \quad & \varepsilon \\ \text{s.t.} \quad & -\varepsilon \leq (\omega \cdot x_i) + b - y_i \leq \varepsilon, \quad i = 1, \dots, l \end{aligned} \quad (8)$$

可见，只要选定的 $\varepsilon > \varepsilon_{\min}$ ，硬 ε -带超平面就有无限多个，从给定的训练集 T 出发，可以构造正负两类点，

$$\begin{aligned} D^+ &= \{(x_i^T, y_i + \varepsilon)^T, i = 1, \dots, l\} \\ D^- &= \{(x_i^T, y_i - \varepsilon)^T, i = 1, \dots, l\} \end{aligned} \quad (9)$$

则回归问题转化为线性二分类问题。由定理 4 可知硬 ε -带超平面与线性划分之间的关系。



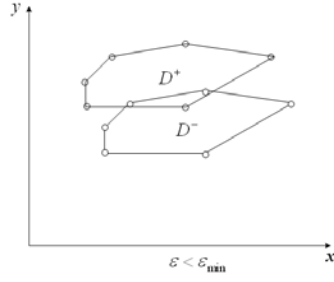


图4 三种情况下 D^+ 和 D^- 的凸包分布

定理4 设上述给定的训练集 T 和 $\varepsilon > 0$, 考虑由(9)定义的集合 D^+ 和 D^- 。则一个超平面 $y = f(x) = (\omega \cdot x) + b$ 是硬 ε -带超平面的充要条件是 D^+ 和 D^- 的凸包分别位于该超平面的两侧。

证明: ①必要性

对给定的 $\varepsilon > 0$, 若 $y = f(x) = (\omega \cdot x) + b$ 是硬 ε -带超平面, 据其定义可知

$$-\varepsilon \leq (\omega \cdot x_i) + b - y_i \leq \varepsilon, \quad i = 1, \dots, l$$

上式可以简单表示为

$$\begin{aligned} (\tilde{y} + \varepsilon e) - A\omega - be &\geq 0 \\ (\tilde{y} - \varepsilon e) - A\omega - be &\leq 0 \end{aligned} \quad (10)$$

其中 $\tilde{y} = (y_1, \dots, y_l)^T$, $e = (1, \dots, 1)^T$, $A = (x_1, \dots, x_l)^T$, D^+ 的凸包中的任意一点都可以表示为

$$t = (t_x^T, t_y^T)^T = \sum_{i=1}^l u_i (x_i^T, y_i + \varepsilon)^T$$

其中 $u = (u_1, \dots, u_l)^T$, 且满足 $\sum_{i=1}^l u_i = 1$, $u_i \geq 0$, $i = 1, \dots, l$, 由(10)的第一式可知

$$t_y - (\omega \cdot t_x) - b = (\hat{y} + \varepsilon e)^T u - \omega^T \cdot (A^T u) - b \geq 0. \quad (11)$$

类似的, D^- 的凸包的任意一点都可以表示为

$$s = (s_x^T, s_y^T)^T = \sum_{i=1}^l v_i (x_i^T, y_i - \varepsilon)^T$$

其中 $v = (v_1, \dots, v_l)^T$, 且满足 $\sum_{i=1}^l v_i = 1$, $v_i \geq 0$, $i = 1, \dots, l$, 由(10)的第二式可知

$$s_y - (\omega \cdot s_x) - b = (\hat{y} - \varepsilon e)^T v - \omega^T \cdot (A^T v) - b \geq 0 \quad (12)$$

由(11)和(12)可知 D^+ 和 D^- 的凸包分别位于该超平面的两侧。

②充分性:

任取满足

$$e^T u = 1, \quad u \geq 0$$

的凸组合系数向量 u , 则可以得到 D^+ 中的一点 $((A^T u)^T, (\tilde{y} + \varepsilon e)^T u)^T$, 与此对应, 选择满足

$$e^T v = 1, \quad A^T u = A^T v$$

的凸组合系数向量 \mathbf{v} ，则可以得到 D^- 中的一点

$$((A^T \mathbf{v})^T, (\tilde{\mathbf{y}} - \varepsilon \mathbf{e})^T \mathbf{v})^T = ((A^T \mathbf{u})^T, (\tilde{\mathbf{y}} - \varepsilon \mathbf{e})^T \mathbf{v})^T$$

将该点与上述 D^+ 中的点比较，他们的前 n 个分量相同，同时超平面

$$y = f(x) = (\omega \cdot x) + b$$

分离两个集合 D^+ 和 D^- 的凸包，则可知他们的第 $n+1$ 个分量满足

$$(\tilde{\mathbf{y}} + \varepsilon \mathbf{e})^T \mathbf{u} - (\tilde{\mathbf{y}} - \varepsilon \mathbf{e})^T \mathbf{v} \geq 0$$

由此可见，关于变量 \mathbf{u}, \mathbf{v} 的不等式组

$$\begin{cases} A^T \mathbf{u} = A^T \mathbf{v} \\ \mathbf{e}^T \mathbf{u} = \mathbf{e}^T \mathbf{v} = 1 \\ (\tilde{\mathbf{y}} + \varepsilon \mathbf{e})^T \mathbf{u} - (\tilde{\mathbf{y}} - \varepsilon \mathbf{e})^T \mathbf{v} < 0 \\ \mathbf{u} \geq 0, \mathbf{v} \geq 0 \end{cases}$$

无解，不难证明

$$\begin{cases} A^T \mathbf{u} = A^T \mathbf{v} \\ \mathbf{e}^T \mathbf{u} = \mathbf{e}^T \mathbf{v} = 1 \\ (\tilde{\mathbf{y}} + \varepsilon \mathbf{e})^T \mathbf{u} - (\tilde{\mathbf{y}} - \varepsilon \mathbf{e})^T \mathbf{v} = -1 \\ \mathbf{u} \geq 0, \mathbf{v} \geq 0 \end{cases}$$

无解，由择一定理知不等式组

$$(\tilde{\mathbf{y}} + \varepsilon \mathbf{e}) - A\omega - b\mathbf{e} \geq 0$$

$$(\tilde{\mathbf{y}} - \varepsilon \mathbf{e}) - A\omega - b\mathbf{e} \leq 0$$

有解，即超平面 $y = f(x) = (\omega \cdot x) + b$ 是硬 ε -带超平面。

定理 4 说明寻找训练集 T 的最优硬 ε -带超平面等价于集合 D^+ 和 D^- 的最优分类超平面。于是可以得到如下的最优化问题

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i - ((\omega \cdot x_i) + b) \leq \varepsilon, \quad i = 1, \dots, l \\ & ((\omega \cdot x_i) + b) - y_i \geq \varepsilon, \quad i = 1, \dots, l \end{aligned} \quad (13)$$

其中 $\omega \in R^n$ ， $b \in R$ 。(13) 的解存在的前提是 $\varepsilon > \varepsilon_{\min}$ ，而 ε_{\min} 就为最优解。定理 5 给出了两者之间的联系。

定理 5 若 $\varepsilon > \varepsilon_{\min}$ ，且 ε_{\min} 是优化问题 (8) 的最优值，则 (13) 的解存在，并且对于 ω 的解唯一，即若 (ω_1^*, b_1^*) ， (ω_2^*, b_2^*) 都是解，则 $\omega_1^* = \omega_2^*$ 。

证明：若 $\varepsilon > \varepsilon_{\min}$ ，则问题 (13) 的可行域非空，又因为该问题是凸的，所以一定有解，解的存在性即证。

下证解的唯一性。设 (ω_1^*, b_1^*) ， (ω_2^*, b_2^*) 是问题 (13) 的两个最优解，显然

$$\|\omega_1^*\| = \|\omega_2^*\| = r$$

其中 r 是一个常数, 则有 $\omega_1^* = \lambda \omega_2^*$, $|\lambda| = 1$ 。令 $\omega = \frac{\omega_1^* + \omega_2^*}{2}$, $b = \frac{b_1^* + b_2^*}{2}$, 则 (ω, b) 是问题 (13) 的可行解, 从而有

$$r \leq \|\omega\| \leq \frac{1}{2} \|\omega_1^*\| + \frac{1}{2} \|\omega_2^*\| = r$$

上式说明 $\|\omega\| = r$, 当 $\lambda = -1$ 时, $\omega = 0$, 即 $r = \|\omega\| = 0$, 则有 $\omega_1^* = \omega_2^* = \mathbf{0}$; 若 $\lambda = 1$, 则有 $\omega_1^* = \omega_2^*$, 故解的唯一性得证。

引入 Lagrange 函数

$$L(\omega, \alpha^{(*)}, b) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^l \alpha_i (\varepsilon + y_i - (\omega \cdot x_i) - b) - \sum_{i=1}^l \alpha_i^* (\varepsilon - y_i + (\omega \cdot x_i) + b)$$

其中 $\alpha^{(*)} = (\alpha_1, \alpha_1^*, \dots, \alpha_l, \alpha_l^*)^T \geq 0$ 为 Lagrange 乘子, 根据 Wolfe 对偶的定义, 分别对 ω , b 求偏导置零得到

$$\begin{aligned} \frac{\partial L}{\partial \omega} &= \omega - \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i = 0 \\ \frac{\partial L}{\partial b} &= \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \end{aligned}$$

带入 Lagrange 函数, 得到 (13) 的对偶问题如下

$$\begin{aligned} \min_{\alpha^{(*)} \in R^{2l}} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(x_i \cdot x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* - \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ & \alpha^{(*)} = (\alpha_1, \alpha_1^*, \dots, \alpha_l, \alpha_l^*)^T \geq 0 \end{aligned} \quad (14)$$

对于给定的训练样本集 T , 首先选择适当的精度 $\varepsilon > 0$, 构造并求解优化问题 (14), 得到最优解 $\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T$, 其中只有支持向量对应的 $\bar{\alpha}_i$ 或 $\bar{\alpha}_i^*$ 非零, 计算

$$\bar{\omega} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i$$

选择 $\bar{\alpha}^{(*)}$ 的正分量 $\bar{\alpha}_j > 0$, 计算

$$\bar{b} = y_j - (\bar{\omega} \cdot x_j) + \varepsilon$$

或选择 $\bar{\alpha}^{(*)}$ 的正分量 $\bar{\alpha}_j^* > 0$, 计算

$$\bar{b} = y_j - (\bar{\omega} \cdot x_j) - \varepsilon$$

构造决策函数 $y = (\bar{\omega} \cdot x) + \bar{b} = \sum_{i=1}^l (\alpha_i^* - \alpha_i)(x_i \cdot x) + \bar{b}$, 即可对未知样本进行预测。

2.2 ε -支持向量回归机

当构造的两类样本 D^+ , D^- 线性不可分时, 通过对硬间隔的软化和引入罚参数, 同样可以考虑在高维空间构造超曲面来达到回归的目的。

对于给定的训练样本集 T , 选择合适的精度参数 $\varepsilon > 0$, 罚参数 $C > 0$ 以及合适的

核函数 $K(x, x')$ 。

构造并求解优化问题

$$\begin{aligned} \min_{\alpha^{(*)} \in R^{2l}} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* - \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\ & 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{l}, \quad i=1, \dots, l \end{aligned} \quad (15)$$

得到最优解 $\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_l, \bar{\alpha}_l^*)^T$ ，选择 $\bar{\alpha}^{(*)}$ 的正分量 $\bar{\alpha}_j > 0$ ，计算

$$\bar{b} = y_j - \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i, x_j) + \varepsilon$$

或选择 $\bar{\alpha}^{(*)}$ 的正分量 $\bar{\alpha}_j^* > 0$ ，计算

$$\bar{b} = y_j - \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i, x_j) - \varepsilon$$

构造决策函数 $y = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i, x) + \bar{b}$ ，即可对未知样本进行预测。

如下定理给出了样本点与 Lagrange 乘子之间的关系。

定理 6 若样本 (x_i, y_i) 在 ε -带内部，则 $\alpha_i^* = \alpha_i = 0$ 。

若样本 (x_i, y_i) 在 ε -带的边界上，则 $\alpha_i^* \in [0, \frac{C}{l}]$ ， $\alpha_i = 0$ ；或 $\alpha_i^* = 0$ ， $\alpha_i \in [0, \frac{C}{l}]$ 。

若样本 (x_i, y_i) 在 ε -带的外部，则 $\alpha_i^* = \frac{C}{l}$ ， $\alpha_i = 0$ ；或 $\alpha_i^* = 0$ ， $\alpha_i = \frac{C}{l}$ 。

可见，在 ε -带内部的样本点都不是支持向量，它们对决策函数没有贡献，并且它们在样本集中的比例较大。

§ 3 乳腺癌的诊断

3.1 问题的提出

乳腺肿瘤通过穿刺采样进行分析可以确定其为良性的或为恶性的。医学研究发现乳腺肿瘤病灶组织的细胞核显微图像的 10 个量化特征：细胞核直径、质地、周长、面积、光滑度、紧密度、凹陷度、凹陷点数、对称度、断裂度与该肿瘤的性质有密切的关系。现试图根据已获得的实验数据建立起一种诊断乳腺肿瘤是良性还是恶性的方法。数据来自已确诊的 500 个病例，每个病例的一组数据包括采样组织中各细胞核的这 10 个特征量的平均值、标准差和最坏值共 30 个数据，并将这种方法用于另外 69 名已做穿刺采样分析的患者。

这个问题实际上属于模式识别问题。什么是模式呢？广义地说，在自然界中可以观察的事物，如果我们能够区别它们是否相同或是否相似，都可以称之为模式。人们为了掌握客观事物，按事物相似的程度组成类别。模式识别的作用和目的就在于面对某一具体事物时将其正确地归入某一类别。

模式识别的方法很多，如数理统计方法、聚类分析方法等。

3.2 支持向量机的分类模型

支持向量机 (Sport Vector Machine , 以下简称 SVM) 是一种基于统计学习理论的模式识别方法。在模式识别等领域获得了广泛的应用。其主要思想是这样: 找到一个超平面, 使得它能够尽可能多的将两类数据点正确地分开, 同时使分开的两类数据点距离分类面最远, 如图 5 (b)。

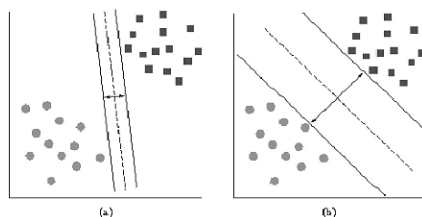


图 5 最佳超平面示意图

记 n (这里 $n = 500$) 个已知观测样本为 $(t_1, g_1), (t_2, g_2), \dots, (t_n, g_n)$ 。其中 $t_i \in R^{30}$, $g_i = 1$ 为良性肿瘤, $g_i = -1$ 为恶性肿瘤。

(1) 样本线性或非线性可分

我们要找一个最优分类面 $w^T x + b = 0$, 其中 $w, x \in R^{30}$, $b \in R$, w, b 待定, 满足如下条件:

$$\begin{cases} w^T t_i + b \geq 1, & g_i = 1 \text{ 时} \\ w^T t_i + b \leq -1, & g_i = -1 \text{ 时} \end{cases}$$

即有 $g_i(w^T t_i - b) \geq 1$, 其中, 满足方程 $w^T t_i + b = \pm 1$ 的样本称为支持向量。

要使两类总体到分类面的距离最大, 则有

$$\max \frac{2}{\|w\|} \Rightarrow \min \frac{1}{2} \|w\|^2$$

于是建立 SVM 的数学模型如下:

模型 1

$$\begin{aligned} \min & \quad \frac{1}{2} \|w\|^2 \\ \text{s. t.} & \quad g_i(w^T t_i + b) \geq 1, \quad i = 1, 2, \dots, n \end{aligned}$$

求得最优值对应的 w^*, b^* , 可得分类函数:

$$g(x) = \text{sgn}(w^{*T} x + b^*)$$

模型 1 是一个二次规划模型。下面把模型 1 化为其对偶问题。

定义广义拉格朗日函数

$$L(w, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^n \alpha_i [1 - g_i(w^T t_i + b)]$$

其中 $\alpha = (\alpha_1, \dots, \alpha_n)^T$, $\alpha_i \geq 0$ 。

由 Karush-Kuhn-Tucker 互补条件, 通过对 w 和 b 求偏导可得

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i g_i t_i = 0$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i g_i = 0$$

得 $w = \sum_{i=1}^n \alpha_i g_i t_i$, $\sum_{i=1}^n \alpha_i g_i = 0$, 代入原始拉格朗日函数得

$$L = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j g_i g_j (t_i \cdot t_j)$$

其中 $(x_i \cdot x_j)$ 表示向量的内积

于是模型 1 可以化为

模型 2

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j g_i g_j (t_i \cdot t_j) \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^n \alpha_i g_i = 0 \\ 0 \leq \alpha_i, i=1, 2, \dots, n \end{cases} \end{aligned}$$

解此二次规划得到最优解 α^* , 从而得权重向量 $w^* = \sum_{i=1}^n \alpha_i^* g_i t_i$ 。

由 KKT 互补条件知

$$\alpha_i^* [1 - g_i ((w^* \cdot t_i) + b^*)] = 0$$

这意味着仅仅是支持向量（距离分类超平面为 1）的输入点 t_i , 使得 α_i^* 为正, 所有其它样本对应的 α_i^* 均为零。选择 α^* 的一个正分量 α_j^* , 并以此计算

$$b^* = g_j - \sum_{i=1}^n g_i \alpha_i^* (t_i \cdot t_j)$$

最终的分类函数表达式如下:

$$g(x) = \text{sgn}(w^{*T} x + b^*)$$

实际上, 模型 2 中的 $(t_i \cdot t_j)$ 是核函数的线性形式。核函数可以将原样本空间线性不可分的向量转化到高维特征空间中线性可分的向量。以下是核函数的一些常用形式:

- 线性内核 $K(x, y) = x \cdot y$
- 径向基函数内核 $K(x, y) = [(x \cdot y) + 1]^q$
- 多项式内核 $K(x, y) = \exp\{-\frac{|x - y|^2}{\sigma^2}\}$
- S 形内核 $K(x, y) = \tanh(v(x \cdot y) + c)$

将模型 2 换成一般的核函数 $K(x, y)$, 可得一般的模型:

模型 3

$$\max \quad \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j g_i g_j K(t_i, t_j)$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=1}^n \alpha_i g_i = 0 \\ 0 \leq \alpha_i, i = 1, 2, \dots, n \end{cases}$$

分类函数表达式:

$$g(x) = \text{sgn}(\sum_{i=1}^n \alpha_i^* g_i K(t_i, x) + b^*)$$

(2) 样本不可分时的软间隔处理。

引入松弛变量 ξ ，其作用在于允许较少数量的样本被错误划分。这种处理方法称为软间隔。控制这个错误程度的参数为 c ，于是模型 1 修改为
模型 4

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \begin{cases} g_i(w^T t_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned}$$

求得最优值对应的 w^*, b^* ，可得分类函数:

$$g(x) = \text{sgn}(w^{*T} x + b^*)$$

对应地，模型 3 改造得到
模型 5

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j g_i g_j K(t_i, t_j) \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^n \alpha_i g_i = 0 \\ 0 \leq \alpha_i \leq c, i = 1, 2, \dots, n \end{cases} \end{aligned}$$

注意到其只是在模型 3 的基础上约束 $0 \leq \alpha_i \leq c$ 。分类函数表达式:

$$g(x) = \text{sgn}(\sum_{i=1}^n \alpha_i^* g_i K(x_i, x) + b^*)$$

应用支持向量机时，可以按照线性可分、线性不可分、非线性可分、非线性不可分的顺序求解。

3.3 分类模型的求解

对于给定的 500 个训练样本，利用 Matlab 的优化工具箱，求解模型 1，得到最优解 w^*, b^* 。说明两类样本点是线性可分的，且对已知样本点，支持向量机的分类方法错误率为 0。

把 69 个测试样本 $\tilde{t}_j, j = 1, 2, \dots, 69$ ，代入分类函数 $g(x) = \text{sgn}(w^{*T} x + b^*)$ ，按如下规则分类:

$g(\tilde{t}_j) = 1$ ，第 j 个样本点为良性肿瘤

$g(\tilde{t}_j) = -1$ ，第 j 个样本点为恶性肿瘤

求解结果见表 1。

表 1 分类结果

良 性									恶 性								
1	3	5	6	7	8	9	11	12	2	4	10	13	15	17	18	22	29
14	16	19	20	21	23	24	25	26	34	36	37	42	63	64	65	66	67
27	28	30	31	32	33	35	38	39	68								
40	41	43	44	45	46	47	48	49									
50	51	52	53	54	55	56	57	58									
59	60	61	62	69													

注：这里的数字代表病例序号。

计算的 Matlab 程序如下：

%原始数据中的 B 替换成 1，M 替换成-1，X 替换成 2，删除了分割符*

```
clc,clear
```

```
load cancerdata.txt;
```

```
can=cancerdata;
```

```
can(:,1)=[]; %删除第一列病例号
```

```
gdb=find(can(:,1)==1); %读出良性肿瘤的序号
```

```
n1=length(gdb); %计算良性肿瘤样本点的个数
```

```
bdb=find(can(:,1)==-1); %读出恶性肿瘤的序号
```

```
n2=length(bdb); %计算恶性肿瘤样本点的个数
```

```
gd=can(gdb,2:end); %提出良性肿瘤的数据
```

```
bd=can(bdb,2:end); %提出恶性肿瘤的数据
```

```
xd=can([501:569],[2:end]); %提出待分类数据
```

```
H=eye(31);H(31,31)=0; %w,b 总共 31 个待定参数，b 对应最后一个参数
```

```
ga=[gd,ones(n1,1)]; %良性样本点对应的系数矩阵
```

```
ba=[bd,ones(n2,1)]; %恶性样本点对应的系数矩阵
```

```
a=[-ga;ba]; %线性不等号约束的系数矩阵
```

```
b=-ones(500,1); %线性不等号约束的右端项
```

```
[wb,fval,flag]=quadprog(H,[],a,b) %调用二次规划命令
```

```
xdf=[xd,ones(69,1)];
```

```
g=xdf*wb; %计算  $w^*x+b$  的值
```

```
g1=find(g>0)' %良性样本序号
```

```
g2=find(g<0)' %恶性样本序号
```

模型 2 的 Matlab 计算程序如下：

```
function svm1
```

```
clc,clear
```

```
global N
```

```
load cancerdata.txt;
```

```
can=cancerdata;
```

```
can(:,1)=[]; %删除第一列病例号
```

```
gdb=find(can(:,1)==1); %读出良性肿瘤的序号
```

```
n1=length(gdb); %计算良性肿瘤样本点的个数
```

```
bdb=find(can(:,1)==-1); %读出恶性肿瘤的序号
```

```
n2=length(bdb); %计算恶性肿瘤样本点的个数
```

```
gd=can(gdb,2:end); %提出良性肿瘤的数据
```



```

bd=can(bdb,2:end); %提出恶性肿瘤的数据
xd=can([501:569],[2:end]); %提出待分类数据
X=[gd;bd]; %重新排序后的训练样本
N=size(X,1);
gb=[ones(n1,1);-ones(n2,1)]; %g_i 的取值向量
a0=eps*ones(N,1); %eps 为 Matlab 中最小的正数
C=1; %人为地设定参数的取值范围
[a,fval,flag]=fmincon(@qfun,a0,[],[],gb',0,zeros(N,1),C*ones(N,1),[],[],X,gb)
w=X'(a.*gb) %计算  $w'*x+b=0$  中的 w
b=sum(diag(a)*(X*w0-gb))/sum([a>10*eps]) %计算  $w'*x+b=0$  中的 b
save dwb w b
function y=qfun(a,X,gb); %定义非线性规划的目标函数
global N
y=-ones(1,N)*a+0.5*a'*diag(gb)*X*X'*diag(gb)*a;

```

模型 3 的 Matlab 计算程序如下：

```

%原始数据中的 B 替换成 1，M 替换成-1，X 替换成 2，删除了分割符*
clc,clear
load cancerdata.txt;
can=cancerdata;
can(:,1)=[]; %删除第一列病例号
gdb=find(can(:,1)==1); %读出良性肿瘤的序号
n1=length(gdb); %计算良性肿瘤样本点的个数
bdb=find(can(:,1)==-1); %读出恶性肿瘤的序号
n2=length(bdb); %计算恶性肿瘤样本点的个数
gd=can(gdb,2:end); %提出良性肿瘤的数据
bd=can(bdb,2:end); %提出恶性肿瘤的数据
xd=can([501:569],[2:end]); %提出待分类数据
hd=[gd;bd]; %重新排序后的训练样本
N=size(hd,1);
basefun=@(x,y) (x*y'+1).^2; %定义多项式核函数的匿名函数
gb=[ones(n1,1);-ones(n2,1)]; %g_i 的取值向量
H=gb*gb'.*basefun(hd,hd); %定义二次规划目标函数的二次项系数矩阵
f=-ones(N,1); %定义二次规划目标函数的一次项系数列向量
aeq=gb'; %定义二次规划的线性等号约束矩阵
[alpha,fval,flag]=quadprog(H,f,[],[],aeq,0,zeros(N,1),ones(N,1))

```

模型 2 和模型 3 的 Matlab 程序计算时间很长，必须寻找一些更好的算法。

习题三十一

1. 蠓虫分类问题：生物学家试图对两种蠓虫（Af 与 Apf）进行鉴别，依据的资料是触角和翅膀的长度，已经测得了 9 支 Af 和 6 支 Apf 的数据如下：

Af: (1.24,1.27), (1.36,1.74), (1.38,1.64), (1.38,1.82), (1.38,1.90), (1.40,1.70),
(1.48,1.82), (1.54,1.82), (1.56,2.08)

Apf: (1.14,1.82), (1.18,1.96), (1.20,1.86), (1.26,2.00), (1.28,2.00), (1.30,1.96)

现在的问题是：

(i) 根据如上资料，如何制定一种方法，正确地区分两类蠓虫。

(ii) 对触角和翼长分别为(1.24,1.80)，(1.28,1.84)与(1.40,2.04)的 3 个标本，用所得到的方法加以识别。

2. 考虑下面的优化问题

$$\begin{aligned} \min \quad & \|w\|^2 + c_1 \sum_{i=1}^n \xi_i + c_2 \sum_{i=1}^n \xi_i^2 \\ \text{s.t.} \quad & g_i((w \cdot x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, n \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, n \end{aligned}$$

讨论参数 c_1 和 c_2 变化产生的影响。导出对偶表示形式。