软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(7):1551−1560 [doi: 10.3724/SP.J.1001.2011.03856]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

支撑向量数据域描述优化问题最优解理论分析*

王晓明+， 王士同

(江南大学 信息工程学院,江苏 无锡 214122)

Theoretical Analysis for the Optimization Problem of Support Vector Data Description

WANG Xiao-Ming+， WANG Shi-Tong

(School of Information, Jiangnan University, Wuxi 214122, China)

+ Corresponding author: E-mail: wxmwm@yahoo.cn

Wang XM, Wang ST. Theoretical analysis for the optimization problem of support vector data description. *Journal of Software*, 2011,22(7):1551−1560. http://www.jos.org.cn/1000-9825/3856.htm

Abstract: A majority of previous research done on Support Vector Data Description (SVDD), which is one of the excellent and applied widely to kernel methods, were directed toward efficient implementations and practical applications. However, very few research attempts have been directed toward studying the properties of SVDD solutions. In this work, the primal optimization of SVDD is first transformed into a convex constrained optimization problem, and the uniqueness of the centre of ball is proved while the non-uniqueness of the radius is investigated. This paper also investigates the property of the centre and radius from the perspective of the dual optimization problem, and suggests a method to calculate the radius. The results of this paper complete the SVDD theory, and contribute to further theoretical study and extensive applications.
Key words: kernel method; support vector data description; convex optimization; uniqueness

摘 要: 支撑向量数据域描述(support vector data description,简称SVDD)作为一种已经得到广泛应用的核方法,目前研究主要集中在其性能和效率的提高上,然而该算法优化问题最优解性质的理论性质却没有得到足够的关注.为此,首先把 SVDD 定义的原始优化问题等价转化为一个凸约束二次优化问题,然后从理论上证明了其构建的超球圆心具有唯一性,然而超球半径在一定条件下却存在不唯一性,并且给出了半径存在不唯一性的充分必要条件.还从对偶优化问题的角度分析了超球的圆心和半径性质,并且给出了 SVDD 算法中在根据优化问题最优解构建超球半径不唯一情况下计算超球半径的方法.完善了该算法的理论和方法体系,从而为其更深入的研究和应用奠定了理论基础.
关键词: 核方法;支撑向量数据域描述;凸优化;唯一性
中图法分类号: TP181 **文献标识码**: A

近年来,核方法[1]作为一种优秀的学习方法得到了广泛的研究和发展,并取得了显著成果.支撑向量机(support vector machines,简称 SVMs)[2]是核方法中一类典型的代表性算法.对于 SVMs 算法,不但其性能和效率得到了广泛的关注,而且其定义的优化问题最优解的性质也得到了深入的研究,如文献[3−7].

* 基金项目: 国家自然科学基金(60773206, 60903100, 60975027, 90820002); 江苏省自然科学基金(BK2009067)
收稿时间: 2009-06-08; 修改时间: 2009-11-26, 2010-03-05; 定稿时间: 2010-04-27

支撑向量数据域描述(support vector data description,简称 SVDD)[8]是核方法中另一类重要的算法.该算法主要是在特征空间中通过构造一个超球来描述一类数据,在超球内的样本点被视为正常(normal)点,然而在超球外面的样本点则被认为是反常(abnormal)点或新颖(novel)点.SVDD 算法在例外点检测[8]和降噪[9]方面表现出了优秀的性能.SVDD 算法还被广泛应用于支撑向量聚类(support vector clustering,简称 SVC)[10]和分类[11],并且表现出了良好的性能.目前,SVDD 算法的另一个研究热点方向是通过 Core-Sets 来实现该算法的快速几何逼近求解[12],而且 SVMs 算法的优化问题通过一定的修改也可以转换成为一种特殊的 SVDD 算法——最小包含球(minimum enclosing ball,简称 MEB)问题(也被称为 hard-margin SVDD[13]),从而实现了 SVMs 算法的快速逼近求解[13,14].

然而,尽管 SVDD 算法已经得到了广泛的研究和应用,但是目前的大量文献都主要集中在其性能和效率的提高以及应用的拓展上[15−17],而缺少对其优化问题最优解性质的理论分析,忽视了根据其优化问题最优解构建超球半径时可能会存在不唯一性.我们认为,其中一个至关重要的原因在于:SVDD 算法定义的原始优化问题并不是一个凸约束优化问题,即该优化问题的约束函数并不都是凸函数,因而不能确保 KKT(Karush-Kuhn-Tucker)[18]条件是该优化问题最优解的充分必要条件[19],从而也使得不能像 SVMs 一样直接应用凸约束优化问题的相关理论来分析 SVDD 的优化问题.针对上述问题,概括地讲,本文主要进行了以下几方面的工作:(1) 把 SVDD 算法定义的原始非凸约束优化问题等价改写为凸约束二次优化问题;(2) 理论上证明了根据 SVDD 算法定义的优化问题的最优解所构建的超球圆心具有唯一性;(3) 给出和证明了根据 SVDD 算法定义的优化问题的最优解所得到的超球半径具有不唯一性的充分必要条件;(4) 从对偶优化问题的角度分析和证明了超球的圆心和半径性质,并且给出了 SVDD 算法中在根据优化问题最优解构建超球半径不唯一情况下新的计算超球半径的方法.所以,本文的工作为 SVDD 算法的深入研究和应用奠定了理论基础.此外,由于 One-Class SVM[20]可以被视为一种特殊的 SVDD 算法[8],因而本文的结论显然容易推广到 One-Class SVM 中.

本文第 1 节简要回顾 SVDD 算法.第 2 节用一个例子来说明 SVDD 算法中存在的问题.第 3 节首先把 SVDD 算法的原始优化问题改写为一个凸约束二次优化问题,然后证明构建的超球圆心具有唯一性,给出和证明超球半径存在不唯一性的充分必要条件.第 4 节从对偶优化问题的角度进行讨论,并且给出新的构建超球半径的方法.第 5 节进一步给出一个真实数据的例子.最后,第 6 节进行总结.

1 支撑向量数据域描述算法回顾

假定有一个含有 N 个样本的训练数据集 $D=\{\boldsymbol{x}_i|\boldsymbol{x}_i\in\boldsymbol{R}^d,i=1,\dots,N\}$,SVDD 算法就是要构建一个包含所有正常样本点的最小超球.在线性情况下定义了如下优化问题:

$$\begin{aligned}&\min R^2+C\sum_{i=1}^{N}\xi_i\\&\text{s.t. } \|\boldsymbol{x}_i-\boldsymbol{b}\|^2\leqslant R^2+\xi_i,\ \xi_i\geqslant 0,\ i=1,\dots,N\end{aligned}\tag{1}$$

其中,$\boldsymbol{b}$ 表示超球中心,R 便是超球半径,C 表示正则化系数(或惩罚系数).一般情况下,通过采用 Lagrange 法[18]把上面的优化问题(1)转化为如下的对偶优化问题来求解:

$$\begin{aligned}&\max -\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j\boldsymbol{x}_i^T\boldsymbol{x}_j+\sum_{i=1}^{N}\alpha_i\boldsymbol{x}_i^T\boldsymbol{x}_i\\&\text{s.t. } \sum_{i=1}^{N}\alpha_i=1,\ 0\leqslant\alpha_i\leqslant C,\ i=1,\dots,N\end{aligned}\tag{2}$$

其中,α_i 称为 Lagrange 乘子.假定优化问题(2)的最优解为 $\boldsymbol{\alpha}^*=[\alpha_1^*,\dots,\alpha_N^*]^T$,如果 $\alpha_i^*>0\,|\,\forall i\in\{1,\dots,N\}$,则定义 $\boldsymbol{x}_i$ 为支撑向量.用 D_{SV} 表示支撑向量集,N_{SV} 表示支撑向量个数,通过利用 KKT 条件可以得到原始问题(1)关于 $\boldsymbol{b}$ 和 R 的最优解:

$$\boldsymbol{b}^*=\sum_{\boldsymbol{x}_i\in D_{SV}}\alpha_i^*\boldsymbol{x}_i\tag{3}$$

$$R^{*2}=\boldsymbol{x}_k^T\boldsymbol{x}_k-\sum_{x_i\in D_{SV}}\alpha_i^*\boldsymbol{x}_i^T\boldsymbol{x}_k+\sum_{x_i\in D_{SV}}\sum_{x_j\in D_{SV}}\alpha_i^*\alpha_j^*\boldsymbol{x}_i^T\boldsymbol{x}_j \tag{4}$$

其中,$\boldsymbol{x}_k$ 为 Lagrange 乘子 $0<\alpha_k^*<C\,|\,\forall k\in\{1,...,N\}$ 所对应的样本.依据优化问题(2)的最优解 $\boldsymbol{\alpha}^*=[\alpha_1^*,...,\alpha_N^*]^T$,可以把训练数据集的样本划分为 3 类:

1. $\alpha_i^*=0\Rightarrow\xi_i^*=0,\|\boldsymbol{x}_i-\boldsymbol{b}^*\|<R^*$,样本点 $\boldsymbol{x}_i$ 在超球里面;
2. $0<\alpha_i^*<C\Rightarrow\xi_i^*=0,\|\boldsymbol{x}_i-\boldsymbol{b}^*\|=R^*$,样本点 $\boldsymbol{x}_i$ 在超球表面;
3. $\alpha_i^*=C\Rightarrow\xi_i^*>0,\|\boldsymbol{x}_i-\boldsymbol{b}^*\|>R^*$,样本点 $\boldsymbol{x}_i$ 在超球外面.

注意,SVDD 算法与其他核方法一样,易于实现核化,只需把上面的 $\boldsymbol{x}_i^T\boldsymbol{x}_j$ 简单替换成 $k(\boldsymbol{x}_i,\boldsymbol{x}_j)$即可.这里,$k(\cdot,\cdot)$为一个合适的核函数.

2 一个关于支撑向量数据域描述超球半径计算问题的例子

这里,我们将用一个例子来说明在 SVDD 算法中传统构造超球半径的方法所存在的问题.我们人工生成了 8 个样本点:$\boldsymbol{x}_1=[-2,0]$,$\boldsymbol{x}_2=[2,0]$,$\boldsymbol{x}_3=[0,-2]$,$\boldsymbol{x}_4=[0,2]$,$\boldsymbol{x}_5=[-1,0]$,$\boldsymbol{x}_6=[1,0]$,$\boldsymbol{x}_7=[0,-1]$,$\boldsymbol{x}_8=[0,1]$.设定惩罚系数为 $C=0.25$,并且采用了线性核的方式,通过求解 SVDD 算法的对偶优化问题(2),可以得到其最优解为

$$\boldsymbol{\alpha}^*=[\alpha_1^*,...,\alpha_8^*]=[0.25,0.25,0.25,0.25,0,0,0,0].$$

根据上述讨论可知,由于:

1. $\alpha_i^*=0.25(i=1,...,4)$, 所以样本点 $\boldsymbol{x}_i(i=1,\ldots,4)$在超球外面;
2. $\alpha_i^*=0(i=5,...,8)$, 所以样本点 $\boldsymbol{x}_i(i=5,\ldots,8)$在超球里面.

所以,根据公式(3)可以得到超球的圆心为 $\boldsymbol{b}^*=\sum_{i=1}^{8}\alpha_i^*\boldsymbol{x}_i=[0,0]$;然而却不能根据公式(4)来构建半径 R,因为不存在 $i\in\{1,\ldots,8\}$使得 $0<\alpha_i^*<0.25$ 成立,进而可知没有样本点在超球面上.事实上,通过后面的分析我们将会明白,该例子实际上对于半径 R 的最优解为一个区间.

这个例子说明,当实际应用 SVDD 算法时,仅仅依据文献[8]中的理论和方法仍然不够,还需要对该算法进行深入的分析和探索.下面分别从 SVDD 算法的原始优化问题和对偶优化问题两个角度对其进入深入分析.

3 支撑向量数据域描述原始优化问题的理论分析

注意,由对偶优化问题(2)可知,要使得问题有解需 $C\geqslant 1/N$.当 $C=1/N$ 时,则有 $\alpha_i^*=1/N$, 表示有的样本点都在超球外面,失去了意义.所以下面的分析假定了 $C>1/N$.为了清晰起见,本文的分析基于线性情况,但其结论容易推广到非线性情况.

3.1 支撑向量数据域描述优化问题的等价改写

首先,对于 SVDD 的原始优化问题(1)有如下性质.

引理 1. 考虑如下优化问题:

$$\begin{aligned}&\min\ r+C\sum_{i=1}^{N}\xi_i\\&\text{s.t. }\|\boldsymbol{x}_i-\boldsymbol{b}\|^2\leqslant r+\xi_i,\ \xi_i\geqslant 0,\ i=1,...,N\end{aligned} \tag{5}$$

并且假定 $r=R^2$,则$\{\boldsymbol{b}^*,r^*,\boldsymbol{\xi}^*\}$为优化问题(5)的最优解的充分必要条件是$\{\boldsymbol{b}^*,R^*,\boldsymbol{\xi}^*\}$为优化问题(1)的最优解.

证明:略(由于该定理的证明较为简单,所以略去). □

通过设定 $s=\boldsymbol{b}^T\boldsymbol{b}-r$,即 $r=\boldsymbol{b}^T\boldsymbol{b}-s$,进一步可以得到有如下结论.

引理 2. 考虑如下优化问题:

$$\begin{aligned}&\min\ \boldsymbol{b}^T\boldsymbol{b}-s+C\sum_{i=1}^N\xi_i\\&\text{s.t. } s-2\boldsymbol{b}^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i\leqslant 0,\ \xi_i\geqslant 0,\ i=1,\dots,N\end{aligned}\tag{6}$$

并且假定 $s=\boldsymbol{b}^T\boldsymbol{b}-r$,则$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$为优化问题(6)的最优解的充分必要条件是$\{\boldsymbol{b}^*,r^*,\boldsymbol{\xi}^*\}$为优化问题(5)的最优解.

证明:首先应该注意,优化问题(5)和优化问题(6)都是凸约束优化问题.

(1) 充分性.$\{\boldsymbol{b}^*,r^*,\boldsymbol{\xi}^*\}$是优化问题(5)的最优解的充要条件为以下式子(KKT 条件)成立[18]:

$$\begin{aligned}&\alpha_i^*(\boldsymbol{b}^{*T}\boldsymbol{b}^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-r^*-\xi_i^*)=0,\ C-\alpha_i^*-\beta_i^*=0,\ \beta_i^*\xi_i^*=0\\&\boldsymbol{b}^{*T}\boldsymbol{b}^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-r^*-\xi_i^*\leqslant 0,\ \xi_i^*\geqslant 0,\ \alpha_i^*\geqslant 0,\ \beta_i^*\geqslant 0,\ i=1,\dots,N\\&\boldsymbol{b}^*=\sum_{i=1}^N\alpha_i^*\boldsymbol{x}_i,\ \sum_{i=1}^N\alpha_i^*=1\end{aligned}\tag{7}$$

由于 $s^*=\boldsymbol{b}^{*T}\boldsymbol{b}^*-r^*$,根据上式则有

$$\begin{aligned}&\alpha_i^*(s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i^*)=0,\ C-\alpha_i^*-\beta_i^*=0,\ \beta_i^*\xi_i^*=0\\&s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i^*\leqslant 0,\ \xi_i^*\geqslant 0,\ \alpha_i^*\geqslant 0,\ \beta_i^*\geqslant 0,\ i=1,\dots,N\\&\boldsymbol{b}^*=\sum_{i=1}^N\alpha_i^*\boldsymbol{x}_i,\ \sum_{i=1}^N\alpha_i^*=1\end{aligned}\tag{8}$$

公式(8)说明$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$满足优化问题(6)最优解的充分必要条件(KKT 条件),故其为优化问题的(6)的最优解.

(2) 必要性.由于$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解,则公式(8)必定成立.由于 $s^*=\boldsymbol{b}^{*T}\boldsymbol{b}^*-r^*$,把其代入公式(8)可得优化问题(5)最优解的充分必要条件公式(7).所以,$\{\boldsymbol{b}^*,r^*,\boldsymbol{\xi}^*\}$为优化问题(5)的最优解.

综合上述两方面的论证,可得命题成立. □

利用上面的引理 1 和引理 2,则可以直接建立 SVDD 原始优化问题(1)和优化问题(6)之间的关系.对此,有如下定理.

定理 1. 假定 $s=\boldsymbol{b}^T\boldsymbol{b}-R^2$,则$\{\boldsymbol{b}^*,R^*,\boldsymbol{\xi}^*\}$是 SVDD 算法的优化问题(1)的最优解的充分必要条件是$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$为优化问题(6)的最优解.

证明:根据引理 1,$\{\boldsymbol{b}^*,R^*,\boldsymbol{\xi}^*\}$是 SVDD 算法的原始优化问题(1)的最优解的充要条件是$\{\boldsymbol{b}^*,r^*,\boldsymbol{\xi}^*\}$为优化问题(5)的最优解.根据引理 2,$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解的充要条件是$\{\boldsymbol{b}^*,r^*,\boldsymbol{\xi}^*\}$是优化问题(5)的最优解,所以命题成立. □

3.2 圆心唯一性证明

通过上面的分析可以看出,研究 SVDD 的原始优化问题(1)可以转化为对优化问题(6)的研究.对于优化问题(6)有如下结论.

引理 3. 优化问题(6)关于 $\boldsymbol{b}$ 的最优解具有唯一性,即若$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$和$\{\hat{\boldsymbol{b}}^*,\hat{s}^*,\hat{\boldsymbol{\xi}}^*\}$都是优化问题(6)的最优解,则有 $\boldsymbol{b}^*=\hat{\boldsymbol{b}}^*$.

证明:首先,定义向量 $\boldsymbol{z}=[\boldsymbol{b},s,\boldsymbol{\xi}]$,$\boldsymbol{z}^*=[\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*]$和 $\hat{\boldsymbol{z}}^*=[\hat{\boldsymbol{b}}^*,\hat{s}^*,\hat{\boldsymbol{\xi}}^*]$,则优化问题(6)的目标函数可以表示为

$$F(\boldsymbol{z})=F([\boldsymbol{b},s,\boldsymbol{\xi}])=\boldsymbol{b}^T\boldsymbol{b}-s+\sum_{i=1}^N\xi_i\tag{9}$$

为了证明命题,只需证明:若 $\boldsymbol{z}^*=[\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*]$与 $\hat{\boldsymbol{z}}^*=[\hat{\boldsymbol{b}}^*,\hat{s}^*,\hat{\boldsymbol{\xi}}^*]$不同,则必有

$$\boldsymbol{b}^*=\hat{\boldsymbol{b}}^*\tag{10}$$

由于优化问题(6)为凸约束二次优化问题,故其最优解集为凸集,而且任意局部最优解都是全局最优解[18].既然 $\boldsymbol{z}^*$和 $\hat{\boldsymbol{z}}^*$ 都是其最优解,所以可得 $\boldsymbol{z}_t=t\boldsymbol{z}^*+(1-t)\hat{\boldsymbol{z}}^*$ $(\forall 0\leqslant t\leqslant 1)$也是其最优解,进而有 $F(\boldsymbol{z}_t)=F(\boldsymbol{z}^*)=F(\hat{\boldsymbol{z}}^*)$.即

$$F(\boldsymbol{z}_t)-F(\boldsymbol{z}^*)=0\tag{11}$$

由于

$$\boldsymbol{z}_t=t\boldsymbol{z}^*+(1-t)\hat{\boldsymbol{z}}^*=[t\boldsymbol{b}^*+(1-t)\hat{\boldsymbol{b}}^*,ts^*+(1-t)\hat{s}^*,\boldsymbol{\xi}^*+(1-t)\hat{\boldsymbol{\xi}}^*]\tag{12}$$

所以,根据公式(11)和公式(12)有对 t 的恒等式:

$$(t\boldsymbol{b}^*+(1-t)\hat{\boldsymbol{b}}^*)^T(t\boldsymbol{b}^*+(1-t)\hat{\boldsymbol{b}}^*)-(ts^*+(1-t)\hat{s}^*)+\sum_{i=1}^{N}(t\xi_i^*+(1-t)\hat{\xi}_i^*)-F(z^*)=0 \tag{13}$$

对上式两端对 t 求二次导数可得

$$2(\boldsymbol{b}^*-\hat{\boldsymbol{b}}^*)^T(\boldsymbol{b}^*-\hat{\boldsymbol{b}}^*)=0 \tag{14}$$

所以公式(10)成立.命题得证. □

利用上面的引理 3,进一步有如下定理.

定理 2. SVDD 算法原始优化问题(1)的最优解构建的超球圆心具有唯一性.

证明:由于 $\boldsymbol{b}$ 在 SVDD 算法中的几何意义为超球圆心,并根据定理 1 和引理 3,易知命题成立. □

3.3 半径不唯一性分析

定义 1. 假定$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解,则定义集合:

$$\begin{aligned}D_1&=\{i\,|\,s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i>0,\ i=1,\dots,N\}\\D_2&=\{i\,|\,s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i=0,\ i=1,\dots,N\}\\D_3&=\{i\,|\,s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i<0,\ i=1,\dots,N\}\end{aligned} \tag{15}$$

并记 N_1,N_2 和 N_3 分别为集合 D_1,D_2 和 D_3 中所含元素个数.

引理 4. 假定$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解,则根据定义 1 有

$$\xi_i^*=\begin{cases}s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i>0, & \forall i\in D_1\\0, & \forall i\in D_2\\0, & \forall i\in D_3\end{cases} \tag{16}$$

并且优化问题(6)目标函数的最小值为

$$\boldsymbol{b}^{*T}\boldsymbol{b}^*-s^*+C\sum_{i\in D_1}\xi_i^* \tag{17}$$

证明:首先,由于$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解,所以下面式子(KKT 条件)成立:

$$\begin{aligned}&\alpha_i^*(s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i^*)=0,\ C-\alpha_i^*-\beta_i^*=0,\ \beta_i^*\xi_i^*=0,\ i=1,\dots,N\\&s^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i^*\leqslant 0,\ \xi_i^*\geqslant 0,\ i=1,\dots,N\end{aligned} \tag{18}$$

所以,根据定义 1 和公式(18)易得命题第 1 部分成立,并进一步利用该结论有

$$\boldsymbol{b}^{*T}\boldsymbol{b}^*-s^*+C\sum_{i=1}^{N}\xi_i^*=\boldsymbol{b}^{*T}\boldsymbol{b}^*-s^*+C\sum_{i\in D_1}\xi_i^* \tag{19}$$

从而命题得证. □

利用上面的引理 4,进一步可以得到如下引理.

引理 5. 假定$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解,则优化问题(6)关于 s 最优解不唯一的充分必要条件是

$$N_1C=1 \tag{20}$$

其中,N_1 为定义 1 中集合 D_1 的元素个数.

证明:首先证明充分性,然后证明必要性.

1. 充分性.首先设定 $\delta=\min_{i\in D_1}\xi_i^*$,并根据引理 4 有$\delta>0$.令 $s'=s^*-\delta,\boldsymbol{b}'=\boldsymbol{b}^*$,并且

$$\xi_i'=\begin{cases}\xi_i^*-\delta, & \forall i\in D_1\\\xi_i'=\xi_i^*, & \forall i\in D_2\\\xi_i'=\xi_i^*, & \forall i\in D_3\end{cases} \tag{21}$$

则根据引理 4,对于$\{\boldsymbol{b}',s',\boldsymbol{\xi}'\}$有:

(1) 对于$\forall i\in D_1$ 有 $\xi_i'=\xi_i^*-\delta\geqslant 0$, 并且 $s'-2\boldsymbol{b}'^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i'=0$;

(2) 对于$\forall i\in D_2$ 有 $\xi_i'=0$, 并且 $s'-2\boldsymbol{b}'^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i'<0$;

(3) 对于$\forall i\in D_3$有 $\xi_i'=0$, 并且 $s'-2\boldsymbol{b}'^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-\xi_i'<0$.

故$\{\boldsymbol{b}',s',\boldsymbol{\xi}'\}$满足优化问题(6)的约束条件,为可行解.又根据 N_1C=1 和公式(16),有

$$\boldsymbol{b}'^T\boldsymbol{b}'-s'+C\sum_{i=1}^{N}\xi_i'=\boldsymbol{b}^{*T}\boldsymbol{b}^*-s^*+C\sum_{i\in D_1}\xi_i^* \tag{22}$$

故$\{\boldsymbol{b}',s',\boldsymbol{\xi}\}$也是优化问题(6)的最优解.

2. 必要性.首先假定$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$和$\{\boldsymbol{b}',s',\boldsymbol{\xi}\}$都是优化问题(6)的最优解,其中,$\boldsymbol{b}'=\boldsymbol{b}^*,s'=s^*-\delta,\delta'>0$.这里,我们假定了 $s^*>s'$,但这并不影响证明的一般性.

由于优化问题(6)是凸约束优化问题,可知其最优解为凸集[18].所以,$s''=s^*-\delta''$也是其关于 s 的最优解.即,$\{\boldsymbol{b}'',s'',\boldsymbol{\xi}''\}$也是最优解.这里,$0\leqslant\delta''\leqslant\delta'$,$\boldsymbol{b}''=\boldsymbol{b}^*$.对于最优解$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$,根据定义 1,总可以选择充分小的$\delta''$,使得:

(1) 对于$\forall i\in D_1$.有 $s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i>0$;

(2) 对于$\forall i\in D_2$.有 $s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i=-\delta''\leqslant 0$;

(3) 对于$\forall i\in D_3$.有 $s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i<0$.

再根据定义 1,可以得到关于最优解$\{\boldsymbol{b}'',s'',\boldsymbol{\xi}''\}$的集合:

$$\begin{aligned}D_1''&=\{i\mid s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i>0,\ i=1,\dots,N\}\\D_2''&=\{i\mid s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i=0,\ i=1,\dots,N\}\\D_3''&=\{i\mid s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i<0,\ i=1,\dots,N\}\end{aligned} \tag{23}$$

所以,根据上面的结论则有 $D_1=D_1''$, $D_2\cup D_3=D_2''\cup D_3''$,进而有 $N_1''=N_1$, $N_2''+N_3''=N_2+N_3$.再根据引理 4,有

$$\xi_i''=\begin{cases}s''-2\boldsymbol{b}''^T\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i=\delta^*-\delta'', & \forall i\in D_1''\\0, & \forall i\in D_2''\cup D_3''\end{cases} \tag{24}$$

所以,对于$\{\boldsymbol{b}'',s'',\boldsymbol{\xi}''\}$的目标函数值为

$$\boldsymbol{b}''^T\boldsymbol{b}''-s''+C\sum_{i\in D_1''}\xi_i''=\boldsymbol{b}^{*T}\boldsymbol{b}^*-s^*+C\sum_{i\in D_1}\xi_i^*-(1-N_1C)\delta'' \tag{25}$$

显然,要使得目标函数值保持最小不变,必有 N_1C=1.

综合步骤 1 和步骤 2,从而命题得证. □

由上面的引理容易得到如下定理.

定理 3. 若假定$\{\boldsymbol{b}^*,R^*,\boldsymbol{\xi}^*\}$是优化问题(1)的最优解,则优化问题(1)关于$R$的最优解不唯一的充分必要条件是

$$N_1C=1 \tag{26}$$

其中,N_1为集合 D_1 的元素个数.这里,D_1定义为

$$D_1=\{i\mid \boldsymbol{b}^{*T}\boldsymbol{b}^*-2\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i-R^{*2}>0,\ i=1,\dots,N\} \tag{27}$$

证明:根据引理 5 和定理 1,易得命题成立. □

4 支撑向量数据域描述对偶优化问题的理论分析及超球半径计算

4.1 对偶优化问题的理论分析

首先应该注意,优化问题(6)与 SVDD 算法的优化问题(1)具有相同的对偶优化问题(2),对此有如下结论.

引理 6. 假定$\boldsymbol{\alpha}^*=[\alpha_1^*,\dots,\alpha_N^*]^T$是对偶优化问题(2)的最优解,如果不存在分量$\alpha_i^*$满足$\alpha_i^*\in(0,C)$, 则原始优化问题(6)关于 s 的最优解为一个区间.即,假定$\{\boldsymbol{b}^*,s^*,\boldsymbol{\xi}^*\}$是优化问题(6)的最优解,则

$$s^*\in[s^L,s^U] \tag{28}$$

其中,

$$s^L=\max_{i\in D_1}\{2\boldsymbol{b}^{*T}\boldsymbol{x}_i-\boldsymbol{x}_i^T\boldsymbol{x}_i\},\ s^U=\min_{i\in D_2}\{2\boldsymbol{b}^{*T}\boldsymbol{x}_i-\boldsymbol{x}_i^T\boldsymbol{x}_i\} \tag{29}$$

这里,

$$D_1=\{i\mid\alpha_i^*=C,\ i=1,...,N\},\ D_2=\{i\mid\alpha_i^*=0,\ i=1,...,N\},\ \boldsymbol{b}^*=\sum_{i=1}^{N}\alpha_i^*\boldsymbol{x}_i^T \tag{30}$$

证明:首先,既然 $\boldsymbol{\alpha}^*=[\alpha_1^*,...,\alpha_N^*]^T$ 是对偶优化问题(2)的最优解,则如下式子(KKT 条件)成立:

$$\begin{aligned}&2\sum_{j=1}^{N}\alpha_j^*\boldsymbol{x}_i^T\boldsymbol{x}_j-\boldsymbol{x}_i^T\boldsymbol{x}_i-s^*-\gamma_i^*+\xi_i^*=0\\&\xi_i^*(\alpha_i^*-C)=0,\ \gamma_i^*\alpha_i^*=0,\ 0\leqslant\alpha_i^*\leqslant C,\ \xi_i^*\geqslant 0,\ \gamma_i^*\geqslant 0,\ i=1,...,N\end{aligned} \tag{31}$$

由命题可知不存在 $\alpha_i^*\in(0,C)$,所以 $\alpha_i^*=0$ 或 $\alpha_i^*=C$.进而有:

1. 当 $\alpha_i^*=C$ 时,由 $\gamma_i^*\alpha_i^*=0$ 可得 $\gamma_i^*=0$,从而进一步根据公式(31)有,

$$s^*=2\sum_{j=1}^{N}\alpha_j^*\boldsymbol{x}_i^T\boldsymbol{x}_j-\boldsymbol{x}_i^T\boldsymbol{x}_i+\xi_i^*\geqslant 2\sum_{j=1}^{N}\alpha_j^*\boldsymbol{x}_i^T\boldsymbol{x}_j-\boldsymbol{x}_i^T\boldsymbol{x}_i=2\boldsymbol{b}^{*T}\boldsymbol{x}_i-\boldsymbol{x}_i^T\boldsymbol{x}_i.$$

2. 当 $\alpha_i^*=0$ 时,由 $\xi_i^*(\alpha_i^*-C)=0$ 可得 $\xi_i^*=0$,从而进一步根据公式(31)有,

$$s^*=2\sum_{j=1}^{N}\alpha_j^*\boldsymbol{x}_i^T\boldsymbol{x}_j-\boldsymbol{x}_i^T\boldsymbol{x}_i-\gamma_i^*\leqslant 2\sum_{j=1}^{N}\alpha_j^*\boldsymbol{x}_i^T\boldsymbol{x}_j-\boldsymbol{x}_i^T\boldsymbol{x}_i=2\boldsymbol{b}^{*T}\boldsymbol{x}_i-\boldsymbol{x}_i^T\boldsymbol{x}_i.$$

所以命题成立. □

定理 4. 假定 $\boldsymbol{\alpha}^*=[\alpha_1^*,...,\alpha_N^*]^T$ 是对偶优化问题(2)的最优解,如果不存在分量 α_i^* 满足 $\alpha_i^*\in(0,C)$,则原始优化问题(1)关于 R 的最优解为一个区间.即,如果$\{\boldsymbol{b}^*,R^*,\boldsymbol{\xi}^*\}$为 SVDD 原始优化问题(1)的最优解,则

$$R^*\in[R^L,R^U] \tag{32}$$

其中,

$$\begin{aligned}R^L&=\sqrt{\max_{i\in D_2}\{\boldsymbol{b}^{*T}\boldsymbol{b}^*-2*\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i\}}=\max_{i\in D_2}\|\boldsymbol{b}^*-\boldsymbol{x}_i\|\\R^U&=\sqrt{\min_{i\in D_1}\{\boldsymbol{b}^{*T}\boldsymbol{b}^*-2*\boldsymbol{b}^{*T}\boldsymbol{x}_i+\boldsymbol{x}_i^T\boldsymbol{x}_i\}}=\min_{i\in D_1}\|\boldsymbol{b}^*-\boldsymbol{x}_i\|\end{aligned} \tag{33}$$

这里,

$$D_1=\{i\mid\alpha_i^*=C,i=1,...,N\},\ D_2=\{i\mid\alpha_i^*=0,i=1,...,N\},\ \boldsymbol{b}^*=\sum_{i=1}^{N}\alpha_i^*\boldsymbol{x}_i^T \tag{34}$$

证明:根据引理 6 和定理 1,易知命题成立. □

4.2 超球半径计算

通过前面的分析可以看出,当根据 SVDD 算法的原始优化问题(1)或对偶优化问题(2)来寻找包含正常类样本点的最小超球时,在一定情况下并不能按照公式(4)来计算超球的半径,从而不能应用传统的 SVDD 算法.因此,在 SVDD 算法中,传统的计算半径方法存在不足,需要进一步加以完善.

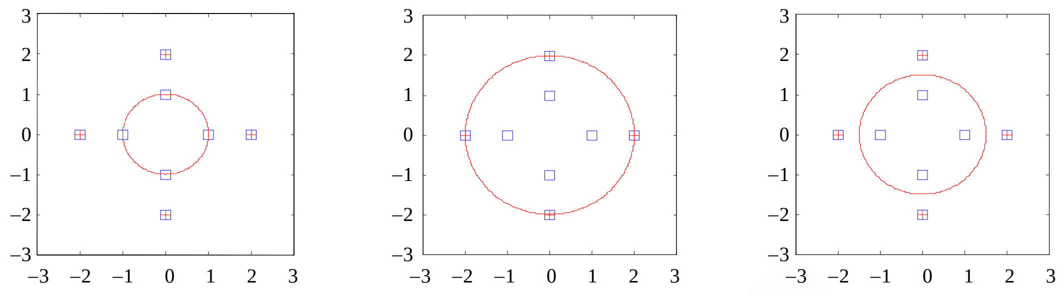
事实上,根据定理 4 可以确定优化问题(1)关于 R 的最优解的取值区间,所以我们可以从该范围中任意选择一个值来作为超球半径,如可以选择 R 最优解取值区间的上界 $R^*=R^L$,下界 $R^*=R^U$ 或者上下界中间值 $R^*=(R^L+R^U)/2$.然而,根据 SVDD 算法的几何意义,我们希望找到的超球既满足优化问题(1),又能使超球最小,即半径 R 最小,因此,选取 R 取值范围的下界来作为最后所求超球半径更加合理,即

$$R^*=R^L \tag{35}$$

这里再来回顾第 2 节中的例子.首先,通过求解对偶优化问题(2)可以得到超球圆心 $\boldsymbol{b}^*=[0,0]$,然而却不能直接得到半径,因为没有 $i\in\{1,…,8\}$使得 $0<\alpha_i^*<0.25$ 成立.但是,根据定理 4 得到关于超球半径 R 最优解的区间为

$$R^*\in[R^L,R^U]=[1,2] \tag{36}$$

图 1(a)~图 1(c)分别描述了当超球半径分别取 R 最优解区间的下界 $R^*=1$、上界 $R^*=2$ 以及上下界中间值 $R^*=1.5$ 时所得超球情况,其中,小方块内有‘+’的是支撑向量.可以看出,当取下界时超球更小,更加体现了 SVDD 算法的几何意义.



(a) Sphere radius R is set to the lower bound of the interval
(a) 超球半径取 R 最优解区间的下界

(b) Sphere radius R is set to the upper bound of the interval
(b) 超球半径取 R 最优解区间的上界

(c) Sphere radius R is set to the middle-value of the interval
(c) 超球半径取 R 最优解区间的上下界中间值

Fig.1 Experimental results on the artificial dataset which is used in Section 2
图 1 第 2 节中人造数据集的实验结果

下面简要概括修改后的 SVDD 算法的步骤.

算法 1. 修改后的 SVDD 算法.

输入:训练数据集 $D=\{\boldsymbol{x}_i|\boldsymbol{x}_i\in\boldsymbol{R}^d,i=1,\dots,N\},C>1/N$;

Step 1. 求解优化问题(2),得到其最优解为$\boldsymbol{\alpha}^*$;

Step 2. 根据公式(3)计算超球圆心 $\boldsymbol{a}^*$;

Step 3. If $\exists i\in\{1,\dots,N\}$使得 $0<\alpha_i^*<C$ 成立

　　根据公式(4)计算超球的半径 R^*;

Else

　　根据公式(35)计算超球的半径 R^*.

End

5 一个真实数据例子

本节我们将从 UCI[21]选取一个数据集——Iris 数据集来进行测试.该数据集由 150 个样本组成,每个样本包含 4 个连续型数值属性.为了使实验具有更强的直观性,我们选取了该数据集第 1 类的前两个特征属性来组成实验数据,共 50 个样本.

实验中我们采用了高斯核函数,即核函数为 $\exp(-(\boldsymbol{u}-\boldsymbol{v})^T(\boldsymbol{u}-\boldsymbol{v})/2\sigma^2)$.其中:$\sigma$为核参数,我们设定为$\sigma=2$;而对于惩罚系数设定为 $C=0.1$.利用实验数据,通过求解其 SVDD 算法的对偶优化问题(2)可以发现,共有 10 个支撑向量,每一个支撑向量对应的 Lagrange 乘子为 $\alpha_i^*=0.1$,$i\in\{9,14,15,16,19,33,34,39,42,43\}$,如图 2(a)所示.其中,小圆圈内有'+'的是支撑向量.根据公式(3),可以计算得到在特征空间中超球圆心向量的二范数平方为$||\boldsymbol{b}^*||^2=0.8323$,然而却不能依据公式(4)来计算特征空间中的超球半径,因为没有 $i\in\{1,\dots,50\}$使得 $0<\alpha_i^*<0.1$ 成立.

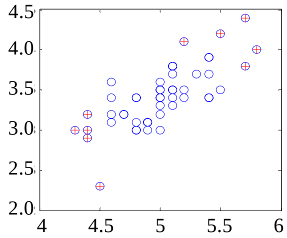
但是,依据定理 4,可以计算得到优化问题(1)关于 R 最优解的区间为

$$R^*\in[R^L,R^U]=[0.2896,0.3160] \tag{37}$$

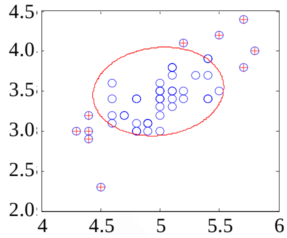
图 2(b)~图 2(d)分别描述了当超球半径分别取 R 最优解区间的下界 $R^*=0.2896$、上界 $R^*=0.3160$ 以及上下界的中间值 $R^*=0.3028$ 时所得到的超球情况.可以看出,当取下界时更为合理.此时,该球把正常类数据包得更为紧凑.

该实验说明:在 SVDD 算法中,依据传统方法来计算超球半径时在一定情况下的不足,也验证了文中理论的正确性,并且证明了根据文中提供的方法,即在原始优化问题关于 R 最优解存在不唯一性时也能够找到一个合

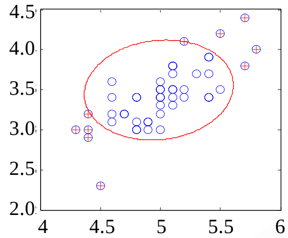
适的超球来描述数据.



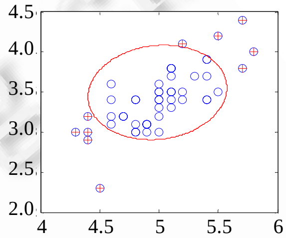
(a) Illustration of the experiment data and the support vectors
(a) 实验数据及支撑向量可视化



(b) Sphere radius R is set to the lower bound of the interval
(b) 超球半径取 R 最优解区间的下界



(c) Sphere radius R is set to the upper bound of the interval
(c) 超球半径取 R 最优解区间的上界



(d) Sphere radius R is set to the middle-value of the interval
(d) 超球半径取 R 最优解区间的上下界中间值

Fig.2 Experimental results on the Iris dataset
图 2 Iris 数据实验结果

6 总 结

近年来,尽管 SVDD 算法已经得到了广泛的研究和发展,但是该算法的优化问题最优解性质的理论分析却缺少关注.本文从理论上证明了在 SVDD 算法中根据其定义的优化问题的最优解来构建超球的圆心具有唯一性,然而其半径在一定情况下却存在不唯一性.本文还从对偶优化问题的角度分析了超球的圆心和半径性质,并且提出了计算超球半径的方法.文中还结合一个人造数据集和一个真实数据集来说明该算法依据传统构建超球方法的不足以及本文提供的超球半径计算方法的可行性.由于 One-Class SVM 可以视为一种特殊的 SVDD 算法,所以显然,本文结论易推广到 One-Class SVM 中.本文的工作完善了 SVDD 算法的理论和方法体系,从而为该算法的深入研究和应用奠定了理论基础.

References:

[1] Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. IEEE Trans. on Neural Networks, 2001,12(2):181–201. [doi: 10.1109/72.914517]

[2] Burges CJC. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998,2(2):121–167. [doi: 10.1023/A:1009715923555]

[3] Burges CJC, Crisp DJ. Uniqueness of the SVM solution. In: Solla SA, Leen TK, Muller KR, eds. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2000. 223–229.

[4] Burges CJC, Crisp DJ. Uniqueness theorems for kernel methods. Neurocomputing, 2003,55(1-2):187–220. [doi: 10.1016/S0925-2312(03) 00434-X]

[5] Schölkopf B, Smola AJ, Williamson RC, Bartlett PL. New support vector algorithms. Neural Computation, 2000,12(5):1207–1245. [doi: 10.1162/089976600300015565]

[6] Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing, 2004,14(3):199–222. [doi: 10.1023/B:STCO.0000035301.49549.88]

[7] Zhang JR, Chiu SY, Lan LS. Non-Uniqueness of solutions of 1-norm support vector classification in dual form. In: Proc. of the IEEE Int'l Joint Conf. on Neural Networks (IJCNN). 2008. 3058−3061. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4634230&tag=1 [doi: 10.1109/IJCNN.2008.4634230]

[8] Tax DMJ, Duin RPW. Support vector data description. Machine Learning, 2004,54(1):45−66. [doi: 10.1023/B:MACH.0000008084.60811.49]

[9] Park J, Kang D, Kim J, Kwok JT, Tsang IW. SVDD-Based pattern denoising. Neural Computation, 2007,19(7):1919−1938. [doi: 10.1162/neco.2007.19.7.1919]

[10] Hur AB, Horn D, Siegelmann HT, Vapnik V. Support vector clustering. Journal of Machine Learning Research, 2001,2:125−137.

[11] Mu T, Nandi AK. Multiclass classification based on extended support vector data description. IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics, 2009,39(5):1206−1216. [doi: 10.1109/TSMCB.2009.2013962]

[12] Chu CS, Tsang IW, Kwok JT. Scaling up support vector data description by using core-sets. In: Proc. of the Int'l Joint Conf. on Neural Networks (IJCNN). 2004. 425−430. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1379943 [doi: 10.1109/IJCNN.2004.1379943]

[13] Tsang IW, Kwok JT, Cheung PM. Core vector machines: Fast SVM training on very large data sets. Journal of Machine Learning Research, 2005,6:363−392.

[14] Tsang IWH, Kwok JTY, Zurada JA. Generalized core vector machines. IEEE Trans. on Neural Networks, 2006,17(5):1126−1140. [doi: 10.1109/TNN.2006.878123]

[15] Chen B, Feng AM, Chen SC, Li B. One-Cluster clustering based data description. Chinese Journal of Computers, 2007,30(8):1325−1332 (in Chinese with English abstract).

[16] Zhao F, Zhang JY, Liu J. An optimizing kernel algorithm for improving the performance of support vector domain description. Acta Automatica Sinica, 2008,34(9):1122−1127 (in Chinese with English abstract).

[17] Wu MR, Ye JP. A small sphere and large margin approach for novelty detection using training data with outliers. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(11):2088−2092. [doi: 10.1109/TPAMI.2009.24]

[18] Boyd S, Vandenberghe L. Convex Optimization. New York: Cambridge University Press, 2004. http://www.stanford.edu/~boyd/cvxbook/

[19] Chang CC, Tsai HC, Lee YJ. A minimum enclosing balls labeling method for support vector clustering. 2009. http://dmlab1.csie.ntust.edu. tw/downloads/papers/SVC_MEB.pdf

[20] Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Computation, 2001,13(7):1443−1471. [doi: 10.1162/089976601750264965]

[21] Asuncion A, Newman DJ. UCI repository of machine learning databases. 2007. http://archive.ics.uci.edu/ml/

附中文参考文献:

[15] 陈斌,冯爱民,陈松灿,李斌.基于单簇聚类的数据描述.计算机学报,2007,30(8):1325−1332.

[16] 赵峰,张军英,刘敬.一种改善支撑向量域描述性能的核优化算法.自动化学报,2008,34(9):1122−1127.

王晓明(1977－),男,四川简阳人,博士生,主要研究领域为机器学习,模式识别,图像处理.

王士同(1964－),男,教授,博士生导师,主要研究领域为模式识别,模糊系统,生物信息学.