

线性规划支撑矢量机

周伟达, 张 莉, 焦李成

(西安电子科技大学雷达信号处理重点实验室, 陕西西安 710071)

摘 要: 本文通过对统计学习理论中一些重要结论, 特别是线性函数 VC 维数的分析, 得到了一种线性规划支撑矢量机, 包括线性规划线性支撑矢量机和线性规划非线性支撑矢量机. 在线性规划支撑矢量机中, 对其 VC 维数界作了适当的放宽. 文中最后对人工和实际样本进行了实验, 结果说明了线性规划支撑矢量机在推广能力上较好地逼近了原支撑矢量机, 而在计算复杂度上明显低于原支撑矢量机.

关键词: 统计学习理论; VC 维数; 支撑矢量机; 推广能力; 线性规划

中图分类号: TP391.4 文献标识码: A 文章编号: 0372-2112(2001)11-1507-05

Linear Programming Support Vector Machines

ZHOU Wei da, ZHANG Li, JIAO Li cheng

(National Key Lab. for Radar Signal Processing, Xidian University, Xi'an, Shanxi 710071, China)

Abstract: Based on analysis of the conclusions in the statistical learning theory, especially the VC dimension of linear functions, linear programming SVMs are presented, including linear programming linear and nonlinear SVMs. In linear programming linear SVMs, the bound of the VC dimension is loosened properly. Simulation results for both artificial and real data show the generalization performance of our method is a good approximation of SVMs and the computation complexity is largely reduced by our method.

Key words: statistical learning theory; VC dimension; support vector machines; generalization performance; linear programming

1 引言

由 Vapnik 等人发展的统计学习理论^[1-3], 通过引入结构风险实现对目标函数集的控制, 从而使学习机在经验风险和过拟合之间取一个折衷, 获得了较好的推广能力. 此基础上产生了一种优秀的机器学习模型——支撑矢量机^[2, 4, 5]. 目前该领域因其较完善的理论和较好的实用性, 已获得越来越广泛的关注.

在统计学习理论中, 粗略地说“ m_{Δ} 边界”分离超平面^[3]就是以 2 倍的 m_{Δ} 间距分离样本的线性超平面, 其 VC 维数有如下的结论: 样本 $x \in X$, 位于半径为 R 的超球体内, 则“ m_{Δ} 边界”分离超平面函数集的 VC 维数有如下的上界:

$$h \leq \min(\lfloor R^2/m_{\Delta}^2 \rfloor, n) + 1 \quad (1)$$

其中 n 为样本的维数, $\lfloor \cdot \rfloor$ 为取小于等于该数的整数. 由式(1)可以看出, 当目标函数集的 VC 维数 h 的上界小于 n 时, 它与分离边界 m_{Δ}^2 成反比, 即在样本分布一定时, 边界越大, 目标函数的 VC 维数 h 越小. 这样对于“ m_{Δ} 边界”分离超平面函数集而言, 最优目标函数 $f(x, w_{\phi}, b_0)$ 可通过最小化经验风险和最大化分离边界 m_{Δ} 来得到, 这就是 Vapnik 在支撑矢量机中引入的结构风险:

$$R_{\text{structure}}(\alpha) = C \cdot R_{\text{emp}}(\alpha) + 1/m_{\Delta}^2 \quad (2)$$

其中 C 为两者之间的一个权重因子. 线性模式识别支撑矢量

机引入结构风险如下^[4]:

$$\min \frac{1}{2} \|w\|_2^2 + C \cdot \sum_{i=1}^l \xi_i \quad (3)$$

约束于 $y_i((w \cdot x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, l, C$ 为惩罚因子, 可取一正常数. 式(3)中第一项 $\frac{1}{2} \|w\|_2^2$ 为最大化边界, 起假设空间容量控制作用, 避免学习机的过拟合; 第二项为最小化经验风险. 求此二次规划的 Wolfe 对偶^[4], 最终可得到下面的结果:

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (4)$$

约束于 $\sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \in [0, C], i = 1, \dots, l$. 通过在线性支撑矢量机中引入核函数 $(x_i \cdot x_j) = K(x_i, x_j)$ 可生成非线性支撑矢量机^[4].

支撑矢量机的求解为一线性约束凸二次规划, 其计算的时间和空间复杂度均很大. 本文从统计学习理论出发, 提出了一种线性规划支撑矢量机, 大大降低了求解的复杂度.

2 线性规划模式识别支撑矢量机

2.1 线性规划线性模式识别支撑矢量机

引理 1^[6] n 维线性空间 V^n 上的向量范数的等价性:

设 $\|x\|_\alpha$ 和 $\|x\|_\beta$ 为有限维线性空间 V^n 上的任意向量范数, 它们包括 l_p 范数, 加权范数等所有向量范数, 则存在两个与 x 无关的正常数 $0 < c_1 < +\infty, 0 < c_2 < +\infty$, 使得下面不等式成立

$$c_1 \|x\|_\beta \leq \|x\|_\alpha \leq c_2 \|x\|_\beta, \forall x \in V^n$$

定理 1 给定样本集 $((x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l)), (x, y) \in (R^n, R)$ 分布于最小半径为 R 的超球体内, 则对于“ m_Δ 边界”分离超平面目标函数集 $f(x, w, b) = w^T x + b$ 和样本 x , 机器学习过程采用如下判决函数:

$$y = \begin{cases} 1, & w^T x + b \geq \Delta \\ -1, & w^T x + b \leq -\Delta \end{cases}; \quad \Delta \geq 0$$

存在一正常数 $c, 0 < c < +\infty$ 且与样本 (x, y) 及目标函数集的参数 w, b 无关, 使得“ m_Δ 边界”分离超平面目标函数集的 VC 维数 h 的上界有下面的不等式成立

$$h \leq \min\left(\left\lceil \frac{c^2 \cdot R^2 \cdot \|w\|_\beta}{\Delta^2} \right\rceil, n\right) + 1 \quad (5)$$

其中 $\lceil \cdot \rceil$ 为取小于等于该数的最大整数, $\|w\|_\beta$ 为任意向量范数. 定理证明见附录 1.

由该定理直接可以导出如下的数学规划: 给定样本集 $((x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l)), (x, y) \in (R^n, R)$ 和线性目标函数集 $f(x, w, b) = w^T x + b$, 根据定理 1 引入结构风险, 先考虑经验风险为 0, 即无误差的情况:

$$\begin{aligned} \max \quad & L = r/c \|w\|_\beta \quad (6) \\ \text{s. t.} \quad & y_i(w^T x_i + b) \geq r, \quad i = 1, \dots, l; \\ & r \geq 0 \end{aligned}$$

其中 $\|w\|_\beta$ 为任意向量范数, c 为一常数, 当 $\|w\|_\beta$ 为欧氏范数时, $c = 1$. 现考虑 l_∞ 范数 $\|x\|_\infty$, 当然也存在一个正常数 c , 且 c 与样本集和函数集参数无关, 使得

$$h \leq \min\left(\left\lceil c^2 \cdot R^2 \cdot \|w\|_\infty^2 / \Delta^2 \right\rceil, n\right) + 1$$

由此, 式(6)可得:

$$L = r/c \|w\|_\infty \quad (7)$$

把上式 $L = r/c \|w\|_\infty$ 中的分母化为约束, 其中由于常数 c 不影响线性规划的求解, 可直接省去. 这样得到如下的规划:

$$\begin{aligned} LP1: \quad & \max \quad L = r \quad (8) \\ \text{s. t.} \quad & y_i(w^T x_i + b) \geq r, \quad i = 1, \dots, l; \\ & r \geq 0 \end{aligned} \quad (9)$$

$$-1 \leq w_j \leq 1, j = 1, \dots, n;$$

判决函数: $f(x, w, b) = w^T x + b$ (10)

模式识别输出为: $y = \text{sgn}(f(x, w, b))$ (11)

其中 $\text{sgn}(\cdot)$ 为符号函数.

关于允许一定误差存在的情况, 通过引入误差量 ξ_i , 可得到如下的线性规划:

$$LP2: \quad \min \quad L = -r + C \sum_{i=1}^l \xi_i \quad (12)$$

$$\text{s. t.} \quad y_i(w^T x_i + b) \geq r - \xi_i, \quad i = 1, \dots, l; \quad (13)$$

$$r \geq 0$$

$$-1 \leq w_j \leq 1, j = 1, \dots, n;$$

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

C 为一正常数.

判决函数和输出与无误差时的式(10)、(11)相同.

上述线性规划线性支撑矢量机, 是在目标函数集为 $f(x, w, b) = w^T x + b$ 的假设下得到的. 对于高维样本(样本数小于等于样本维数), 可采用如下的目标函数: $f(x) = \sum_{i=1}^l \alpha_i y_i(x_i \cdot x) + b$, 就可得到如下的线性规划支撑矢量机:

$$LP3: \quad \min \quad L = -r + C \sum_{i=1}^l \xi_i \quad (14)$$

$$\begin{aligned} \text{s. t.} \quad & yf(x_i) = y_i \left(\sum_{j=1}^l \alpha_j y_j (x_j \cdot x_i) + b \right) \geq r - \xi_i, \quad i = 1, \dots, l; \\ & r \geq 0 \\ & -1 \leq \alpha_j \leq 1, j = 1, \dots, l; \\ & \xi_i \geq 0, \quad i = 1, \dots, l \end{aligned} \quad (15)$$

2.2 线性规划非线性支撑矢量机

从与支撑矢量机通过引入核函数导出非线性支撑矢量机^[1-4]的方法得到启发, 在此也通过引入核函数 $K(x_i, x_j)$ 直接实现从线性到非线性的转换^[7]. 在式(9)和(13)中, 关于 w 和 x , 仅存在内积的形式($w \cdot x$). 现在首先假设把样本从低维空间映射到一个高维欧氏空间 H

$$\Phi: R^n \rightarrow H$$

当然算法中出现的仍是高维欧氏空间 H 中的内积形式($\Phi(x_i) \cdot \Phi(x_j)$). 如果一个核函数能分解成 $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ 就可以符合算法的要求. 实际在算法中仅出现核函数的形式, 而无需知道映射 Φ .

定理 2 给定样本集 $((\Phi(x_1), y_1), (\Phi(x_i), y_i), \dots, (\Phi(x_l), y_l)), (\Phi(x), y) \in (H, R)$ 分布于最小半径为 R 的超球体内, 和目标函数集:

$$f(x) = \sum_{i=1}^l \alpha_i y_i \Phi(x_i) \cdot \Phi(x) + b = \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \quad (16)$$

且对于样本 x , 机器学习过程采用如下决策

$$y = \begin{cases} 1, & f(x) \geq \Delta \\ -1, & f(x) \leq -\Delta \end{cases}; \quad \Delta \geq 0$$

则存在一正常数 $c, 0 < c < +\infty$ 且与样本 (x, y) 及目标函数集的参数 α, b 无关, 使得目标函数集的 VC 维数 h 的上界有下面的不等式成立

$$h \leq \min\left(\left\lceil c^2 \cdot R^2 \cdot \|\alpha\|_\beta / \Delta^2 \right\rceil, n\right) + 1 \quad (17)$$

其中 $\alpha = (\alpha_1 \dots \alpha_l)^T, \|\alpha\|_\beta$ 为任意向量范数, 证明见附录 2.

类似线性规划线性支撑矢量机, 考虑 $\|\alpha\|_\infty$, 得到如下的规划:

$$LP4: \quad \max \quad L = r \quad (18)$$

$$\begin{aligned} \text{s. t.} \quad & yf(\alpha_i) = y_i \left(\sum_{j=1}^l \alpha_j y_j K(x_j, x_i) + b \right) \geq r, \quad i = 1, \dots, l; \\ & r \geq 0 \\ & -1 \leq \alpha_j \leq 1, j = 1, \dots, l; \end{aligned} \quad (19)$$

$$\text{判决函数:} \quad f(x) = \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \quad (20)$$

模式识别输出为: $y = \text{sgn}(f(x))$ (21)

关于允许一定误差存在的情况, 通过引入误差量 ξ_i , 可得到如下的线性规划:

$$LP5: \quad \min \quad L = -r + C \sum_{i=1}^l \xi_i \quad (22)$$

$$\begin{aligned} \text{s. t.} \quad & y_i f(x_i) = y_i \left(\sum_{j=1}^l \alpha_j y_j K(x_j, x_i) + b \right) \geq r - \xi_i, \quad i = 1, \dots, l; \\ & r \geq 0 \\ & -1 \leq \alpha_j \leq 1, \quad j = 1, \dots, l; \\ & \xi_i \geq 0, \quad i = 1, \dots, l \end{aligned} \quad (23)$$

其中 C 为一正常数, 判决函数和输出与无误差时的式(20)(21)相同.

3 实验仿真

本文总共做了三个实验: 对于线性规划线性支撑矢量机, 用人工数据做了一个实验, 主要验证一下其原理的合理性; 对于线性规划非线性支撑矢量机, 用双螺旋线分类数据和手写体数字数据进行了实验. 在本文的三个实验中, 为了使训练时间具有可比性, 不管是二次规划支撑矢量机和线性规划支撑矢量机, 都采用最常用的二次规划和线性规划算法(程序来自 Matlab Optimization Toolboxes)来完成.

实验 1: 人工线性可分样本实验

随机产生了两组线性可分样本, 共 300 个, 100 个用于训练, 200 个用于检验. 表 1 和图 1 中给出了实验结果.

表 1 线性规划支撑矢量机对人工线性可分样本的实验结果

学习模型	训练样本数	检验样本数	训练时间 (s)	平均错误率 (%)	VC 界
					$R^2 \cdot \ W\ _2^2$ $\Delta^2 \Delta$
二次规划支撑矢量机	100	200	16.2	0.5	7820.2
线性规划支撑矢量机	100	200	0.8	0.5	7820.2

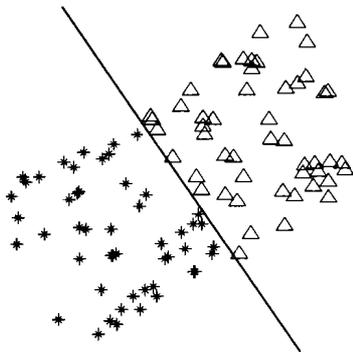


图 1 给出了线性规划支撑矢量机对两类线性可分样本的分类情况, 实线为 $f(x) = 0$ 虚线为 $f(x) = \pm r$

从实验 1 的结果可以看出, 在分类性能上和 VC 界上, 线性规划支撑矢量机与二次规划支撑矢量机完全一致, 而训练时间比二次规划支撑矢量机快了一个数量级以上.

实验 2: 人工双螺旋线分类

双螺旋线分类一直是模式识别领域公认的一个相当有难度的问题, 它也因其难度而经常被用作检验模式识别算法性能的试金石^[8]. 本文用它对线性规划支撑矢量机作了实验. 表 2 和图 2 中给出了实验的结果. 在本实验中, 选用了径向基核函数: $K(x_i, x_j) = \exp(-|x_i - x_j|^2 / 2p^2)$, 其中 $p = 8$.

表 2 线性规划支撑矢量机对双螺旋线分类的实验结果

学习模型	训练样本数	检验样本数	训练时间 (s)	平均错误率 (%)	VC 界
					$R^2 \cdot \ W\ _2^2$ $\Delta^2 \Delta$
二次规划支撑矢量机	252	628	4103.5	0	192.7
线性规划支撑矢量机	252	628	199.7	0	679.1

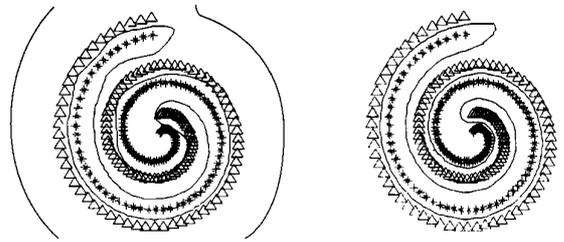


图 2 绘出了线性规划(a)和二次规划(b)支撑矢量机在同条件下对双螺旋线分类的情况, 实线为 $f(x) = 0$, 虚线为 $f(x) = \pm r$. 从图中, 可以看出两种学习模型分类情况都不错, 这在表 2 中的检验误差指标也有反映, 同时二次规划支撑矢量机比线性规划支撑矢量机的推广能力好, 这是为赢得算法效率而必须付出的性能代价, 可喜的是这个代价并不是很大.

从实验中可以看出, 线性规划支撑矢量机在训练时间上比二次规划支撑矢量机仍然占有较大的优势, 但其 VC 界明显比二次规划支撑矢量机大, 理论上其推广能力肯定应该不如二次规划支撑矢量机. 表 2 中平均错误率指标并没有反映出这个结果, 是因为样本由人工产生, 不含噪声的缘故. 图 2 中就可以明显地看出线性规划支撑矢量机比二次规划支撑矢量机推广能力的下降. 这是为赢得算法效率而付出的性能代价, 可喜的是这种性能的下降并不大, 在表 2 检验误差率指标中几乎是看不出这种下降.

实验 3: 手写体数字识别

本实验数据来源于美国 AT&T Bell 实验室的 MNIST 数据库^①, 该数据已在许多模式识别和支撑矢量机的文献中被用作实验数据^[9~11], 在此对数据的提供者表示感谢. 该手写体数字库中包含了 60000 个训练样本, 10000 个检验样本, 每个样本为 28×28 的灰度图点阵. 由于该手写体数字库比较大, 只取其中识别比较困难的两类“6”和“9”进行实验并对数据进行了标准化. 实验中仍然采用径向基核函数, 其中 $p = 30$. 表 3 中给出了实验结果.

① 可从 <http://www.research.att.com/~yann/ocr/mnist> 得到.

表3 线性规划支撑矢量机对手写数字识别的实验结果

训练样本数	检验样本数	线性规划支撑矢量机			二次规划支撑矢量机		
		训练时间 (s)	平均错误率 (%)	VC 界 $R^2 \cdot \ w\ _2^2 / \Delta^2$	训练时间 (s)	平均错误率 (%)	VC 界 $R^2 \cdot \ w\ _2^2 / \Delta^2$
100 + 100	1009 + 958	117.0	1.4	175.4	187.0	1.0	47.3
200 + 200	1009 + 958	369.3	1.2	311.9	2117.4	1.1	60.4
300 + 300	1009 + 958	546.1	0.7	323.4	5978.1	1	73.2
400 + 400	1009 + 958	823.3	1.0	355.8	18509.5	0.875	87.7
500 + 500	1009 + 958	1076.0	0.8	402.0	100878.7	0.7	101.8

从表3中可以看出,对于大样本时,二次规划支撑矢量机的训练时间随样本数的增加而指数增长,线性规划支撑矢量机随样本数的增加成近似线性增长,线性规划支撑矢量机的训练时间比二次规划支撑矢量机的训练时间快了一个数量级以上。

4 讨论

本文根据统计学习理论,特别是关于容量控制,VC维数,结构风险和线性目标函数集的VC界的理论,推导出线性规划支撑矢量机。线性规划支撑矢量机对VC界作了一定的放宽,这使得学习机在推广能力上有一定的下降,从实验结果看这些下降是微小的。线性规划的求解简单于二次规划,特别是对大规模的问题更是如此。本文采用线性规划来取代二次规划求解支撑矢量机,使得支撑矢量机的求解算法效率有了近一个数量级以上的提高。所以可以看出本文的方法通过牺牲少许的推广能力来换取算法效率是非常有效的,在实际运用中也是非常合理的。目前二次规划支撑矢量机已经出现多种快速算法,Chunking 算法^[12], Osuna 算法^[13], SMO 算法^[14], CDRM-SVM 算法^[15]等,而线性规划支撑矢量机是否有专门的针对支撑矢量机特性的快速算法,需要在以后的工作中进一步研究。

附录1 定理1证明

证明: 给定样本集 $((x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l))$, $(x, y) \in (R^n, R)$ 分布于最小半径为 R 的超球体内, " m_Δ 边界" 分离超平面目标函数集 $f(x, w, b) = w^T x + b$ 。对于样本 x , 机器训练过程采用如下判决函数:

$$y = \begin{cases} 1, & w^T x + b \geq \Delta \\ -1, & w^T x + b \leq -\Delta \end{cases}; \Delta \geq 0$$

则对于 " m_Δ 边界" 分离超平面函数集, 由式(1)可得, 其 VC 维数的上界为:

$$h \leq \min([R^2 \cdot \|w\|_2^2 / \Delta^2], n) + 1 \quad (附1)$$

又由引理1可得, 存在一个正常数 c , 满足 $0 < c < +\infty$ 且与向量 w 无关, 使得

$$\|w\|_2 \leq c \|w\|_\beta, \text{ 其中 } \|w\|_\beta \text{ 为任意向量范数。}$$

(附2)

由式(附1)和(附2)得 " m_Δ 边界" 分离超平面函数集的 VC 维数上界:

$$h \leq \min([R^2 \cdot \|w\|_2^2 / \Delta^2], n) + 1 \\ \leq \min([c^2 \cdot R^2 \cdot \|w\|_\beta^2 / \Delta^2], n) + 1$$

原命题得证。

附录2 定理2证明

证明 给定样本集 $((\Phi(x_1), y_1), \dots, (\Phi(x_i), y_i), \dots, (\Phi(x_l), y_l)), (\Phi(x), y) \in (H, R)$ 分布于最小半径为 R 的超球体内, 和目标函数集:

$$f(x) = \sum_{i=1}^l \alpha_i y_i \Phi(x_i) \cdot \Phi(x) + b = \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \quad (附3)$$

且对于样本 x , 机器学习过程采用如下决策函数

$$y = \begin{cases} 1, & f(x) + b \geq \Delta \\ -1, & f(x) + b \leq -\Delta \end{cases}; \Delta \geq 0$$

该目标函数集相当于在空间 H 中的一个超平面:

$$f(x) = w \cdot \Phi(x) + b \quad (附4)$$

其中

$$w = \sum_{i=1}^l \alpha_i y_i \Phi(x_i) \quad (附5)$$

在空间 H 中, 由式(附5)得

$$\|w\|_2^2 = w \cdot w = \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \Phi(x_i) \cdot \Phi(x_j) \\ = \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (附6)$$

令 K 为 $l \times l$ 的矩阵, $K(i, j) = y_i y_j K(x_i, x_j)$, $\alpha = (\alpha_1, \dots, \alpha_l)^T$, 则上式(附6):

$$\|w\|_2^2 = \alpha^T K \alpha \quad (附7)$$

由式(6), (7)可得, 改目标函数集的 VC 维数的上界为:

$$h \leq \min([R^2 \cdot \|w\|_2^2 / \Delta^2], n) + 1 \\ = \min([R^2 \cdot \alpha^T K \alpha / \Delta^2], n) + 1 \quad (附8)$$

下面再引入加权范数的概念:

定义1 加权范数^[9]

设 A 为任一 n 阶正定矩阵, 列向量 $x \in R^n$, 则函数 $\|x\|_A = (x^T A x)^{1/2}$ 是一种向量范数, 此范数被称为加权范数, 也叫椭圆范数。

由定义1可得, 由于式(附7)中 K 为已知, 对于参数相量 α :

$$\|w\|_2^2 = \alpha^T K \alpha = \| \alpha \|_K^2 \quad (附9)$$

因要求矩阵 K 为正定矩阵, 所以在此核函数的容许条件与二次规划支撑矢量机中的容许条件一致^[5]。由引理1得: 存在一正常数 c , 满足 $0 < c < +\infty$ 且与向量 α 和矩阵 K 无关, 使得下面不等式成立

$$\| \alpha \|_K \leq c \| \alpha \|_\beta \quad (附10)$$

其中 $\| \alpha \|_\beta$ 为任一向量范数, 由式(附8), (附9)和(附10)得目标函数集的 VC 维数 h 有如下的上界:

$$h = \min\left[\left[\frac{R^2 \cdot \alpha^T K \alpha}{\Delta^2}\right], n\right] + 1 = \min\left[\left[\frac{R^2 \cdot \| \alpha \|_K^2}{\Delta^2}\right], n\right] + 1 \\ \leq \min\left[\left[\frac{c^2 \cdot R^2 \cdot \| \alpha \|_\beta^2}{\Delta^2}\right], n\right] + 1$$

原命题得证。

参考文献:

- [1] Vapnik V. Statistical Learning Theory [M]. New York: Wiley Inter-science Publication, 1998.
- [2] Vapnik V. The Nature of Statistical Learning Theory [M]. New York: Springer Verlag, 1995.
- [3] Vapnik V. An overview of statistical learning theory [J]. IEEE Transactions on Neural Networks, 1999, 10(5): 988– 998.
- [4] Burges C J C. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 2(2): 1– 47.
- [5] Smola A J, Schölkopf B. A tutorial on support vector regression [R]. NeuroCOLT Technical Report NC TR-98-030, Royal Holloway College, University of London, UK, 1998.
- [6] 程云鹏, 张凯院, 徐仲, 等. 矩阵理论 [M]. 西安: 西北工业大学出版社, 1989.
- [7] Wahba G. An introduction to model building with reproducing kernel hilbert spaces [R]. Technical Report 1020, University of Wisconsin–Madison, Statistics Dept., 2000.
- [8] Lang K J, Witbrock M J. Learning to tell two spirals apart [A]. In Proc. 1989 Connectionist Models Summer School [C], 1989: 52– 61.
- [9] Schölkopf B, Burges C, Vapnik V. Extracting support data for a given task [A]. In M Fayyad, R Uthurusamy (Eds.), Proceedings, First International Conference on Knowledge Discovery & Data Mining [C], Menlo Park. AAAI Press, 1995.
- [10] Schölkopf B, Burges C, Vapnik V. Incorporating invariances in support vector learning machines [A]. In C. von der Malsburg, W. von Seelen, J. C. Vorberggen, and B. Sendhoff (Eds.), Artificial Neural Networks—ICANN' 96, Berlin: Springer Lecture Notes in Computer Science [C], 1996, 1112: 47– 52.
- [11] Burges C, Schölkopf B. Improving the accuracy and speed of support vector learning machines [A]. In M Mozer, M Jordan, T Petsche (Eds.), Advances in Neural Information Processing Systems 9 [C], Cambridge, MA: MIT Press, 1997, 375– 381.
- [12] Vapnik V. Estimation of Dependencies Based on Empirical Data [in Russian] [M]. Moscow: Nauka, 1979. (English translation, New York: Springer Verlag, 1982).
- [13] Osuna E, Freund R, Girosi G. Improved training algorithm for support vector machines [A]. Proc. IEEE NNSP' 97 [C]. Amelia Island, 1997.
- [14] Platt J. Fast training of support vector machines using sequential minimal optimization [A]. In B Schölkopf, C J C Burges, A J Smola, editors, Advances in Kernel Methods—Support Vector Learning [C], Cambridge, MA: MIT Press., 1999: 185– 208.
- [15] 焦李成, 张莉, 周伟达. 支撑矢量预选取的中心距离比值法 [J]. 电子学报, 2001, 29(3): 383– 386.

作者简介:



周伟达 男. 1974 年出生于浙江省海宁市. 西安电子科技大学在读博士. 研究领域包括: 机器学习、统计学习理论和智能信号处理.



张莉 女. 1975 年出生于贵州省绥阳县. 西安电子科技大学在读博士. 主要研究方向有数据挖掘、模式识别和神经网络.