

有效的 ν 支持向量回归机的 ν 解路径算法*

顾彬^{1,2,3+}, 王建东³

¹(江苏省网络监控中心(南京信息工程大学), 江苏 南京 210044)

²(南京信息工程大学 计算机与软件学院, 江苏 南京 210044)

³(南京航空航天大学 计算机科学与技术系, 江苏 南京 210016)

Effective ν -Path Algorithm for ν -Support Vector Regression

GU Bin^{1,2,3+}, WANG Jian-Dong³

¹(Jiangsu Engineering Center of Network Monitoring (Nanjing University of Information Science and Technology), Nanjing 210044, China)

²(School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China)

³(Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

+ Corresponding author: E-mail: jsgubin@163.com

Gu B, Wang JD. Effective ν -path algorithm for ν -support vector regression. *Journal of Software*, 2012, 23(10): 2643–2654 (in Chinese). <http://www.jos.org.cn/1000-9825/4153.htm>

Abstract: The ν -support vector regression (ν -SVR) proposed by Schölkopf, *et al.*, has the advantage of using the parameter ν to control the number of support vectors and margin errors, however, compared to ε -SVR, its formulation is more complicated. Until now, there have been no effective methods used to compute the ν -path for it. This paper proposes a new solution path algorithm, which is designed based on a modified formulation of ν -SVR and traces the solution path with respect to the parameter ν . Through theoretical analysis and experiments, results can show that the algorithm can avoid the infeasible updating path, and fit the entire ν -path in finite steps.

Key words: model selection; solution path; ν -support vector regression; support vector machine; machine learning

摘要: 由 Schölkopf 等人提出的 ν 支持向量回归机具有通过参数 ν 控制支持向量和错误向量个数的优点,然而与标准的支持向量机相比,其形式更为复杂,迄今为止仍没有有效的算法计算 ν 解路径.基于 ν 支持向量回归机的修改形式,提出了一种新的解路径算法,它能够追踪参数 ν 对应的所有解,并通过理论分析和实验,说明了该算法能够尽可能地避免不可行的更新路径,并在有限步内拟合出所有的 ν 解路径.

关键词: 模型选择;解路径; ν 支持向量回归机;支持向量机;机器学习

中图法分类号: TP18 **文献标识码:** A

由于良好的泛化能力,支持向量机(support vector machines,简称 SVMs)^[1]已成为机器学习中最流行的方法之一.然而,仍有一些悬而未决的问题需要解决,其中一个重要的问题就是模型选择^[2-4].即如何调整支持向量机

* 基金项目: 国家自然科学基金重点项目(61139002); 国家自然科学基金青年科学基金(61202137); 江苏高校优势学科建设工程资助项目; 南京信息工程大学科研启动费(20110433)

收稿时间: 2010-01-08; 修改时间: 2011-01-20, 2011-07-21; 定稿时间: 2011-11-02

的参数以达到最佳的泛化性能.

一般的模型选择方法是指定一些候选参数值,然后应用交叉验证的方法^[4]在这些候选值中做出最佳的选择.如果搜索空间较大,即意味着需要在不同的参数下训练 SVM 多次,因此,对于最优化参数的全面且广泛的搜索应用得很少.为了克服这个困难,Hastie 等人^[5]和 Zhu 等人^[6]提出了一种新奇的方法(SvmPath 算法),它可以拟合二范数 SVM^[1]和一范数 SVM^[6]正则化参数的整个解路径,而不需要多次训练 SVM 模型;在此基础上,Gunter 和 Zhu^[7,8]以及 Wang 等人^[9]将 SvmPath 算法扩展到了支持向量回归机(ε -SVR).这些算法都是基于解路径线性分段的属性而提出来的.除此之外,Wang 等人^[10]还提出了一种非线性算法,该算法可以近似追踪 C-SVC 关于核参数对应的解路径.值得注意的是,Loosli 等人^[11]应用 SvmPath 给出相应 ν -SVR 解路径算法,但他们没有注意到,简单地应用 SvmPath 于 ν -SVR 将导致不可更新路径的问题(参见第 2.2 节),本文将针对该问题提出一种有效的 ν -SVR 解路径算法.

ν 支持向量回归机(ν -SVR)^[12]是一种有趣的支持向量机形式.给定训练样本集合 $T=\{(x_1,y_1),\dots,(x_l,y_l)\}$ (其中, $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$), ν -SVR 考虑如下的原始问题:

$$\begin{aligned} \min_{w, \varepsilon, b, \xi^{(*)}} \quad & \frac{1}{2} \langle w, w \rangle + C \cdot \left(\nu \varepsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^{*}) \right) \\ \text{s.t.} \quad & (\langle w, \phi(x_i) \rangle + b) - y_i \leq \varepsilon + \xi_i, \quad y_i - (\langle w, \phi(x_i) \rangle + b) \leq \varepsilon + \xi_i^{*} \\ & \xi_i^{(*)} \geq 0, \quad \varepsilon \geq 0 \end{aligned} \quad (1)$$

其中, $(*)$ 是一种有或无星号变量的简写.通过转换函数 ϕ ,训练样本 x_i 被映射到一个高维再生核 Hilbert 空间(reproducing kernel Hilbert space,简称 RKHS); C 是正则化常数; ν 是引进的比例参数($0 \leq \nu \leq 1$),通过 ν 可以控制支持向量和错误向量的数量.更准确地说, ν 是错误向量比例的上界值,是支持向量比例的下界值.此外, ν 以概率 1 渐近地等于错误向量和支持向量的比例值.

虽然公式(1)有上述优点,但其对偶问题比 ε -SVR 更为复杂,如下所示:

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

其中, $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, $\langle \cdot, \cdot \rangle$ 表示在 RKHS 中的内积.相对于 ε -SVR 的对偶形式,可以很清楚地看出,公式(2)多出了一个额外的不等式约束.

关于 ν -SVM 的早期研究主要集中在参数 ν 渐近的最佳选择上^[12,13].由于上述额外的不等式约束,到目前为止,仍没有有效的方法计算 ν -SVR 的 ν 解路径.Loosli 等人^[11]虽然应用 SvmPath 给出相应的 ν -SVR 解路径算法,但他们并未注意到,简单地应用 SvmPath 算法将导致不可更新路径的问题(参见第 2.2 节).本文将在我们前期工作^[4]的基础上,致力于给出有效的 ν -SVR 的解路径算法.首先,本文将引入一个 ν -SVR 的修改形式,它将公式(2)中的不等式约束转化成等式约束;另一个需要解决的问题是直接将 SvmPath 算法应用到 ν -SVR 的修改形式不能保证总是生成一个可行的更新路径.针对该问题,本文提出了一种新的解路径算法(统称为 ν -SvrPath),它可以看作是 SvmPath 算法的一种扩展.通过理论分析以及实验可以说明, ν -SvrPath 将尽可能地避免不可行的更新路径,并在有限步内拟合出全部的解路径.

本文第 1 节给出一种修改的 ν -SVR 表示形式及其 KKT(Karush-Kuhn-Tucker)条件.第 2 节提出 ν -SvrPath 算法.第 3 节证明 ν -SvrPath 方法的可行性.第 4 节描述实验结果.最后对本文内容作总结展望.

1 ν -SVR 的修改形式及其 KKT 条件

为了使 ν -SVR 形式更加简单,本文首先介绍下面的定理^[15],该定理说明了在公式(2)中对于任意给定的 ν , 总存在最优解满足等式 $\sum_{i=1}^l (\alpha_i^* + \alpha_i) = C\nu$.

定理 1(Chang 和 Lin,2002^[15]). 对于对偶问题(2),其中, $0 \leq \nu \leq 1$,总存在最优解使得等式 $\sum_{i=1}^l (\alpha_i^* + \alpha_i) = C\nu$

满足.

根据该定理,在公式(2)中的不等式约束 $\sum_{i=1}^l (\alpha_i^* + \alpha_i) \leq C\nu$ 可以视为等式约束 $\sum_{i=1}^l (\alpha_i^* + \alpha_i) = C\nu$, 这意味着本文可以考虑如下的最小化问题来代替对偶问题(2):

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i, x_j) - \sum_{i=1}^l (\alpha_i^* - \alpha_i)y_i \\ \text{s.t.} \quad & \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, \sum_{i=1}^l (\alpha_i^* + \alpha_i) = C\nu, 0 \leq \alpha_i^{(*)} \leq \frac{C}{l}, i=1, \dots, l \end{aligned} \quad (3)$$

此外,为了使该最小化问题形式上更加紧凑,本文引入一个扩展的训练样本集 S , 它被定义为 $S = S^- \cup S^+$, 其中, $S^- = \{(x_i, y_i, z_i = -1)\}_{i=1}^l, S^+ = \{(x_i, y_i, z_i = +1)\}_{i=1}^{2l}, z_i$ 是训练样本 (x_i, y_i) 的标签. 在此基础上,对应的最小化问题(3)可以被改写为

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i,j=1}^{2l} \alpha_i \alpha_j Q_{ij} - \sum_{i=1}^{2l} \alpha_i y_i \\ \text{s.t.} \quad & \sum_{i=1}^{2l} z_i \alpha_i = 0, \sum_{i=1}^{2l} \alpha_i = C\nu, 0 \leq \alpha_i \leq \frac{C}{l}, i=1, \dots, 2l \end{aligned} \quad (4)$$

这里, Q 是半正定矩阵, 其中, $Q_{ij} = z_i z_j K(x_i, x_j)$.

根据凸优化理论^[16], 通过优化如下的凸二次目标函数, 同样可以得到最小化问题(4)的解:

$$\min_{0 \leq \alpha_i \leq \frac{C}{l}} W_{\nu} = \frac{1}{2} \sum_{i,j=1}^{2l} \alpha_i \alpha_j Q_{ij} - \sum_{i=1}^{2l} z_i y_i \alpha_i + b' \left(\sum_{i=1}^{2l} z_i \alpha_i \right) + \varepsilon' \left(\sum_{i=1}^{2l} \alpha_i - C\nu \right) \quad (5)$$

其中, b' 和 ε' 都是拉格朗日乘子.

根据 KKT 定理^[16], 类似文献[17]中的推导可以获得如下的 KKT 条件:

$$\forall (x_i, y_i, z_i) \in S: g_i = \frac{\partial W_{\nu}}{\partial \alpha_i} = \sum_{j=1}^{2l} \alpha_j Q_{ij} - z_i y_i + z_i b' + \varepsilon' \begin{cases} g_i \geq 0, \text{ for } \alpha_i = 0 \\ g_i = 0, \text{ for } 0 < \alpha_i < \frac{C}{l} \\ g_i \leq 0, \text{ for } \alpha_i = \frac{C}{l} \end{cases} \quad (6)$$

$$\frac{\partial W_{\nu}}{\partial b'} = \sum_{i=1}^{2l} z_i \alpha_i = 0 \quad (7)$$

$$\frac{\partial W_{\nu}}{\partial \varepsilon'} = \sum_{i=1}^{2l} \alpha_i - C\nu = 0 \quad (8)$$

根据函数 g_i 的值, 扩展的训练样本集 S 可分为 3 个相互独立的集合(如图 1 所示):

- 集合(a): S_S 包含严格在 ε 带边界上的支持向量;
- 集合(b): S_E 包含在 ε 带外的错误支持向量;
- 集合(c): S_R 包含被 ε 带覆盖的剩余向量.

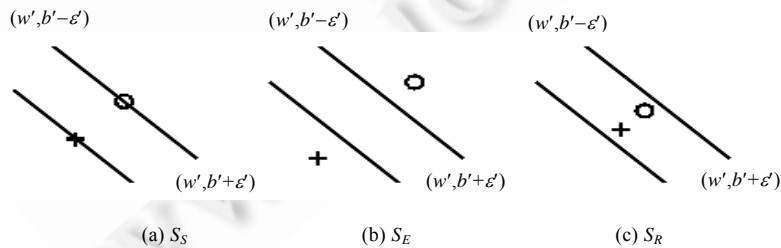


Fig.1 Partition of the extended training samples S into three independent sets
图 1 ν 支持向量回归机扩展的训练样本集 S 划分为 3 个相互独立的集合

2 ν -SVR 的 ν 解路径算法

本节针对最小化问题(4)提出 ν -SvrPath 算法(见算法 1).该算法主要包含两部分:一是建立最小化问题(4)的初始解,二是针对参数 $\nu(0 \leq \nu \leq 1)$ 搜索相应的解路径.

算法 1. ν -SvrPath 算法.

输入: $S, \Delta \nu$;

输出: $\alpha(\nu), g(\nu): 0 \leq \nu \leq 1$.

1. 初始化 $\nu \leftarrow 0$
2. Repeat
3. 初始化 $\alpha, g, b', \varepsilon', S_S, S_E$ 和 $S_R, \text{exception} \leftarrow \text{true}$ (见第 2.1 节).
4. While $\nu > 0$ & exception 为真时执行:
 5. 根据公式(16)和公式(17)计算 $\hat{\beta}$ 和 $\hat{\gamma}$,
 6. 根据公式(18)计算最小调整量 $\Delta \zeta^{\min}$ (见第 2.2.2 节),
 7. 更新 $\nu, \alpha, g, b', \varepsilon', S_S, S_E$ 和 S_R ,
 8. 根据文献[14]中的相关规则更新逆矩阵 \hat{R} .
 9. 如果 $|\sum_{i \in S_S} z_i| = |S_S|$ 并且 $S_E^{-Z_{S_S}} = \emptyset$, 则
 10. 更新 $\text{exception} \leftarrow \text{false}$
 11. 更新 $\nu \leftarrow \nu + \Delta \nu$
 12. 结束
 13. 结束
14. Until $\nu < 1$

2.1 初始化

ν -SvrPath 算法第 1 步需要建立最小化问题(4)的初始解.因为在 $\nu=0$ 时,问题(4)不需要求解二次规划问题(见引理 2),所以 ν -SvrPath 算法将以 $\nu=0$ 建立初始解.事实上,该算法也可以从任意中间解(例如,任意可行的 $\nu(0 \leq \nu \leq 1)$ 对应的问题(4)的最优解)开始.

引理 2. 如果 $\nu=0$,则最小化问题(4)的最优解为所有 $\alpha_i=0$.

证明:如果 $\nu=0$,明显有 $\alpha_i=0$ 仅是最小化问题(4)的唯一可行解.引理得证. \square

当所有的 $\alpha_i=0$ 时,由公式(6)我们可以得到:

$$\forall i \in S: g_i = -z_i y_i + z_i b' + \varepsilon' \geq 0 \quad (9)$$

令 S_+ 表示 $z_i=+1$ 的扩展的训练样本集合, S_- 表示 $z_i=-1$ 的扩展的训练样本集合,选择 $i_+ = \arg \max_{i \in S_+} y_i$ 和 $i_- = \arg \min_{i \in S_-} y_i$ (本文假设 i_+ 和 i_- 是唯一的).那么,公式(9)可以转化为

$$\varepsilon' \geq \frac{y_{i_+} - y_{i_-}}{2} \quad (10)$$

$$b' \in [y_{i_-} - \varepsilon', y_{i_+} + \varepsilon'] \quad (11)$$

当 $\varepsilon' = \frac{y_{i_+} - y_{i_-}}{2}$ 时,则有 $b' = \frac{y_{i_+} + y_{i_-}}{2}$.这就构成了 ε' 和 b' 的初始状态,该状态意味着集合 S_S 将从两个或更多个样本开始.

2.2 ν 解路径

获得问题(4)的初始解后, ν 解路径算法可在保持所有样本满足 KKT 条件的情况下,通过逐渐增加 ν (亦即 $\frac{\sum_{i \in S} \alpha_i}{C}$) 来搜索对应的解路径;当 $\sum_{i \in S} \alpha_i = C$ 时,算法终止.

此外,如果 ν -SvrPath 算法从中间解开始,也可以相似地得到解路径.

2.2.1 ν 的绝缘增量调整

对于 ν 值的每一次调整,为了保持所有的样本满足 KKT 条件, S_S 中的样本对应的权值,拉格朗日乘子 b' 和 ε' 都应作相应的调整.根据公式(6)~公式(8),可得如下线性方程组:

$$\Delta g_i = \sum_{j \in S_S} \Delta \alpha_j Q_{ij} + z_i \Delta b' + \Delta \varepsilon' = 0, \forall i \in S_S \tag{12}$$

$$\sum_{j \in S_S} z_j \Delta \alpha_j = 0 \tag{13}$$

$$\sum_{j \in S_S} \Delta \alpha_j + \epsilon \Delta \varepsilon' + \Delta \zeta = 0 \tag{14}$$

在上面的式子中, $\Delta \zeta$ 是引进的调整 $\sum_{i \in S} \alpha_i$ 的变量, ϵ 是任意的负数.类似文献[14]中的处理, $\epsilon \Delta \varepsilon'$ 在公式(14)中作为额外项.当 $\left| \sum_{i \in S_S} z_i \right| = |S_S|$ 时,由于 $\Delta \zeta$ 的变化在 $\sum_{i \in S_S} z_i \Delta \alpha_i = 0$ 和 $\sum_{i \in S_S} \Delta \alpha_i = -\Delta \zeta$ 之间存在如表 1 所示的两种冲突情形,因此使用该项可以有效地阻止在调整参数 ν 的过程中所发生的冲突(具体见第 3 节的分析与第 4 节的实验).

Table 1 Two cases of conflicts on the adiabatic incremental adjustment

表 1 绝缘增量调整中的两种冲突情形

集合 S_S 样本的标签		冲突
+1	-1	是/否
✓	✓	否
✓		是
	✓	是

如果定义 $e_{S_S} = [1, \dots, 1]^T$ 为 $|S_S|$ 维的列向量,并且令 $z_{S_S} = [z_1, \dots, z_{|S_S|}]^T$,那么,线性方程组(12)~(14)可以改写为

$$\underbrace{\begin{bmatrix} 0 & 0 & z_{S_S}^T \\ 0 & \epsilon & e_{S_S}^T \\ z_{S_S} & e_{S_S} & Q_{S_S, S_S} \end{bmatrix}}_{\hat{Q}} \cdot \underbrace{\begin{bmatrix} \Delta b' \\ \Delta \varepsilon' \\ \Delta \alpha_{S_S} \end{bmatrix}}_{\Delta h} = - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Delta \zeta \tag{15}$$

根据定理 4, \hat{Q} 矩阵存在一个逆矩阵 \hat{R} ,那么可以得到 Δh 和 $\Delta \zeta$ 之间的线性关系,如下所示:

$$\Delta h = \begin{bmatrix} \Delta b' \\ \Delta \varepsilon' \\ \Delta \alpha_{S_S} \end{bmatrix} = -\hat{R} \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Delta \zeta \stackrel{\text{def}}{=} \underbrace{\begin{bmatrix} \hat{\beta}_{b'} \\ \hat{\beta}_{\varepsilon'} \\ \hat{\beta}_{S_S} \end{bmatrix}}_{\hat{\beta}} \Delta \zeta \tag{16}$$

根据公式(16),有 $\sum_{i \in S} \Delta \alpha_i = -(1 + \epsilon \hat{\beta}_{\varepsilon'}) \Delta \zeta$ (根据定理 6,在 $\epsilon < 0$ 的条件下有 $1 + \epsilon \hat{\beta}_{\varepsilon'} \geq 0$),这意味着,对 $\sum_{i \in S} \alpha_i$ 的调整控制可以通过 $\Delta \zeta$ 来实现.

最后,把公式(16)代入公式(12),可以得到 Δg_i 和 $\Delta \zeta$ 之间的线性关系,如下所示:

$$\forall i \in S: \Delta g_i = \left(\sum_{j \in S_S} \hat{\beta}_j Q_{ij} + z_i \hat{\beta}_{b'} + \hat{\beta}_{\varepsilon'} \right) \Delta \zeta = \hat{\gamma}_i \Delta \zeta \tag{17}$$

在公式(17)中,很明显地,对于任意的 $i \in S_S$,有 $\hat{\gamma}_i = 0$.

2.2.2 计算最小调整量 $\Delta \zeta^{\min}$

绝缘增量调整并不能直接给出所有的解路径,问题在于,随着 ν 值的增加,会改变集合 S_S, S_R 和 S_E 的组成成分.其解决方法是捕获集合 S_S, S_R 和 S_E 的最小变化,即计算最小调整量 $\Delta \zeta^{\min}$,使得正好某一样本在集合 S_S, S_R 和 S_E 间要进行迁移.为了更好地说明集合间组成成分的变化,类似文献[14],需考虑如下 3 种情形:

1. S_S 中样本对应的权值 α_i 到达边界(上界或者下界).这样,在集合 S_S 中的某个样本移动到集合 S_R 或 S_E

- 之前存在一个最小可能更新值 $\Delta\zeta^{S_S}$;
- 在 S_R 或 S_E 中某个样本的 g_i 函数到达 0. 这样, 在集合 S_R 或 S_E 里的某个样本移动到 S_S 集合之前存在一个最小可能更新值 $\Delta\zeta^{S_R, S_E}$;
 - 当 $\sum_{i \in S} \alpha_i$ 到达 C 时, 也即终止条件满足. 那么, 在 ν 解路径算法遇到 ν 终止条件之前存在最小调整量:

$$\Delta\zeta^\omega = \frac{\sum_{i \in S} \alpha_i - C}{1 + \epsilon \hat{\beta}_{\epsilon'}}.$$

最后, 3 个值中的最大值:

$$\Delta\zeta^{\min} = \max \{ \Delta\zeta^{S_S}, \Delta\zeta^{S_R, S_E}, \Delta\zeta^\omega \} \quad (18)$$

将构成 $\Delta\zeta$ 的最小调整量. 值得注意的是, 一旦 S_S 发生改变, 逆矩阵 \hat{R} 可以类似文献[14]进行相应更新.

最小调整量 $\Delta\zeta^{\min}$ 被确定之后, ν 可以按照下面的规则进行更新:

$$\nu \leftarrow \nu - \frac{1 + \epsilon \hat{\beta}_{\epsilon'}}{C} \Delta\zeta^{\min} \quad (19)$$

3 可行性分析

通过理论分析, 本文证明了 ν -SvrPath 算法将尽可能地避免不可行更新路径, 并且能够在有限步内拟合出整个解路径.

本节首先定义能量函数 \mathcal{E} , 如下所示:

$$\mathcal{E}(\alpha) = \frac{1}{2} \alpha^T Q \alpha - \sum_{i=1}^{2l} z_i y_i \alpha_i + b' \sum_{i=1}^{2l} z_i \alpha_i + \epsilon' \left(\sum_{i=1}^{2l} \alpha_i - C \right) \quad (20)$$

然后证明能量函数 \mathcal{E} 在绝缘增量调整中是严格单调增加的 (见定理 10), 随后展示了能量函数 \mathcal{E} 在有限步内将收敛到 $\min_{0 \leq \alpha_i \leq C/l} \mathcal{E}$ 的最优解 (见定理 11), 这意味着 ν -SvrPath 算法可以在有限步内拟合出整个解路径.

假设 1 (Gu 和 Wang 2008^[18]). $\forall i, j \in S_S, i \neq j$, 如果 $x_i \neq \pm x_j$, 那么, 向量集合 $\{z_i \phi(x_i), \dots, z_{|S_S|} \phi(x_{|S_S|})\}$ 在 RKHS 里是线性无关的.

引理 3. $\forall i, j \in S_S, i \neq j$, 如果 $x_i \neq \pm x_j$, 矩阵 Q_{S_S, S_S} 是正定的.

这里证明略去, 可参考文献[18]的相关内容.

定理 4. 在绝缘增量调整中, 如果 $\epsilon < 0$, 那么, \hat{Q} 矩阵的行列式总大于 0.

证明: 为了证明该定理, 首先, 我们定义 \hat{Q} 矩阵为

$$\begin{bmatrix} 0 & 0 & z_{S_S}^T \\ 0 & 0 & e_{S_S}^T \\ z_{S_S} & e_{S_S} & Q_{S_S, S_S} \end{bmatrix}.$$

根据行列式因式分解^[19], 我们有

$$\begin{aligned} \det(\tilde{Q}) &= \det(-[y_{S_S}, e_{S_S}]^T Q_{S_S, S_S}^{-1} [y_{S_S}, e_{S_S}]) \det(Q_{S_S, S_S}) \\ &= \det \left(- \begin{bmatrix} y_{S_S}^T Q_{S_S, S_S}^{-1} y_{S_S} & y_{S_S}^T Q_{S_S, S_S}^{-1} e_{S_S} \\ e_{S_S}^T Q_{S_S, S_S}^{-1} y_{S_S} & e_{S_S}^T Q_{S_S, S_S}^{-1} e_{S_S} \end{bmatrix} \right) \det(Q_{S_S, S_S}) \\ &= (y_{S_S}^T Q_{S_S, S_S}^{-1} y_{S_S} \cdot e_{S_S}^T Q_{S_S, S_S}^{-1} e_{S_S} - y_{S_S}^T Q_{S_S, S_S}^{-1} e_{S_S} \cdot e_{S_S}^T Q_{S_S, S_S}^{-1} y_{S_S}) \det(Q_{S_S, S_S}). \end{aligned}$$

因为 Q_{S_S, S_S} 是一个正定矩阵, 所以 $\det(Q_{S_S, S_S}) > 0$, 根据 Cauchy-Schwarz 不等式^[19]:

$$z_{S_S}^T Q_{S_S, S_S}^{-1} z_{S_S} \cdot e_{S_S}^T Q_{S_S, S_S}^{-1} e_{S_S} - z_{S_S}^T Q_{S_S, S_S}^{-1} e_{S_S} \cdot e_{S_S}^T Q_{S_S, S_S}^{-1} z_{S_S} \geq 0.$$

因此, 我们有

$$\det(\tilde{Q}) \geq 0 \quad (21)$$

类似地, 我们也有

$$\det(\tilde{Q}_{\setminus \epsilon'}) = \det(0 - z_{S_S}^T Q_{S_S, S_S}^{-1} z_{S_S}) \cdot \det(Q_{S_S, S_S}) < 0 \quad (22)$$

根据拉普拉斯展开 \tilde{Q} 的行列式^[19],我们有

$$\det(\hat{Q}) = \det(\tilde{Q}) + \epsilon \det(\tilde{Q}_{\setminus e'e'}) \quad (23)$$

基于公式(21)、公式(22)和前提 $\epsilon < 0$, 结论 $\det(\hat{Q}) > 0$ 成立. □

定理 5. 在绝缘增量调整中,如果 $\epsilon < 0$, 则 $1 + \epsilon \hat{\beta}_{e'} \geq 0$ 当且仅当 $|\sum_{i \in S_S} z_i| = |S_S|$ 等式成立.

证明:为了证明该定理,首先,我们定义 \tilde{Q} 矩阵为

$$\begin{bmatrix} 0 & 0 & z_{S_S}^T \\ 0 & 0 & e_{S_S}^T \\ z_{S_S} & e_{S_S} & Q_{S_S S_S} \end{bmatrix}$$

由公式(16),我们有

$$1 + \epsilon \hat{\beta}_{e'} = 1 - \frac{\epsilon \det(\tilde{Q}_{\setminus e'e'})}{\det(\hat{Q})} = \frac{\det(\tilde{Q})}{\det(\hat{Q})}$$

如果 $|\sum_{i \in S_S} z_i| = |S_S|$, 即集合 S_S 的样本只有一种标签, 则有 $e_{S_S} = \pm z_{S_S}$, 这意味着 \tilde{Q} 矩阵的行列式等于 0. 那么, 根据定理 4, 我们有 $1 + \epsilon \hat{\beta}_{e'} = 0$.

如果 $|\sum_{i \in S_S} z_i| < |S_S|$, 类似于定理 4 的证明, 易证 \tilde{Q} 的行列式大于 0. 因此, 根据定理 4, 我们有 $1 + \epsilon \hat{\beta}_{e'} > 0$. □

引理 6. 在绝缘增量调整中, $\forall t \in S_S$, 有 $\hat{\beta}_t \neq \pm\infty$, 并且 $\forall t \in S - S_S$, 有 $\hat{\gamma}_t \neq \pm\infty$.

此处证明略去, 可参考文献[18]的相关内容.

引理 7. 在绝缘增量调整中, 如果 $|\sum_{i \in S_S} z_i| = |S_S|$, 即集合 S_S 的样本有相同的标签(这里, 集合 S_S 的标签被定义为 z_{S_S}), 那么, 对 $\forall t \in S_S$, 有 $\hat{\beta}_t = 0$, 对于 $\forall t \in S_{z_{S_S}}$, 有 $\hat{\gamma}_t = 0$, 对于 $\forall t \in S_{-z_{S_S}}$, 有 $\hat{\gamma}_t = -\frac{2}{\epsilon}$.

证明:由公式(16),对于 $\forall t \in S_S$, 我们有

$$\hat{\beta}_t = -\hat{R}_{te'} = -\frac{(-1)^{i_{e'}+i_t} \det(\hat{Q}_{\setminus e't})}{\det(\hat{Q})}$$

如果 $|\sum_{i \in S_S} z_i| = |S_S|$, 容易验证矩阵 $\hat{Q}_{\setminus e't}$ 的行列式等于 0, 那么 $\hat{\beta}_t = 0$.

进一步地, 根据定理 4 证明中的分析, 如果 $|\sum_{i \in S_S} z_i| = |S_S|$, 我们有 $\det(\hat{Q}) = \epsilon \det(\hat{Q}_{\setminus e'e'})$, 那么,

$$\begin{aligned} \hat{\beta}_{e'} &= -\hat{R}_{e'e'} = -\frac{\det(\hat{Q}_{\setminus e'e'})}{\epsilon \det(\hat{Q}_{\setminus e'e'})} = -\frac{1}{\epsilon}, \\ \hat{\beta}_{b'} &= -\hat{R}_{b'e'} = -\frac{-\det(\hat{Q}_{\setminus e'b'})}{\epsilon \det(\hat{Q}_{\setminus e'e'})} = \frac{z_{S_S}}{\epsilon}. \end{aligned}$$

因此, 根据公式(17), $\forall t \in S_{z_{S_S}}$, 有 $\hat{\gamma}_t = z_{S_S} \frac{z_{S_S}}{\epsilon} - \frac{1}{\epsilon} = 0$, $\forall t \in S_{-z_{S_S}}$, 有 $\hat{\gamma}_t = -z_{S_S} \frac{z_{S_S}}{\epsilon} - \frac{1}{\epsilon} = -\frac{2}{\epsilon}$. 引理得证. □

假设 2. 在绝缘增量调整中, 如果 $|\sum_{i \in S_S} z_i| = |S_S|$, 即集合 S_S 的样本有相同的标签(这里, 集合 S_S 的标签被定义为 z_{S_S}), 那么, 集合 $S_E^{-z_{S_S}}$ (它包含了所有的 $z_i = -z_{S_S}$ 错误支持向量)总是非空的.

实际上, 在绝缘增量调整中, $|\sum_{i \in S_S} z_i| = |S_S|$ 并且 $S_E^{-z_{S_S}} = \emptyset$ (以下称这种情形为异常)是非常罕见的, 实验部分将对发生异常的可能性进行更详细的讨论.

推论 8. 对于每一次绝缘增量调整, $-\infty < \Delta \zeta^{\min} < 0$.

证明:根据第 2.2.2 节的表述, 该推理很容易从定理 5、引理 6、引理 7 以及假设 2 得出. □

定理 9. 在绝缘增量调整中, 能量函数 \mathcal{E} 是严格单调增加函数.

证明:在绝对增量调整中, 假设前一次调整用 k 进行索引, 下一次调整用 $k+1$ 进行索引, 并令 $\hat{\beta}_{S_R} = 0, \hat{\beta}_{S_E} = 0$, 那么有,

$$\begin{aligned}
 \mathcal{E}(\alpha^{[k+1]}) &= \frac{1}{2} \sum_{i_1, i_2 \in S} (\alpha_{i_1}^{[k]} + \hat{\beta}_{i_1}^{[k]} \Delta \zeta^{[k]})(\alpha_{i_2}^{[k]} + \hat{\beta}_{i_2}^{[k]} \Delta \zeta^{[k]}) Q_{i_1 i_2} - \sum_{i \in S} z_i y_i (\alpha_i^{[k]} + \hat{\beta}_i^{[k]} \Delta \zeta^{[k]}) + \\
 &\quad (b^{[k]} + \hat{\beta}_b^{[k]} \Delta \zeta^{[k]}) \sum_{i \in S} y_i (\alpha_i^{[k]} + \hat{\beta}_i^{[k]} \Delta \zeta^{[k]}) + (\varepsilon^{[k]} + \hat{\beta}_{\varepsilon'}^{[k]} \Delta \zeta^{[k]}) \left(\sum_{i \in S} \alpha_i^{[k]} + \hat{\beta}_i^{[k]} \Delta \zeta^{[k]} - C \right) \\
 &= \mathcal{E}^{[k]} + \sum_{i \in S} g_i^{[k]} \hat{\beta}_i^{[k]} \Delta \zeta^{[k]} + \frac{1}{2} \sum_{i \in S} \hat{\gamma}_i^{[k]} \hat{\beta}_i^{[k]} (\Delta \zeta^{[k]})^2 + \frac{1}{2} \hat{\beta}_{\varepsilon'}^{[k]} \sum_{i \in S} \hat{\beta}_i^{[k]} (\Delta \zeta^{[k]})^2 + \hat{\beta}_{\varepsilon'}^{[k]} \Delta \zeta^{[k]} \left(\sum_{i \in S} \alpha_i^{[k]} - C \right) \\
 &= \mathcal{E}^{[k]} + \frac{1}{2} \hat{\beta}_{\varepsilon'}^{[k]} \sum_{i \in S} \hat{\beta}_i^{[k]} (\Delta \zeta^{[k]})^2 + \hat{\beta}_{\varepsilon'}^{[k]} \Delta \zeta^{[k]} \left(\sum_{i \in S} \alpha_i^{[k]} - C \right) \\
 &= \left(\sum_{i \in S} \left(\alpha_i^{[k]} - C - \frac{1}{2} (1 + \xi \hat{\beta}_{\varepsilon'}^{[k]}) \Delta \zeta^{[k]} \right) \right) \hat{\beta}_{\varepsilon'}^{[k]} \Delta \zeta^{[k]} + \mathcal{E}^{[k]}.
 \end{aligned}$$

换句话说, $\mathcal{E}^{[k+1]} - \mathcal{E}^{[k]} = \left(\sum_{i \in S} \left(\alpha_i^{[k]} - C - \frac{1}{2} (1 + \xi \hat{\beta}_{\varepsilon'}^{[k]}) \Delta \zeta^{[k]} \right) \right) \hat{\beta}_{\varepsilon'}^{[k]} \Delta \zeta^{[k]}$. 根据推论 8 和定理 5, 容易验证:

$$\left(\sum_{i \in S} \left(\alpha_i^{[k]} - C - \frac{1}{2} (1 + \xi \hat{\beta}_{\varepsilon'}^{[k]}) \Delta \zeta^{[k]} \right) \right) \Delta \zeta^{[k]} > 0.$$

另外, 根据定理 4 得到 $\hat{\beta}_{\varepsilon'}^{[k]} > 0$, 所以 $\mathcal{E}^{[k+1]} - \mathcal{E}^{[k]} > 0$, 定理得证. □

定理 10. 在绝缘增量调整中, 能量函数 \mathcal{E} 在有限步内将收敛到 $\min_{0 \leq \alpha_i \leq C/\gamma} \mathcal{E}$ 的最优解.

证明: 如果定义 $(\mathcal{E}^{[1]}, \mathcal{E}^{[2]}, \mathcal{E}^{[3]}, \dots)$ 为在绝缘增量调整中生成的能量函数序列, 那么, 为了证明该定理, 我们只需证明该序列 $(\mathcal{E}^{[1]}, \mathcal{E}^{[2]}, \mathcal{E}^{[3]}, \dots)$ 是一个有限序列并收敛到 $\min_{0 \leq \alpha_i \leq C/\gamma} \mathcal{E}$ 的最优解.

假设 $(\mathcal{E}^{[k]})$ 序列是一个无限序列, 根据定理 9, 有 $\Delta \mathcal{E}^{[k]} = \mathcal{E}^{[k+1]} - \mathcal{E}^{[k]} > 0$, 那么, $(\mathcal{E}^{[k]})$ 必然是一个无限的单调递增序列. 如果 \mathcal{E}^* 是 $\min_{0 \leq \alpha_i \leq C/\gamma} \mathcal{E}$ 的最优解, 那么容易验证这个序列有一个上限值 \mathcal{E}^* , 即 $\lim_{k \rightarrow \infty} \mathcal{E}^{[k]} \leq \mathcal{E}^*$. 所以, 序列 $(\mathcal{E}^{[k]})$ 是收敛的, 那么我们有 $\lim_{k \rightarrow \infty} \Delta \mathcal{E}^{[k]} = 0$. 这与 $\Delta \mathcal{E}^{[k]} > 0$ 矛盾. 因此, 序列 $(\mathcal{E}^{[k]})$ 是一个有限的序列, 表示成 $(\mathcal{E}^{[1]}, \mathcal{E}^{[2]}, \dots, \mathcal{E}^{[l]})$.

接下来, 假设 $(\mathcal{E}^{[1]}, \mathcal{E}^{[2]}, \dots, \mathcal{E}^{[l]})$ 不收敛于 $\min_{0 \leq \alpha_i \leq C/\gamma} \mathcal{E}$ 的最优解, 根据凸二次规划问题强对偶理论^[16], 第 l 步调整将不满足 $\min_{0 \leq \alpha_i \leq C/\gamma} W_l$ 的 KKT 条件, 而这与绝缘增量调整终止在第 l 步的事实不符.

因此, 序列 $(\mathcal{E}^{[1]}, \mathcal{E}^{[2]}, \dots, \mathcal{E}^{[l]})$ 必收敛到 $\min_{0 \leq \alpha_i \leq C/\gamma} \mathcal{E}$ 的最优解. □

4 实验

本文实现了 ν 解路径算法 (也就是 ν -SvrPath 算法), 并在一些标准数据集上说明了该算法能够尽可能地避免不可行的更新路径, 并在有限步内拟合出整个解路径. 此外, 我们还与文献[8]提出的解路径算法 (本文简称为 SvrPath) 进行了比较, 说明了 ν -SvrPath 算法因可以自动调整 ε -SVR 中的参数 ε , 所以能得到比 SvrPath 更小的最小回归均方误差.

所有实验在 3.2GHz, Pentium-4, 具有 1GB 的内存和 MATLAB 7.5 平台的机器上进行. 表 2 总结了在实验中使用的 3 个标准数据集的特性 (这些数据集可以从 <http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/> 获得). 因为算法的重点是非线性核, 所以实验中统一使用了高斯核 $K(x_i, y_j) = \exp\left(-\frac{\|x_i - y_j\|^2}{2\sigma^2}\right)$. 而参数 ε 的任何负值的改变并不会影响 ν 解路径**, 所以在整个绝缘增量调整过程中, 参数 ε 都固定在 -1. 如果在实验中遇到异常, 则算法使用 $\nu \leftarrow \nu + 0.0001$ 重新建立初始解. 在与 SvrPath 的比较实验中, SvrPath 算法的参数 ε 固定在 1, 模型选择的标准为 GCV 标准^[8].

** 容易验证, $\hat{\beta}_i, \hat{\beta}_b$ 和 $\hat{\beta}_{\varepsilon'}$ 有共同的分母 $\det(\hat{Q})$, 且 ε 仅与 $\det(\hat{Q})$ 相关, 所以 ε 能够决定 $\Delta \zeta^{\min}$ 的大小, 但与 S_S, S_R 和 S_E 结构的改变无关.

Table 2 Benchmark datasets used in the experiments

表 2 实验中使用的标准数据集

数据集	最大训练集大小	属性集	验证数据集大小	测试数据集大小
Triazines	186	60	60	60
MPG	392	7	80	80
Housing	506	13	120	120

4.1 ν -SvrPath算法验证实验

A) Triazines 数据

Triazines 数据集的任务是学习一个回归方程去预测描述性结构属性的活动.该数据集具有 60 个连续属性和 186 个实例.

为了说明 ν -SvrPath 算法可以在有限步内拟合整个解路径,实验在训练集大小分别为 40,80,120,160(如表 3 中“大小”列所示)、核参数 σ 选择分别为 7.071 1,2.236 1,0.707 1 的条件下,统计了绝缘增量调整的迭代次数(如表 3 中“步数”列所示).此外,实验也研究了冲突(见表 1 中“冲突”列)和异常(参考假设 1,见表 1 中“异常”列)发生的次数.从表中可以看出,异常发生的次数明显少于冲突发生的次数,并且发生异常的频率是较少的.所以说, ν -SvrPath 算法将尽可能地避免不可行的更新路径.图 2 展示了在数据集大小等于 40、核参数 σ 选择分别为 7.071 1,2.236 1,0.707 1 时边界支持向量集随着 ν 解路径发生变化的情形.

Table 3 Results of ν -SvrPath on Triazines dataset

表 3 ν -SvrPath 算法在 Triazines 数据集上的结果

大小	$\sigma=7.0711$			$\sigma=2.2361$			$\sigma=0.7071$		
	步数	冲突	异常	步数	冲突	异常	步数	冲突	异常
40	73.7	0.6	0.2	74.1	0.6	0.4	72.2	0.5	0.2
80	87.9	1.7	0.3	88.8	1.7	0.2	125.5	1.4	0.3
120	132.7	2.0	0.1	146.5	1.9	0.4	176.8	2.3	0.2
160	187.5	2.4	0.1	184.5	2.1	0.0	157.2	2.4	0.0

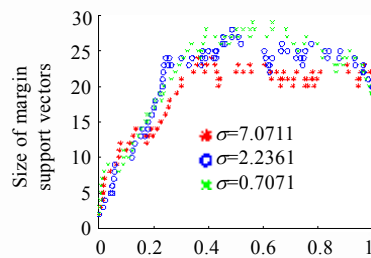


Fig.2 Size of margin support vectors along the ν -SvrPath on Triazines dataset

图 2 Triazines 数据集上边界支持向量集随着 ν 解路径的变化结果

B) MPG 数据

该数据集来自于卡内基梅隆大学的 StatLib 图书馆,该数据依据 3 个离散的属性和 5 个连续的属性去预测城市循环耗油量(英里/加仑).

为了说明 ν -SvrPath 能够尽可能地避免不可行更新路径,并在有限步内拟合出整个解路径,实验在训练数据大小分别为 80,160,240,320、核参数 σ 选择分别为 7.071 1,2.236 1,0.707 1 的条件下,统计了绝缘增量调整的迭代、冲突以及异常的次数(见表 4).从表 4 中可以发现,不仅异常的发生是极少的,而且它还能避免大量的如表 1 所示冲突的发生.除此之外,在数据大小是 80、核参数 σ 选择分别为 7.071 1,2.236 1,0.707 1 时,图 3 显示了边界支持向量集随着 ν 解路径发生变化的情形.

Table 4 Results of ν -SvrPath on MPG dataset

表 4 ν -SvrPath 算法在 MPG 数据集的结果

大小	$\sigma=7.0711$			$\sigma=2.2361$			$\sigma=0.7071$		
	步数	冲突	异常	步数	冲突	异常	步数	冲突	异常
80	53.1	20.5	1.1	42.7	18.2	1.1	36.0	16.1	1.2
160	79.2	18.0	1.0	132.7	41.7	1.0	81.02	25.7	1.1
240	134.5	18.5	0.8	189.6	29.2	1.0	132.1	14.5	0.9
320	36.1	1.7	0.7	17.9	1.8	0.8	6.2	1.1	0.7

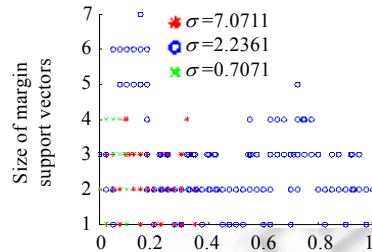


Fig.3 Size of margin support vectors along the ν -SvrPath on MPG dataset

图 3 MPG 数据集上边界支持向量集随着 ν 解路径的变化结果

C) Housing 数据

该数据集从卡内基梅隆大学的统计数字图书馆获得,它关注的是在波士顿郊区的房价,样本中有 12 个连续属性和 1 个二进制值的属性。

为了说明 ν -SvrPath 算法能够尽可能地避免不可行更新路径,并在有限步内拟合整个解路径,实验在训练数据集大小分别为 120,240,360,480、核参数 σ 选择分别为 7.071 1,2.236 1,0.707 1 的条件下,统计了绝缘增量调整的迭代次数、冲突和异常的次数(见表 5)。从表 5 中很明显地看出,异常的发生是极少的。在数据大小为 120、核参数 σ 选择分别为 7.071 1,2.236 1,0.707 1 时,边界支持向量集随着 ν 解路径的变化情形如图 4 所示。

Table 5 Results of ν -SvrPath on Housing dataset

表 5 ν -SvrPath 算法在 Housing 数据集的结果

大小	$\sigma=7.0711$			$\sigma=2.2361$			$\sigma=0.7071$		
	步数	冲突	异常	步数	冲突	异常	步数	冲突	异常
120	53.1	20.5	1.1	42.7	18.2	1.1	36.0	16.1	1.2
240	79.2	18.0	1.0	132.7	41.7	1.0	81.02	25.7	1.1
360	134.5	18.5	0.8	189.6	29.2	1.0	132.1	14.5	0.9
480	36.1	1.7	0.7	17.9	1.8	0.8	6.2	1.1	0.7

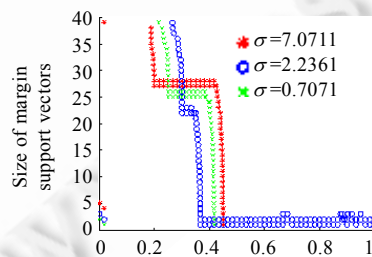


Fig.4 Size of margin support vectors along the ν -SvrPath on Housing dataset

图 4 Housing 数据集上边界支持向量集随着 ν 解路径的变化结果

4.2 与SvrPath算法比较实验

进一步地,为了说明 ν -SvrPath 算法的优越性,本文在验证数据集大小如表 1 所示、核参数选择分别为 7.071 1, 2.236 1, 0.707 1 的条件下,对 ν -SvrPath 算法与 SvrPath 算法给出的解路径根据 GCV 标准进行了模型选择;并在测试数据集大小如表 1 所示的基础上,比较了最小回归均方误差的结果(见表 6)。从表 6 中可以看出, ν -SvrPath 算法因可以自动调整参数 ε ,所以具有更小的最小回归均方误差。

Table 6 Minimum regression error between ν -SvrPath and SvrPath

表 6 ν -SvrPath 与 SvrPath 的最小回归误差结果

数据集	$\sigma=7.0711$		$\sigma=2.2361$		$\sigma=0.7071$	
	SvrPath	ν -SvrPath	SvrPath	ν -SvrPath	SvrPath	ν -SvrPath
Triazines	14.9	13.5	10.7	9.7	11.2	9.9
MPG	9.6	7.8	7.8	6.6	8.0	7.1
Housing	12.8	10.9	9.5	8.7	9.8	8.9

5 结 论

为了对 ν -SVR 设计一种有效的 ν 解路径算法,本文首先引入了一个 ν -SVR 的修改形式,在此基础上提出了 ν -SvrPath 算法。通过理论分析和实验,说明了 ν -SvrPath 算法能够尽可能地避免不可行的更新路径,并在有限步内拟合出全部解路径。

理论上讲, ν -SvrPath 算法也可以扩展到更广泛的一类核学习机上(含有多个等式约束),例如 ν 支持向量分类机^[12]和排列支持向量机^[20]等。

References:

- [1] Vapnik V. Statistical Learning Theory. New York: Wiley, 1998.
- [2] Keerthi S, Sindhvani V, Chapelle O. An efficient method for gradient-based adaptation of hyperparameters in SVM models. In: Schölkopf B, Platt J, Hoffman T, eds. Advances in Neural Information Processing Systems 19. Cambridge: MIT Press, 2007. 673–680.
- [3] Bousquet O, Elisseeff A. Stability and generalization. Journal of Machine Learning Research, 2002,2(Mar.):499–526. [doi: 10.1162/153244302760200704]
- [4] Anguita D, Ridella S, Riviccio F. K-Fold generalization capability assessment for support vector classifiers. In: Proc. of the Int'l Joint Conf. on Neural Networks. Montreal, 2005. 855–858. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1555964&isnumber=33090> [doi: 10.1109/IJCNN.2005.1555964]
- [5] Hastie T, Rosset S, Tibshirani R, Zhu J. The entire regularization path for the support vector machine. Journal of Machine Learning Research, 2004,5:1391–1415.
- [6] Zhu J, Rosset S, Hastie T, Tibshirani R. 1-Norm support vector machines. In: Advances in Neural Information Processing Systems. 2003. http://books.nips.cc/papers/files/nips16/NIPS2003_AA07.pdf
- [7] Gunter L, Zhu J. Efficient computation and model selection for the support vector regression. Neural Computation, 2007,19(6): 1633–1655. [doi:10.1162/neco.2007.19.6.1633]
- [8] Gunter L, Zhu J. Computing the solution path for the regularized support vector regression. In: Advances in Neural Information Processing Systems, Vol.18. Cambridge: MIT Press, 2005. 483–490. http://books.nips.cc/papers/files/nips18/NIPS2005_0360.pdf
- [9] Wang G, Yeung DY, Lochoovsky FH. A new solution path algorithm in support vector regression. IEEE Trans. on Neural Networks, 2008,19(10):1753–1767. [doi: 10.1109/TNN.2008.2002077]
- [10] Wang G, Yeung DY, Lochoovsky FH. A kernel path algorithm for support vector machines. In: Proc. of 24th Int'l Conf. on Machine Learning. 2007. 951–958. <http://www.machinelearning.org/proceedings/icml2007/papers/60.pdf> [doi: 10.1145/1273496.1273616]
- [11] Loosli G, Gasso G, Canu S. Regularization paths for ν -SVM and ν -SVR. In: Advances in Neural Networks Int'l Symp. on Neural Networks (ISNN). 2007. 486–496. <http://www.springerlink.com/content/11j52x77q0h15333>

- [12] Schölkopf B, Smola AJ, Williamson RC, Bartlett PL. New support vector algorithms. *Neural Computation*, 2000,12(5):1207–1245. [doi:10.1162/089976600300015565]
- [13] Steinwart I. On the optimal parameter choice for ν -support vector machines. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003,25(10):1274–1284. [doi: 10.1109/TPAMI.2003.1233901]
- [14] Gu B, Wang JD, Yu YC, Zheng GS, Wang LN. An effective regularization path for ν -support vector classification. In: *Proc. of the 2009 3rd Int'l Symp. on Intelligent Information Technology Application*, Vol.02. Washington, 2009. 90–93. <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05370506>
- [15] Chang CC, Lin CJ. Training ν -support vector regression: Theory and algorithms. *Neural Computation*, 2002,14(8):1959–1977. [doi: 10.1162/089976602760128081]
- [16] Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge: Cambridge University Press, 2004.
- [17] Osuna E, Freund R, Girosi F. Training support vector machines: An application to face detection. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 1997. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=609310 [doi: 10.1109/2fCVPR.1997.609310]
- [18] Gu B, Wang JD, Chen HY. On-Line off-line ranking support vector machine and analysis. In: *Proc. of the Int'l Joint Conf. on Neural Networks (IJCNN 2008)*. New York: IEEE Press, 2008. 1365–1370. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4633975 [doi: 10.1109/IJCNN.2008.4633975]
- [19] Householder AS. *The Theory of Matrices in Numerical Analysis*. New York: Dover Publications, Inc., 1974.
- [20] Shashua A, Levin A. Taxonomy of large margin principle algorithms for ordinal regression problems. In: *Advances in Neural Information Processing Systems*, Vol.15. MIT Press, 2002. <http://www.cs.huji.ac.il/~shashua/papers/k-planes-long.pdf>



顾彬(1983—),男,江苏南通人,博士,讲师,CCF 会员,主要研究领域为数据挖掘,机器学习.



王建东(1945—),男,教授,博士生导师,主要研究领域为数据挖掘,机器学习与知识工程,人工智能.