文章编号：1001-0920(2009)11-1657-06

两种求解非正定核 Laplace-SVR 的 SMO 算法

周晓剑，马义中
（南京理工大学 管理科学与工程系，南京 210094）

摘　要：提出 2 种用于求解非正定核 Laplace-SVR 的序列最小最优化(SMO)算法. 第 1 种算法仅针对 Laplace-SVR 而设计；第 2 种算法将 Laplace-SVR 作为所要解决问题的一种特殊情况，使算法更具通用性. 所提出的算法在保证收敛的前提下，使非正定 Laplace-SVR 能够达到比较理想的回归精度，具有一定的理论意义和实用价值.

关键词：非正定核；序列最小最优化算法；支持向量回归机
中图分类号：TP18　　　**文献标识码**：A

Two types of SMO algorithms for solving Laplace-SVR with non-positive kernels

ZHOU Xiao-jian, *MA Yi-zhong*
(Department of Management Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China. Correspondent: ZHOU Xiao-jian, E-mail: xjzhou2008@yahoo.com.cn)

Abstract: Two types of sequential minimal optimization(SMO) algorithms applied in solving Laplace-SVR with non-positive kernels are proposed. The first algorithm is only designed for Laplace-SVR, and the second one regarding Laplace-SVR as a special case is done for a general purpose. Because of the difficulty of solving SVR with non-positive kernels, the presented algorithms have a certain theoretical and practical significance.

Key words: Non-positive kernel; SMO algorithm; SVR

1 引　　言

当采用支持向量机(SVM)解决实际分类或回归问题时，棘手之处在于如何提高计算效率. 针对这一问题，文献 [1]提出了专用算法——选块算法. 但是，随着算法迭代次数的增多，计算速度变得十分缓慢. 针对选块算法存在的不足，文献[2]提出了适用于 SVC(Support vector classification)的“分解”算法，并由文献[3]将其扩展到支持向量回归机(SVR).“分解”算法有效地克服了选块算法的不足.

序列最小最优化(SMO)算法是分解算法中工作集的元素个数为 2 的一种特殊情形. 现有的 SMO 算法分为求解 SVC 的 SMO 算法和求解 SVR 的 SMO 算法. 求解 SVC 的 SMO 算法最早出现在文献[4]中，同年由 Smola 等[5]将其扩展到求解 SVR. 以上这两类求解大型支持向量机的 SMO 算法都假定支持向量机的核函数是正定的且满足 Mercer 条件. 但是，在某些特定的应用领域，所使用的核函数并不能保证为正定核函数，如神经网络中的 Sigmoid 核在某些条件下就是非正定的，这样现有的 SMO 算法的应用便受到了限制. 文献[6]提出的用于解决非正定核的 SVC 优化问题的方法，可以保持 SVC 的泛化性能. 然而有关非正定核 SVR 的求解方法，目前尚未见重要的研究成果公开发表. 本文将文献[6]的思想引入 Laplace-SVR 模型求解，以解决非正定核 Laplace-SVR 求解问题.

2 Laplace-SVR 的 SMO 算法的两种基本形式

2.1 Laplace-SVR 模型

给定数据集 $\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_l, y_l)\}$ 及核矩阵 $K_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$，若损失函数取 Laplace 函数

$$L(f(\boldsymbol{x}) - y) = |f(\boldsymbol{x}) - y|, \tag{1}$$

则构成 Laplace-SVR，其模型如下：

$$\min \Phi(\boldsymbol{w}, \xi) = \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + C\sum_{i=1}^{l}(\xi_i^- + \xi_i^+). \tag{2}$$

收稿日期：2008-12-21；**修回日期**：2009-03-14.
基金项目：国家自然科学基金项目(70672088)；国家自然科学基金国际交流项目(70711140386)；国家自然科学基金重点项目(70931002).

作者简介：周晓剑(1979—)，男，江西吉安人，博士生，从事人工神经网络、智能质量控制的研究；马义中(1964—)，男，河南泌阳人，教授，博士生导师，从事质量工程、质量管理等研究.

$$\text{s.t.}\ \ f(\boldsymbol{x}_i)-y_i\leqslant\xi_i^+;$$
$$y_i-f(\boldsymbol{x}_i)\leqslant\xi_i^-,i=1,\cdots,l;$$
$$\xi_i^-,\xi_i^+\geqslant 0.$$

该模型的 Lagrange 对偶模型为

$$\min_{\boldsymbol{\alpha}^{(*)}}\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i. \tag{3}$$
$$\text{s.t.}\ \ 0\leqslant\alpha_i,\ \alpha_i^*\leqslant C,\ i=1,\cdots,l;$$
$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0.$$

下面对上述对偶问题采用两种方式进行处理：

第 1 种处理方式如下：采取等量代换，令 $\bar{\boldsymbol{\alpha}}=\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_i^*$，则式(3) 可写成

$$\min_{\bar{\boldsymbol{\alpha}}}\frac{1}{2}\sum_{i,j=1}^{l}\bar{\alpha}_i\bar{\alpha}_jK(\boldsymbol{x}_i,\boldsymbol{x}_j)-\sum_{i=1}^{l}\bar{\alpha}_iy_i. \tag{4}$$
$$\text{s.t.}\ \ -C\leqslant\bar{\alpha}_i\leqslant C,\ i=1,\cdots,l;$$
$$\sum_{i=1}^{l}\bar{\alpha}_i=0.$$

再将上述模型写成如下矩阵的形式：

$$\min_{\bar{\boldsymbol{\alpha}}}F(\bar{\boldsymbol{\alpha}})=\frac{1}{2}\bar{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{K}\bar{\boldsymbol{\alpha}}-\boldsymbol{y}^{\mathrm{T}}\bar{\boldsymbol{\alpha}}. \tag{5}$$
$$\text{s.t.}\ \ -C\leqslant\bar{\alpha}_i\leqslant C,\ i=1,\cdots,l;$$
$$\sum_{i=1}^{l}\bar{\alpha}_i=0.$$

第 2 种处理方式如下：将式(3) 展开，则式(3) 可写成如下矩阵的形式：

$$\min_{\tilde{\boldsymbol{\alpha}}}F(\tilde{\boldsymbol{\alpha}})=\frac{1}{2}\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{Q}\tilde{\boldsymbol{\alpha}}+\tilde{\boldsymbol{y}}\tilde{\boldsymbol{\alpha}}. \tag{6}$$
$$\text{s.t.}\ \ 0\leqslant\tilde{\alpha}_t\leqslant C,\ t=1,\cdots,2l;$$
$$\boldsymbol{z}\tilde{\boldsymbol{\alpha}}=0.$$

其中

$$\tilde{\boldsymbol{y}}=[-\boldsymbol{y};\boldsymbol{y}]^{\mathrm{T}},\ \tilde{\boldsymbol{\alpha}}=\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^*\end{bmatrix},$$
$$\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{K} & -\boldsymbol{K}\\ -\boldsymbol{K} & \boldsymbol{K}\end{bmatrix},\ \boldsymbol{z}=[\boldsymbol{e}\quad -\boldsymbol{e}]. \tag{7}$$

$\boldsymbol{e}$ 是 $1\times l$ 的单位向量.

下面基于式(5) 和(6)，给出相应的两种 Laplace-SVR 的 SMO 模型.

2.2 第 1 种 Laplace-SVR 的 SMO 模型(简称模型 1)

通过简单推导，易将式(5) 写成如下模型：

$$\min_{\bar{\alpha}_i,\bar{\alpha}_j}\frac{1}{2}[\bar{a}_i\quad\bar{a}_j]\begin{bmatrix}K_{ii} & K_{ij}\\ K_{ji} & K_{jj}\end{bmatrix}\begin{bmatrix}\bar{a}_i\\ \bar{a}_j\end{bmatrix}-\left[\begin{bmatrix}\boldsymbol{y}_i\\ \boldsymbol{y}_j\end{bmatrix}+\boldsymbol{K}_{BN}\tilde{\boldsymbol{\alpha}}_N\right]^{\mathrm{T}}\begin{bmatrix}\bar{a}_i\\ \bar{a}_j\end{bmatrix}. \tag{8}$$
$$\text{s.t.}\ \ -C\leqslant\bar{a}_k\leqslant C,\ k\in\{i,j\};$$
$$\bar{\alpha}_i+\bar{\alpha}_j=-\sum_{k\neq i,j}^{l}\bar{\alpha}_k=\text{常数}.$$

2.3 第 2 种 Laplace-SVR 的 SMO 模型(简称模型 2)

同理，可将式(6) 写成如下模型：

$$\min_{\tilde{\alpha}_i,\tilde{\alpha}_j}\frac{1}{2}[\tilde{\alpha}_i\quad\tilde{\alpha}_j]\begin{bmatrix}Q_{ii} & Q_{ij}\\ Q_{ji} & Q_{jj}\end{bmatrix}\begin{bmatrix}\tilde{\alpha}_i\\ \tilde{\alpha}_j\end{bmatrix}+\left[\begin{bmatrix}\tilde{\boldsymbol{y}}_i\\ \tilde{\boldsymbol{y}}_j\end{bmatrix}+\boldsymbol{Q}_{BN}\tilde{\boldsymbol{\alpha}}_N\right]^{\mathrm{T}}\begin{bmatrix}\tilde{\alpha}_i\\ \tilde{\alpha}_j\end{bmatrix}. \tag{9}$$
$$\text{s.t.}\ \ 0\leqslant\tilde{\alpha}_k\leqslant C,\ k\in\{i,j\};$$
$$\tilde{\alpha}_i+\tilde{\alpha}_j=-\sum_{k\neq i,j}^{l}\tilde{\alpha}_k=\text{常数}.$$

3 工作集的选取及停机准则的设定

工作集的选取是 SMO 算法中最为重要的环节之一，工作集选取得好坏将直接影响算法的收敛速度. 对于模型 1，可采取直接方式进行选取，具体过程见 3.1 节. 但是，对于模型 2，则需要进行展开，若不这样做，而直接采用文献[7] 的方法，势必要同时考虑 4 个变量，即 α_i，α_j 及 α_i^*，α_j^*，从而不可避免地需要进行大量繁琐的判断，以确定可行域在哪个象限. 本文对文献[7] 中变量的判断方式进行改进，只需选择 2 个变量即可.

停机准则从不同的角度一般可为分 3 种，具体内容可参见文献[8]，本文则通过观察 KKT (Karush-Kuhn-Tucker) 条件的违背程度来判断是否可以停机. 下面分别给出模型 1 和模型 2 中工作集的选取步骤以及停机准则的推导过程.

3.1 模型 1 工作集的选取及停机准则

令

$$w(\bar{\boldsymbol{\alpha}})=\sum_i\bar{\alpha}_iK_i. \tag{10}$$

其中：$K_i=\phi(x_i)$，$K(x_i,x_j)=\phi(x_i)^{\mathrm{T}}\phi(x_j)$. 引入 Lagrange 函数

$$\bar{L}=\frac{1}{2}w(\bar{\boldsymbol{\alpha}})\cdot w(\bar{\boldsymbol{\alpha}})-\sum_{i=1}^{l}y_i\bar{\alpha}_i-\sum_{i=1}^{l}\delta_i(\bar{\alpha}_i+C)+\sum_{i=1}^{l}\mu_i(\bar{\alpha}_i-C)+\beta\sum_{i=1}^{l}\bar{\alpha}_i, \tag{11}$$

则

$$\frac{\partial\bar{L}}{\partial\bar{\alpha}_i}=w(\bar{\boldsymbol{\alpha}})K_i-y_i-\delta_i+\mu_i+\beta=0,\quad i=1,\cdots,l. \tag{12}$$

对于式(12)，可分为以下 3 种互不相交的情形加以考虑：

1) $\bar{\alpha}_i=-C$，则

$$u_i=0,\delta_i\geqslant 0,w(\bar{\boldsymbol{\alpha}})K_i-y_i+\beta\geqslant 0,$$

即

$$\beta \geqslant -w(\bar{\boldsymbol{\alpha}})K_i + y_i; \tag{13a}$$

2) $-C < \bar{\alpha}_i < C$，则

$$u_i = 0, \delta_i = 0, w(\bar{\boldsymbol{\alpha}})K_i - y_i + \beta = 0,$$

即

$$\beta = -w(\bar{\boldsymbol{\alpha}})K_i + y_i; \tag{13b}$$

3) $\bar{\alpha}_i = C$，则

$$u_i \geqslant 0, \delta_i = 0, w(\bar{\boldsymbol{\alpha}})K_i - y_i + \beta \leqslant 0,$$

即

$$\beta \leqslant -w(\bar{\boldsymbol{\alpha}})K_i + y_i. \tag{13c}$$

若定义

$$I_0 = \{i \mid -C < \bar{\alpha}_i < C\},$$
$$I_1 = \{i \mid \bar{\alpha}_i = -C\},$$
$$I_2 = \{i \mid \bar{\alpha}_i = C\},$$

并令

$$\bar{F}_i = -w(\bar{\boldsymbol{\alpha}})K_i + y_i,\ i \in I_1 \cup I_0, \tag{14}$$
$$\tilde{F}_i = -w(\bar{\boldsymbol{\alpha}})K_i + y_i,\ i \in I_2 \cup I_0. \tag{15}$$

则归纳上述 3 种情形，可得

$$\beta \leqslant \tilde{F}_i,\ \forall i \in I_0 \cup I_2;$$
$$\beta \geqslant \bar{F}_i,\ \forall i \in I_0 \cup I_1. \tag{16}$$

若再令

$$b_{\mathrm{up}} = \min\{\tilde{F}_i : i \in I_0 \cup I_2\},$$
$$b_{\mathrm{low}} = \max\{\bar{F}_i : i \in I_0 \cup I_1\}, \tag{17}$$

则可以得到工作集的选取方法及停机准则.

令

$$i = \arg\max_{i \in I_{\mathrm{up}}(\bar{\alpha}, C)} \bar{F}_i,\ j = \arg\min_{j \in I_{\mathrm{low}}(\bar{\alpha}, C)} \bar{F}_i. \tag{18}$$

其中

$$I_{\mathrm{up}}(\bar{\alpha}, C) = \{i \mid i \in I_0 \cup I_1\},$$
$$I_{\mathrm{low}}(\bar{\alpha}, C) = \{j \mid j \in I_0 \cup I_2\}.$$

则工作集取为 $B = \{i, j\}$，停机准则取为 $b_{\mathrm{low}} \leqslant b_{\mathrm{up}}$. 因为在数值解中不可能完全达到最优，因此可用下式代替：

$$b_{\mathrm{low}} \leqslant b_{\mathrm{up}} + 2\mathrm{tol},$$

其中 tol 表示误差所能容忍的范围.

3.2　模型 2 工作集的选取及停机准则

令

$$w(\tilde{\boldsymbol{\alpha}}) = \sum_i \tilde{\alpha}_i Q. \tag{19}$$

其中

$$Q_i = \phi(x_i),\ Q(x_i, x_j) = \phi(x_i)^{\mathrm{T}} \phi(x_j),$$
$$Q = [K \quad -K; -K \quad K].$$

引入 Lagrange 函数

$$\bar{L} = \frac{1}{2} w(\tilde{\boldsymbol{\alpha}}) \cdot w(\tilde{\boldsymbol{\alpha}}) + \sum_{i=1}^{2l} \tilde{y}_i \tilde{\alpha}_i - \sum_{i=1}^{2l} \delta_i \tilde{\alpha}_i + \sum_{i=1}^{2l} \mu_i (\tilde{\alpha}_i - C) + \beta \sum_{i=1}^{2l} \tilde{\alpha}_i z_i, \tag{20}$$

其中 z 在式(7) 中已有定义. 则

$$\frac{\partial \bar{L}}{\partial \tilde{\alpha}_i} = w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i - \delta_i + \mu_i + \beta z_i = 0,$$
$$i = 1, \cdots, 2l. \tag{21}$$

对于上式，可分为以下 6 种互不相交的情形加以考虑：

1) $i \leqslant l$, $\tilde{\alpha}_i = 0$，则

$$u_i = 0, \delta_i \geqslant 0, w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i + \beta \geqslant 0,$$

即

$$\beta \geqslant -(w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i); \tag{22a}$$

2) $i \leqslant l, 0 < \tilde{\alpha}_i < C$，则

$$u_i = 0, \delta_i = 0, w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i + \beta = 0,$$

即

$$\beta = -(w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i); \tag{22b}$$

3) $i \leqslant l, \tilde{\alpha}_i = C$，则

$$u_i \geqslant 0, \delta_i = 0, w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i + \beta \leqslant 0,$$

即

$$\beta \leqslant -(w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i); \tag{22c}$$

4) $l < i \leqslant 2l, \tilde{\alpha}_i = 0$，则

$$u_i = 0, \delta_i \geqslant 0, w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i - \beta \geqslant 0,$$

即

$$\beta \leqslant w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i; \tag{22d}$$

5) $l < i \leqslant 2l, 0 < \tilde{\alpha}_i < C$，则

$$u_i = 0, \delta_i = 0, w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i - \beta = 0,$$

即

$$\beta = w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i; \tag{22e}$$

6) $l < i \leqslant 2l, \tilde{\alpha}_i = C$，则

$$u_i \geqslant 0, \delta_i = 0, w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i - \beta \leqslant 0,$$

即

$$\beta \geqslant w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i. \tag{22f}$$

若定义

$$I_{0a} = \{i \mid i \leqslant l, 0 < \tilde{\alpha}_i < C\},$$
$$I_{0b} = \{i \mid l < i \leqslant 2l, 0 < \tilde{\alpha}_i < C\},$$
$$I_1 = \{i \mid i \leqslant l, \tilde{\alpha}_i = 0\},$$
$$I_2 = \{i \mid i \leqslant l, \tilde{\alpha}_i = C\},$$
$$I_3 = \{i \mid l < i \leqslant 2l, \tilde{\alpha}_i = 0\},$$
$$I_4 = \{i \mid l < i \leqslant 2l, \tilde{\alpha}_i = C\},$$

并定义 $I_0 = I_{0a} \cup I_{0b}$. 令

$$\bar{F}_i = \begin{cases} -(w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i),\ i \in I_1 \cup I_{0a}; \\ w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i,\ i \in I_4 \cup I_{0b}; \end{cases} \tag{23}$$

$$\tilde{F}_i = \begin{cases} -(w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i),\ i \in I_2 \cup I_{0a}; \\ w(\tilde{\boldsymbol{\alpha}})Q_i + \tilde{y}_i,\ i \in I_3 \cup I_{0b}. \end{cases} \tag{24}$$

则归纳上述 6 种情形可得

$$\beta \leqslant \tilde{F}_i,\ \forall i \in I_0 \cup I_2 \cup I_3;$$
$$\beta \geqslant \bar{F}_i,\ \forall i \in I_0 \cup I_1 \cup I_4. \tag{25}$$

若再令

$$b_{\rm up} = \min\{\widetilde{F}_i : i \in I_0 \cup I_2 \cup I_3\},$$
$$b_{\rm low} = \max\{\bar{F}_i : i \in I_0 \cup I_1 \cup I_4\}, \quad (26)$$

则同理可以得到工作集的选取方法及停机准则. 令

$$i = \arg\max_{i\in I_{\rm up}(\bar{\alpha},C)} \bar{F}_i,\ j = \arg\max_{j\in I_{\rm low}(\bar{\alpha},C)} \widetilde{F}_i. \quad (27)$$

其中

$$I_{\rm up}(\bar{\boldsymbol{\alpha}},C) = \{i \mid i \in I_0 \cup I_1 \cup I_4\},$$
$$I_{\rm low}(\bar{\boldsymbol{\alpha}},C) = \{j \mid j \in I_0 \cup I_2 \cup I_3\}.$$

则工作集取 $B=\{i,j\}$,停机准则为 $b_{\rm low} \leqslant b_{\rm up} + 2{\rm tol}$.

从上述分析可以看出,模型 1 与模型 2 的求解过程是殊途同归. 模型 1 通过等量代换对问题进行了有效的简化,但这种简化只针对 Laplace-SVR 这一特殊情形,求解过程并不具有通用性,即可移植性不强. 模型 2 的求解思想更具有普遍意义,因为当损失函数不像 Laplace 函数那么简单时,同样可以采用本文提出的求解思路. 这里并不想将简单问题复杂化,只是试图提出一种更具普遍意义的解决办法. 为此,下面主要针对模型 2 进行讨论,模型 1 可类似加以处理.

4 求解非正定核 Laplace-SVR 的 SMO 算法

当核函数为非正定时,式(6) 中的核矩阵 $\boldsymbol{Q}$ 中的分块矩阵 $\boldsymbol{K}$ 不能写成如下内积的形式:

$$K(x_i,x_j) \equiv \phi(x_i)^{\rm T}\phi(x_j), \quad (28)$$

所以得不到式(2) 的那种形式. 因此,从式(6) 中解出 $\tilde{\boldsymbol{\alpha}}$ 后,不能确定它就是式(2) 的解. 换言之,式(2) 与(6) 的对偶关系不再成立,从而无法得到有效的回归函数. 对此,考虑如下新的二次规划问题:

$$\min \Phi(\boldsymbol{w},\xi) = \frac{1}{2}\boldsymbol{\alpha}^{\rm T}\boldsymbol{Q}\boldsymbol{\alpha} + C\sum_{i=1}^{l}(\xi_i^- + \xi_i^+). \quad (29)$$
$$\text{s.t.}\ (\boldsymbol{Q}^*\boldsymbol{\alpha})_i - y_i \leqslant \xi_i^+;$$
$$y_i - (\boldsymbol{Q}^*\boldsymbol{\alpha})_i \leqslant \xi_i^-,\ i=1,\cdots,l;$$
$$\xi_i^-,\ \xi_i^+ \geqslant 0.$$

其中:$\boldsymbol{Q} = [\boldsymbol{Q}^*; -\boldsymbol{Q}^*]$,$\boldsymbol{Q}^* = [\boldsymbol{K} \quad -\boldsymbol{K}]$.

上式是将式(19) 代入(2) 得到的. 注意到式(29) 中的 $\boldsymbol{\alpha}$ 可能是负的. 这样做的目的是为了研究用式(6) 求出 $\tilde{\boldsymbol{\alpha}}$ 后,用 $\tilde{\boldsymbol{\alpha}}$ 来构建回归函数的误差是否会很大. 为此,有如下定理:

定理 1 如果 $\hat{\tilde{\boldsymbol{\alpha}}}$ 是式(6) 的稳定点,则它是式(29) 的可行点.

证明 假设 $\hat{\tilde{\boldsymbol{\alpha}}}$ 是式(6) 的稳定点,则它满足 KKT 条件,即存在 Lagrange 乘子向量 $\boldsymbol{\delta},\boldsymbol{\mu},\boldsymbol{\beta}$ 满足

$$\boldsymbol{Q}\hat{\tilde{\boldsymbol{\alpha}}} - \tilde{\boldsymbol{y}} - \boldsymbol{\delta} + \boldsymbol{\mu} - \boldsymbol{\beta} = \boldsymbol{0},$$
$$\boldsymbol{\delta} \geqslant \boldsymbol{0},\ \boldsymbol{\mu} \geqslant \boldsymbol{0},\ \boldsymbol{\delta}\tilde{\boldsymbol{\alpha}} = \boldsymbol{0},$$
$$\boldsymbol{\mu}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{C}) = \boldsymbol{0}, \quad (30)$$

其中 $\boldsymbol{\beta} = \beta[\boldsymbol{e} \quad -\boldsymbol{e}]^{\rm T}$.

由 $\boldsymbol{\mu} \geqslant \boldsymbol{0}$ 得 $\boldsymbol{Q}\hat{\tilde{\boldsymbol{\alpha}}} - \tilde{\boldsymbol{y}} \leqslant \boldsymbol{\delta} + \boldsymbol{\beta}$. 当 $i < l$ 时,有

$$(\boldsymbol{Q}\hat{\tilde{\boldsymbol{\alpha}}} - \tilde{\boldsymbol{y}})_i = (\boldsymbol{Q}^*\hat{\tilde{\boldsymbol{\alpha}}})_i - y_i;$$

当 $l \leqslant i \leqslant 2l$ 时,有

$$(\boldsymbol{Q}\hat{\tilde{\boldsymbol{\alpha}}} - \tilde{\boldsymbol{y}})_i = -(\boldsymbol{Q}^*\hat{\tilde{\boldsymbol{\alpha}}})_{i-l} + y_{i-l}.$$

因此,综合以上两种情况,可写成

$$(\boldsymbol{Q}\hat{\tilde{\boldsymbol{\alpha}}} - \tilde{\boldsymbol{y}}) = [(\boldsymbol{Q}^*\hat{\tilde{\boldsymbol{\alpha}}}) - \boldsymbol{y};\ -(\boldsymbol{Q}^*\hat{\tilde{\boldsymbol{\alpha}}}) + \boldsymbol{y}].$$

所以 $[\boldsymbol{\alpha};\ \xi_i^+;\ \xi_i^-] = [\hat{\tilde{\boldsymbol{\alpha}}};\ \boldsymbol{\delta}+\boldsymbol{\beta}]$ 即为式(29) 的可行解. □

从定理 1 可以看出,如果式(6) 的稳定点 $\hat{\tilde{\boldsymbol{\alpha}}}$ 不存在太多的非零元素,即使 $\boldsymbol{Q}$ 是非正定的,训练误差也不会太大. 基于以上分析,本文提出用于求解非正定核 Laplace-SVR 的 SMO 算法. 为便于问题的求解及收敛性的判定,作如下变换:

在式(9) 中,令

$$\tilde{\alpha}_i = \tilde{\alpha}_i^k + d_i,\ \tilde{\alpha}_j = \tilde{\alpha}_j^k + d_j, \quad (31)$$

则含两变量的子最优化问题可化为

$$\min_{\tilde{\alpha}_i,\tilde{\alpha}_j} \frac{1}{2}[d_i \quad d_j]\begin{bmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj}\end{bmatrix}\begin{bmatrix} d_i \\ d_j\end{bmatrix} +$$
$$[\nabla F(\tilde{\alpha}_i^k) \quad \nabla F(\tilde{\alpha}_j^k)]\begin{bmatrix} d_i \\ d_j\end{bmatrix}. \quad (32)$$
$$\text{s.t.}\ 0 \leqslant \tilde{\alpha}_i^k + d_i;\ \tilde{\alpha}_j^k + d_j \leqslant C;$$
$$m_i d_i + m_j d_j = 0;$$
$$m_k = \begin{cases} 1,\ 0 \leqslant k \leqslant l; \\ -1,\ l \leqslant k \leqslant 2l; \\ k \in \{i,j\}. \end{cases}$$

对于式(32),可分为以下 2 种情形加以讨论:

第 1 种情形:$i \leqslant l, j \leqslant l$ 或 $l < i \leqslant 2l, l < j \leqslant 2l$. 由 $(\tilde{\alpha}_i^k + d_i) + (\tilde{\alpha}_j^k + d_j) = \tilde{\alpha}_i^k + \tilde{\alpha}_j^k$ 得 $d_i = -d_j$,并令

$$L = \max(-\tilde{\alpha}_j^k, \tilde{\alpha}_i^k - C),$$
$$H = \min(\tilde{\alpha}_i^k, C - \tilde{\alpha}_j^k), \quad (33)$$

则式(32) 可化为

$$\min_{d_j} \frac{1}{2}[Q_{ii} - 2Q_{ij} + Q_{jj}]d_j^2 +$$
$$(-\nabla F(\tilde{\alpha}_i^k) + \nabla F(\tilde{\alpha}_j^k))d_j, \quad (34)$$
$$\text{s.t.}\ L \leqslant d_j \leqslant H.$$

第 2 种情形:$i \leqslant l, l < j \leqslant 2l$ 或 $l < i \leqslant 2l, j \leqslant l$. 由 $(\tilde{\alpha}_i^k + d_i) - (\tilde{\alpha}_j^k + d_j) = \tilde{\alpha}_i^k - \tilde{\alpha}_j^k$ 得 $d_i = d_j$. 同理,令

$$L = \max(-\tilde{\alpha}_i^k, -\tilde{\alpha}_j^k),$$
$$H = \min(C - \tilde{\alpha}_i^k,\ C - \tilde{\alpha}_j^k), \quad (35)$$

则式(32) 可化为

$$\min_{d_j} \frac{1}{2}(Q_{ii} + 2Q_{ij} + Q_{jj})d_j^2 +$$

$$(\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k))d_j, \tag{36}$$

s. t. $L\leqslant d_j\leqslant H$.

这样便将优化问题转化为在给定区间求抛物线的极值点问题，从而可以解析求解式(2).

下面给出 Laplace-SVR 的 SMO 算法.

输入：样本集$\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_l,y_l)\}$，惩罚系数 $C>0$，核函数 K，精度 tol >0 及 $\tilde{\boldsymbol{\alpha}}$；

输出：已优化的 $\hat{\tilde{\boldsymbol{\alpha}}}$.

Step1：设定精度 tol，$\tilde{\boldsymbol{\alpha}}^0=\boldsymbol{0}$，令 $k=0$，并初始化 b_{low} 和 b_{up}.

Step2：while($b_{\text{low}}>b_{\text{up}}+2\text{tol}$).

Step2.1：按式(27)中的准则选取工作集 $B=\{i,j\}$；

Step2.2：如果是第1种情形，则求解式(34). 若 $Q_{ij}-2Q_{ij}+Q_{ij}>0$，则

$$\hat{d}_j=\max\left(-\frac{(-\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k))}{Q_{ii}-2Q_{ij}+Q_{jj}},\ L\right);$$

否则，如果 $i\leqslant l,j\leqslant l$，则

$$\hat{d}_j=L; \tag{37}$$

否则，即如果 $l<i\leqslant 2l,l<j\leqslant 2l$，则

$$\hat{d}_j=H; \tag{38}$$

计算

$$\hat{d}_i=-\hat{d}_j;$$

否则，求解式(36). 若 $Q_{ii}+2Q_{ij}+Q_{jj}>0$，则

$$\hat{d}_j=\min\left\{\max\left(-\frac{\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k)}{Q_{ii}+2Q_{ij}+Q_{jj}},L\right),\ H\right\};$$

否则，若 $\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k)>0$，则

$$\hat{d}_j=L; \tag{39}$$

否则，若 $\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k)<0$，则

$$\hat{d}_j=H; \tag{40}$$

否则

$$\hat{d}_j=\arg\min_{j\in(L,H)}\{f(L),f(H)\}, \tag{41}$$

计算

$$\hat{d}_i=\hat{d}_j.$$

Step2.3：计算

$$\tilde{\alpha}_i^{k+1}=\tilde{\alpha}_i^k+\hat{d}_i,\ \tilde{\alpha}_j^{k+1}=\tilde{\alpha}_j^k+\hat{d}_j.$$

Step2.4：令 $k=k+1$，并计算 b_{low} 和 b_{up}.

endwhile.

Step3：取近似最优解 $\hat{\tilde{\boldsymbol{\alpha}}}=\tilde{\boldsymbol{\alpha}}^{k+1}$，算法结束.

从以上算法可以看出，当 $Q_{ii}-2Q_{ij}+Q_{jj}\geqslant 0$ 或 $Q_{ii}+2Q_{ij}+Q_{jj}\geqslant 0$ 时，式(34)和(36)的目标函数为凸的抛物线或直线. 有关正定核情况下的SMO算法的收敛性证明可以参见文献[9]，这里将着重讨论 $Q_{ii}-2Q_{ij}+Q_{jj}<0$ 或 $Q_{ii}+2Q_{ij}+Q_{jj}<0$ 时的情形，并给出收敛性证明.

定理2　存在 $\xi>0$，使得对于任意的自然数 k，有

$$F(\tilde{\boldsymbol{\alpha}}^{k+1})-F(\tilde{\boldsymbol{\alpha}}^k)\leqslant-\frac{\xi}{2}\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|^2. \tag{42}$$

证明　分2种情形加以证明.

1) 当 $i\leqslant l,j\leqslant l$ 或 $l<i\leqslant 2l,l<j\leqslant 2l$ 时，结合式(23)，(24)和(27)易得 $(-\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k))\hat{d}_j\leqslant 0$. 并且从 $\hat{d}_i=-\hat{d}_j$ 可得 $\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|^2=2\hat{d}_j^2$. 因此，由式(34)的目标函数可得

$$\begin{aligned}&F(\tilde{\boldsymbol{\alpha}}^{k+1})-F(\tilde{\boldsymbol{\alpha}}^k)\leqslant\\&\frac{1}{2}(Q_{ii}-2Q_{ij}+Q_{jj})\hat{d}_j^2=\\&\frac{1}{4}(Q_{ii}-2Q_{ij}+Q_{jj})\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|^2\leqslant\\&-\frac{\xi}{2}\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|^2,\end{aligned} \tag{43}$$

其中

$$\xi=-\max_{i,j}\left\{\frac{Q_{ii}-2Q_{ij}+Q_{jj}}{2}\mid Q_{ii}-2Q_{ij}+Q_{jj}<0\mid\right\}.$$

2) 当 $i\leqslant l,l<j\leqslant 2l$ 或 $l<i\leqslant 2l,j\leqslant l$ 时，从工作集的选择策略无法判定 $\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k)$ 的符号，但通过步长选取策略(见式(39)和(40))可得$(\nabla F(\tilde{\alpha}_i^k)+\nabla F(\tilde{\alpha}_j^k))\hat{d}_j\leqslant 0$. 并结合 $\hat{d}_i=\hat{d}_j$(此时 $\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|^2=2\hat{d}_j^2$)，类似于式(43)可得

$$\begin{aligned}&F(\tilde{\boldsymbol{\alpha}}^{k+1})-F(\tilde{\boldsymbol{\alpha}}^k)\leqslant\\&\frac{1}{2}(Q_{ii}+2Q_{ij}+Q_{jj})\hat{d}_j^2\leqslant\\&-\frac{\xi}{2}\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|^2,\end{aligned}$$

其中

$$\xi=-\max_{i,j}\left\{\frac{Q_{ii}+2Q_{ij}+Q_{jj}}{2}\mid Q_{ii}+2Q_{ij}+Q_{jj}<0\mid\right\}.$$

□

从以上定理可以看出，可以考虑用式(37)～(40)及(41)作为步长取值策略的理由. 这样做的目的是要得到 $\lim_{k\to\infty}\|\tilde{\boldsymbol{\alpha}}^{k+1}-\tilde{\boldsymbol{\alpha}}^k\|=0$，因而也就保证了算法的收敛性.

虽然这里提出的求解非正定核的SMO算法是收敛到稳定点而不是全局最优点，但当稳定点上的支持向量并不多时，或者说稳定点的非零元素并不多时，用非正定核训练时误差并不大. 以下的实验说明了这一点.

5　实验及其结果

5.1　实验设计

取函数 $y=\sin(\exp(x))(x\in[-4,2])$ 为基准信号，令间隔为 0.1 产生训练样本. 其中：非正定核取组合高斯核

$$\exp\left(\frac{-\|s-t\|^2}{\sigma_1}\right)+\exp\left(\frac{-\|s-t\|^2}{\sigma_2}\right)-$$
$$\exp\left(\frac{-\|s-t\|^2}{\sigma_3}\right)^{[10]};$$

标准差分别取为 $\sigma_1 = 0.8, \sigma_2 = 1.2, \sigma_3 = 4$;正定核取 Gauss 径向基核函数 $\exp(-\|s-t\|^2/2\sigma^2)$,其中 $\sigma = 1$.

5.2 实验结果与分析

本实验中参数 C 取为 $C = 10$,有关参数的选择方法可参见文献[11,12];预设精度为 tol = 0.05. 回归的结果如图 1 和图 2 所示. 图中:符号"*"为样本点,实线为回归线. 同时,从以下 3 个方面对实验的回归结果进行定量比较(结果见表 1):

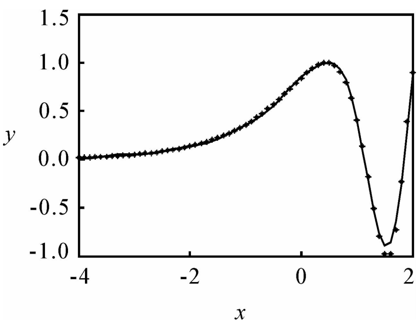

图 1 用非正定核对 $y = \sin(\exp(x))$ 回归的结果

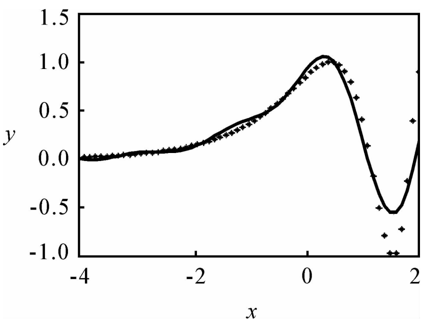

图 2 用正定核对 $y = \sin(\exp(x))$ 回归的结果

1) 最大绝对误差

MAE(Max absolute error) =

$$\max|y_i - \hat{y}|,\ i = 1,\cdots,n,$$

其中 n 为样本容量;

2) 平均绝对误差

AAE(Average absolute error) =

$$\sum_{i=1}^{n_{\mathrm{error}}}|y_i - \hat{y}|/n_{\mathrm{error}};$$

表 1 用正定核和非正定核回归的误差比较

误 差	用正定核回归	用非正定核回归
MAE	0.6170	0.1478
AAE	0.0784	0.0151
RMSE	0.1588	0.0334

3) 均方根误差

RMSE(Root mean square error) =

$$\sqrt{\sum_{i=1}^{n_{\mathrm{error}}}(y_i - \hat{y})^2/n_{\mathrm{error}}}.$$

从图1,图2及表1的误差比较分析可以看出,对于如 $\sin(\exp(x))$ 之类的比较难以回归的函数,采用非正定核进行回归比采用正定核能够达到更理想的回归精度. 这说明在某些情况下采用非正定核的必要性以及本文算法的有效性.

6 结 论

本文提出了 2 种用于求解非正定核 Laplace-SVR 的思路. 对于模型 1 并没有过多论述,因为模型 1 可视为模型 2 的一种特殊情形. 本文主要针对更为普遍的情形加以阐述,并设计了相应的 SMO 算法,这样做的目的是使求解的思路更具通用性. 对于非正定核,如 Sigmoid 核函数(包括本文所使用的核函数),因为只能保证解是局部最优的,而不一定是全局最优,因此如何克服它们的不足并发挥它们的优势,仍然是一个有意义的研究课题.

参考文献(References)

[1] Vapnik V N. Estimation of dependences based on empirical data [M]. Berlin: Springer-Verlag, 1982.

[2] Osuna E, Freund R, Girosi F. Improved training algorithm for support vector machines [C]. Neural Networks and Signal Processing. Amelia Island, FL, 1997: 276-285.

[3] Osuna E. Support vector machines: Training and applications [M]. Cambridge, MA: Massachusetts Institute of Technology, 1998.

[4] Platt J C. Fast training of support vector machines using sequential minimal optimization — Advances in kernel methods: Support vector machines [M]. Cambridge MA: MIT Press, 1998: 185-208.

[5] Smola A J, Schölkopf B. A tutorial on support vector regression [R]. London: Royal Holloway College, University of London, 1998.

[6] Lin H-T, Lin C-J. A study on sigmoid kernels for SVM and the training of non-psd kernels by SMO-type methods[R]. Taibei: Department of Computer Science and Information Engineering, National Taiwan University, 2003.

[7] Keerthi S, Shevade S, Bhattacharyya C, et al. Improvements to SMO algorithm for SVM regression [J]. IEEE Trans on Neural Networks, 2000, 11(5): 1188-1193.

(下转第 1672 页)