第36卷 第15期 计 算 机 工 程 2010年8月
Vol.36 No.15 Computer Engineering August 2010
·人工智能及识别技术· 文章编号：1000—3428(2010)15—0185—03 文献标识码：A 中图分类号：TP18

ε-不敏感的光滑支持向量回归机的收敛性

陈 勇，徐建敏
(东莞理工学院计算机学院，东莞 523808)

摘 要： ε-不敏感的光滑支持向量回归机采用快速的迭代方法进行求解，使回归性能及效率得到了提高，但并没有考虑该回归机的收敛性。针对该问题，采用集合论等方法，通过相关的理论推导，证明该光滑支持向量回归机对任意给定的惩罚参数都是全局收敛的，并给出它的收敛上界，为该光滑支持向量机提供了基本的理论支持。

关键词： ε-不敏感损失函数；回归；支持向量机；光滑；收敛

Convergence of ε-insensive Smooth Support Vector Regression

CHEN Yong, XU Jian-min
(Computer College, Dongguan University of Technology, Dongguan 523808)

【**Abstract**】The ε-insensive Smooth Support Vector Regression(ε-SSVR) is solved by a fast iterative algorithm, which improves the performance and efficiency of regression, but the problem of convergence remains still. In the present paper, ε-SSVR's global convergence for an arbitrary given penalty parameter is proved by the method of set theory, and the upper bound of the convergence is worked out, which provides the smooth support vector machine with a basic theoretical support.

【**Key words**】ε-insensive loss function; regression; support vector machine; smooth; convergence

1 概述

支持向量机是针对分类问题提出的一种机器学习方法，它在解决小样本、非线性及高维模式识别中表现出许多特有的优势，并可推广到回归问题[1-2]。很多研究者从不同角度对分类问题的支持向量机进行了改进，其中与收敛性有关且影响较大的算法有如下几种：(1)文献[3]提出了支持向量机的SMO(Sequential Minimal Optimization)算法，但只是用实验结果说明了该算法可以改进支持向量机的分类效果，而没有证明该算法的收敛性。鉴于收敛性是 SMO 算法的一个基本理论问题，文献[4]证明了该算法的收敛性，从理论上完善了该算法。(2)文献[5]提出了支持向量机的分块算法，也只是从实验的角度验证了该算法的有效性，没有证明分块算法的收敛性，文献[6]对该算法的收敛性进行了证明。(3)文献[7]提出了光滑支持向量机(Smooth Support Vector Machine, SSVM)，用实验结果表明 SSVM 的有效性，并证明了该模型的收敛性。(4)文献[8]提出了多项式光滑的支持向量机(Polynomial Smooth Support Vector Machine, PSSVM)，一方面用实验结果表明 PSSVM 的有效性，另一方面还证明了该模型的收敛性。(5)文献[9-10]提出了一类新的多项式光滑函数，接着提出了多项式光滑的支持向量机在分类问题中的一般模型 dPSSVM(dth-order Polynomial Smooth Support Vector Machine)，并证明了该一般模型的收敛性。文献[11]对光滑的支持向量机进行了综述，指出了当前存在的问题和今后可能的研究方向。

以上对支持向量机收敛性的研究有力地促进了支持向量机理论的发展，但这些研究都只是针对分类问题的，而对回归问题的支持向量机的收敛性研究却没有见到。

支持向量机用于回归问题的研究始于 1997 年，文献[12]将支持向量机用于分类问题的一套方法推广到回归问题，提出支持向量回归机(SVR)，但求解该支持向量回归机不涉及收敛性问题。文献[13]指出支持向量机用于回归问题后，仍然具有它在分类问题中的诸多优点。

文献[14]使用光滑技术，用 Sigmoid 积分函数的复合函数作为光滑函数，提出ε-不敏感的光滑支持向量回归机ε-SSVR(以下简称光滑支持向量回归机)，并从实验的角度验证了它的有效性。然而，文献[14]只对当惩罚参数为 1 时该光滑支持向量回归机的收敛性进行了简单说明，没有给出证明，更未对当惩罚参数为一般情形时的收敛性进行分析和证明，因此，ε-SSVR 的收敛性问题没有得到解决。其实对于这个光滑支持向量回归机，收敛性是一个前提条件和重要问题，因为如果不收敛于原支持向量回归机，则谈不上进一步研究其性能和效率，这种光滑支持向量回归机将毫无意义。经分析，该光滑支持向量回归机还存在如下 2 个尚未解决的重要问题：(1)当惩罚参数为一般情形时其收敛性如何；(2)如果收敛，其收敛上界是什么。针对这些理论问题，本文用集合论等方法给出当惩罚参数为一般情形时ε-SSVR 收敛的证明过程，并得出它的收敛上界。

2 ε-不敏感的光滑支持向量回归机

2.1 ε-不敏感损失函数

设数据集(训练集)为

$$S=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}\subseteq R^n\times R$$

目的是利用该数据集求回归函数 $y=f(\boldsymbol{x})$。一般用损失函数衡量回归函数 $y=f(\boldsymbol{x})$ 偏离 $y_i=f(\boldsymbol{x}_i)$ 的程度，若记损失

基金项目： 广东省自然科学基金资助项目(9151170003000017)
作者简介： 陈 勇(1964－)，男，工程师、硕士，主研方向：人工智能，数据挖掘；徐建敏，讲师、硕士
收稿日期： 2010-03-15 **E-mail：** chenyongdg@126.com

函数为 $c(\boldsymbol{x},y,f)$，则回归问题变为经验风险的极小化问题：

$$\min_{f\in F}\sum_{i=1}^{m}c(\boldsymbol{x}_i,y_i,f(\boldsymbol{x}_i))$$

由此解得回归函数 $y=f(\boldsymbol{x})$。若 $c(\boldsymbol{x},y,f)$ 采用平方损失函数，则为最小二乘法。

解这个极小化问题首先遇到的问题就是如何选择回归函数的类别，即函数类集合为 F。因为一方面不能把 F 限定为非常狭小的函数类，如线性函数，否则对非线性情形会导致回归的误差太大；另一方面也不能把 F 取得过大，否则会使回归函数毫无意义。例如，当 F 是所有实函数的集合时，会得到：

$$f(\boldsymbol{x})=\begin{cases}y_i & \boldsymbol{x}=\boldsymbol{x}_i\\ 0 & \boldsymbol{x}\neq\boldsymbol{x}_i\end{cases}\qquad i=1,2,\cdots,m$$

显然这个回归函数符合给定的数据集，却很不合理。

为解决这个问题，Vapnik 等人提出采用ε-不敏感损失函数。因为ε-不敏感损失函数存在一个不为目标函数提供任何失值的不敏感区域，即ε-带，该带内样本点的信息不会出现在回归函数中，使回归函数显得稀疏和简洁[1]。

回归函数 $f(\boldsymbol{x})$ 与观测值 y 的ε-不敏感损失函数为

$$|y-f(\boldsymbol{x})|_{\varepsilon}=\max\{0,\ |y-f(\boldsymbol{x})|-\varepsilon\}$$

其中，ε 为某一给定的小正数。对于线性情形，$f(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b$，可用图 1 表示。图中 2 条虚线构成的带子称为 ε-带，只有训练点位于这个点之外，才有损失，该损失为 $|y-f(\boldsymbol{x})|-\varepsilon$。

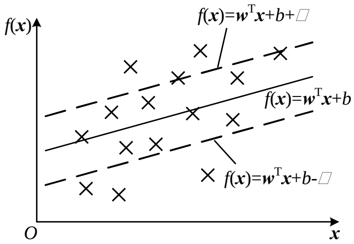

图 1　ε-不敏感的线性回归函数

2.2　ε-不敏感的光滑支持向量回归机的形式

设 $\boldsymbol{x}_i$ 是一个 n 维向量，每个 $\boldsymbol{x}_i$ 对应的观测值为 y_i。令 $\boldsymbol{A}=[\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_m]$，则数据集 $S=\{(A_i,y_i)\mid A_i\in R^n, y_i\in R\}$，$i$=1, 2, $\cdots$, m，其中，A_i 是矩阵 A 的第 i 个向量。回归的目的是利用给定的数据集 S，采用ε-不敏感损失函数，训练出一个回归函数 $f(x)$。对于线性回归的情形，回归函数 $f(x)=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b$，其中，$\boldsymbol{w}\in R^n$ 是一个待定向量；b 是一个待定常量[1]。

文献[12]采用ε-不敏感损失函数，将支持向量机在分类问题上的一套方法推广到回归问题，把回归问题表示为一个有约束条件的最优化问题，提出支持向量回归机。文献[14]在此基础上，把上述回归问题表示为无约束条件的最优化问题：

$$\min_{(\boldsymbol{w},b)\in R^{n+1}}\frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+b^2)+\frac{C}{2}\sum_{i=1}^{m}|\boldsymbol{A}_i\boldsymbol{w}+b-y_i|_{\varepsilon}^2 \tag{1}$$

其中，C 为大于 0 的惩罚参数，此式即为无约束条件的支持向量回归机。

显然，式(1)中目标函数的光滑性等价于函数 $|x|_{\varepsilon}^2$ 的光滑性。由文献[14]知：

$$|x|_{\varepsilon}^2=(x-\varepsilon)_+^2+(-x-\varepsilon)_+^2 \tag{2}$$

因此，式(1)中目标函数的光滑性就转换成函数 x_+^2 的光滑性问题了。由文献[13]知 x_+^2 只具有一阶光滑性，因此无约束的支持向量回归机也只具有一阶光滑性，导致该支持向量回归机不能用快速的 Newton-Armijo 算法进行求解[1]，使得回归效果受到影响。

为适用快速的迭代方法求解，文献[14]提出 p_{ε}^2-函数：

$$p_{\varepsilon}^2(x,\alpha)=(p(x-\varepsilon,\alpha))^2+(p(-x-\varepsilon,\alpha))^2 \tag{3}$$

其中，Sigmoid 函数的积分函数 $p(x,\alpha)=x+\dfrac{1}{\alpha}\ln(1+\mathrm{e}^{-\alpha x})$，$\alpha$ 为光滑因子，$\alpha>0$，将这个 p_{ε}^2-函数作为光滑函数，逼近式(1)中的 $|\bullet|_{\varepsilon}^2$，提出$\varepsilon$-不敏感的光滑支持向量回归机$\varepsilon$-SSVR：

$$\begin{aligned}\min_{(\boldsymbol{w},b)\in R^{n+1}}\phi_{\varepsilon,\alpha}(\boldsymbol{w},b):=&\min_{(\boldsymbol{w},b)\in R^{n+1}}\frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+b^2)+\\&\frac{C}{2}\sum_{i=1}^{m}p_{\varepsilon}^2(\boldsymbol{A}_i\boldsymbol{w}+b-y_i,\alpha)=\\&\min_{(\boldsymbol{w},b)\in R^{n+1}}\frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+b^2)+\\&\frac{C}{2}\hat{\boldsymbol{1}}^{\mathrm{T}}p_{\varepsilon}^2(\boldsymbol{A}\boldsymbol{w}+\hat{\boldsymbol{1}}b-y,\alpha)\end{aligned} \tag{4}$$

其中，C 为某个大于 0 的惩罚参数；$\hat{\boldsymbol{1}}$表示元素全为 1 的列向量。

由于 $p(x,\alpha)$ 具有无穷阶光滑性[1]，由式(3)易知 p_{ε}^2-函数也具有无穷阶光滑性，因此光滑支持向量回归机ε-SSVR 也具有无穷阶光滑性，可以用快速的 Newton-Armijo 算法进行求解。文献[14]用 Newton-Armijo 算法对ε-SSVR 做了数值实验，验证了ε-SSVR 对回归问题的有效性，但只对ε-SSVR 当惩罚参数为 1 时的收敛性进行了简单说明，没有给出证明过程，更未对当惩罚参数为一般情形时的收敛性进行说明和证明，因此ε-SSVR 的收敛性问题依然存在。该收敛性问题包括：当惩罚参数为一般情形时ε-SSVR 的是否收敛，如果收敛，其收敛上界是什么。本文参照文献[7-8]的基本思路，用集合论等方法给出当惩罚参数为一般情形时ε-SSVR 收敛的证明过程，并得出它的收敛上界。

3　收敛性证明

3.1　相关引理

将式(1)的目标函数写成：

$$h_{\varepsilon}(\boldsymbol{x})=\frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+b^2)+\frac{C}{2}\sum_{i=1}^{m}|\boldsymbol{A}_i\boldsymbol{w}+b-y_i|_{\varepsilon}^2 \tag{5}$$

将式(4)的目标函数写成：

$$g_{\varepsilon}(\boldsymbol{x},\alpha)=\frac{1}{2}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+b^2)+\frac{C}{2}\hat{\boldsymbol{1}}^{\mathrm{T}}p_{\varepsilon}^2(\boldsymbol{A}\boldsymbol{w}+\hat{\boldsymbol{1}}b-y,\alpha) \tag{6}$$

为便于证明，把式(5)和式(6)中的 $\boldsymbol{w}$ 和 b 合成一个向量 $\boldsymbol{x}$，即 $\boldsymbol{x}=[\boldsymbol{w}\ \ b]^{\mathrm{T}}$。显然，$\boldsymbol{x}\in R^{n+1}$，记 $\tilde{\boldsymbol{A}}=[\boldsymbol{A}\vdots\hat{\boldsymbol{1}}]$，显然 $\tilde{\boldsymbol{A}}\in R^{m\times(n+1)}$。所以，式(5)和式(6)中的目标函数可分别写成：

$$h_{\varepsilon}(\boldsymbol{x})=\frac{1}{2}C\sum_{i=1}^{m}|\tilde{\boldsymbol{A}}_i\boldsymbol{x}-y_i|_{\varepsilon}^2+\frac{1}{2}\|\boldsymbol{x}\|_2^2 \tag{7}$$

$$g_{\varepsilon}(\boldsymbol{x},\alpha)=\frac{1}{2}C\sum_{i=1}^{m}p_{\varepsilon}^2(\tilde{\boldsymbol{A}}_i\boldsymbol{x}-y_i,\alpha)+\frac{1}{2}\|\boldsymbol{x}\|_2^2 \tag{8}$$

文献[14]中只说明了当惩罚参数为 1 时 $g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解收敛于 $h_{\varepsilon}(\boldsymbol{x})$ 的解，未对当惩罚参数为一般情形时 $g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解是否收敛于 $h_{\varepsilon}(\boldsymbol{x})$ 的解进行分析和证明，本文对这个问题进行研究。为此先研究 3 个相关的引理：

引理 1　$|x|_{\varepsilon}^2$ 和 $p_{\varepsilon}^2(x,\alpha)$ 分别由式(2)和式(3)给出，则 $p_{\varepsilon}^2(x,\alpha)>|x|_{\varepsilon}^2$。

证明：由文献[8]知，$p(x,\alpha)>x_+$，由式(2)和式(3)知：

$$|x|_{\varepsilon}^{2}-p_{\varepsilon}^{2}(x,\alpha)=[(x-\varepsilon)_{+}^{2}-p(x-\varepsilon,\alpha)^{2}]+[(-x-\varepsilon)_{+}^{2}-p(-x-\varepsilon,\alpha)^{2}]<0$$

所以，$p_{\varepsilon}^{2}(x,\alpha)>|x|_{\varepsilon}^{2}$。证毕。

引理 2 目标函数 $h_{\varepsilon}(\boldsymbol{x})$ 和 $g_{\varepsilon}(\boldsymbol{x},\alpha)$ 分别定义如式(7)和式(8)，则优化问题 $\min\limits_{x} h_{\varepsilon}(\boldsymbol{x})$ 和 $\min\limits_{x} g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解均存在且唯一。

证明：以下分存在性和唯一性 2 步进行证明：

(1)存在性

对任意的 $\mu>0$，若 $g_{\varepsilon}(\boldsymbol{x},\alpha)\leqslant\mu$，由引理 1：$p_{\varepsilon}{}^{2}(\boldsymbol{x},\alpha)\geqslant|\boldsymbol{x}|_{\varepsilon}{}^{2}$，则有：

$$\mu\geqslant g_{\varepsilon}(\boldsymbol{x},\alpha)=\frac{C}{2}\sum_{i=1}^{m}p_{\varepsilon}{}^{2}(\tilde{\boldsymbol{A}}_{i}\boldsymbol{x}-y_{i},\alpha)+\frac{1}{2}\|\boldsymbol{x}\|_{2}^{2}\geqslant$$
$$\frac{C}{2}\sum_{i=1}^{m}|\tilde{\boldsymbol{A}}_{i}\boldsymbol{x}-\boldsymbol{y}_{i}|_{\varepsilon}^{2}+\frac{1}{2}\|\boldsymbol{x}\|_{2}^{2}=h_{\varepsilon}(\boldsymbol{x})$$

即 $h_{\varepsilon}(\boldsymbol{x})\leqslant\mu$，所以，水平集 $L_{\mu}(g_{\varepsilon}(\boldsymbol{x},\alpha))$ 和 $L_{\mu}(h_{\varepsilon}(\boldsymbol{x}))$ 的关系为：$L_{\mu}(g_{\varepsilon}(\boldsymbol{x},\alpha))\subseteq L_{\mu}(h_{\varepsilon}(\boldsymbol{x}))$。由 $g_{\varepsilon}(\boldsymbol{x},\alpha)\leqslant\mu$ 和 $h_{\varepsilon}(\boldsymbol{x})\leqslant\mu$ 知：

$$2\mu\geqslant g_{\varepsilon}(\boldsymbol{x},\alpha)+h_{\varepsilon}(\boldsymbol{x})$$

结合式(7)和式(8)，得：

$$2\mu\geqslant\frac{C}{2}\sum_{i=1}^{m}p_{\varepsilon}{}^{2}(\tilde{\boldsymbol{A}}_{i}\boldsymbol{x}-y_{i},\alpha)+\frac{C}{2}\sum_{i=1}^{m}|\tilde{\boldsymbol{A}}_{i}\boldsymbol{x}-y_{i}|_{\varepsilon}^{2}+\|\boldsymbol{x}\|_{2}^{2}>\|\boldsymbol{x}\|_{2}^{2}$$

因此，

$$L_{\mu}(g_{\varepsilon}(x,\alpha))\subseteq L_{\mu}(h_{\varepsilon}(x))\subseteq\{x\,\big|\,\|x\|_{2}^{2}\leqslant2\mu\}$$

表明 $L_{\mu}(g_{\varepsilon}(\boldsymbol{x},\alpha))$ 和 $L_{\mu}(h_{\varepsilon}(\boldsymbol{x}))$ 都是 R^{n} 中的紧子集，所以，$\min\limits_{x} h_{\varepsilon}(\boldsymbol{x})$ 和 $\min\limits_{x} g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解都是存在的。

(2)因 $\|\bullet\|_{2}^{2}$ 为严格凸函数，而 $|\bullet|_{\varepsilon}^{2}$ 为凸函数，由式(7)和式(8)知，$h_{\varepsilon}(\boldsymbol{x})$ 和 $g_{\varepsilon}(\boldsymbol{x},\alpha)$ 皆为严格凸函数，故它们的解是唯一的。

证毕。

引理 3 $h_{\varepsilon}(\boldsymbol{x})$ 和 $g_{\varepsilon}(\boldsymbol{x},\alpha)$ 分别定义如式(7)和式(8)，记优化问题 $\min\limits_{x} h_{\varepsilon}(\boldsymbol{x})$ 和 $\min\limits_{x} g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解分别为 $\bar{\boldsymbol{x}}$ 和 $\bar{\boldsymbol{x}}_{\alpha}$，则

(1) $h_{\varepsilon}(\boldsymbol{x})-h_{\varepsilon}(\bar{\boldsymbol{x}})\geqslant\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{2}^{2}$

(2) $g_{\varepsilon}(\boldsymbol{x},\alpha)-g_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha},\alpha)\geqslant\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}_{\alpha}\|_{2}^{2}$

证明：

(1)在引理 2 的证明过程中知，$h_{\varepsilon}(\boldsymbol{x})$ 是严格凸函数，因此有[1]：

$$h_{\varepsilon}(\boldsymbol{x})-h_{\varepsilon}(\bar{\boldsymbol{x}})\geqslant h_{\varepsilon}^{'}(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})+\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{2}^{2}\tag{9}$$

代入一阶最优性条件 $h_{\varepsilon}^{'}(\bar{\boldsymbol{x}})=0$，得：

$$h_{\varepsilon}(\boldsymbol{x})-h_{\varepsilon}(\bar{\boldsymbol{x}})\geqslant\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{2}^{2}\tag{10}$$

(2)证明过程与(1)类似，略。

证毕。

3.2 收敛性定理及其证明

有了以上 3 个引理，就可以证明下面的收敛性定理：

定理 $h_{\varepsilon}(\boldsymbol{x})$ 和 $g_{\varepsilon}(\boldsymbol{x},\alpha)$ 分别定义如式(7)和式(8)，则对任意惩罚参数 $C>0$，优化问题 $\min\limits_{x} g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解全局收敛于优化问题 $\min\limits_{x} h_{\varepsilon}(\boldsymbol{x})$ 的解。

证明：由引理 2 知，优化问题 $\min\limits_{x} h_{\varepsilon}(\boldsymbol{x})$ 和 $\min\limits_{x} g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解是存在且唯一的。

下面来证明 $\min\limits_{x} g_{\varepsilon}(\boldsymbol{x},\alpha)$ 的解 $\bar{\boldsymbol{x}}_{\alpha}$ 收敛于 $\min\limits_{x} h_{\varepsilon}(\boldsymbol{x})$ 的解 $\bar{\boldsymbol{x}}$，即证明 $\lim\limits_{\alpha\to+\infty}\bar{\boldsymbol{x}}_{\alpha}=\bar{\boldsymbol{x}}$。

由引理 3 知：

$$h_{\varepsilon}(\boldsymbol{x})-h_{\varepsilon}(\bar{\boldsymbol{x}})\geqslant\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|_{2}^{2}$$

令 $\boldsymbol{x}=\bar{\boldsymbol{x}}_{\alpha}$，得：

$$h(\bar{\boldsymbol{x}}_{\alpha})-h(\bar{\boldsymbol{x}})\geqslant\frac{1}{2}\|\bar{\boldsymbol{x}}_{\alpha}-\bar{\boldsymbol{x}}\|_{2}^{2}\tag{11}$$

由引理 3 知：

$$g_{\varepsilon}(\boldsymbol{x},\alpha)-g_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha},\alpha)\geqslant\frac{1}{2}\|\boldsymbol{x}-\bar{\boldsymbol{x}}_{\alpha}\|_{2}^{2}$$

令 $\boldsymbol{x}=\bar{\boldsymbol{x}}$，得：

$$g_{\varepsilon}(\bar{\boldsymbol{x}},\alpha)-g_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha},\alpha)\geqslant\frac{1}{2}\|\bar{\boldsymbol{x}}-\bar{\boldsymbol{x}}_{\alpha}\|_{2}^{2}\tag{12}$$

式(11)加式(12)，得：

$$\|\bar{\boldsymbol{x}}_{\alpha}-\bar{\boldsymbol{x}}\|_{2}^{2}\leqslant(g_{\varepsilon}(\bar{\boldsymbol{x}},\alpha)-h_{\varepsilon}(\bar{\boldsymbol{x}}))-(g_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha},\alpha)-h_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha}))\tag{13}$$

式(8)减式(7)，并用 $\boldsymbol{x}=\bar{\boldsymbol{x}}_{\alpha}$ 代入，得：

$$g_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha},\alpha)-h_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha})=\frac{C}{2}\sum_{i=1}^{m}(p_{\varepsilon}{}^{2}(\tilde{\boldsymbol{A}}_{i}\bar{\boldsymbol{x}}_{\alpha}-b_{i},\alpha)-|\tilde{\boldsymbol{A}}_{i}\bar{\boldsymbol{x}}_{\alpha}-b_{i}|_{\varepsilon}^{2})$$

上式结合引理 1，$p_{\varepsilon}{}^{2}(\boldsymbol{x},\alpha)\geqslant|\boldsymbol{x}|_{\varepsilon}{}^{2}$，可得 $g_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha},\alpha)-h_{\varepsilon}(\bar{\boldsymbol{x}}_{\alpha})\geqslant0$，因此由式(13)，可得：

$$\|\bar{\boldsymbol{x}}_{\alpha}-\bar{\boldsymbol{x}}\|_{2}^{2}\leqslant g_{\varepsilon}(\bar{\boldsymbol{x}},\alpha)-h_{\varepsilon}(\bar{\boldsymbol{x}})=\frac{C}{2}\sum_{i=1}^{m}(p_{\varepsilon}{}^{2}(\tilde{\boldsymbol{A}}_{i}\bar{\boldsymbol{x}}-y_{i},\alpha)-|\tilde{\boldsymbol{A}}_{i}\bar{\boldsymbol{x}}-y_{i}|_{\varepsilon}^{2})\tag{14}$$

其中，$\tilde{A}_{i}$ 是 $\tilde{A}$ 的第 i 个行向量；y_{i} 是 y 的第 i 个分量。文献[14]已证明：$p_{\varepsilon}^{2}(x,\alpha)-|x|_{\varepsilon}^{2}\leqslant2(\frac{\ln2}{\alpha})^{2}+\frac{2\rho}{\alpha}\ln2$，其中，$\rho$ 为一给定的正数，将上式代入式(14)，得：

$$\|\bar{\boldsymbol{x}}_{\alpha}-\bar{\boldsymbol{x}}\|_{2}^{2}\leqslant mC[(\frac{\ln2}{\alpha})^{2}+\frac{\rho}{\alpha}\ln2]\tag{15}$$

其中，m 为训练样本数目；C 为某个任意大于 0 的惩罚参数；α 为光滑因子。显然 $\lim\limits_{\alpha\to+\infty}\bar{\boldsymbol{x}}_{\alpha}=\bar{\boldsymbol{x}}$。证毕。

注意到式(4)对应于 $g_{\varepsilon}(\boldsymbol{x},\alpha)$，式(1)对应于 $h_{\varepsilon}(\boldsymbol{x})$，由定理可知，当 α 趋于正无穷时，对任意大于 0 的惩罚参数，光滑支持向量回归机的解都全局收敛于原支持向量回归机的解。需要指出的是，Lee 等人只简单说明了光滑支持向量回归机当惩罚参数为 1 时是收敛的，并未进行证明[14]。显然定理的证明也包含了惩罚参数为 1 时的情形，换言之，本文同时证明了当惩罚参数为 1 时，光滑支持向量回归机也是全局收敛的。

由定理可得到关于光滑支持向量回归机的收敛上界：

推论 光滑支持向量回归机的收敛上界等于 $0.692\ 7mC/\alpha^{2}$。

证明：定理证明了光滑支持向量回归机收敛，令 $\rho=\frac{1}{\alpha}$ [8]，得 $(\frac{\ln2}{\alpha})^{2}+\frac{\rho}{\alpha}\ln2=0.692\ 7/\alpha^{2}$

代入式(15)可得：$\|\bar{\boldsymbol{x}}_{\alpha}-\bar{\boldsymbol{x}}\|_{2}^{2}\leqslant0.692\ 7mC/\alpha^{2}$

即收敛上界等于 $0.692\ 7mC/\alpha^{2}$。证毕。

4 结束语

ε-不敏感的光滑支持向量回归机只用数值实验验证了它的有效性，未对当惩罚参数为一般情形时的收敛性进行分析和证明。鉴于收敛性是光滑支持向量回归机的一个重要理论问题，本文用集合论等方法证明了这个问题，并得出了它的收敛上界。注意到在收敛性定理及其证明过程中，没有对惩罚参数 C 作任何限制，即，只要惩罚参数大于 0，这个光滑支持向量回归机都是全局收敛的，从而突破了文献[14]中设定惩罚参数为 1 的限制，解决了概述中提出的 2 个问题，得

（下转第 190 页）