

求解非线性回归问题的 Newton 算法

韩 敏 王亚楠

(大连理工大学电子信息与电气工程学部 辽宁 大连 116024)

摘 要 针对大规模非线性回归问题,提出基于静态储备池的 Newton 算法.利用储备池搭建高维特征空间,将原始问题转化成与储备池维数相关的线性支持向量回归问题,并应用 Newton 算法求解.鲁棒损失函数的应用可抑制异常点对预测结果的干扰.通过与 SVR(Support Vector Regression)及储备池 Tikhonov 正则化方法比较,验证了所提方法的快速性、较高的预测精度和较好的鲁棒性.

关键词 支持向量回归;静态储备池;Newton 算法;鲁棒性

中图法分类号 TP18 **DOI 号**: 10.3724/SP.J.1016.2010.00841

A Newton Algorithm for Nonlinear Regression

HAN Min WANG Ya-Nan

(Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning 116024)

Abstract A Newton algorithm is adopted in static reservoir for large scale nonlinear regression in the paper. Based on high-dimension ‘reservoir’ state space which translates the nonlinear regression to linear support vector regression (SVR), the Newton optimization is investigated. Meantime, the robust loss functions are adopted to restrain the interference of outliers. Comparisons with SVR(Support Vector Regression) and ‘reservoir’ Tikhonov regularization method in experiment, the results demonstrate the proposed algorithm has a fast operation speed, high prediction accuracy and good robustness.

Keywords support vector regression; static reservoir; Newton algorithm; robustness

1 引 言

回声状态网络(Echo State Network, ESN)是一种发展迅速的新型递归网络,具有结构简单、训练速度快、可提前确定稳定性等优势. ESN 由输入层、中间的递归网络层和输出层构成,各层之间通过权值相连.输入和输出间的部分又称储备池,网络的许多特性都通过储备池实现.储备池状态本质上为输入信号的高维显现,表征输入信号的特性^[1].当储备池的内部连接权值全部为零时,ESN 演化成为一种前向

网络——极限学习机(Extreme Learning Machine, ELM)^[2]. ELM 的结构和性质与 ESN 极其相似,但较 ESN 更适用于处理静态问题.

文献[3]在 ESN 网络储备池的基础上,提出 SVESMs(Support Vector Echo State Machines)模型,利用 ESN 网络储备池搭建特征空间,并在其中运用线性支持向量回归(Support Vector Regression, SVR)技术,对混沌时间序列进行预测.但是,对于如何结合储备池的特性对模型进行优化求解,文章并未进行深入分析.而目前得以广泛应用的 SVM 分解算法,如 SMO(Sequential Minimal Opti-

mization)算法对线性 SVM 的训练速度较慢,且训练时间随超参数 C 的增大成比例增长^[4],无法体现出储备池将非线性问题线性化的优势. Newton 算法具有至少二阶的超线性收敛速度,适合求解线性优化问题. Mangasarian 等人将 Newton 算法用于线性分类 SVM,显示出很强的运算优势^[5]. Chapelle 提出了针对非线性分类问题的原始 SVM 递归有限 Newton 算法,算法的运算复杂度与常规非线性 SVM 分类方法相同^[6]. 薄列峰等在已有研究之上,将 Newton 算法推广到非线性 SVR 问题的求解上,得到和常规分解算法相似的结果^[7].

本文在前人研究的基础上,将支持向量回归思想应用于静态储备池中,并结合 Newton 算法对非线性回归问题求解. 该方法首先通过储备池将复杂非线性回归问题转化成与储备池维数相关的线性优化问题,然后运用 Newton 算法求解,并采用 Wolfe-Powell 法搜索 Newton 迭代步长. 构造一个分段多项式函数,将它和开平方函数一起作为静态储备池的鲁棒损失函数. 数值实验结果表明,所提算法在训练速度上优于常规 SVM 方法,并具有较好的鲁棒性.

2 静态储备池模型

ELM 的网络方程如下:

$$x_i = \text{sigmoid}(\mathbf{W}_{in} \cdot \mathbf{u}_i + \mathbf{b}_{in}) \quad (1)$$

$$y_i = \mathbf{w}^T \mathbf{x}_i + b \quad (2)$$

其中, \mathbf{u}, \mathbf{x} 分别为输入向量和储备池状态向量, \mathbf{W}_{in} 是输入权值矩阵, \mathbf{b}_{in}, b 代表偏置, \mathbf{w} 为输出权值向量, y 为网络输出, 下标 $i = 1, \dots, N$ 表示第 i 个样本, N 为样本总数. 输入矩阵 \mathbf{W}_{in} 和偏置向量 \mathbf{b}_{in} 的值在初始时刻随机给定, 网络训练的目的是求输出权值 \mathbf{w} (含偏置 b). 令 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$, 期望输出为 $\mathbf{y}_d = [y_{d1}, y_{d2}, \dots, y_{dN}]^T$, 取代价函数为

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^d} L_p(\mathbf{w}) &= C \sum_{i=1}^N (\mathbf{w}^T \mathbf{x}_i - y_{di})^2 + \|\mathbf{w}\|^2 \\ &= C \|\mathbf{X}\mathbf{w} - \mathbf{y}_d\|^2 + \|\mathbf{w}\|^2 \end{aligned} \quad (3)$$

其中, $C \in \mathbb{R}^+$ 是正则化系数, d 为储备池维数. 可用 Tikhonov 方法求得输出权值为^[3]

$$\mathbf{w}^* = (\mathbf{X}^T \mathbf{X} + C^{-1} \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}_d \quad (4)$$

运用对偶原理, 可得式(3)的对偶式为^[6]

$$\max_{\boldsymbol{\alpha} \in \mathbb{R}^N} 2C^{-1} \boldsymbol{\alpha}^T \mathbf{y}_d - \boldsymbol{\alpha}^T (\mathbf{X}\mathbf{X}^T + C^{-1} \mathbf{I}) \boldsymbol{\alpha} \quad (5)$$

相应的对偶解为

$$\boldsymbol{\alpha}^* = 2C^{-1} (\mathbf{X}\mathbf{X}^T + C^{-1} \mathbf{I})^{-1} \mathbf{y}_d \quad (6)$$

不论是通过求取代价函数得到最优输出权值 \mathbf{w} , 还是求其对偶形式得到最优 Lagrange 乘子 $\boldsymbol{\alpha}$, 最终得到相同的预测结果. 在传统的 SVM 中, 由于 $\mathbf{X} = [\Phi(\mathbf{u}_1), \Phi(\mathbf{u}_2), \dots, \Phi(\mathbf{u}_N)]^T$, $\Phi(\cdot)$ 是将输入从原始空间映射至 Hilbert 空间中的一未知映射. 由于 \mathbf{X} 状态未知, SVM 需通过求内积, 即应用核函数计算出 Gram 矩阵 $\mathbf{X}\mathbf{X}^T$, 得到式(6)中的 $\boldsymbol{\alpha}$, 进而求得预测输出. 选择不同的核函数, 相当于选择不同的映射 $\Phi(\cdot)$ 或不同的内积. 由于式(6)中 $\boldsymbol{\alpha}$ 的运算复杂度高达 $O(N^3)$, 计算量随样本规模增大而变大. 这就是在数据规模较大时, SVM 运算速度慢的原因.

在 ELM 中, \mathbf{X} 为储备池状态矩阵, 可通过式(1)计算得到. 相比于 SVM, 可将 ELM 网络的储备池映射看作一种特殊的映射 $\Phi(\cdot)$, 并在储备池空间运用线性支持向量回归技术. 这正是本文算法应用的基本思想. 在储备池中, 由于储备池状态可计算得到, 能够通过计算输出权值 \mathbf{w} 来得到预测输出. 从式(4)可以看出, 计算 \mathbf{w} 的运算复杂度为 $O(d^3)$ (d 一般在 100~1000 之间). 因此, 在数据规模很大时, 储备池方法仍具有较快的训练速度.

3 静态储备池模型 Newton 算法

本节将 Newton 算法应用于静态储备池的求解中. 将式(3)推广可得代价函数的广义形式如下

$$L_p(\mathbf{w}, b) = C \sum_{i=1}^N L[x_i, y_{di}, f] + \|\mathbf{w}\|^2 \quad (7)$$

其中, f 表示任意损失函数.

在式(3)中, f 为二次损失函数. 此时, 运用 Tikhonov 算法, 通过矩阵 Cholesky 分解快速求得最优权值. 但是, 当数据含有噪声特别是存在异常点时, 二次函数会带来误差的超线性增长, 对训练产生较大干扰, 对预测结果具有不利影响. 为了提高模型的抗干扰能力, 需选用新的鲁棒损失函数.

3.1 鲁棒损失函数

为降低异常点对预测结果的干扰, 所选损失函数在误差较大时应为线性或近似线性的增长. 同时, Newton 算法要求损失函数为二次光滑的凸函数. 而损失函数本身应为非负偶函数. 综上所述, 所选用的损失函数应满足以下条件:

(1) $f(x)$ 为偶函数, 且恒有 $f(x) \geq 0$;

(2) $f(x)$ 是二次光滑条件的凸函数;

(3) 在误差较大时, $f(x)$ 呈线性或近似线性增长.

遵照上面 3 个要求, 本文选用一个二次分段光滑多项式函数和开平方函数一起作为静态储备池模型的损失函数.

(1) 多项式函数

在文献[8]的基础上, 构造出一个分段二次光滑多项式函数, 如下式

$$f(x) = \begin{cases} x - \frac{1}{2h}, & x \geq \frac{1}{h} \\ -\frac{h^3}{2}x^4 + h^2x^3, & 0 < x < \frac{1}{h} \\ -\frac{h^3}{2}x^4 - h^2x^3, & -\frac{1}{h} < x \leq 0 \\ -x - \frac{1}{2h}, & x \leq -\frac{1}{h} \end{cases} \quad (8)$$

其中, $h > 0$. 式(8)显然满足条件(1)和(3), 下面, 证明其为二次光滑凸函数.

引理 1. 若 $f(x)$ 在开凸集 S 上具有二阶连续偏导数, 则 $f(x)$ 为 S 上凸函数的必要条件是: $f(x)$ 的黑塞矩阵 $\nabla^2 f(x)$ 在 S 上处处半正定^[9].

命题 1. 式(8)所示的函数 $f(x)$ 一定是二次光滑凸函数.

证明. 由于函数 $f(x)$ 在各分断区间内均为二次光滑, 且在分段点 $\pm \frac{1}{h}, 0$ 上, 经计算有

$$\begin{cases} f\left(\pm \frac{1}{h}\right) = \frac{1}{2h}, f(0) = 0 \\ \nabla f\left(-\frac{1}{h}\right) = -1, f(0) = 0, \nabla f\left(\frac{1}{h}\right) = 1. \\ \nabla^2 f\left(\pm \frac{1}{h}\right) = \nabla^2 f(0) = 0 \end{cases}$$

可知 $f(x)$ 在 $(-\infty, +\infty)$ 为二次光滑. 又因为

$$\nabla^2 f(x) = \begin{cases} 0, & x \geq \frac{1}{h} \\ -6h^3x^2 + 6h^2x, & 0 < x < \frac{1}{h} \\ -6h^3x^2 - 6h^2x, & -\frac{1}{h} < x \leq 0 \\ 0, & x \leq -\frac{1}{h} \end{cases}$$

且当 $-\frac{1}{h} < x \leq 0$ 时, $\nabla^2 f(x) = -6h^3x\left(\frac{1}{h} + x\right) > 0$, 当 $0 < x < \frac{1}{h}$ 时, $\nabla^2 f(x) = 6h^3x\left(\frac{1}{h} - x\right) > 0$.

根据引理 1 可知, 函数 $f(x)$ 定为凸函数.

综上所述, 结论得证.

证毕.

(2) 开平方函数

开平方损失函数定义如下

$$f(x) = \sqrt{x^2 + \tau} \quad (9)$$

$\tau > 0$. 开平方函数是在二次损失函数基础上, 通过添加正的常数项后开方得到的. 容易看出, 它满足本文鲁棒损失函数选择的所有条件, 并在误差较大时呈近似线性增长.

图 1 为 ϵ -不敏感损失函数、开平方函数、分段多项式函数和二次损失函数曲线图. 从图中可知, 在误差较大时, 二次损失函数取值迅速增大, 而其它 3 种函数呈线性(或近似线性)增长. 同时, 由于 ϵ -不敏感损失函数为不连续, 无法用 Newton 算法进行求解.

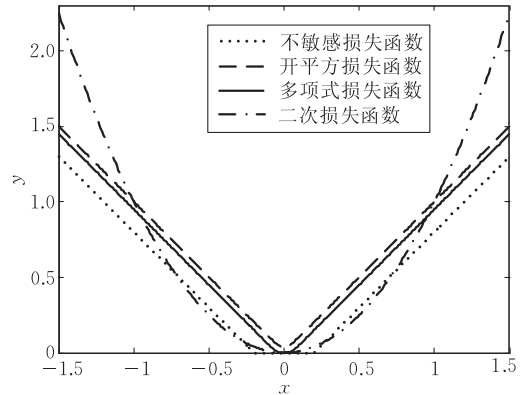


图 1 4 种损失函数曲线

3.2 Newton 算法具体步骤

鲁棒损失函数确定后, 应用 Newton 算法求解模型最优输出权值 w . Newton 下降方向计算公式为

$$d^k = -(\nabla^2 L_p(w^k))^{-1} \nabla L_p(w^k) \quad (10)$$

在式(10)中, $\nabla^2 L_p(w^k)$ 为黑塞矩阵, $\nabla L_p(w^k)$ 为方向导数, k 为迭代步长. Newton 下降方向确定之后, 应用 Wolfe-Powell 线性搜索法计算出迭代步长, 它能够在获取有效步长的同时, 兼顾算法的快速性. 基于静态储备池的 Newton 算法具体步骤如下:

1. 赋初值. 任意给定向量 w^1 及 δ_1, δ_2 , 满足 $0 < \delta_1 < 0.5$, $\delta_1 < \delta_2 < 1$. 同时, 选取适当的 $\epsilon_1 > 0$, 令 $k = 0$.
2. $k := k + 1$, 计算黑塞阵 $\nabla^2 L_p(w^k)$, 其必为对称正定. 用 Cholesky 分解法从式(10)中求得 d^k .
3. 判断 $\|d^k\| \leq \epsilon_1 \cdot \max(1.0, \|w^k\|)$ 是否满足. 满足则求得最终解 $w^* = w^k$, 计算结束. 否则转步 4.
4. 得到 d^k 后, 用 Wolfe-Powell 算法确定迭代步长. 取函数 $\phi(\alpha) = L_p(w^k + \alpha d^k)$, 令 $\alpha_k = 1$, 若满足

$$\begin{cases} \phi(\alpha_k) \leq \phi(0) + \sigma_1 \alpha_k \phi'(0)' \\ \phi(\alpha_k)' \geq \sigma_2 \phi'(0)' \end{cases} \quad (11)$$

$\phi(\alpha_k)' = (\nabla L_p(w^k + \alpha_k d^k))^T d^k$, 转步 8. 不满足则转步 5.

5. 任取 $\beta > 0, \rho, \rho_1 \in (0, 1)$. 先取 $\alpha_k^{(0)}$ 为集合 $\{\beta \rho^i, i = 0, \pm 1, \pm 2, \dots\}$ 中满足式(11)中第 1 个不等式的最大值后, 再

令 $i=0$.

6. 若 $\alpha_k^{(i)}$ 满足式(11)第 2 个不等式, 则步长 $\alpha_k = \alpha_k^{(i)}$, 转步 8; 否则取 $\beta_k^{(i)} = \rho^{-1} \alpha_k^{(i)}$, 转步 7.

7. 取 $\alpha_k^{(i+1)}$ 为集合 $\{\alpha_k^{(i)} + \rho_j^{(i)} (\beta_k^{(i)} - \alpha_k^{(i)}), i = 0, \pm 1, \pm 2, \dots\}$ 中满足式(11)中第 1 个不等式的最大值. 令 $i := i+1$, 转步 6.

8. 令 $w^{k+1} := w^k + \alpha_k d^k$, 转步 2.

4 仿真实例

为验证本文方法的有效性, 将其应用于 10 组基准数据^①的实验仿真中, 数据见表 1 所示. 并与 SVR 和 Tikhonov 方法进行比较, 以验证所提方法的快速性. 同时, 将本文方法和 Tikhonov 方法分别用于两组含异常点的数值实验中, 以验证所提方法的鲁棒性. 实验结果表明, 本文所提方法具有较快的训练速率、较高的预测精度和较好的鲁棒性.

表 1 10 组基准数据列表

数据类型	输入个数		属性维数	
	训练	测试	连续	名义
Auto Price	80	79	14	1
Breast cancer	100	94	32	0
Triazines	100	86	60	0
Machine CPU	100	109	6	0
Abalone	2000	2177	7	1
Delta ailerons	3000	4129	5	0
Cpu activity	4000	4192	12	0
Delta elevators	4000	5517	6	0
Cal housing	8000	12640	8	0
Census	10000	12784	8	0

4.1 基准数据举例

将本文方法应用于十组基准数据仿真中, 并与 SVR^[2] 和 Tikhonov 方法比较. SVR 和静态储备池的模型复杂度对照见表 2 所示.

表 2 SVR 与静态储备池复杂度对照表

数据	支持向量	储备池节点
Auto Price	21.25	20
Breast Cancer	74.3	12
Triazines	48.42	20
Machine CPU	7.8	12
Abalone	309.84	35
Delta ailerons	82.44	50
Cpu activity	64.2	200
Delta elevators	260.38	135
Cal housing	2189.2	150
Census	810.24	200

下面将根据数值实验的结果从效率和预测精度两方面对 3 种算法进行简要评价. 效率主要指训练时间, 所用的时间越短则效率就越高. 而预测精度则

选用均方根误差 E_{RMSE} 进行定量说明

$$E_{RMSE} = \left(\frac{1}{S-1} \sum_{i=1}^S [P_i - O_i]^2 \right)^{1/2} \quad (12)$$

其中, S 为测试样本个数, O_i 是实际观测值, P_i 为对应的预测输出值. E_{RMSE} 越小则精度越高. 3 种方法的实验仿真结果如表 3、表 4 所示.

表 3 Tikhonov, SVR 和 Newton 算法训练时间对照表

(单位: s)

数据	训练时间			
	Tikhonov	SVR	Newton	
			开平方	多项式
Auto Price	0.0012	0.0042	0.0015	0.0018
Breast Cancer	0.0006	0.0064	0.0006	0.0013
Triazines	0.0016	0.0086	0.0019	0.0019
Machine CPU	0.0012	0.0018	0.0024	0.0015
Abalone	0.0115	1.6123	0.0250	0.0435
Delta ailerons	0.0165	0.6726	0.0640	0.0992
Cpu activity	0.1416	1.0149	0.6435	2.0120
Delta elevators	0.0988	1.121	0.3321	0.5058
Cal housing	0.1960	74.184	0.9913	2.2597
Census	0.3518	11.251	1.8370	4.4722

表 4 Tikhonov, SVR 和 Newton 算法测试 E_{RMSE} 对照表

数据	E_{RMSE}			
	Tikhonov	SVR	Newton	
			开平方	多项式
Auto Price	0.0944	0.0937	0.0963	0.0958
Breast Cancer	0.2623	0.2643	0.2622	0.2678
Triazines	0.2004	0.1829	0.1990	0.2000
Machine CPU	0.535	0.0811	0.0533	0.0564
Abalone	0.0768	0.0784	0.0771	0.0768
Delta ailerons	0.0391	0.0429	0.0390	0.0392
Cpu activity	0.0345	0.0470	0.0347	0.0344
Delta elevators	0.0531	0.0540	0.0533	0.0531
Cal housing	0.1269	0.1180	0.1268	0.1274
Census	0.0653	0.0746	0.0655	0.0652

在表 3、表 4 中, 加粗的数字表示性能最好. 从表 3 可以看出, 3 种方法的训练时间随样本规模的增大表现出较大差异. 其中, Tikhonov 算法的训练时间最短, 且随数据样本的增加略微增长. Newton 算法的训练时间也较短, 随着数据样本和储备池维数的增加而增长. SVR 的训练时间最长, 训练速率随着样本数增多而显著变慢. 而从表 4 可以看出, 对不同的数据, 本文所提算法均具有较高的预测精度. 对实验结果进行简要分析. 由于 Tikhonov 方法只需进行一次 Cholesky 分解即得最优解, 所以它的训练速度最快. 而静态储备池 Newton 算法将非线性

① http://www.niaad.liacc.up.pt/~ltorgo/Regression/ds_menu.html

问题转化为与储备池规模相关的线性无约束优化问题,且储备池维数远小于数据样本个数,所以在样本量较大时仍具有很高的训练速度.而 SVR 的 SMO 算法,本质上是对 Lagrange 乘子 α 进行运算,计算复杂度为样本规模数的三次方,因此,当数据样本较多时,SVR 的训练时间显著增长.

4.2 Newton 算法鲁棒性评估

在实际中,由于噪声或外界干扰的出现,经常会遇到训练数据包含异常点的情况.所谓异常点是指与既定模型输出偏离较大的数据点.它的存在会使训练结果与期望值之间产生较大偏差.因此,对异常点抑制能力的大小,实际上反映了算法鲁棒性的强弱.本小节,将本文的 Newton 算法和 Tikhonov 算法分别应用于两组含异常点的数据仿真中,并对结果进行简要评价.

4.2.1 SinC 函数

SinC 函数定义如下^[10]

$$f(x) = \begin{cases} \frac{\sin x}{x}, & x \neq 0 \\ 1, & x = 0 \end{cases} \quad (13)$$

在 $[-15, 15]$ 区间的 SinC 曲线上随机取 300 个数据点作为训练样本,并在相应的输出上添加标准差 0.1 的高斯白噪声.另取 $(-11, 2.5)$ 和 $(0, -1.5)$ 两点作异常点,将它们和前面 300 个数据点一起用作训练样本,如图 2 所示.

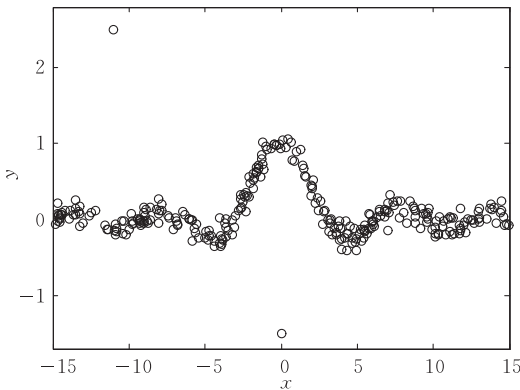


图 2 训练样本图

分别将 Newton 算法和 Tikhonov 正则化方法用于 SinC 函数逼近,并对 3001 个数据点进行测试,得到仿真结果如图 3、图 4 所示.从图中可以看出, Tikhonov 正则化算法的预测曲线在两个异常点附近与期望结果产生了较大偏离,而本文所提算法有效地克服了异常点的干扰,预测结果十分理想.

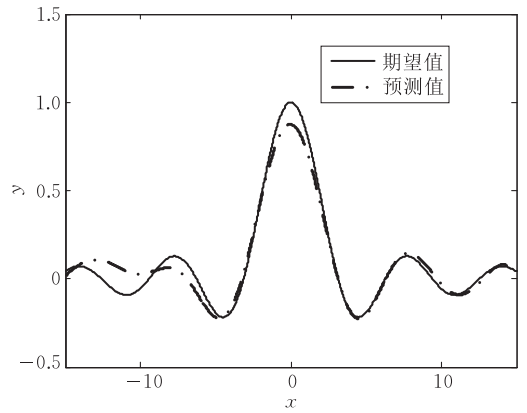


图 3 Tikhonov 法预测输出曲线

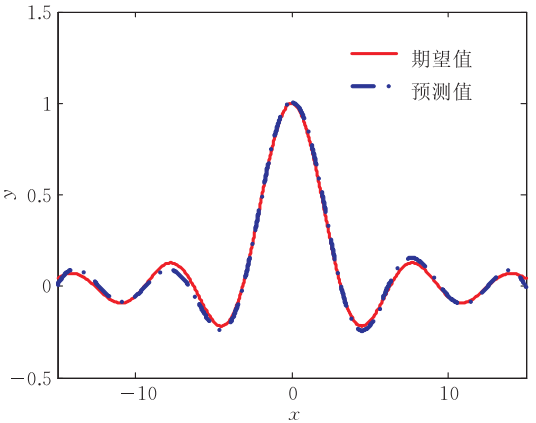


图 4 Newton 法预测输出曲线

4.2.2 Motorcycle 数据

Motocycle 数据是统计学中较为常用的一组实际测试数据.由于其数据分布不规则,具有异方差性,在某种意义上可将其视为含异常点的数据^[10].分别将 Newton 算法和 Tikhonov 算法用于 Motocycle 数据仿真中,训练样本和测试样本均为实测值,得仿真结果如图 5 所示.

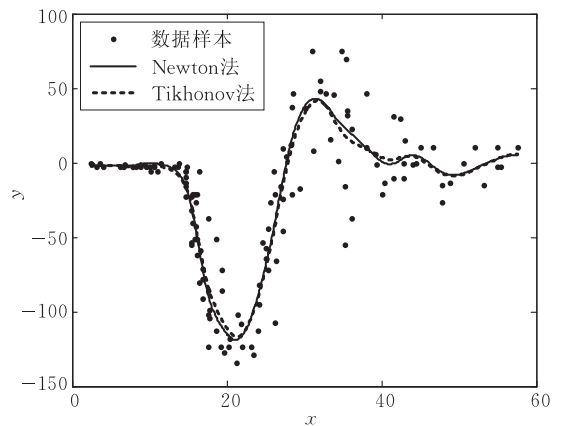


图 5 Motocycle 数据预测输出结果

从图 5 可以看出,在数据分布较密集的始端和末端,两种方法均实现了较为理想的逼近效果.而在数据点分布比较分散、不规则的中间阶段,Tikhonov 方法的预测曲线受异常点的拉动影响较大,在 $[30,40]$ 区间上出现一段明显的下移.而 Newton 方法的预测输出受异常点的影响相对较小.

从 SinC 函数和 Motorcycle 数据仿真结果中可以看出,鲁棒损失函数的选用有效地降低了异常点对训练过程的影响,使得本文算法具有较好的鲁棒性.

5 结 论

本文将鲁棒损失函数应用于静态储备池中,并运用 Newton 算法对复杂非线性回归问题进行求解.通过储备池将原空间的非线性数据关系转化为储备池空间的线性数据关系,由于储备池状态可计算得到,可运用最优化算法对储备池输出权值进行优化求解.鲁棒损失函数的运用可有效抑制异常点的干扰,使得本文方法具有较强的鲁棒性.将所提方法应用于十组基准数据的实验仿真中,并和 Tikhonov 方法及 SVR 进行比较.实验结果表明本文所提方法在处理大规模数据问题时,具有较高的训练速率和预测精度.同时,通过对两组含异常点的数据进行实验仿真,验证了所提方法较好的鲁棒性能.

致 谢 作者感谢各位审稿人提出的宝贵意见及为此文付出劳动的编辑部同志们!



HAN Min, born in 1959, professor, Ph.D. supervisor. Her current research interests include neural network and chaos and their applications to control and identification.

Background

This work is supported by National Natural Science Foundation of China (60674073), National High Technology Research and Development Program (863 Program) of China (2007AA04Z158), National Key Technology R&D Program of China (2006BAB14B05) and National Basic Research Program (973 Program) of China (2006CB403405). They aim to efficiently solve the large scale nonlinear regression problem.

参 考 文 献

- [1] Jaeger H. Reservoir riddles: Suggestions for echo state network research//Proceedings of the IEEE International Joint Conference on Neural Networks. Montreal, QC, Canada, 2005, 8: 1460-1462
- [2] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications. *Neurocomputing*, 2006, 70 (1-3): 489-501
- [3] Shi Z W, Han M. Support vector echo-state machine for chaotic time-series prediction. *IEEE Transactions on Neural Networks*, 2007, 18(2): 359-372
- [4] Kao W C, Chuang K M, Sun C L et al. Decomposition methods for linear support vector machines. *Neural Computation*, 2004, 16(8): 1689-1704
- [5] Fung G M, Mangasarian O L. A feature selection Newton method for support vector machine classification. *Computational Optimization and Applications*, 2004, 28(2): 185-202
- [6] Chapelle O. Training a support vector machine in the primal. *Neural Computation*, 2007, 19(5): 1155-1178
- [7] Bo L F, Wang L, Jiao L C. Recursive finite Newton algorithm for support vector regression in the primal. *Neurocomputing*, 2007, 19(4): 1082-1096
- [8] Yuan Yu-Bo, Yan Jie, Xu Cheng-Xian. Polynomial smooth support vector machine(PSSVM). *Chinese Journal of Computers*, 2005, 28(1): 9-17(in Chinese)
(袁玉波, 严杰, 徐成贤. 多项式光滑的支撑向量机. *计算机学报*, 2005, 28(1): 9-17)
- [9] Li Dong-Hui, Dong Xiao-Jiao, Wan Zhong. *Numerical Optimization*. Beijing: Science Press, 2005(in Chinese)
(李董辉, 董小娇, 万中. 数值最优化. 北京: 科学出版社, 2005)
- [10] Suykens J A K, Brabanter J D, Lukas L et al. Weighted least squares support vector machines: Robustness and sparse approximation. *Neurocomputing*, 2002, 48(1-4): 85-105

WANG Ya-Nan, born in 1985, M. S. candidate. Her current research interests include neural network and multivariate time series analysis.

By replacing the “kernel trick” of SVM with the “reservoir trick” of ESN, the authors have proposed a kernel-free SVMs. Subject to the constraints of kernel function, the advantages of many optimization methods are not reflected. In the paper, a Newton algorithm is adopted in static kernel-free SVMs, which has greatly improved the operation speed and robust loss functions are adopted to restrain the interference of outliers.