Lagrange 双支撑向量回归机

郑逢德 张鸿宾
(北京工业大学计算机学院 北京 100124)

摘 要 提出一种快速的支撑向量回归算法。首先将支撑向量回归的带有两组约束的二次规划问题转化为两个小的分别带有一组约束的二次规划问题,而每一个小的二次规划问题又采用一种快速迭代算法求解,该迭代算法能从任何初始点快速收敛,避免了二次优化问题求解,因此能显著提高训练速度。在多个标准数据集上的实验表明,该算法比传统支撑向量机快很多,同时具有良好的泛化性能。

关键词 支撑向量回归,Langrage 支撑向量机,双支撑向量回归,迭代算法
中图法分类号 TP181 **文献标识码** A

Lagrange Twin Support Vector Regression

ZHENG Feng-de ZHANG Hong-bin
(College of Computer Science, Beijing University of Technology, Beijing 100124, China)

Abstract This paper proposed a fast support vector regression algorithm. This algorithm converts the quadratic programming problems(Qpps) with pair groups of linear inequality constraints to two small size Qpps with only one group of linear inequality constraints. Each of the small size Qpps is solved by an iterative algorithm. The iterative algorithm converges from any starting point and does not need any quadratic optimization packages. Thus this algorithm is fast. The experimental results on several benchmark datasets demonstrate the effectiveness of the proposed algorithm.

Keywords Support vector regression, Lagrange support vector machine, Twin support vector regression, Iterative algorithm

1 引言

Vapnik 引入 ϵ 不敏感损失函数后,支撑向量机成功应用于回归问题[1-3]。但由于支撑向量回归(SVR)需要求解二次规划问题,需要较长的训练时间[4],因此快速性一直是支撑向量回归的一个主要研究内容。近来提出了许多 SVR 的改进算法,如光滑支撑向量回归(SSVR)[5]、最小二乘支撑向量回归(LSSVR)[6]、临近支撑向量回归(PSVR)[7]和双支撑向量回归(TSVR)[8]。TSVR 寻找两个非平行函数,每一个函数决定 ϵ 不敏感带的下界或上界,因此 TSVR 将 SVR 的一个大的二次规划问题转化为两个小的二次规划问题[8]。另外,SVR 具有两组约束,而 TSVR 的每个二次规划问题只有一组约束。减少约束可使问题规模变小,从而缩短运算时间。解决两个小的二次规划问题的策略使得 TSVR 效率比 SVR 高。Lagrange 支撑向量机(LSVM)[9-11]将二次规划问题采用一个迭代算法求解,主要运算量在于对一个维数等于样本数的矩阵求逆,避免了求解线性优化或二次优化问题。

本文结合 LSVM 和 TSVR 提出 Lagrange 双支撑向量回归(LTSVR),首先将支撑向量回归的带有两组约束的二次规划问题转化为两个小的分别带有一组约束的二次规划问题,而每一个小的二次规划问题又采用一种快速迭代算法求解,避免了求解二次优化问题,因此本文方法能显著提高训练速度。在多个标准数据集上的实验也表明了本文算法的有效性。

2 基本概念

2.1 支撑向量回归机

考虑训练样本$\{(x_i,y_i),i=1,\cdots,m\}$,这里 $x_i\in R^n$ 是输入向量,y_i 是相应的目标向量,m 是训练样本数目。SVR 算法解决如下带约束的二次优化问题:

$$\begin{aligned}&\min_{w,b,\xi,\xi^*}\ \frac{1}{2}w^{\mathrm T}w+C(e^{\mathrm T}\xi+e^{\mathrm T}\xi^*)\\ &s.t.\quad Y-(Aw+be)\leqslant\epsilon e+\xi\\ &\qquad (Aw+be)-Y\leqslant\epsilon e+\xi^*\\ &\qquad \xi,\xi^*\geqslant 0,i=1,\cdots,m\end{aligned}\tag{1}$$

式中,$A=(A_1,A_2,\cdots,A_m)$是数据矩阵,其第 i 行 A_i 为向量 $x_i^{\mathrm T}$,$Y=(y_1,y_2,\cdots,y_m)$为相应训练样本的目标值,$C>0$ 为正则化参数,该值在拟合错误和线性回归函数的平滑行之间取得平衡,ξ 和 ξ^* 是松弛向量,e 是相应维数的单位向量。该问题可以通过求解其对偶问题求解。

2.2 双支撑向量回归机

双支撑向量回归(TSVR)得到两个非平行平面,对两个

到稿日期:2011-01-05 返修日期:2011-04-18 本文受国家自然科学基金项目(60775011)资助。
郑逢德(1980—),男,博士生,主要研究方向为核方法、机器学习,E-mail:zheng_fd@163.com;张鸿宾 男,教授,博士生导师,主要研究方向为模式识别、神经网络、数字水印等。

小的二次规划问题求解[6]。SVR 中有两组约束，而在 TSVR 中只有一组约束。TSVR 寻找两个函数 $f_1(x)=w_1^T x+b_1$ 和 $f_2(x)=w_2^T x+b_2$，分别决定 ε 不敏感带的下边界或上边界。特别地给定训练数据(A,Y)，函数 $f_1(x)$决定 ε_1 不敏感带的下边界，$f_2(x)$决定 ε_2 不敏感带的上边界。线性 TSVR 解决式(2)和式(3)两个二次规划问题，每一个决定一个相应的边界函数。

$$\begin{aligned}&\min \frac{1}{2}\|Y-e\varepsilon_1-(Aw_1+eb_1)\|^2+C_1 e^T\xi\\&s.t. Y-(Aw_1+eb_1)\leqslant e\varepsilon_1-\xi,\xi\geqslant 0\end{aligned}\tag{2}$$

$$\begin{aligned}&\min \frac{1}{2}\|Y+e\varepsilon_2-(Aw_2+eb_2)\|^2+C_2 e^T\eta\\&s.t. (Aw_2+eb_2)-Y\leqslant e\varepsilon_2-\eta,\eta\geqslant 0\end{aligned}\tag{3}$$

式中，$C_1,C_2>0$，$\varepsilon_1,\varepsilon_2$ 是参数，ξ,η是松弛向量。

3 Lagrange 双支撑向量回归机

采用 Lagrange 支撑向量机的方法[7-9]，将问题(2)和(3)中松弛变量 ξ 和 η 的 1 范数改成 2 范数平方，使得约束 $\xi,\eta\geqslant 0$ 没有必要，从而可以去掉。线性双支撑向量回归的约束最小化问题可以写成

$$\begin{aligned}&\min \frac{1}{2}\|Y-e\varepsilon_1-(Aw_1+eb_1)\|^2+\frac{1}{2}C_1\xi^T\xi\\&s.t. Y-(Aw_1+eb_1)\geqslant e\varepsilon_1-\xi\end{aligned}\tag{4}$$

$$\begin{aligned}&\min \frac{1}{2}\|Y+e\varepsilon_2-(Aw_2+eb_2)\|^2+\frac{1}{2}C_2\eta^T\eta\\&s.t. (Aw_2+eb_2)-Y\geqslant e\varepsilon_2-\eta\end{aligned}\tag{5}$$

式中，$C_1,C_2>0$，$\varepsilon_1,\varepsilon_2$ 是参数，ξ,η是松弛向量。

通过引入拉格朗日乘子 a，式(4)对应的拉格朗日函数为

$$\begin{aligned}L(w_1,b_1,\xi,a)=&\frac{1}{2}\|Y-e\varepsilon_1-(Aw_1+eb_1)\|^2+\frac{C_1}{2}\xi^T\xi-\\&a^T(Y-(Aw_1+eb_1)-e\varepsilon_1+\xi)\end{aligned}\tag{6}$$

KKT 优化条件为

$$\frac{\partial L}{\partial w_1}=-A^T(Y-e\varepsilon_1-(Aw_1+eb_1))+A^Ta=0\tag{7}$$

$$\frac{\partial L}{\partial b_1}=-e^T(Y-e\varepsilon_1-(Aw_1+eb_1))+e^Ta=0\tag{8}$$

$$\frac{\partial L}{\partial \xi}=C_1\xi-a=0\tag{9}$$

$$Y-(Aw_1+eb_1)\geqslant e\varepsilon_1-\xi\tag{10}$$

$$a^T(Y-(Aw_1+eb_1)-e\varepsilon_1-\xi)=0,a\geqslant 0\tag{11}$$

定义

$$G=[A\quad e],f=Y-e\varepsilon_1,u_1=[w_1^T\quad b_1]^T\tag{12}$$

合并式(7)和式(8)，可得

$$-\begin{bmatrix}A^T\\e^T\end{bmatrix}((Y-e\varepsilon_1)-[A\quad e]\begin{bmatrix}w_1\\b_1\end{bmatrix})+\begin{bmatrix}A^T\\e^T\end{bmatrix}a=0\tag{13}$$

由上式可得

$$u_1=(G^TG)^{-1}G^T(f-a)\tag{14}$$

如果需要，这里 G^TG 可以加上一个正则项 $\delta I,\delta>0$，I 是相应维数的单位矩阵，这样可避免 G^TG 不可求逆的出现。相应的 u_1 为

$$u_1=(G^TG+\delta I)^{-1}G^T(f-a)\tag{15}$$

将式(14)及其他 KKT 条件式带入式(6)并去除常量项，可得优化问题(4)的对偶形式为

$$\min_{0\leqslant a}\frac{1}{2}a^TQ_1a-r_1^Ta\tag{16}$$

其中

$$Q_1=\frac{I}{C_1}+G(G^TG)^{-1}G^T\tag{17}$$

$$r_1=f^T+f^TG(G^TG)^{-1}G^T\tag{18}$$

同样地可以得到优化问题(5)的对偶形式。定义

$$h=Y+e\varepsilon_2\tag{19}$$

可以得到

$$u_2=(G^TG)^{-1}G^T(h+\gamma)\tag{20}$$

优化问题(5)的对偶形式为

$$\min_{0\leqslant\gamma}\frac{1}{2}\gamma^TQ_2\gamma-r_2^T\gamma\tag{21}$$

其中

$$Q_2=\frac{I}{C_2}+G(G^TG)^{-1}G^T\tag{22}$$

$$r_2=h^T+h^TG(G^TG)^{-1}G^T\tag{23}$$

上述线性支撑向量回归可以扩展到非线性形式。非线性情况下，两个边界函数分别为 $f_1(x)=K(x^T,A^T)w_1+b_1$ 和 $f_2(x)=K(x^T,A^T)w_2+b_2$。相应的优化问题为

$$\begin{aligned}&\min \frac{1}{2}\|Y-e\varepsilon_1-(K(A,A^T)w_1+eb_1)\|^2+\frac{1}{2}C_1\xi^T\xi\\&s.t. Y-(K(A,A^T)w_1+eb_1)\geqslant e\varepsilon_1-\xi\end{aligned}\tag{24}$$

$$\begin{aligned}&\min \frac{1}{2}\|Y+e\varepsilon_2-(K(A,A^T)w_2+eb_2)\|^2+\frac{1}{2}C_2\eta^T\eta\\&s.t. (K(A,A^T)w_2+eb_2)-Y\geqslant e\varepsilon_2-\eta\end{aligned}\tag{25}$$

式中，$C_1,C_2>0$，$\varepsilon_1,\varepsilon_2$是参数，$\xi,\eta$是松弛向量。非线性 LTSVR 对偶问题可以通过相似的方式得到，只需要重新定义 G。

$$G=\lfloor K(A,A^T)e\rfloor\tag{26}$$

线性或者非线性情况下的对偶问题都可以写成

$$\min_{0\leqslant a}\frac{1}{2}a^TQa-r^Ta\tag{27}$$

该问题的 KKT 充分必要优化条件为

$$0\leqslant a\perp(Qa-r)\geqslant 0\tag{28}$$

该优化条件可以写成如下等价形式，对任意 $p>0$，

$$(Qa-r)=((Qa-r)-pa)_+\tag{29}$$

其中，加号函数$((Qa-r)-pa)_+$将所有的负向量$(Qa-r)-pa$ 置为零。

为得到上述问题的解，可以采用如下的迭代算法。

$$a^{i+1}=Q^{-1}(r+((Qa^i-r)-pa^i)_+)\tag{30}$$

求得 a 后，就可得到 u_1 和 u_2，也就得到了回归函数

$$f(x)=\frac{1}{2}(w_1+w_2)^TK(A,x)+\frac{1}{2}(b_1+b_2)\tag{31}$$

Lagrange 双支撑向量回归算法流程为：

1)给定参数 $C_1,C_2,\varepsilon_1,\varepsilon_2$ 和核函数参数，构造核矩阵 $K(A,A^T)$。

2)根据式(26)定义 G，按照式(17)、式(18)、式(22)、式(23)求出 Q_1、r_1、Q_2、r_2。

3)给出任意 $p>0$，从任一初始值 a(如 $a=0.2$)按照式(30)迭代求解 a。

4)按照式(12)、式(19)求出 f、h，然后按照式(15)、式(20)求出 u_1、u_2，采用式(31)构造回归函数 $f(x)$。

4 数值实验

为了检验本文算法的有效性，将本文算法同 SVR，LS-SVR[4]和 TSVR[6]在标准数据集上进行实验对比。所有的实

验在 MATLAB 2009b，酷睿 2 CPU 1.8GHz，2 GB 内存平台上完成。均方根误差用于评价预测值和真实值的偏差，定义为

$$RMSE=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i-f_i)^2} \tag{32}$$

式中，y_i 是真实值，f_i 是预测值。采用 RBF 核函数 $f(u,v)=\exp(-\frac{\|u-v\|^2}{2\delta^2})$，参数 δ 和 SVR，LS-SVR，TSVR（$C_1=C_2$）和 LTSVR（$C_1=C_2$）参数 C 采用交叉验证确定参数。TSVR 和 LTSVR 中参数 ε_1 和 ε_2 都为 10^{-7}。LTSVR 停机准则中的容差为 10^{-5} 或者迭代 10^4 次。

测试的标准数据集有 Boston housing，Bodyfat，AutoMPG，Mackey-Glass，Servo，Machine CPU，Wiscoin breast cancer datasets，Concrete Compressive Strength，Auto price，forest fires 和 trizaines。这些都是机器学习算法中的常用数据集。其中 Bodyfat 数据集来自 Statlib collections[12]，Mackey-Glass 数据集来自文献[13]，trizaines 数据集来自文献[14]，其他的数据集都来自 UCI 数据库[15]。Boston Housing 数据集包含 506 个样本，每个样本有 13 个特征。随机选取 400 个为训练样本，106 个为测试样本。Bodyfat 数据集包含 252 个样本，每个样本有 14 个特征，随机选取 150 个为训练样本，102 个为测试样本，AutoMPG 数据集包含 392 个样本，每个样本有 7 个特征，随机选取 250 个为训练样本，142 个为测试样本。Mackey-Glass 数据集包含 800 个样本，随机选取 585 个为训练样本，其余的为测试样本。Servo 数据集包含 120 个训练样本和 47 个测试样本，每个样本有 4 个特征。Machine CPU 数据集包含 150 个训练样本和 59 个测试样本，每个样本有 6 个特征。Wiscoin breast cancer 数据集包含 194 个样本，每个样本有 32 个特征，随机选取 150 个为训练样本，剩下的 44 个为测试样本。Concrete Compressive Strength 数据集包含 1030 个样本，每个样本有 9 个特征，随机选取 500 个作为训练样本，其余 530 个为测试样本。Auto price 数据集包含 159 个 15 维的样本，随机选取 120 个为训练样本，其余 39 个为测试样本。Forest fires 包含 517 个 12 维的样本，随机选取 400 个为训练样本，其余 117 个为测试样本。trizaines 数据集包含 186 个 58 维的样本，随机选取 140 个为训练样本，其余 46 个为测试样本。

每个实验都运行 10 次。表 1 为各种算法的实验结果，从中可以看出本文提出的 LTSVR 算法基本上在每个数据集都取得了和 SVR，TSVR 相当的拟合能力，但运算速度却比 SVR，TSVR 快很多，同时本文的 LTSVR 的泛化能力比 LSSVR 算法的强，因此提出的 LTSVR 算法是一种有效的回归算法。

表 1　本文算法与其他回归算法实验对比结果

Dataset	Samples train test	SVR MSE Time(s)	LS-SVR MSE Time(s)	TSVR MSE Time(s)	LTSVR MSE Time(s)
Boston housing	400	0.0739±0.0191	0.1800±0.0137	0.0719±0.076	0.0716±0.0206
	106	327.6141	0.2028	96.0124	5.8313
Bodyfat	200	0.0040±0.0024	0.0175±0.0012	0.0073±0.0018	0.0037±0.0020
	52	46.7441	0.0343	13.5565	1.4882
Auto-mpg	250	0.0643±0.0050	0.1613±0.0069	0.0604±0.0038	0.0617±0.0041
	140	81.7195	0.2090	25.9772	2.4492
Mackey-Glass	800	0.0844±0.0035	0.1614±0.0023	0.0802±0.0064	0.0790±0.0014
	585	2.6143e+003	3.7721	777.9115	23.5936
Servo	120	0.1588±0.0333	0.2147±0.0302	0.0970±0.0349	0.1057±0.0319
	47	9.5442	0.0250	3.3509	0.5803
Machine CPU	150	0.0614±0.0287	0.1530±0.0436	0.0853±0.0415	0.0857±0.0333
	57	18.5641	0.0842	6.1090	0.8549
Wisconsin B. C.	150	0.1178±0.0134	0.2873±0.0225	0.2464±0.00591	0.1474±0.0258
	44	0.7051	0.0468	6.1339	0.8486
Concrete CS	500	0.1008±0.0020	0.1571±0.0072	0.0872±0.0061	0.0933±0.0041
	530	628.7402	3.7846	187.6973	9.3830
Auto price	120	0.0614±0.0077	0.1455±0.0388	0.0589±0.0123	0.0674±0.0129
	39	9.8874	0.0842	3.4226	0.5834
Forest fires	400	0.0494±0.0421	0.0493±0.0413	0.0800±0.0248	0.0580±0.0346
	117	346.8039	0.2402	96.4523	6.0372
Trizaines	140	0.0227±0.0150	0.2758±0.0055	0.0219±0.0143	0.0203±0.0148
	46	16.3395	0.0562	4.9951	0.8143

结束语　本文提出了一种快速回归算法：Lagrange 双支撑向量回归算法。它首先将支撑向量回归的带有两组约束的二次规划问题转化为两个小的分别带有一组约束的二次规划问题，而每一个小的二次规划问题又采用一种快速迭代算法求解。由于二次规划问题求解是支撑向量机中最耗费时间的步骤，因此本文的方法将二次规划问题转化为两个小的二次规划问题，然后采用迭代算法求解。其主要运算量在于矩阵求逆，避免了求解线性优化或二次优化问题，能显著提高训练速度。在多个标准数据集上的实验表明，本文算法是一种有效的回归算法。

参 考 文 献

[1] Smola A J，Schölkopf B. A tutorial on support vector regression [J]. Statistics and Computing，2004，14(3)：199-222

[2] Vapnik V N. Statistical learning theory [M]. New York：Wiley Interscience，1998

[3] 邓乃扬，田英杰. 支持向量机——理论、算法与拓展[M]. 北京：科学出版社，2009：170-187

[4] 文益民，王耀南，吕宝粮，等. 支持向量机处理大规模问题算法综述书名[J]. 计算机科学，2009，36(7)：20-25

（下转第 254 页）

每幅图像以空间中点的形式输出;图 9 分别以图像和空间中点的形式显示了 ISOMAP 算法在该数据集上的实验效果,可明显发现 ISOMAP 算法只能大致发现人脸面朝左边到右边的变化趋势,即沿图 10(a)的横轴方向。但图 10(a)的左半部分在纵轴方向上出现了过分收缩,即没有发现人脸从仰视到俯视的姿势变化趋势。

结束语 本文针对 ISOMAP 算法的不足,提出了一种适合于被间隔的流形学习算法 G-ISOMAP。该算法不仅对现有 ISOMAP 算法所适合的数据集有效,更重要的是它对于抽样于同一流形但中间有间隔的数据集也能得到比其他算法更满意的结果,所以本文的 G-ISOMAP 算法比 ISOMAP 算法更灵活,应用面更广。

不过需要提及的是:①如果被间隔的子流形间最短的欧氏距离对应的数据点恰好位于间隔区域的边缘,那么本文算法能够成功发现内在低维流形的结构,然而,所发现的数据点离间隔区域越远,效果就越不佳。②本文的假设是被间隔的数据是采样于一个流形的。如果数据集来源于多个类,每个类的数据集抽样于一个单独的流形,那么如何将本文的方法扩展到该情形。③目前主流的 ISOMAP 算法及其变种大多属于无监督范畴,如果完成了第②个工作,那么在此基础上,应用神经网络找到原始的高维数据到低维嵌入之间的映射函数,最后应用现有的分类器(如 K 近邻分类器等)来完成分类任务,从而可以使 ISOMAP 算法适合有监督框架。这 3 点也许是未来值得研究的问题。

参 考 文 献

[1] Seung H S,Lee D D. Manifold ways of perception[J]. Science, 2000,290(5500):2268-2269

[2] Tenenbaum J B,de Silva V,Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000,290(5500):2319-2323

[3] Roweis S T,Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science,2000,290(5500):2323-2326

[4] Donoho D,Grimes C. Hessian Eigenmaps:new locally linear embedding techniques for high-dimensional data[J]. PNAS,2003, 100(10):5591-5596

[5] Belkin M,Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation,2003, 15(6):1373-1396

[6] He X,Niyogi P. Locality preserving projections[C]// Advances in Neural Information Processing Systems. Cambridge: MIT Press,2004:327-334

[7] Zhang Z,Zha H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment[J]. Journal of Scientific Computing,2005,26(1):313-338

[8] Weinberger K Q, Saul L K. Unsupervised learning of image manifolds by semidefinite programming[J]. International Journal of Computer Vision,2006,70(1):77-90

[9] de Silva V, Tenenbaum J. Global versus local methods in nonlinear dimensionality reduction[C]// Becker S,Thrun S,Obermayer K,eds. Proc. NIPS. 2003,15:721-728

[10] Choi H Y,Choi S J. Robust kernel isomap[J]. Pattern Recognition,2007,40(3):853-862

[11] Geng X, Zhan D C, Zhou Z H. Supervised nonlinear dimensionality reduction for visualization and classification[J]. IEEE Trans. on Systems,Man and Cybernetics,2005,35(6):1098-1107

[12] 邵超,黄厚宽,赵连伟. 一种更具拓扑稳定性的 ISOMAP 算法[J]. 软件学报,2007,18(4):869-877

[13] 周志华,曹存根. 神经网络及其应用[M]. 北京:清华大学出版社,2004

[14] Bernstein M,de Silva V,Langford J C,et al. Graph app-roximations to geodesics on embedded manifolds[R]. Stanford University,2000

[15] Balasubramanian M, Schwartz E L. The ISOMAP algorit-hm and topological stability[J]. Science,2002,295(5552):7

[16] http://waldron. stanford. edu/~isomap/datasets. html

(上接第 249 页)

[5] Lee Y J, Hsieh W F, Huang C M. e-SSVR: a smooth support vector machine for e-insensitive regression[J]. IEEE Transactions on Knowledge and Data Engineering,2005,17(5):678-685

[6] Johan A K S,Tony V G,Jos D B,et al. Least Squares Support Vector Machine [M]. World Scientific,2002:182-183

[7] 杜喆,胡廷锋,刘三阳. 用于回归的临近支持向量机[J]. 计算机科学,2009,36(1):126-128

[8] Peng X J. TSVR:An efficient Twin Support Vector Machine for Regression[J]. Neural Networks,2010,23:365-372

[9] Mangasarian O L,Musicant D R. Lagrangian support vector machine[J]. Journal of Machine Learning Research,2001,1:161-177

[10] Mangasarian O L,Musicant D R. Finite Newton method for Lagrangian support vector machine[J]. Neurocomputing,2003,55: 39-55

[11] Balasundaram S K. On Lagrangian support vector regression [J]. Expert Systems with Applications,2010,37:8784-8793

[12] Bodyfat DataSets [EB/OL]. http://lib. stat. cmu. edu/datasets/

[13] Mackey-Glass DataSets [EB/OL]. http://www. cse. ogi. edu/~ericwan

[14] Trizaines DataSets [EB/OL]. http://www. liaad. up. pt/~ltorgo/Regression/DataSets. html

[15] UCIdatabases[EB/OL]. http://www. ics. uci. edu/_mlearn/MLRepository. html