

增量和减量式标准支持向量机的分析*

顾彬^{1,2}, 郑关胜^{1,2,3}, 王建东³

¹(南京信息工程大学 江苏省网络监控中心, 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)

通讯作者: 顾彬, E-mail: jsgubin@163.com

摘要: 当训练数据每次发生改变时,例如增加或者删除部分数据,标准支持向量机的批处理算法就需要重新进行训练,这不适合在线环境的计算.为了克服这个问题,Cauwenberghs 和 Poggio 提出了增量和减量式标准支持向量机算法(C&P 算法).通过理论分析,证明 C&P 算法的可行性和有限收敛性.可行性证明确保了 C&P 算法的每步调整都是可靠的,有限收敛性证明确保了 C&P 算法通过有限步调整最终收敛到问题的最优解.在此基础上,进一步通过实验结果验证了所给出的理论分析的结果.

关键词: 支持向量机;增量式学习;减量式学习;可行性分析;收敛性分析

中图法分类号: TP181 **文献标识码:** A

中文引用格式: 顾彬,郑关胜,王建东.增量和减量式标准支持向量机的分析.软件学报,2013,24(7):1601-1613. <http://www.jos.org.cn/1000-9825/4327.htm>

英文引用格式: Gu B, Zheng GS, Wang JD. Analysis for incremental and decremental standard support vector machine. Ruan Jian Xue Bao/Journal of Software, 2013,24(7):1601-1613 (in Chinese). <http://www.jos.org.cn/1000-9825/4327.htm>

Analysis for Incremental and Decremental Standard Support Vector Machine

GU Bin^{1,2}, ZHENG Guan-Sheng^{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: GU Bin, E-mail: jsgubin@163.com

Abstract: Batch implementations of standard support vector machine (SVM) are inefficient on an online setting because they must be retrained from scratch every time the training set is modified (i.e., adding or removing some data samples). To solve this problem, Cauwenberghs and Poggio propose an incremental and decremental support vector classification algorithm (C&P algorithm). This paper proves the feasibility and finite convergence of the C&P algorithm through theoretical analysis. The feasibility ensures that each adjustment step in the C&P algorithm is reliable, and the finite convergence ensures that the C&P algorithm can converge to the optimal solution within finite steps. Finally, the conclusions of the theoretical analysis are verified by the experimental results.

Key words: support vector machine; incremental learning; decremental learning; feasibility analysis; convergence analysis

由 Vapnik 等人提出的支持向量机(support vector machine,简称 SVM)^[1]目前已成为解决分类、回归和其他

* 基金项目: 国家自然科学基金重点项目(61139002, 61232016); 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: 2012-05-21; 修改时间: 2012-07-16; 定稿时间: 2012-09-29; jos 在线出版时间: 2013-01-17

CNKI 网络优先出版: 2013-01-17 15:26, <http://www.cnki.net/kcms/detail/11.2560.TP.20130117.1526.002.html>

统计学习问题的一种流行的技术,其中,支持向量机的训练需要求解一个与问题样本数目同样多参数的凸二次规划问题,特别是对于大规模数据集的训练存在着一定的困难.目前已经提出了一些实用的批处理方法,比如工作集选择法^[2]、样本对迭代计算法^[3]、闭包球法^[4]、梯度下降法^[5,6]、类感知器法^[7]、切平面算法^[8]、束捆计算法^[9,10]等.

而在实际的问题环境中(例如网络数据安全分析、传感器网络数据分析、金融数据分析等),训练样本数据大多为在线环境下增量提供.以网络数据安全分析为例,给定的网络数据流一般不够稳定,亦即数据分布随着时间的变化也会跟着变化,那么如果这样的异常分布数据出现在训练样本中,并且训练样本集很大,批处理算法将会因重新计算而失效.增量和减量式算法对这种情形更适用,因为增量式(减量式)算法的优点正是能够吸收(剔除)额外的数据而不需要对所有数据重新计算,所以,有必要对支持向量机设计相应的增量和减量式算法.

2001年,Cauwenberghs和Poggio^[11]针对标准的支持向量机给出了一种精确的增量和减量式学习算法(C&P算法).为了对C&P算法有更深入的认识,2006年,Laskov等人^[12]对该算法增量式部分给出了部分理论分析结果(也就是本文可行性分析中的定理1和定理4),并将它扩展到了一分类支持向量机,设计了增量式一分类支持向量机算法;2008年,顾彬等人^[13]将该算法扩展到排列支持向量机上,并给出了相应算法的理论分析,包括可行性以及收敛性分析;2010年,Karasuyama等人^[14]扩展了C&P算法,使之能同时处理多个样本的增量式学习;2012年,顾彬等人^[15]扩展了C&P算法,使之能处理 ν 支持向量分类机.本文试图在上述工作的基础上,从理论分析与实验验证的角度,全面讨论C&P算法的可行性及有限收敛性.

本文第1节介绍增量和减量式标准支持向量机算法.第2节证明该算法的可行性,该分析将确保C&P算法的每步调整都是可靠的.第3节证明算法的有限收敛性,该分析将确保C&P算法通过有限步调整最终收敛到问题的最优解.第4节的实验数据进一步验证上述理论分析的结果.最后对本文内容进行总结和展望.

1 增量和减量式标准支持向量机算法

当训练集增加新样本或删除原始的某些样本时,为了获得更新后的训练样本集的最优解,增量式(减量式)学习算法需要设计相应的增量(减量)更新法则,使得更新后的训练样本集仍然满足KKT等价条件^[16].

而增量式标准支持向量机算法更新法则的基本思想是:令新增样本为 (x_c, y_c) ,在严格保持原先样本集 $S = \{(x_1, y_1), \dots, (x_l, y_l)\}$ 中所有样本满足KKT等价条件的情况下,逐步增加新增样本 (x_c, y_c) 的权值 α_c ,如此反复,直至新增样本也满足KKT条件.

相似地,减量式标准支持向量机算法更新法则的基本思想是:令待删除样本为 (x_c, y_c) ,在严格保持 $S - \{(x_c, y_c)\}$ 中所有样本满足KKT等价条件的情况下,逐步减少待删样本 (x_c, y_c) 的权值 α_c ,如此反复,直至 $\alpha_c = 0$.

1.1 KKT等价条件

根据凸优化理论^[17],通过优化问题(1)的求解,同样可以得到标准支持分类机^[1]的解:

$$\min_{0 \leq \alpha_i \leq C} W = \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^l \alpha_i + b \left(\sum_{i=1}^l y_i \alpha_i \right) \quad (1)$$

其中, b 是拉格朗日乘子.

根据KKT定理^[16],对 W 求一阶偏导,可以导出下列的KKT等价条件:

$$g_i = \sum_j \alpha_j y_i y_j K(x_i, x_j) + y_i b - 1 \begin{cases} \geq 0, & \text{对于 } \alpha_i = 0 & S_R \\ = 0, & \text{对于 } C > \alpha_i > 0 & S_S \\ \leq 0, & \text{对于 } \alpha_i = C & S_E \end{cases} \quad (2)$$

$$\sum_{i=1}^l y_i \alpha_i = 0$$

公式(2)即称为标准支持分类机的KKT等价条件.

相应地,公式(2)将训练集 S 划分成3个子集: S_R, S_S 和 S_E ^[11],其中, S_S 为边缘支持向量集, S_E 为错分支持向量集, S_R 为剩余向量集.

1.2 绝缘更新法则

当向训练集 $S = \{(x_1, y_1), \dots, (x_l, y_l)\}$ 中增加一个新样本 (x_c, y_c) , 并且 α_c 和 g_c 不满足 KKT 条件时, 那么需要相应地调整 α_c 和 α_S , 以找到扩增样本集的最优解. 相似地, 当从训练集 S 中删除样本 (x_c, y_c) , 且 $\alpha_c > 0$ 时, 那么相应地也需要调整 α_c 和 $\alpha_{S - \{(x_c, y_c)\}}$, 以找到收缩样本集的最优解. 其中, $\alpha_S = [\alpha_1, \dots, \alpha_l]^T$ 表示训练集 S 中所有样本对应的 α 乘子向量; 相应地, $\alpha_{S - \{(x_c, y_c)\}}$ 表示集合 $S - \{(x_c, y_c)\}$ 中所有样本对应的 α 乘子向量. 本节将针对调整 α_c 介绍绝缘更新法则, 其基本原则是: 在 α_c 的调整过程中, 保证集合 S 或 $S - \{(x_c, y_c)\}$ 内的所有样本都满足 KKT 等价条件.

在对 α_c 的调整过程中, 为了保证集合 S 或 $S - \{(x_c, y_c)\}$ 内的所有样本都满足 KKT 等价条件, 那么仅有边缘支持向量集 $S_S = \{i \in S : g_i = 0 \wedge \alpha_i > 0 \wedge \alpha_i > C\}$ 中的 $|S_S|$ 个样本对应的 α 乘子以及拉格朗日乘子 b 可以调整. 根据 KKT 条件 (2), 则可以获得如下 $|S_S| + 1$ 个线性方程**:

$$\left. \begin{aligned} \Delta g_i &= \sum_{j \in S_S} \Delta \alpha_j Q_{ij} + y_i \Delta b + \Delta \alpha_c Q_{ic} = 0 \\ \sum_{j \in S_S} y_j \Delta \alpha_j + y_c \Delta \alpha_c &= 0 \end{aligned} \right\} \quad (3)$$

其中, $Q_{ij} = y_i y_j K(x_i, x_j)$.

可以看出, 线性方程组 (3) 含有一个自由度.

更进一步地, 线性方程组 (3) 可以变换为

$$Q' \cdot \begin{bmatrix} \Delta b \\ \Delta \alpha_{S_S} \end{bmatrix} = - \begin{bmatrix} y_c \\ Q_{S_S c} \end{bmatrix} \Delta \alpha_c \quad (4)$$

其中, $Q' = \begin{bmatrix} 0 & y_{S_S}^T \\ y_{S_S} & Q_{S_S S_S} \end{bmatrix}$.

这样, 可以得到 α_c 和 α_{S_S} 与 b 的线性变化关系. 其中, α_{S_S} 表示边缘支持向量集 S_S 中的所有样本对应的 α 乘子向量:

$$\left. \begin{aligned} \Delta \alpha_{S_S} &= \beta_{S_S}^c \Delta \alpha_c \\ \Delta b &= \beta_b^c \Delta \alpha_c \end{aligned} \right\} \quad (5)$$

其中, $\begin{bmatrix} \beta_b^c \\ \beta_{S_S}^c \end{bmatrix} = -R \cdot \begin{bmatrix} y_c \\ Q_{S_S c} \end{bmatrix}, R = Q'^{-1}$.

相应地, 也可以得到 α_c 与 g_i 的线性变化关系:

$$\Delta g_i = \gamma_i^c \Delta \alpha_c \quad (6)$$

其中, $\gamma_i^c = Q_{ic} + \sum_{j \in S_S} Q_{ij} \beta_j^c + y_i \beta_b^c, \forall i \in S$.

1.3 增量式标准支持向量机算法

绝缘更新法则不能直接给出问题的最优解, 问题在于: 随着 α_c 值的增加, 会改变集合 S_S, S_R 和 S_E 的组成成分. 解决方法是捕获集合 S_S, S_R 和 S_E 的最小变化, 即计算最大调整量 $\Delta \alpha_c^{\max}$, 使得正好某一样本在集合 S_S, S_R 和 S_E 间要进行迁移. 具体需考虑如下 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).

首先, 计算集合 $I_+^{S_S} = \{i \in S_S : \beta_i^c > 0\}, I_-^{S_S} = \{i \in S_S : \beta_i^c < 0\}$.

这样, 可能的权更新是:

** 公式 (3) 对于减量式调整也成立. 在减量式调整过程中, 如果待删除样本 (x_c, y_c) 在调整前属于边缘支持向量集 S_S , 而在调整过程中, 根据引理 2 可知, 样本 (x_c, y_c) 将不再满足 KKT 等价条件, 那么 (x_c, y_c) 在调整过程中将不属于边缘支持向量集 S_S ; 如果待删除样本 (x_c, y_c) 在调整前属于边缘支持向量集 S_E , 相应地也可知道, 在调整过程中, (x_c, y_c) 也将不属于边缘支持向量集 S_S , 所以在公式 (3) 中, $\Delta \alpha_c$ 将不会被重复计算.

$$\Delta\alpha_i^{\max} = \begin{cases} C - \alpha_i, & i \in I_+^{S_S} \\ -\alpha_i, & i \in I_-^{S_S} \end{cases} \quad (7)$$

那么,在集合 S_S 中的某个样本移动到集合 S_R 或 S_E 之前存在一个最大可能更新值:

$$\Delta\alpha_c^{S_S} = \min_{i \in I_+^{S_S} \cup I_-^{S_S}} \frac{\Delta\alpha_i^{\max}}{\beta_i^c}.$$

(2) 在 S_R 或 S_E 中的某个样本的 g_i 函数到达 0.

计算集合 $I_+^{S_E} = \{i \in S_E : \gamma_i^c > 0\}$, $I_-^{S_R} = \{i \in S_R : \gamma_i^c < 0\}$.

那么,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大可能更新值:

$$\Delta\alpha_c^{S_{R,E}} = \min_{i \in I_+^{S_E} \cup I_-^{S_R}} \frac{-g_i}{\gamma_i^c}.$$

(3) 新增样本 (x_c, y_c) 满足 KKT 条件.

那么,可以在样本 (x_c, y_c) 达到 KKT 条件之前存在一个最大可能更新值:

$$\Delta\alpha_c^{\alpha, g} = \min \left\{ \frac{-g_c}{\gamma_c^c}, C - \alpha_c \right\}.$$

最后,3 个值中的最小值:

$$\Delta\alpha_c^{\max} = \min \{ \Delta\alpha_c^{S_S}, \Delta\alpha_c^{S_{R,E}}, \Delta\alpha_c^{\alpha, g} \} \quad (8)$$

将构成 $\Delta\alpha_c$ 的最大调整量.

基于 $\Delta\alpha_c$ 的最大调整量的计算法则,当一个新样本 (x_c, y_c) 增加到集合 S 时,扩增样本集的最优解可以通过算法 1 求解获得.

算法 1. 增量式标准支持向量机算法.

输入: $S, (x_c, y_c), \alpha_S$;

输出: α_S, α_c .

1. 初始化新样本 (x_c, y_c) 的 Lagrange 乘子为 0;
2. 如果样本 (x_c, y_c) 满足 KKT 条件,那么算法结束;
3. 否则,应用最大调整量 $\Delta\alpha_c^{\max}$,使得恰好有一个样本在 S_R, S_S 和 S_E 间移动;
4. 更新矩阵 R ,如果样本 (x_c, y_c) 满足 KKT 条件,那么跳到步骤 2;否则,跳到步骤 3.

1.4 减量式标准支持向量机算法

与增量式标准支持向量机算法相似,减量式支持向量机算法需要计算最小调整量 $\Delta\alpha_c^{\min}$,使得正好某一样本在集合 S_S, S_R 和 S_E 间要进行迁移,相应地,也需考虑如下 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).

首先,计算集合 $I_-^{S_S} = \{i \in S_S : \beta_i^c < 0\}$, $I_+^{S_S} = \{i \in S_S : \beta_i^c > 0\}$.

这样,可能的权更新是:

$$\Delta\alpha_i^{\max} = \begin{cases} C - \alpha_i, & i \in I_-^{S_S} \\ -\alpha_i, & i \in I_+^{S_S} \end{cases} \quad (9)$$

那么,在集合 S_S 中的某个样本移动到集合 S_R 或 S_E 之前存在一个最大可能更新值:

$$\Delta\alpha_c^{S_S} = \max_{i \in I_+^{S_S} \cup I_-^{S_S}} \frac{\Delta\alpha_i^{\max}}{\beta_i^c}.$$

(2) 在 S_R 或 S_E 中的某个样本的 g_i 函数到达 0.

计算集合 $I_-^{S_E} = \{i \in S_E : \gamma_i^c < 0\}$, $I_+^{S_R} = \{i \in S_R : \gamma_i^c > 0\}$.

那么,在集合 S_R 或 S_E 中的某个样本移动到 S_S 集合之前存在一个最大可能更新值:

$$\Delta\alpha_c^{S_{R,E}} = \max_{i \in I^{S_E} \cup I_c^{S_R}} \frac{-g_i}{\gamma_i^c}$$

(3) 待删除样本 (x_c, y_c) 的系数 $\alpha_c=0$.

那么,可以在样本 (x_c, y_c) 的系数 α_c 到达 0 之前存在一个最大可能更新值:

$$\Delta\alpha_c^{\alpha,g} = \max \left\{ \frac{g_c}{\gamma_c^c}, -\alpha_c \right\}$$

最后,3 个值中的最小值:

$$\Delta\alpha_c^{\min} = \max \{ \Delta\alpha_c^{S_S}, \Delta\alpha_c^{S_{R,E}}, \Delta\alpha_c^{\alpha,g} \} \tag{10}$$

将构成 $\Delta\alpha_c$ 的最小调整量.

基于 $\Delta\alpha_c$ 的最小调整量的计算法则,当一个样本 (x_c, y_c) 从训练集合 S 中删除时,收缩样本集的最优解可以通过算法 2 求解获得.

算法 2. 减量式标准支持向量机算法.

输入: $S, (x_c, y_c), \alpha_S$;

输出: $\alpha_{S-\{(x_c, y_c)\}}$.

1. 如果待删除样本 (x_c, y_c) 的系数 $\alpha_c=0$,那么算法结束;
2. 否则,应用最小调整量 $\Delta\alpha_c^{\min}$,使得恰好有一个样本在 S_R, S_S 和 S_E 间移动;
3. 更新矩阵 R ,如果 $\alpha_c=0$,那么跳到步骤 1;否则,跳到步骤 2.

1.5 矩阵 R 的更新

α_c 的调整将改变边缘支持向量集 S_S ,所以矩阵 R 相应地也会发生调整.对于集合 S_S 的变化有以下两种情形:

第 1 种是增加一个样本到集合 S_S ;

第 2 种是从集合 S_S 移出一个样本.

对于第 1 种情形,矩阵 R 根据如下规则扩展:

$$R^* \leftarrow \begin{bmatrix} R & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\gamma_t^t} \begin{bmatrix} \beta_b^t \\ \beta_{S_S}^t \\ 1 \end{bmatrix} \begin{bmatrix} \beta_b^t \\ \beta_{S_S}^t \\ 1 \end{bmatrix}^T,$$

其中, t 代表增加到集合 S_S 的一个样本.

对于第 2 种情形,矩阵 R 根据如下规则进行收缩:

$\forall i, j \in S_S$, 并且 $i \neq t, j \neq t$, 有

$$R_{ij}^- \leftarrow R_{ij} - R_{it}^{-1} R_{it} R_{tj},$$

其中, t 代表集合 S_S 中一个移出的样本.

2 可行性分析

为了说明 C&P 算法(增量式支持向量机算法和减量式支持向量机算法)的每步计算是可靠的,需要回答绝缘更新过程中一些关键的问题,这些问题具体包括:

- 1) 在调整 α_c 过程中,边缘支持向量集 S_S 始终不为空(定理 1);
- 2) 矩阵 Q' 始终存在逆矩阵(定理 2);
- 3) 对矩阵 R 的更新策略是正确的(定理 3);
- 4) 向边缘支持向量集 S_S 新加入(移出)一个样本 t 后,在紧接着的调整中, t 不会反向移出(加入)边缘支持向量集 S_S (定理 4 和定理 5).

需要说明的是:首先,增量式标准支持向量机算法(算法 1)和减量式标准支持向量机算法(算法 2)使用相同的绝缘更新法则,故本节在证明过程中并没有明显区分这两种算法,并且本节所证明的结论对于这两种算法都

成立;其次,定理 1 和定理 4 在文献[12]的第 4 节中已有证明,本文为了全面探讨 C&P 算法的可行性,重新组织了其证明过程;最后,除了定理 1 和定理 4,其余结论皆为本文的工作。

定理 1. 在 α_c 的调整过程中,边缘支持向量集 S_S 始终不为空。

证明:如果 $S_S = \emptyset$,可以改变 b 的值使一个样本移入集合 S_S ,从而使 S_S 不为空.这样,需要考虑改变 b 值的方向,使得移入 S_S 的样本在接下来的调整中不会移出.假设 $S_S = \{(x_i, y_i)\}$,根据公式(5)可以证明, β_i^c 的正负号由 $-y_i y_c$ 决定;再结合公式(2),可得 b 值的改变方向必须为 y_c 才能保证移入 S_S 的样本在接下来的调整中不会移出.据此,又定义可行性条件 $I_+^{S_R} = \{i \in S_R : -y_i y_c > 0\}$, $I_-^{S_E} = \{i \in S_E : -y_i y_c < 0\}$.那么,在 S_R 或 S_E 中的一个样本移入 S_S 之前,最大可能的 Δb 的绝对值可以表示为 $|\Delta b^{S_R, S_E}| = \min_{i \in I_+^{S_R} \cup I_-^{S_E}} |g_i|$;在候选样本 c 移入 S_S 之前,最大可能的 Δb 的绝对值可以表示为 $|\Delta b^c| = -g_c$,那么 b 的最后改变值为 $\Delta b^* = y_c \min\{|\Delta b^{S_R, S_E}|, |\Delta b^c|\}$.定理得证. \square

引理 1. 矩阵 Q 是半正定的。

证明:令 $\alpha = (\alpha_1, \dots, \alpha_n) \neq 0$,那么有

$$\alpha^T Q \alpha = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j Q_{ij} = \left(\sum_{i=1}^n \alpha_i y_i \phi(x_i) \right) \cdot \left(\sum_{i=1}^n \alpha_i y_i \phi(x_i) \right) \geq 0.$$

引理得证. \square

假设 1:矩阵 Q 是正定的。

如引理 1 的证明,当且仅当 $\{y_1 \phi(x_1), \dots, y_n \phi(x_n)\}$ 是线性无关时,矩阵 Q 才是正定的.而实际上, $\{y_1 \phi(x_1), \dots, y_n \phi(x_n)\}$ 是线性相关的可能性是相当小的^[2].例如,如果 $K(x_i, x_j) = e^{-\|x_i - x_j\|^2}$,其中 $x_i \neq x_j$,那么矩阵 Q 便是正定的,因为矩阵 Q 的行数和列数相比较被核函数影射到的空间维数是相当小的,所以矩阵 Q 趋向于正定,故假设 1 一般情况下成立^[2].

定理 2. 矩阵 Q' 始终存在逆矩阵。

证明:根据定理 1 可知,在调整 α_c 的过程中,集合 S_S 始终不为空,则不失一般性,假设 S_S 存在 n 个样本,那么,

$$Q' = \begin{bmatrix} 0 & y_1 & \cdots & y_n \\ y_1 & Q_{11} & \cdots & Q_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ y_n & Q_{n1} & \cdots & Q_{nn} \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & y^T \\ y & Q \end{bmatrix}.$$

通过假设 1 可知 Q 正定,那么必存在 Q^{-1} ,并且 Q^{-1} 也是正定的。

$$\text{因为} \begin{bmatrix} 1 & -y^T Q^{-1} \\ 0 & I_n \end{bmatrix} \cdot \begin{bmatrix} 0 & y^T \\ y & Q \end{bmatrix} = \begin{bmatrix} -y^T Q^{-1} y^T & 0 \\ y & Q \end{bmatrix},$$

所以有 $\det(Q') = \det(Q) \cdot (-y^T Q^{-1} y) \neq 0$,

那么易知,矩阵 Q' 始终存在逆矩阵. \square

引理 2. 当向集合 S_S 增加一个样本 t 时,相应地, $\gamma_t^c > 0$ 。

证明:首先证明结论 $y_t \phi(x_t) + \sum_{j \in S_S} \beta_j^t y_j \phi(x_j) \neq 0$ 。

假设结论不成立,那么有 $y_t \phi(x_t) = -\sum_{j \in S_S} \beta_j^t y_j \phi(x_j)$,把它带入公式(6),可以得到 $\gamma_t^c = 0$,这与前提条件向集合 S_S 增加一个样本 t 相矛盾.基于此,进一步有

$$\begin{aligned}
 \gamma'_t &= Q_u + \sum_{j \in S_S} Q_{ij} \beta_j^c + y_t \beta_b^t = y_t \phi(x_t) \cdot \left(y_t \phi(x_t) + \sum_{j \in S_S} \beta_j^t y_j \phi(x_j) \right) + \beta_b^t y_t \\
 &= y_t \phi(x_t) \cdot \left(y_t \phi(x_t) + \sum_{j \in S_S} \beta_j^t y_j \phi(x_j) \right) - \beta_b^t \sum_{j \in S_S} \beta_j^t y_j \\
 &= y_t \phi(x_t) \cdot \left(y_t \phi(x_t) + \sum_{j \in S_S} \beta_j^t y_j \phi(x_j) \right) - \sum_{j \in S_S} \beta_j^t \left(-y_j \phi(x_j) \cdot \left(y_t \phi(x_t) + \sum_{j \in S_S} \beta_j^t y_j \phi(x_j) \right) \right) \\
 &= \left(y_t \phi(x_t) + \sum_{j \in S_S} \beta_j^t y_j \phi(x_j) \right)^2 > 0.
 \end{aligned}$$

引理 3. 当向集合 S_S 移出一个样本 t 时,相应地, $R_{tt} > 0$.

根据逆矩阵的定义和假设 1,可以很容易地得出该结论.

定理 3. 对于集合 S_S 增加一个样本或移出一个样本 t 的两种情形,其相应的矩阵 R 的两种更新策略均是正确的.

证明:对于集合 S_S 增加一个样本 t ,我们可以证明:

$$Q^+ \cdot R^+ = \begin{pmatrix} 0 & y_1 & \cdots & y_n & y_t \\ y_1 & Q_{11} & \cdots & Q_{1n} & Q_{1t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_n & Q_{n1} & \cdots & Q_{nn} & Q_{nt} \\ y_t & Q_{t1} & \cdots & Q_{tn} & Q_{tt} \end{pmatrix} \cdot \left(\begin{bmatrix} R & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 \cdots & 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix}^T \right) = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \gamma'_t \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 1 \end{bmatrix},$$

其中,等号 1 应用了公式(6), I 为相应的单位矩阵.对于集合 S_S 移出一个样本 t ,我们可以证明:

$$Q^- \cdot R^- = Q^- \cdot R_{tt} - \frac{Q^-}{R_{tt}} (R_{*t} \cdot R_{t*})_{tt} = Q^- \cdot R_{tt} + (Q_{*t} \cdot R_{t*})_{tt} = I$$

定理 4. 向集合 S_S 新加入一个样本 t 后,在紧接着的调整中 t 不会反向移出集合 S_S .

证明:假设前面一次调整记为 k ,则下次调整记为 $k+1$,那么有,

$$\begin{aligned}
 \beta^{[k+1]} &= \begin{bmatrix} \beta_b^{[k+1]} \\ \beta_1^{[k+1]} \\ \vdots \\ \beta_n^{[k+1]} \\ \beta_t^{[k+1]} \end{bmatrix} = -R^+ \cdot \begin{bmatrix} y_c \\ Q_{1c} \\ \vdots \\ Q_{nc} \\ Q_{tc} \end{bmatrix} = - \left(\begin{bmatrix} R & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 \cdots & 0 & 0 \end{bmatrix} + \frac{1}{\gamma'_t} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix}^T \right) \cdot \begin{bmatrix} y_c \\ Q_{1c} \\ \vdots \\ Q_{nc} \\ Q_{tc} \end{bmatrix} \\
 &= \begin{bmatrix} \beta_b^{[k]} \\ \beta_1^{[k]} \\ \vdots \\ \beta_n^{[k]} \\ 0 \end{bmatrix} - \frac{1}{\gamma'_t} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix}^T \cdot \begin{bmatrix} y_c \\ Q_{1c} \\ \vdots \\ Q_{nc} \\ Q_{tc} \end{bmatrix} = \begin{bmatrix} \beta_b^{[k]} \\ \beta_1^{[k]} \\ \vdots \\ \beta_n^{[k]} \\ 0 \end{bmatrix} - \frac{\gamma'_c}{\gamma'_t} \begin{bmatrix} \beta_b^t \\ \beta_1^t \\ \vdots \\ \beta_n^t \\ 1 \end{bmatrix}.
 \end{aligned}$$

于是有,

$$\beta_t^{[k+1]} = -\frac{\gamma'_c}{\gamma'_t}.$$

又因为

$$\begin{aligned}
 \gamma_i^{[k+1]} &= Q_{ic} + \sum_{j \in S_S^{[k+1]}} Q_{jt} \beta_j^{[k+1]} + y_t \beta_b^{[k+1]} \\
 &= Q_{ic} + \sum_{j \in S_S} \beta_j^{[k]} Q_{jt} + y_t \beta_b^{[k]} - \frac{\gamma_i^t}{\gamma_i^c} \cdot \left(Q_{it} + \sum_{j \in S_S} \beta_j^t Q_{jt} + y_t \beta_b^t \right) \\
 &= \gamma_i^c - \frac{\gamma_i^t}{\gamma_i^c} \cdot \gamma_i^t \\
 &= \gamma_i^c - \gamma_i^t \\
 &= 0.
 \end{aligned}$$

所以, $\beta_i^{[k+1]} = -\frac{\gamma_i^c}{\gamma_i^t}$.

这意味着向集合 S_S 新加入一个样本 t 后,在紧接着的调整中, t 不会反向移出 S_S .因为对于 t 加入 S_S 有两种方式:一是从 S_R 移向 S_S ,二是从 S_E 移向 S_S .对于第 1 种方式,有 $\alpha_i^k = 0, g_i^k > 0$ 且 $\gamma_i^k < 0$,那么通过引理 2 可得 $\beta_i^{[k+1]} > 0$;对于第 2 种方式,有 $\alpha_i^k = C, g_i^k < 0$ 且 $\gamma_i^k > 0$,那么通过引理 2 可得 $\beta_i^{[k+1]} < 0$.所以在紧接着的调整中, t 不会反向移出 S_S . □

定理 5. 从集合 S_S 移出一个样本 t 后,在紧接着的调整中, t 不会反向加入集合 S_S .

证明:假设前面一次调整记为 k ,则下次调整记为 $k+1$,那么有:

$$\begin{aligned}
 \beta^{[k+1]} &= \begin{bmatrix} \beta_b^{[k+1]} \\ \beta_1^{[k+1]} \\ \vdots \\ \beta_{n-1}^{[k+1]} \end{bmatrix} = -R^- \cdot \begin{bmatrix} y_c \\ Q_{ic} \\ \vdots \\ Q_{(n-1)c} \end{bmatrix} = - \left(R_{n \times n} - \frac{1}{R_{tt}} \begin{bmatrix} R_{1t} R_{t1} & \cdots & R_{1t} R_{tm} \\ \vdots & \ddots & \vdots \\ R_{nt} R_{t1} & \cdots & R_{nt} R_{tm} \end{bmatrix} \right) \cdot \begin{bmatrix} y_c \\ Q_{ic} \\ \vdots \\ Q_{(n-1)c} \end{bmatrix} \\
 &= \begin{bmatrix} \beta_b^{[k]} - R_{1t} Q_{ic} \\ \beta_1^{[k]} - R_{2t} Q_{ic} \\ \vdots \\ \beta_{n-1}^{[k]} - R_{nt} Q_{ic} \end{bmatrix} + \frac{1}{R_{tt}} \begin{bmatrix} -R_{1t} \beta_t^{[k]} + R_{1t} R_{tt} Q_{ic} \\ -R_{2t} \beta_t^{[k]} + R_{2t} R_{tt} Q_{ic} \\ \vdots \\ -R_{nt} \beta_t^{[k]} + R_{nt} R_{tt} Q_{ic} \end{bmatrix},
 \end{aligned}$$

$$\begin{aligned}
 \gamma_i^{[k+1]} &= Q_{ic} + \sum_{j \in S_S^{[k+1]}} Q_{jt} \beta_j^{[k+1]} + y_t \beta_b^{[k+1]} = Q_{ic} + \sum_{j \in S_S^{[k+1]}} Q_{jt} (\beta_j^{[k]} - R_{j+1t} Q_{ic}) + y_t (\beta_b^{[k]} - R_{1t} Q_{ic}) + \frac{y_t}{R_{tt}} (-R_{1t} \beta_t^{[k]} + R_{1t} R_{tt} Q_{ic}) + \\
 &\quad \frac{1}{R_{tt}} \left(\sum_{j \in S_S^{[k+1]}} Q_{jt} (-R_{j+1t} \beta_t^{[k]} + R_{j+1t} R_{tt} Q_{ic}) \right) = \gamma_i^c - \frac{\beta_t^k}{R_{tt}} \cdot \left(Q_{tt} R_{tt} + \sum_{j \in S_S^{[k+1]}} Q_{jt} R_{jt} + y_t R_{1t} \right) = -\frac{\beta_t^k}{R_{tt}},
 \end{aligned}$$

其中,等号 1 应用了 γ_i^c 的定义,等号 2 应用了等式 $Q_{tt} R_{tt} + \sum_{j \in S_S^{[k+1]}} Q_{jt} R_{jt} + y_t R_{1t} = 1$.类似于定理 4 后面的分析,可以得出结论:从集合 S_S 移出一个样本 t 后,在紧接着的调整中 t 不会反向加入集合 S_S . □

3 有限收敛性分析

有限收敛性将确保 C&P 算法(增量式支持向量机算法和减量式支持向量机算法)在有限步后收敛到问题的最优解.在证明过程中,我们首先证明了增量式支持向量机算法在绝缘调整过程中,目标函数 W 是严格单调递减的(见定理 6);而减量式标准支持向量机算法在调整过程中,目标函数 W 是严格单调递增的(见定理 7).在此基础上,证明了增量式支持向量机算法和减量式标准支持向量机算法经过有限步迭代后,将收敛于相应问题的最优解(见定理 7).

引理 4. 在绝缘更新过程中, $\forall i \in S_S, \beta_i^c \neq \pm\infty$, 并且 $\forall i \in S-S_S, \gamma_i^c \neq \pm\infty$.

证明:根据公式(5), $\forall i \in S_S$,我们有,

$$\beta_i^c = -\sum_{k \in S_S} R_{ik} Q_{kc} - R_{i1} y_c = -\frac{1}{\det(Q')} \left(\sum_{k \in S_S} (-1)^{i+k} \det(Q'_{,ki}) Q_{kc} + y_c (-1)^{i+1} \det(Q'_{,i1}) \right).$$

假设存在一个样本 $i \in S_S$, 使得 $\beta_i^c \neq \pm\infty$, 那么必然有 $\det(Q')=0$, 这与定理 2 相矛盾. 所以 $\forall i \in S_S, \beta_i^c \neq \pm\infty$; 并根据公式(6), 也有 $\forall i \in S-S_S, \gamma_i^c \neq \pm\infty$. □

引理 5. 对于增量式标准支持向量机算法, 相应的最大调整量 $\Delta\alpha_c^{\max} > 0$; 对于减量式标准支持向量机算法, 相应的最小调整量 $\Delta\alpha_c^{\min} < 0$.

根据第 1.3 节和第 1.4 节的内容以及引理 4, 较容易推导出引理 5 的结论.

定理 6. 对于增量式标准支持向量机算法, 调整过程中, 目标函数 W 是严格单调递减的.

证明: 假设前面一次调整记为 k , 则下一次调整记为 $k+1$, 并令 $\beta_{S_R} = 0, \beta_{S_E} = 0, \beta_c = 1$, 那么有,

$$\begin{aligned} W^{[k+1]} &= \frac{1}{2} \sum_{i_1, i_2 \in S} (\alpha_{i_1}^{[k]} + \beta_{i_1}^{[k]} \Delta\alpha_c^{[k]})(\alpha_{i_2}^{[k]} + \beta_{i_2}^{[k]} \Delta\alpha_c^{[k]}) Q_{i_1 i_2} + \\ &\quad (b^{[k]} + \beta_b^{[k]} \Delta\alpha_c^{[k]}) \left(\sum_{i \in S} y_i (\alpha_i^{[k]} + \beta_i^{[k]} \Delta\alpha_c^{[k]}) \right) - \sum_{i \in S} (\alpha_i^{[k]} + \beta_i^{[k]} \Delta\alpha_c^{[k]}) \\ &= W^{[k]} + \sum_{i \in S} g_i^{[k]} \beta_i^{[k]} \Delta\alpha_c^{[k]} + \frac{1}{2} \sum_{i \in S} \gamma_i^{[k]} \beta_i^{[k]} (\Delta\alpha_c^{[k]})^2 \\ &= W^{[k]} + g_c^{[k]} \Delta\alpha_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} (\Delta\alpha_c^{[k]})^2 \\ &= W^{[k]} + \left(g_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} \Delta\alpha_c^{[k]} \right) \Delta\alpha_c^{[k]}. \end{aligned}$$

容易验证 $g_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} \Delta\alpha_c^{[k]} < 0$, 因此根据引理 5 可知, $W^{[k+1]} - W^{[k]} < 0$. □

定理 7. 对于减量式标准支持向量机算法, 调整过程中, 目标函数 W 是严格单调递减的.

证明: 假设前面一次调整记为 k , 则下一次调整记为 $k+1$, 与定理 6 的证明相似, 我们可以证明出相似的结论:

$$W^{[k+1]} - W^{[k]} = \left(g_c^{[k]} + \frac{1}{2} \gamma_c^{[k]} \Delta\alpha_c^{[k]} \right) \Delta\alpha_c^{[k]}.$$

在减量式标准支持向量机算法的调整过程中, 有 $g_c^{[k]} \leq 0, \Delta\alpha_c^{[k]} < 0, \gamma_c^{[k]} > 0$.

故 $W^{[k+1]} - W^{[k]} > 0$. 定理得证. □

定理 8. 对于增量式标准支持向量机算法以及减量式标准支持向量机算法, 经过有限步迭代后, 将收敛于相应问题的最优解.

证明: 关于增量式标准支持向量机算法以及减量式标准支持向量机算法有限收敛性的证明, 下文仅针对增量式标准支持向量机算法给出了证明, 相似地, 可以证明减量式标准支持向量机算法的有限收敛性.

如果 $(\alpha_c^{[1]}, \alpha_c^{[2]}, \alpha_c^{[3]}, \dots), (g_c^{[1]}, g_c^{[2]}, g_c^{[3]}, \dots), (W^{[1]}, W^{[2]}, W^{[3]}, \dots)$ 被定义为增量式标准支持向量机算法调整过程中生成的 3 个序列, 那么为了证明定理, 我们只需证明序列 $(\alpha_c^{[1]}, \alpha_c^{[2]}, \alpha_c^{[3]}, \dots)$ 和 $(g_c^{[1]}, g_c^{[2]}, g_c^{[3]}, \dots)$ 是有限序列且收敛到问题对应的 KKT 条件.

假设 $(\alpha_c^{[k]}), (g_c^{[k]})$ 和 $(W^{[k]})$ 序列是 3 个无限序列, 由定理 6 我们有 $\Delta W^{[k]} = W^{[k+1]} - W^{[k]} < 0$, 那么 $(W^{[k]})$ 必须是一个无限的单调递减序列. 进一步地, 易证每个 $W^{[k]}$ 对应着唯一一个边缘支持向量集 S_S , 而边缘支持向量集 S_S 的组合数是有限的, 所以 $(W^{[k]})$ 序列是一个有限序列. 因此 $(\alpha_c^{[k]})$ 和 $(g_c^{[k]})$ 也是有限序列, 它们可以表示为 $(\alpha_c^{[1]}, \alpha_c^{[2]}, \dots, \alpha_c^{[\ell]})$ 和 $(g_c^{[1]}, g_c^{[2]}, \dots, g_c^{[\ell]})$. 根据引理 4 和引理 5, 容易验证 $\Delta\alpha_c^{[k]} > 0$ 和 $\Delta g_c^{[k]} = \gamma_c^{[k]} \Delta\alpha_c^{[k]} > 0$, 那么 $(\alpha_c^{[1]}, \alpha_c^{[2]}, \dots, \alpha_c^{[\ell]})$ 和 $(g_c^{[1]}, g_c^{[2]}, \dots, g_c^{[\ell]})$ 是两个有限单调递增序列.

接下来, 假设 $(\alpha_c^{[1]}, \alpha_c^{[2]}, \dots, \alpha_c^{[\ell]})$ 和 $(g_c^{[1]}, g_c^{[2]}, \dots, g_c^{[\ell]})$ 不收敛到问题对应的 KKT 条件, 也就是 $\alpha_c^{[\ell]} < C$ 并且 $g_c^{[\ell]} < 0$, 而这与绝对增量调整终止在第 ℓ 步的调整事实不符. 因此, 序列 $(\alpha_c^{[1]}, \alpha_c^{[2]}, \dots, \alpha_c^{[\ell]})$ 和 $(g_c^{[1]}, g_c^{[2]}, \dots, g_c^{[\ell]})$ 收

敛到问题对应的 KKT 条件. □

4 实验

本文基于 MATLAB 实现了增量式标准支持向量机算法(incremental support vector machine,简称 ISVM)和减量式标准支持向量机算法(decremental support vector machine,简称 DSVM).实验中,对可行性与有限收敛性分析中的主要结论都分别作了验证,这些结论包括:1) 边缘支持向量集 S_S 不为空;2) 矩阵 Q' 始终存在逆矩阵;3) 向边缘支持向量集 S_S 新加入(移出)一个样本 t 后,在紧接着的调整中 t 不会反向移出(加入)边缘支持向量集;4) ISVM 和 DSVM 在有限步内能收敛到相应问题的最优解.

所有实验在 3.2GHz,Pentium-4,具有 1GB 的内存和 MATLAB7.5 平台的机器上进行.表 1 给出了在实验中使用的 4 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).实验中使用了 3 种核函数,分别为:线性(linear)核 $K(x_i, x_j) = \langle x_i, x_j \rangle$;多项式(polynomial)核 $K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d$,其中, $d=2$;高斯(Gaussian)核 $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2)$,其中, $\sigma=0.707$.

Table 1 Benchmark datasets used in the experiments

表 1 实验中使用的标准数据集

数据集名称	数据集的大小	属性集大小
Sonar	208	60
Ionosphere	351	34
Diabetes	768	8
Breast cancer	683	10

4.1 Sonar数据

Sonar 数据集的任务是区别来自金属圆筒返回的声纳信号还是来自粗糙的圆柱形岩石返回的声纳信号.该数据集包含具有 60 个连续数值属性的 208 个实例样本.

为了说明在 ISVM 和 DSVM 学习中边缘支持向量集 S_S 不为空、矩阵 Q' 存在逆矩阵以及向边缘支持向量集 S_S 新加入(移出)一个样本 t 后,在紧接着的调整中, t 不会反向移出(加入)边缘支持向量集,实验分别在数据量为 50,100,150,200 的基础上,分别尝试 200 次 ISVM 和 DSVM 学习,测试结果均验证了上述的结论.图 1 说明了 ISVM 和 DSVM 在不同训练数据大小情况下的平均迭代次数,数据结果验证了算法的有限收敛性,尤其是算法在高斯核下具有更快速的收敛性.

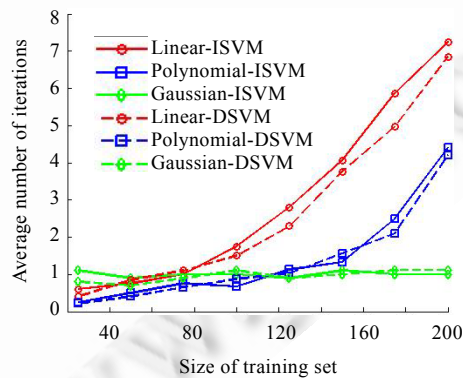


Fig.1 Average numbers of iterations on Sonar dataset

图 1 Sonar 数据集上的平均迭代次数

4.2 Ionosphere数据

Ionosphere 数据集收集来自 16 个高频天线的相控阵的雷达数据.数据的目标是来自大气层的自由电子,数

据的标签是“好”(“好”表示雷达数据反映了大气层的结构)和“坏”.这个数据集包含具有 34 个连续数值属性的 351 个实例样本.

实验分别在数据量为 80,160,240,320 的基础上,分别尝试 200 次 ISVM 和 DSVM 学习,测试结果验证了如下的结论:边缘支持向量集 S_S 不为空、矩阵 Q' 存在逆矩阵以及向边缘支持向量集 S_S 新加入(移出)一个样本 t 后,在紧接着的调整中, t 不会反向移出(加入)边缘支持向量集.图 2 说明了 ISVM 和 DSVM 在不同训练数据大小情况下的平均迭代次数,数据结果验证了算法的有限收敛性.

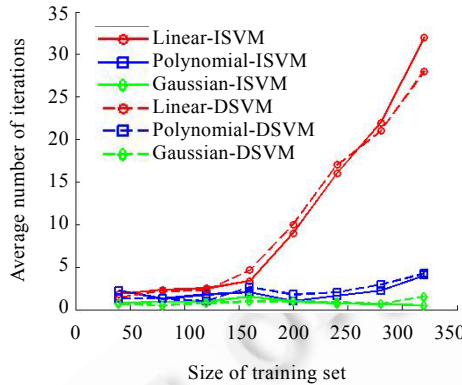


Fig.2 Average numbers of iterations on Ionosphere dataset

图 2 Ionosphere 数据集上的平均迭代次数

4.3 Diabetes 数据

Diabetes 数据集是给 21 岁及以上皮玛族裔本土美国妇女的诊断信息.该数据集由 768 位妇女的诊断信息构成,其中,268 个病人被诊断为患糖尿病,而另 500 人被诊断为非患糖尿病.数据集拥有 8 个特征,其中 6 个特征是各种临床检测的连续数值属性,剩下的 2 个属性是年龄和怀孕次数的离散数值信息.

实验分别在数据量为 170,340,510,680 的基础上,分别尝试 200 次 ISVM 和 DSVM 学习,测试结果验证了如下的结论:边缘支持向量集 S_S 不为空、矩阵 Q' 存在逆矩阵以及向边缘支持向量集 S_S 新加入(移出)一个样本 t 后,在紧接着的调整中, t 不会反向移出(加入)边缘支持向量集.图 3 说明了 ISVM 和 DSVM 在不同训练数据大小情况下的平均迭代次数,数据结果验证了两种算法的有限收敛性.

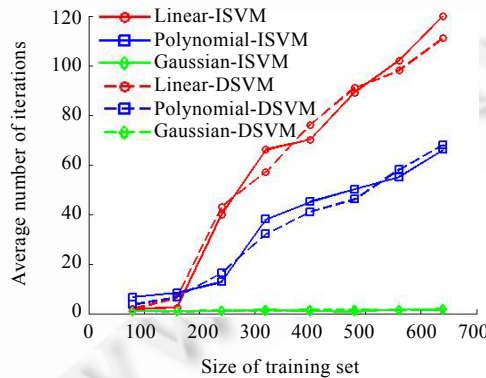


Fig.3 Average numbers of iterations on Diabetes dataset

图 3 Diabetes 数据集上的平均迭代次数

4.4 Breast cancer数据

Breast cancer 数据集来自 Wolberg 博士的临床病例,因此数据反映了病例的时间先后信息.该数据集包含具有 10 个数值属性的 683 个实例样本.

实验分别在数据量为 150,300,450,600 的基础上,分别尝试 200 次 ISVM 和 DSVM 学习,测试结果验证了:边缘支持向量集 S_S 不为空;矩阵 Q' 存在逆矩阵;向边缘支持向量集 S_S 新加入(移出)一个样本 t 后,在紧接着的调整中, t 不会反向移出(加入)边缘支持向量集.图 4 说明了 ISVM 和 DSVM 在不同训练数据大小情况下的平均迭代次数,数据结果验证了算法的有限收敛性.同时,算法在高斯核下也具有更快速的收敛性.

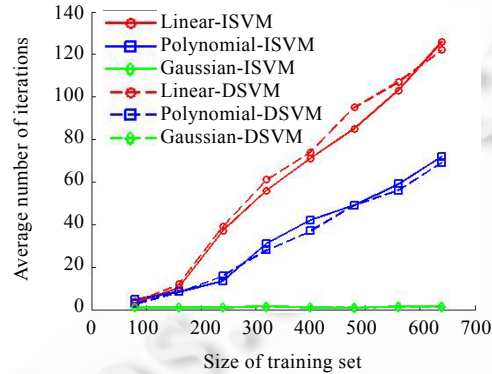


Fig.4 Average numbers of iterations on Breast cancer dataset

图 4 Breast cancer 数据集上的平均迭代次数

5 结论

C&P 算法是一种精确的增量式和减量式支持向量机算法,在留一估计、受限制学习(内存有限)、在线学习等任务中有着广泛的应用前景.为了更好地认识该算法,本文从理论上较为全面地分析了 C&P 算法的可行性与有限收敛性,并通过实验数据验证了理论分析的结果,为全面认识 C&P 算法开展了富有成效的工作.

本文的分析主要针对 C&P 算法,也可以扩展到其他基于 KKT 等价条件的支持向量机算法,例如回归支持向量机的增量式和减量式算法^[18]、标准支持向量机的解路径算法^[19]、回归支持向量机的解路径算法^[20]等.

本文讨论的增量式和减量式标准支持向量机算法(C&P 算法)有一个隐含的假设,即假设每次仅有一个样本到达或离开边界.然而,算法在实际计算过程中是否有多个样本同时到达或离开边界?这个答案是显然的.在进一步的工作中,希望针对该情形给出相应健壮的更新策略,并进一步回答算法的可行性以及有限收敛性.

References:

- [1] Vapnik V. Statistical learning theory. In: Wiley Series on Adaptive and Learning Systems for Signal Processing Communications and Control. John Wiley & Sons, 1998.
- [2] Lin C.J. On the convergence of the decomposition method for support vector machines. IEEE Trans. on Neural Network, 2001,12(6): 1288–1298. [doi: 10.1109/72.963765]
- [3] Frieß TT, Cristianini N, Campbell C. The kernel-adatron algorithm: A fast and simple learning procedure for support vector machines. In: Proc. of the ICML'98. San Francisco: Morgan Kaufmann Publishers, 1998. 188–196. http://www.isn.ucsd.edu/pubs/nips00_inc.pdf
- [4] Tsang IW, Kwok JT, Lai KT. Core vector regression for very large regression problems. In: Proc. of the ICML 2005. 2005. 912–919. [doi: 10.1145/1102351.1102466]
- [5] Zhang T. Solving large scale linear prediction problems using stochastic gradient descent algorithms. In: Proc. of the ICML 2004. 2004. 919–926. [doi: 10.1145/1015330.1015332]

- [6] Lee S, Wright SJ. ASSET: Approximate stochastic subgradient estimation training for support vector machines. In: Proc. of the Int'l Conf. on Pattern Recognition Applications and Methods. 2012. 223–228. http://pages.cs.wisc.edu/~sklee/papers/asset_icpram12.pdf
- [7] Mangasarian OL, Musicant DR. Successive overrelaxation for support vector machines. IEEE Trans. on Neural Networks, 1999,10: 1032–1037. [doi: 10.1109/72.788643]
- [8] Joachims T, Finley T, Yu CNJ. Cutting-Plane training of structural SVMs. Machine Learning, 2009,77(1):27–59. [doi: 10.1007/s10994-009-5108-8]
- [9] Teo CH, Vishwanthan SVN, Smola AJ, Le QV. Bundle methods for regularized risk minimization. The Journal of Machine Learning Research, 2010,11(1):311–365.
- [10] Franc V, Sonnenburg S. Optimized cutting plane algorithm for support vector machines. In: Proc. of the Int'l Conf. on Machine Learning. 2008. 320–327. [doi: 10.1145/1390156.1390197]
- [11] Cauwenberghs G, Poggio T. Incremental and decremental support vector machine learning. In: Advances in Neural Information Processing Systems 13. MIT Press, 2001. http://www.isn.ucsd.edu/pubs/nips00_inc.pdf
- [12] Laskov P, Gehl C, KrAuger S, Müller KR, Bennett K, Parrado-Hern E. Incremental support vector learning: Analysis, implementation and applications. Journal of Machine Learning Research, 2006,7(9):1909–1936.
- [13] 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). IEEE Press, 2008. 1364–1369. [doi: 10.1109/IJCNN.2008.4633975]
- [14] Karasuyama M, Takeuchi I. Multiple incremental decremental learning of support vector machines. IEEE Trans. on Neural Networks, 2010,21(7):1048–1059. [doi: 10.1109/TNN.2010.2048039]
- [15] Gu B, Wang JD, Yu YC, Zheng GS, Huang YF, Xu T. Accurate on-line v-support vector learning. Neural Networks, 2012,27: 51–59. [doi: 10.1016/j.neunet.2011.10.006]
- [16] Kuhn HW, Tucker AW. Nonlinear programming. In: Proc. of the 2nd Berkeley Symp. Berkeley: University of California Press, 1995. 481–492. http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.bsmmsp/1200500249
- [17] Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004.
- [18] Ma JS, Theiler J, Perkins S. Accurate on-line support vector regression. Neural Computation, 2003,15(11):2683–2703. [doi: 10.1162/089976603322385117]
- [19] Hastie T, Rosset S, Tibshirani R, Zhu J. The entire regularization path for the support vector machine. The Journal of Machine Learning Research, 2004,5(10):1391–1415.
- [20] 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]



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.
E-mail: jsgubin@163.com



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.
E-mail: aics@nuaa.edu.cn



郑关胜(1983—),男,副教授,主要研究领域为数据挖掘,机器学习.
E-mail: guanshzh@nuaa.edu.cn