

A Uniformly Smoothing Newton Method for Support Vector Machine

Zhen Wu*, Zhensheng Yu#

College of Science, University of Shanghai for Science and Technology, Shanghai
Email: xv907177937@qq.com, #zhsh-yu@163.com

Received: Jan. 30th, 2020; accepted: Feb. 13th, 2020; published: Feb. 20th, 2020

Abstract

To improve the classification performance of smooth support vector machines, a new uniformly smooth approximation function is introduced to replace the positive sign function. It doesn't only overcome the non-differentiability in the process of solving the support vector machine model, but also is superior to other smooth functions in solving convergence rate. Based on the uniform smooth approximation function, we design a corresponding Newton algorithm and establish the convergence of the proposed algorithm. Finally, numerical simulation shows the superior performance of the function in accuracy, efficiency and generalization adaptability for solving the smooth support vector machine model.

Keywords

Support Vector Machine, Smooth Function, Dual Problem, Newton-Armijo Algorithm

支撑向量机的一致光滑牛顿法

吴 振*, 宇振盛#

上海理工大学理学院, 上海
Email: xv907177937@qq.com, #zhsh-yu@163.com

收稿日期: 2020年1月30日; 录用日期: 2020年2月13日; 发布日期: 2020年2月20日

摘 要

提升光滑支撑向量机分类性能, 本文引入了一种新的一致光滑逼近函数来替代正号函数, 此函数不仅克

*第一作者。

#通讯作者。

服了支撑向量机模型求解过程中出现的不可微性, 也在求解收敛速度上优于其他光滑函数。基于此一致光滑逼近函数我们设计了相应的牛顿算法并证明了该光滑函数的收敛性, 最后通过数值模拟体现了该函数在光滑支撑向量机模型中的求解精度、效率和推广适应性的优越性能。

关键词

支撑向量机, 一致光滑函数, 对偶问题, Newton-Armijo算法

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着智能时代的到来, 机器学习领域越来越广泛受到社会青睐, 成为很多学者的研究热门。2001年 Lee 等人通过对支撑向量机的深入研究首次使用了 sigmoid 积分函数 $p(x, \alpha)$ 对无约束的支撑向量机 (Support Vector Machine, SVM) [1] 进行光滑化, 这是首次在支撑向量机中引入光滑的概念。得出了光滑支撑向量机 (Smooth Support Vector Machine, SSVM) [2] 模型, 然后在 Armijo 线搜索的基础上采用了牛顿算法对此光滑支撑向量机模型进行求解, 结果展现出了比一般的支撑向量机模型更好的分类性能, 同时也提高了其计算的效率。同年 Mangasarian 等人也提出了 RSVM (Reduced Support Vector Machine) [3], 将模型研究的约束凸优化问题改进为无约束强凸优化问题, 从而使相应算法在求解问题上提升了一定的速度, 尽管该模型在实际过程中对于一些大规模数据问题的求解所花费的时间比较长, 且需要保存的信息和计算量也比较多, 在求解思路, 仍是一种主流的进步方法。从此越来越多的学者们逐渐将研究光滑函数作为 SSVM 模型的核心, 并以此开辟了支撑向量机的一个新的研究方向。

2005年, 袁玉波[4]等人提出了两个多项式光滑函数, 通过对支撑向量机模型的无约束目标函数进行光滑化处理, 引出了 PSSVM 模型, 也证明了 PSSVM 相比 SSVM 而言有着更好的分类性能。之后又提出了三阶样条光滑函数[5], 从而引出了新的支撑向量机的 TSSVM 模型。通过分析 TSSVM 模型的收敛性和对应数据实验结果可以说明 TSSVM 有着更优秀的分类效果。2008年, 熊金志等人[6]更是提出了高于三阶样条光滑函数一个数量级的六阶光滑函数。

随着光滑函数逼近正号函数的精度的提高, 显然其函数本身也变得越来越复杂, 这并不利于科学计算。本文主要介绍了一种新的一致光滑逼近函数来替代正号函数。克服了支撑向量机模型求解过程中出现的不可微性, 并基于此一致光滑逼近函数设计了相应算法, 通过理论计算并证明了该光滑函数的收敛性, 最后通过数值模拟体现了该函数在光滑支撑向量机模型中的求解精度、效率和推广适应性的优越性能。

2. 光滑支撑向量机

2.1. 支撑向量机

对于给定一组数据

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$$

其中 $x_i \in \mathbb{R}^d, y_i \in \{-1, 1\}$, 我们的目标是从数据中寻求一个假设函数 $h: \mathbb{R} \rightarrow [-1, 1]$, 使得 $h(x_i) = y_i$, 即

$$\forall i, y_i h(x_i) = 1 \tag{1}$$

更进一步, 线性二分类模型认为假设函数的形式是基于对特征 x_i 的线性组合, 即

$$h(x_i) := \text{sign}(w^T x_i + b) \tag{2}$$

其中 $w_i \in \mathbb{R}^d, b \in \mathbb{R}$ 。

定义 1. 线性二分类模型的目标是找到一组合适的参数 (w, b) , 使得

$$\forall i, y_i (w^T x_i + b) > 0 \tag{3}$$

即, 线性二分类模型希望在特征空间找到一个划分类别的超平面, 将属于不同标记的样本分开。证明

$$\begin{aligned} y_i h(x_i) = 1 &\Leftrightarrow y_i \text{sign}(w^T x_i + b) = 1 \\ &\Leftrightarrow y_i (w^T x_i + b) > 0 \end{aligned} \tag{4}$$

由于能将样本分开的超平面有可能有很多, 故 SVM 希望能找到一个离各个样本都比较远的划分超平面, 即最优超平面。

当面对样本的随机扰动时, 离各个样本都较远的划分超平面抗干扰的能力比较强, 我们希望在优化间隔的同时, 允许分类错误的样本出现, 但这类样本应尽可能少:

$$\begin{cases} \min_{\omega, b} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^m I(\cdot) \\ y_i \neq \text{sign}(\omega^T \varphi(x_i) + b) \\ \text{s.t. } y_i (\omega^T \varphi(x_i) + b) \geq 1 \end{cases} \tag{5}$$

其中, $I(\cdot)$ 是指示函数, C 是个可调节参数, 用于权衡优化间隔和少量分类错误样本这两个目标。但是, 指示函数不连续, 更不是凸函数, 使得优化问题不再是二次规划问题。所以我们需要对其进行简化。

公式(5)难以实际应用的原因在于指示函数只有两个离散取值 0/1, 对应样本分类正确/错误。为了能使用优化问题继续保持为二次规划问题, 我们需要引入一个取值为连续值的变量, 刻画样本满足约束的程度。我们引入松弛变量 ξ_i , 用于度量样本违背约束的程度。当样本违背约束的程度越大, 松弛变量值越大。即

$$\xi_i = \begin{cases} 0, & \text{若 } y_i (\omega^T \varphi(x_i) + b) \geq 1 \\ 1 - y_i (\omega^T \varphi(x_i) + b), & \text{若 } y_i = \text{sign}(\omega^T \varphi(x_i) + b) \end{cases} \tag{6}$$

定义 2. (软间隔支撑向量机基本型)软间隔支撑向量机旨在找到一组合适的参数 (w, b) 使得

$$\begin{cases} \min_{\omega, b} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^m \xi_i \\ \text{s.t. } y_i (\omega^T \varphi(x_i) + b) \geq 1 - \xi_i \\ i = 1, 2, \dots, m, \xi_i \geq 0, i = 1, 2, \dots, m \end{cases} \tag{7}$$

其中, C 是个可调节参数, 用于权衡优化间隔和少量样本违背大间隔约束这两个目标。当 C 比较大时, 我们希望更多的样本满足大间隔约束; 当 C 比较小时, 我们允许有一些样本不满足大间隔约束。

定义 3. (软间隔支持向量机对偶型)软间隔支持向量机的对偶问题等价于找到一组合适的 α , 使得

$$\begin{cases} \min_{\alpha} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \varphi(x_i)^T \varphi(x_j) - \sum_{i=1}^m \alpha_i \\ \text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0, 0 \leq \alpha_i \leq \xi_i, i = 1, 2, \dots, m \end{cases} \tag{8}$$

2.2. SSVM

由软间隔支撑向量机的原始型和对偶形式我们可以对其进行求解得到 α_i^* , 于是分类函数

$$\begin{aligned} f(x) &= \text{sign} \left[\sum_{i=1}^m y_i \alpha_i^* (\varphi(x_i), \varphi(x)) + b^* \right] \\ &= \text{sign} \left[\sum_{i=1}^m y_i \alpha_i^* K(x_i, x) + b^* \right] \end{aligned} \quad (9)$$

其中,

$$b^* = \frac{1}{m} \sum_{i=1}^m \left(y_i - \sum_{j=1}^m y_j \alpha_j^* (\varphi(x_j), \varphi(x_i)) \right)$$

令 $X = (x_1, x_2, \dots, x_m)^T$

$$x_i = (x_{i1}, x_{i2}, \dots, x_{in}) \quad (i = 1, 2, \dots, m)$$

$$e = \text{ones}(m, 1)$$

$$D = \text{diag}(y_1, y_2, \dots, y_m), \quad A = DK(X, X^T)D$$

则问题(7)改写为

$$\begin{cases} \min_{\alpha, b} \frac{1}{2} \alpha^T A \alpha + C e^T \xi \\ \text{s.t. } D(K(X, X^T)D\alpha + eb) - e + \xi \geq 0, \xi \geq 0 \end{cases} \quad (10)$$

对于任何核函数, 上式均是个凸二次规划问题, 不妨取 $A = I$, 同时用 b 衡量分类间隔 $\left(\frac{2}{\| (w, b) \|_2} \right)$,

于是在对偶问题中 $b = -e^T D\alpha$, 同时让松弛变量 $\xi^T \xi$ 最小, 则(10)转化为

$$\begin{cases} \min_{\alpha, b} \frac{1}{2} (\alpha^T \alpha + b^2) + \frac{c}{2} \xi^T \xi \\ \text{s.t. } D(K(X, X^T)D\alpha + eb) - e + \xi \geq 0, \xi \geq 0 \end{cases} \quad (11)$$

由约束条件知,

$$\xi = \left(e - D(K(X, X^T)D\alpha + eb) \right)_+$$

这里是非光滑的。将其代入问题(11), 得到一个强凸无约束优化问题:

$$\min_{\alpha, b} \frac{1}{2} (\alpha^T \alpha + b^2) + \frac{c}{2} \left\| \left(e - D(K(X, X^T)D\alpha + eb) \right)_+ \right\|_2^2 \quad (12)$$

有唯一解。

$$\text{令 } z = \begin{pmatrix} \alpha \\ b \end{pmatrix}, r = (r_1, r_2, \dots, r_i)^T,$$

$$(r_+)_i = \max\{r_i, 0\}, r_i = 1 - H_i z, H_i = y_i [K_i - 1]$$

K_i 是核矩阵 K_i 的第 i 行。则将(13)改写为一般形式为

$$\min_z f(z) = \frac{1}{2} z^T z + \frac{c}{2} \|r_+\|_2^2 \quad (13)$$

用不同的光滑函数来近似逼近正号函数 $(\cdot)_+$, 可得不同的 SSVM 模型, 典型的光滑函数有 sigmoid 函数的积分函数、多项式函数、样条函数和插值函数等等。

这里我们假定一个光滑函数 $p(r,k)$ 来近似 r_+ , 则得到光滑支撑向量机 SSVM:

$$\min_z f_p(z) = \frac{1}{2} z^T z + \frac{c}{2} (p(r,k))^T p(r,k) \tag{14}$$

即为常见的光滑支撑向量机模型。

3. 一致光滑逼近函数及其性质

基于对正号函数的逼近的考虑, 我们寻找了一致光滑逼近函数来光滑目标函数中的不可微部分, 利用一致光滑逼近函数来近似绝对值函数, 从而通过新问题的求解来逼近原问题的解。

给出的一致光滑逼近函数如下所示:

$$h_\lambda(t) = \mu \ln \left(e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}} \right) + (\lambda - 1) \mu \ln 2, \tag{15}$$

$$\mu > 0, \lambda \in [0, 1]$$

当 $\mu \rightarrow 0$ 时, $h_\lambda(t) \rightarrow |t|$ 。

该函数是由 $\phi(t) = |t|$ 的两个一致光滑逼近函数 $\psi_\mu(t)$ 和 $\phi_\mu(t)$ [7] 的凸组合得到, 其中

$$\psi_\mu(t) = \mu \ln \left(e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}} \right), \quad \mu > 0;$$

$$\phi_\mu(t) = \mu \ln \left(\frac{1}{2} e^{\frac{t}{\mu}} + \frac{1}{2} e^{-\frac{t}{\mu}} \right), \quad \mu > 0.$$

利用 pycharm, 采用精度 $\varepsilon = 0.001$, $k = 1000$, 在区间 $\left[-\frac{10}{k}, \frac{10}{k}\right]$ 上一致光滑逼近函数 $\psi_\mu(t)$ 和 $\phi_\mu(t)$ 及其凸组合形式 $p_\lambda(x,k)$ 分别与正号函数 $\max\{x, 0\}$ 的逼近程度见图 1, 从图中可以看出, 其凸组合形式的光滑函数 $p_\lambda(x,k)$ 因其参数的选择不同可以有更好的适应性。

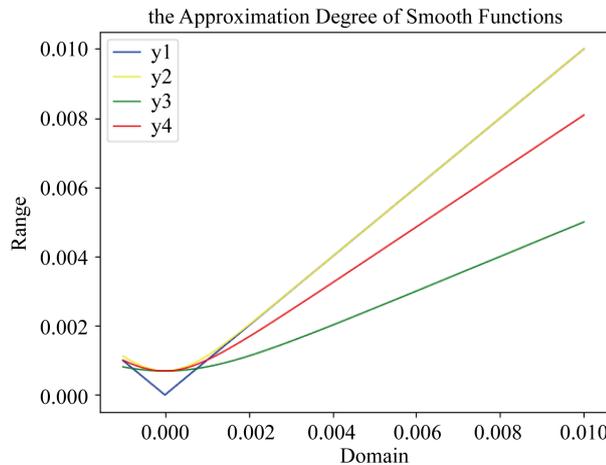


Figure 1. The approximation degree of smooth functions
图 1. 不同光滑函数的逼近程度

记 $\psi_\mu(t) = \mu \ln \left(e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}} \right)$, $\mu > 0$ 为 $\phi(t) = |t|$ 的上方一致光滑逼近函数,

- 则: 1) $0 < \psi_\mu(t) - \phi(t) \leq \mu \ln 2$;
 2) $\psi_\mu(t)$ 的值随参数 μ 的减小而单调下降, 且当 $\mu \rightarrow 0$ 时, $\psi_\mu(t) \rightarrow \phi(t)$;
 3) $\left| \frac{d(\psi_\mu(t))}{dt} \right| < 1$, 且 $\left. \frac{d(\psi_\mu(t))}{dt} \right|_{t=0} = 0$ 。

证明 1) $\psi_\mu(t) - \phi(t) = \mu \ln \left(e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}} \right) - \mu \ln e^{\frac{|t|}{\mu}} = \mu \ln \left(e^{\frac{t-|t|}{\mu}} + e^{-\frac{t-|t|}{\mu}} \right)$

由于 $|t| = \max\{t, -t\}$
 从而 $\begin{cases} t - |t| \leq 0 \\ -t - |t| \leq 0 \end{cases}$ 且 $t - |t|$ 与 $-t - |t|$ 中至少有一个等于 0, 因此

$$\mu \ln 1 < \psi_\mu(t) - \phi(t) = \mu \ln \left(e^{\frac{t-|t|}{\mu}} + e^{-\frac{t-|t|}{\mu}} \right) \leq \mu \ln(1+1)$$

即

$$0 < \psi_\mu(t) - \phi(t) \leq \mu \ln 2$$

2) $\psi_\mu(t)$ 的值随参数 μ 的减小而单调下降的证明可参考于李兴斯等[8] [9]关于凝聚函数的性质推导过程。结合 $0 < \psi_\mu(t) - \phi(t) \leq \mu \ln 2$, 即 $\phi(t) < \psi_\mu(t) \leq \phi(t) + \mu \ln 2$, 这表明 $\psi_\mu(t)$ 有界, 利用单调有界必有极限的定理及夹逼准则, 则有 $\lim_{\mu \rightarrow 0^+} \psi_\mu(t) = \phi(t)$ 。

3) $\frac{d(\psi_\mu(t))}{dt} = \frac{e^{\frac{t}{\mu}} - e^{-\frac{t}{\mu}}}{e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}}} = \frac{e^{\frac{2t}{\mu}} - 1}{e^{\frac{2t}{\mu}} + 1}$, 而 $\left| \frac{e^{\frac{2t}{\mu}} - 1}{e^{\frac{2t}{\mu}} + 1} \right| < \left| \frac{e^{\frac{2t}{\mu}}}{e^{\frac{2t}{\mu}} + 1} \right|$, 即 $\left| \frac{e^{\frac{2t}{\mu}} - 1}{e^{\frac{2t}{\mu}} + 1} \right| < 1$, 因此 $\left| \frac{d(\psi_\mu(t))}{dt} \right| < 1$, 且

$$\left. \frac{d(\psi_\mu(t))}{dt} \right|_{t=0} = \left. \left(\frac{e^{\frac{t}{\mu}} - e^{-\frac{t}{\mu}}}{e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}}} \right) \right|_{t=0} = 0。$$

我们给出了 $\mu = 0.001$ 时该函数与绝对值函数的逼近情况如图 2。

记 $\phi_\mu(t) = \mu \ln \left(\frac{1}{2} e^{\frac{t}{\mu}} + \frac{1}{2} e^{-\frac{t}{\mu}} \right)$, $\mu > 0$ 为 $\phi(t) = |t|$ 的下方一致光滑逼近函数,

- 则: 1) $\mu \ln \frac{1}{2} < \phi_\mu(t) - \phi(t) \leq 0$;
 2) $\phi_\mu(t)$ 的值随参数 μ 的减小而单调递增, 且当 $\mu \rightarrow 0$ 时, $\phi_\mu(t) \rightarrow \phi(t)$;
 3) $\left| \frac{d(\phi_\mu(t))}{dt} \right| < 1$, 且 $\left. \frac{d(\phi_\mu(t))}{dt} \right|_{t=0} = 0$ 。

证明 1) 由于 $\phi_\mu(t) = \psi_\mu(t) + \mu \ln \frac{1}{2}$, 于是 $\phi_\mu(t) - \phi(t) = \mu \ln \frac{1}{2} + \psi_\mu(t) - \phi(t)$, 结合上文中叙述的 $0 < \psi_\mu(t) - \phi(t) \leq \mu \ln 2$, 则有

$$\mu \ln \frac{1}{2} < \phi_\mu(t) - \phi(t) \leq 0$$

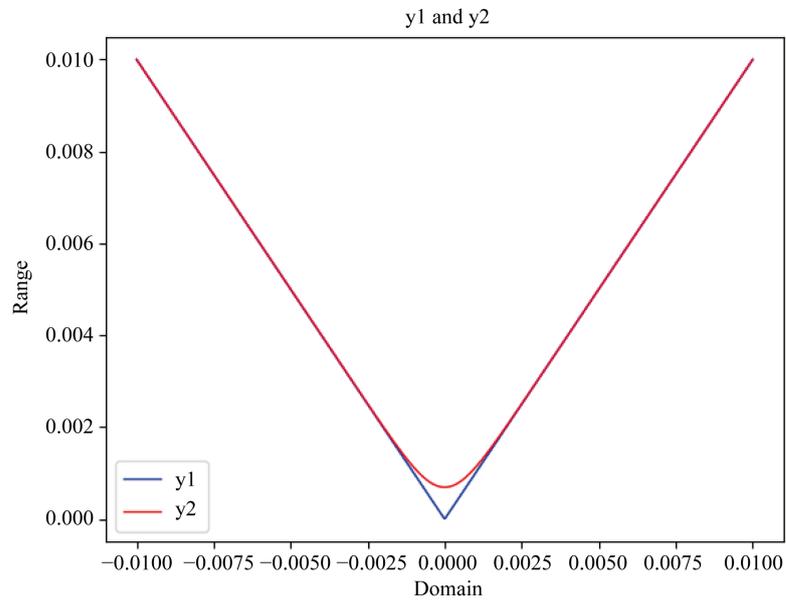


Figure 2. The approximation degree of y1 and y2
图 2. 光滑函数 y2 与绝对值函数 y1 的逼近程度

2) $\psi_\mu(t)$ 的值随参数 μ 的减小而单调递增的证明可参考于杨庆之等人[10] [11]等关于凝聚函数法的文献。

结合 $\mu \ln \frac{1}{2} < \phi_\mu(t) - \phi(t) \leq 0$, 即 $\phi(t) + \mu \ln \frac{1}{2} < \phi_\mu(t) \leq \phi(t)$, 这表明 $\phi_\mu(t)$ 有界, 利用单调有界必有极限的定理及夹逼准则, 则有 $\lim_{\mu \rightarrow 0^+} \phi_\mu(t) = \phi(t)$ 。

$$3) \frac{d(\phi_\mu(t))}{dt} = \frac{e^{\frac{t}{\mu}} - e^{-\frac{t}{\mu}}}{e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}}} = \frac{e^{\frac{2t}{\mu}} - 1}{e^{\frac{2t}{\mu}} + 1}, \text{ 而 } \left| e^{\frac{2t}{\mu}} - 1 \right| < \left| e^{\frac{2t}{\mu}} + 1 \right|, \text{ 即 } \left| \frac{e^{\frac{2t}{\mu}} - 1}{e^{\frac{2t}{\mu}} + 1} \right| < 1, \text{ 因此 } \left| \frac{d(\phi_\mu(t))}{dt} \right| < 1, \text{ 且}$$

$$\left. \frac{d(\phi_\mu(t))}{dt} \right|_{t=0} = \left(\frac{e^{\frac{t}{\mu}} - e^{-\frac{t}{\mu}}}{e^{\frac{t}{\mu}} + e^{-\frac{t}{\mu}}} \right) \Bigg|_{t=0} = 0。$$

我们给出了 $\mu = 0.001$ 时该函数与绝对值函数的逼近情况如图 3。

以上得到的两个一致光滑逼近函数均可以用作近似逼近正号函数 $(\cdot)_+$, 然而其精度并不是很好, 基于凸组合的特殊线性组合原理, 从而将是由 $\phi(t) = |t|$ 的两个一致光滑逼近函数 $\psi_\mu(t)$ 和 $\phi_\mu(t)$ 的凸组合而得到的新的函数:

$$h_\lambda(t) = \lambda \psi_\mu(t) + 1 - \lambda \phi_\mu(t), \lambda \in [0, 1]$$

而 $r_+ = \max\{0, r\} = \frac{1}{2}(r + |r|)$ 于是

$$t_+ = \frac{1}{2}(t + \mu \ln et\mu + e - t\mu + (\lambda - 1)\mu \ln 2), \mu > 0$$

从而相应的光滑函数为

$$p_\lambda(x, k) = \frac{1}{2} \left[x + \frac{1}{k} \ln(e^{kx} + e^{-kx}) + \frac{1}{k} (\lambda - 1) \ln 2 \right], k > 0 \tag{16}$$

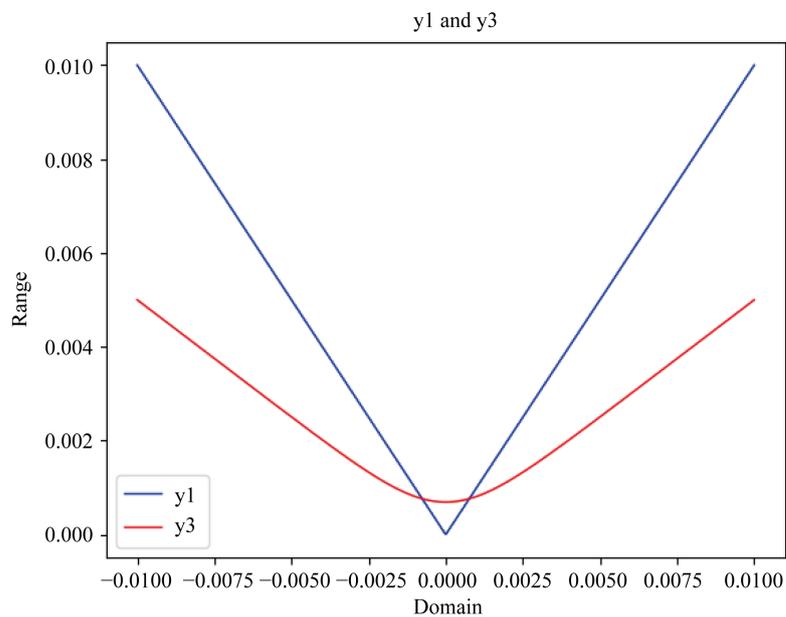


Figure 3. The approximation degree of y_1 and y_3
图 3. 光滑函数 y_3 与绝对值函数 y_1 的逼近程度

对于这样一个新的凸组合光滑逼近函数，将其与绝对值函数进行比较如图 4 所示。
 当 $k = 1000$ ， $\lambda = 0.5$ 时

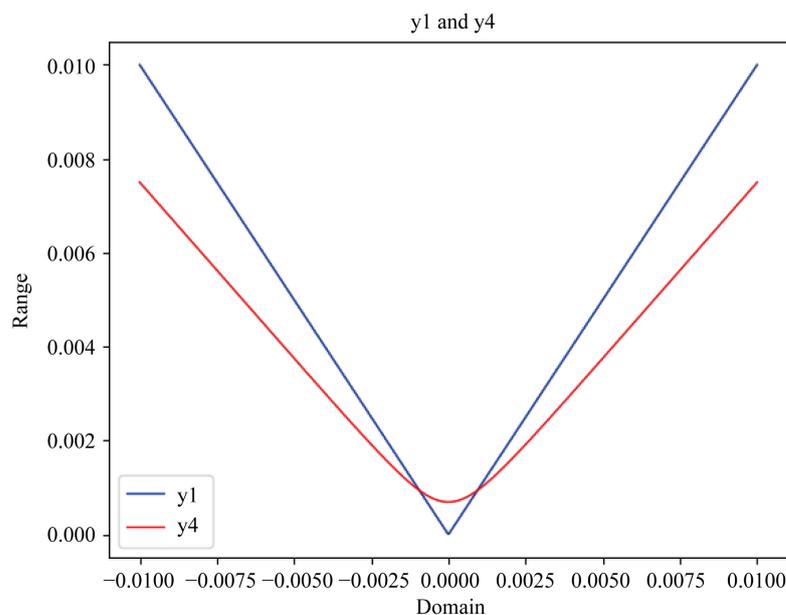


Figure 4. The approximation degree of y_1 and y_4
图 4. 光滑函数 y_4 与绝对值函数 y_1 的逼近程度

当 $k = 1000$ ， $\lambda = 0.618$ 时，见图 5。

对于给定的光滑因子 k ，

当 $|x| < \rho$ 时，逼近程度

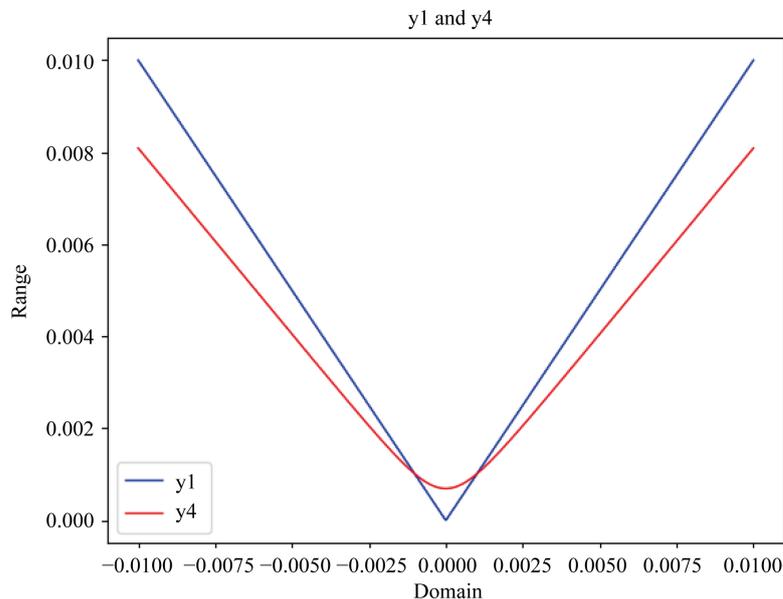


Figure 5. The approximation degree of y1 and y4
图 5. 更新后的光滑函数 y4 与绝对值函数 y1 的逼近程度

$$p_{\lambda}^2(x, k) - x_+^2 \leq \frac{(\lambda \ln 2)^2}{4k^2} + \frac{(\lambda \ln 2)(x + |x|)}{2k}$$

$x \geq 0$ 时, 逼近程度

$$p_{\lambda}^2(x, k) - x_+^2 \leq \frac{(\lambda \ln 2)^2}{4k^2} + \frac{\lambda \ln 2}{k} \rho$$

故当 $\rho = \frac{1}{k}$ 时,

$$p_{\lambda}^2(x, k) - x_+^2 \leq \frac{1}{4k^2} ((\lambda \ln 2)^2 + 4\lambda \ln 2)$$

故当 $\rho = \frac{\lambda}{k}$ 时,

$$p_{\lambda}^2(x, k) - x_+^2 \leq \frac{1}{4k^2} ((\lambda \ln 2)^2 + 4\lambda^2 \ln 2) \approx 0.8133 \frac{\lambda^2}{k^2}$$

故当 $\rho = \frac{\lambda \ln 2}{k}$ 时,

$$p_{\lambda}^2(x, k) - x_+^2 \leq \frac{5(\lambda \ln 2)^2}{4k^2} \approx 0.6006 \frac{\lambda^2}{k^2}$$

$x < 0$ 时, 逼近程度

$$p_{\lambda}^2(x, k) - x_+^2 \leq \frac{(\lambda \ln 2)^2}{4k^2} \approx 0.1201 \frac{\lambda^2}{k^2}$$

显然, 只需取 λ 为趋向于 0 的合适的值我们即可得到较强的逼近精度。对于任意一个 $k > 0$, 问题(14)和(15)的解的逼近精度始终满足于

$$\|\bar{z} - z\|_2^2 \leq \frac{m}{2} \left(\frac{(\lambda \ln 2)^2}{4k^2} + \frac{\lambda \ln 2}{k} \eta \right)$$

其中 $\eta = \max_{1 \leq i \leq m} |r_i|$ 。

要求 SSVM 的解满足精度要求。只要使得 $\|\bar{z} - z\|_2^2 \leq \varepsilon$, 只需有

$$\frac{m}{2} \left(\frac{(\lambda \ln 2)^2}{4k^2} + \frac{\lambda \ln 2}{k} \eta \right) \leq \varepsilon$$

$x \geq 0$ 时,

$$\text{故当 } \rho = \frac{\lambda}{k} \text{ 时, 解得 } k_{opt}(m, \varepsilon) \geq \lambda \sqrt{\frac{0.4066m}{\varepsilon}}$$

$$\text{当 } \rho = \frac{\lambda \ln 2}{k} \text{ 时, 解得 } k_{opt}(m, \varepsilon) \geq \lambda \sqrt{\frac{0.3003m}{\varepsilon}}$$

$$x < 0 \text{ 时, 解得 } k_{opt}(m, \varepsilon) \geq \lambda \sqrt{\frac{0.0601m}{\varepsilon}}$$

4. Newton-Armijo 算法

Newton-Armijo 算法是对于牛顿法的一个改进。采用 Armijo 线搜索, 并利用牛顿法进行求解。记式 (13) 的目标函数梯度为

$$\nabla f(z) = z - CH^T r_+,$$

使用光滑函数 $p_\lambda(r)$ 代替 r_+ 得到: $G_p(z) = z - CH^T p_\lambda(r)$, 其中 $(p_\lambda(r))$ 是 k 取定值时的 $p_\lambda(x, k)$

Hessian 矩阵 $\nabla G_p(z) = I + CH^T \tilde{\Lambda}_p(z) H$, 其中 $\tilde{\Lambda}_p(z) = \text{diag}(p'_\lambda(r_1), p'_\lambda(r_2), \dots, p'_\lambda(r_m))$; 式(15)的目标函数梯度为:

$$g \nabla f_p(z) = z - CH^T p_\lambda(r) p'_\lambda(r),$$

Hessian 矩阵 $G = \nabla^2 f_p(z) = I + CH^T \Lambda_p(z) H$, 其中

$$\Lambda_p(z) = \text{diag}\left(\left(p'_\lambda(r_1)\right)^2 + p_\lambda(r_1) p''_\lambda(r_1), \dots, \left(p'_\lambda(r_m)\right)^2 + p_\lambda(r_m) p''_\lambda(r_m)\right)$$

一般的 Newton-Armijo 算法

目标函数的下降方向 d_k 由方程 $G_k d_k = -g_k$ 求得, 由于使用了 Armijo 一维搜索, 保证了目标函数值在迭代过程中的不断下降, 其具体步骤如下:

Step 1

Initialization. Given the initial point z_0 , let $k = 0$, $\varepsilon > 0$, parameter: C, σ, H generated from the input data;

Step 2

Calculate g_k . If $\|g_k\| < \varepsilon$, stop; or turn to step 3;

Step 3

Calculate Hessian matrix G_k , solve $G_k d_k = -g_k$ and then we get d_k ;

Step 4 (Armijo line search) select

$$\lambda_k = \max\{1, 2^{-1}, 2^{-2}, \dots\}$$

and let

$$f_p(z_k + \lambda_k d_k) \leq f_p(z_k) + \sigma \lambda_k g_k^T d_k$$

Step 5

Let $z_{k+1} = z_k + \lambda_k d_k$, $k = k + 1$, turn to step 2.

该算法每次迭代需要计算 Hessian 矩阵 G_k , 再确定搜索方向, 所需要的工作量相当大, 并且由于数据的特征比较多, 相应的对于 Hessian 矩阵的计算就会变得比较复杂。

改进的 Newton-Armijo 算法

将问题(13)中的目标函数 $\nabla f(z)$ 光滑为 $G_p(z)$, 令 $G_p(z) = 0$, 然后求解该方程, 由于 $\nabla G_p(z)$ 中

$$\tilde{\Lambda}_p(z) = \text{diag}(p'_\lambda(r_1), p'_\lambda(r_2), \dots, p'_\lambda(r_l))$$

比 $\nabla^2 f_p(z)$ 中 $\Lambda_p(z) = \text{diag}((p'_\lambda(r_1))^2 + p_\lambda(r_1)p''_\lambda(r_1), \dots, (p'_\lambda(r_l))^2 + p_\lambda(r_l)p''_\lambda(r_l))$ 简单, 因而这个改进的方法在计算起来速度更快。其具体步骤如下:

Step 1

Initialization. Given the initial point z_0 , let $k = 0$, $\varepsilon > 0$, parameter: C, σ, H generated from the input data;

Step 2

Calculate g_k . If $\|g_k\| < \varepsilon$, stop; or turn to step 3;

Step 3

Calculate Hessian matrix G_k , solve $G_k d_k = -g_k$ and then we get d_k ;

Step 4 (Armijo line search) select

$$\lambda_k = \max\{1, 2^{-1}, 2^{-2}, \dots\}$$

and let

$$h_p(z_k + \lambda_k d_k) \leq h_p(z_k) + \sigma \lambda_k g_k^T d_k$$

其中 $h_p(z) = \frac{1}{2} z^T z + C \sum_{i=1}^m \varphi_p(r_i)$, 而 $\varphi_p(t) = \int_a^t p(u) du$ (a 是 $p(x)$ 的定义域下界), 因为 $\min_z h_p(z)$ 的一阶充分必要条件是 $G_p(z) = 0$;

Step 5

Let $z_{k+1} = z_k + \lambda_k d_k$, $k = k + 1$, turn to step 2.

5. 数值结果和比较

5.1. 一致光滑牛顿法求解

数值实验数据来自用 python 编写的随机线性不可分数据集, 统共样本数据有 1000 样本, 测试样本有 300 个, 每个样本有 30 个特征, 采用了高斯核来将其映射到高维特征空间上得到分类结果如图 6~9。

可以看出在小规模数据集下其分类的正确率都有着良好的性质。基于此结果的影响下, 我们采用了其他数据集, 例如区分数字的分类数据集 USPS, 我们实验选取了 5000 个训练样本, 1800 个测试样本, 每个样本有着 160 个特征, 在精度要求 $\varepsilon = 0.001$ 的情况下使用不同 Newton-Armijo 算法求解算法下分类所需的迭代次数、时间、训练错误率和测试错误率如表 1 所示。

5.2. 分析与比较

由图表中数据进行分析可知, 我们得到一些关于不同光滑函数与算法的结合求解 SVM 的相关结论:

- 1) 一般来说, 随着光滑函数与目标函数的逼近程度的靠近, 求解 SVM 的解的精度会更好, 在迭代次数有要求的情况下, 光滑函数的选取决定了它是否能够满足实验精度要求, 否则只能在最大迭代次数下才能得到满足实验精度的结果。新选取的光滑函数的参数选择也决定了它是否能够很好的契合实际问

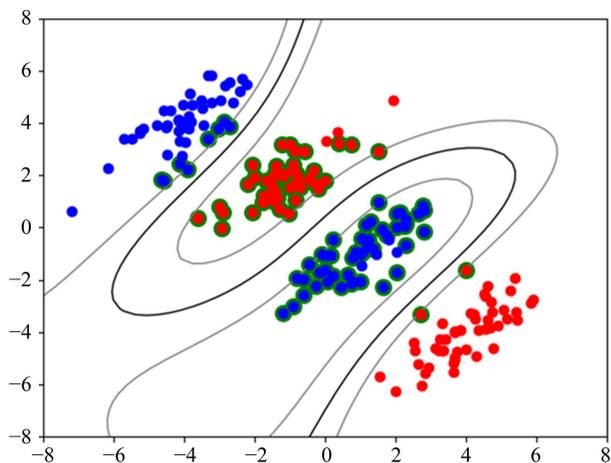


Figure 6. The result (1) of classification

图 6. 分类结果(1)

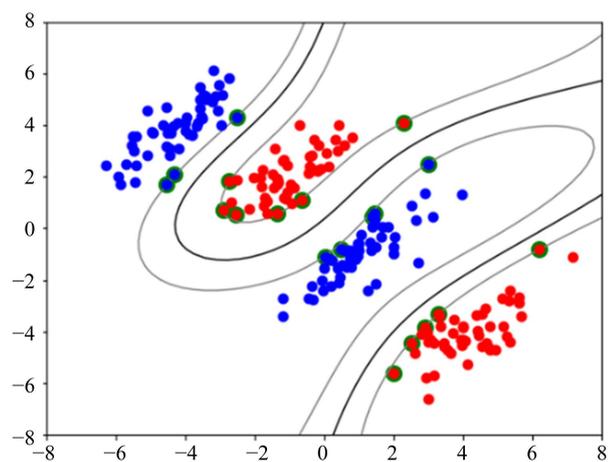


Figure 7. The result (2) of classification

图 7. 分类结果(2)

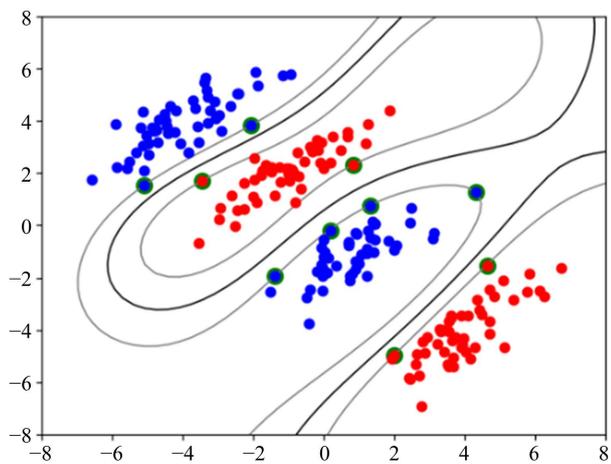


Figure 8. The result (3) of classification

图 8. 分类结果(3)

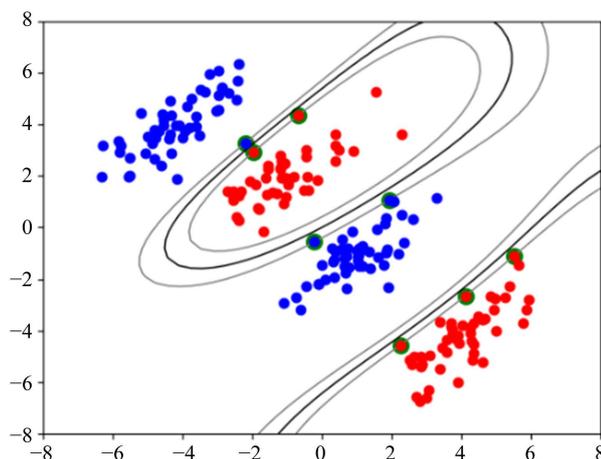


Figure 9. The result (4) of classification
图 9. 分类结果(4)

Table 1. Experimental results of different solving algorithms for different smooth functions
表 1. 不同光滑函数求解算法的实验结果

Algorithm	Smooth function	iterations	time/s	Training error rate	Test error rate	lambda
General Newton-Armijo algorithm	$\psi_{\mu}(t)$	69	6.378173	0.0089	0.0179	
	$\phi_{\mu}(t)$	79	7.001273	0.0056	0.0156	
	$p_{\lambda}(x, k)$	34	4.912734	0.0032	0.0151	0.5
	$p_{\lambda}(x, k)$	32	4.721893	0.0012	0.0149	0.618
Improved Newton-Armijo algorithm:	$\psi_{\mu}(t)$	61	6.001282	0	0.0139	
	$\phi_{\mu}(t)$	75	5.832902	0.0023	0.0187	
	$p_{\lambda}(x, k)$	30	4.027182	0	0.0123	0.5
	$p_{\lambda}(x, k)$	27	3.891276	0	0.0146	0.618

题的解决思路。在这方面，凸组合形式的光滑函数给了我们一些新的想法，在逼近程度上与求解问题精度上做一个很好的平衡是一个好的思路。

2) 一般来说，光滑函数的逼近可以使得训练错误率的下降逐渐减小甚至为 0，但是数据训练的时间也会越来越多。这就要求我们在求解大规模问题中需要寻求一个适应性很好的光滑函数，在光滑函数的逼近程度和函数本身的复杂中间需要我们做好平衡，如何保证在尽量减少计算量的情况下最大化计算效率和求解速度，并在算法参数和光滑参数的结合上做一个好的调整是一件值得考虑的事情。

6. 结论

综上所述，支撑向量机的光滑算法中，光滑函数的不同对于算法的使用有着不同的收敛速度及计算复杂度，很明显改进后的改进的 Newton-Armijo 算法的收敛速度优于一般的 Newton-Armijo 算法；尽管是同样的算法中实验，光滑函数的不同对于结果也有着不同的影响，其影响程度也与光滑函数的逼近程度相关，越逼近正号函数的模型，其解的精确度就越高，训练结果达到的正确率也就越高，但是其训练时间也就随之增加。今后待研究问题方向有：除却 Newton-Armijo 算法之外的 BFGS-Armijo 算法、Newton-PCG 等其他算法在不同光滑函数的模型下求解 SSVM 模型的收敛速度及求解精确度的问题、在

大规模数据中处理数据训练时间和精度的平衡、光滑算法的优化改进等等。

参考文献

- [1] Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1-299. <https://doi.org/10.1007/978-1-4757-2440-0>
- [2] Lee, Y.J. and Mangasarian, O.L. (2001) SSVM: A Smooth Support Vector Machine for Classification. *Computational Optimization and Applications*, **20**, 5-22. <https://doi.org/10.1023/A:1011215321374>
- [3] Lee, Y.J. and Mangasarian, O.L. (2001) RSVM: Reduced Support Vector Machine. *Proceedings of the SIAM International Conference on Data Mining*, Chicago, 5-7 April 2001, 1-17. <https://doi.org/10.1137/1.9781611972719.13>
- [4] 袁玉波, 严杰, 徐成贤. 多项式光滑的支撑向量机[J]. 计算机学, 2005, 28(1): 9-17.
- [5] Yuan, Y., Fan, W. and Pu, D. (2007) Spline Function Smooth Support Vector for Classification. *Journal of Industrial and Management Optimization*, **3**, 529-542.
- [6] 黄仁泰, 熊金志. 一个支持向量机的新光滑函数及其性能分析[J]. 计算机工程与应用, 2008, 44(18): 64-65.
- [7] 雍龙泉. 绝对值函数的一致光滑逼近函数[J]. 数学的实践与认识, 2015, 45(20): 250-255.
- [8] 李兴斯. 解非线性规划的凝聚函数法[J]. 中国科学: A 辑, 1991, 12(12): 1283-1288.
- [9] 李兴斯. 一类不可微优化问题的有效解法[J]. 中国科学: A 辑, 1994, 24(4): 371-377.
- [10] 杨庆之, 杨德庄, 张敏洪. 调节熵函数法[J]. 计算数学, 2001, 23(1): 81-86.
- [11] 杨庆之. 对凝聚函数法的分析[J]. 计算数学, 1996, 18(4): 405-410.