中国理论数学前沿
Frontiers of Chinese Pure Mathematics
2024 年第 02 卷第 03 期
Vol. 02, No. 03, 2024

研究性论文 http://www.scifootprint.com/FCPM/index.html

具有 l_1 正则项的稀疏支持向量机的最优性条件及算法

杨晓辉，高彩霞*

（内蒙古大学数学科学学院，呼和浩特 010021）

摘要：支持向量机（SVM）作为机器学习的主要方法之一，是用于解决分类和回归任务的强大学习工具，在图像分类、模式识别和疾病诊断领域都备受瞩目。在支持向量机模型中，$L_{0/1}$ 损失函数是理想的损失函数，已有的损失函数大多是其代理函数。稀疏优化是研究带有稀疏结构的最优化问题，根据 l_1 范数良好的稀疏性，可以通过特征选择去除冗余特征，本文在 $L_{0/1}$ 软间隔损失的模型基础上，提出一个基于 $L_{0/1}$ 损失的 l_1 范数稀疏支持向量机（简称 $L_{0/1}$-SSVM），证明了模型解的存在性，给出模型的 KKT 点和 P-稳定点，并证明全局最优解与 KKT 点的关系。利用 l_1 范数的近端算子设计 ADMM 算法迭代框架，并对算法进行收敛性分析，证明其收敛于 P-稳定点。

关键词：l_1 范数；$L_{0/1}$ 损失；稀疏支持向量机；最优性条件；ADMM 算法

DOI：10.48014/fcpm.20240411001

引用格式：杨晓辉，高彩霞．具有 l_1 正则项的稀疏支持向量机的最优性条件及算法[J]．中国理论数学前沿，2024，2(3)：16-24.

0 引言

稀疏优化问题的研究兴起于二十世纪九十年代，在本世纪得到了蓬勃的发展，已经广泛应用于机器学习[1]，信号处理与通信[2]，模式识别[3]，统计学中的回归问题[4]，经济学中的投资组合问题[5]以及主成分分析[6]等等。在机器学习领域中稀疏优化是重要的技术手段之一，其目的是可以在训练模型的过程中减少特征的数量，以此提高模型的泛化能力和可解释性，而支持向量机[7]（Support vector machine，SVM）作为机器学习领域中的一种重要方法，对其进行稀疏优化则受到了广泛关注。

支持向量机是建立在统计学习理论的 VC 维理论和结构风险最小化原理[8]基础上的一种机器学习方法，在逻辑回归和神经网络相比，支持向量机在学习复杂的非线性方程时提供了一种更为清晰强大的方式，多用于解决数据挖掘和模式识别领域中数据分类问题，并且它在解决小样本、非线性和高维模式识别问题中表现出许多特有的优势。在实际生活中，支持向量机也广泛应用于各个领域。在图像识别[9]中，SVM 可以实现人脸识别、物体检测等任务；在生物信息学中，SVM 可以用于基因表达数据的分类和预测等任务；在金融市场中，SVM 可以辅助分析股票走势和风险评估等任务。

正则项和损失函数是研究软间隔 SVM 的核心内容，不仅决定模型对训练数据的敏感性，还影响着模型的稀疏性。Cortes 和 Vapnik 在文献[7]中指出，理想的损失函数是 $L_{0/1}$ 损失函数，但 $L_{0/1}$ 损失是非凸非连续的有界函数，具有 NP 难的特点，因此许多学者致力于构造新的代理损失函数，在算法和理论上也取得了丰富的研究成果，如 Bartlett 和

* 通讯作者 **Corresponding author**：高彩霞，smsgcx@imu.edu.cn

收稿日期：2024-04-11； **录用日期**：2024-05-22； **发表日期**：2024-09-28

Wegkamp[10]提出的广义合页损失函数,Jumutc 和 Huang[11]提出的弹球损失函数,沈晓彤[12]等人提出的滑道损失函数等等。文献[13]设法建立了 $L_{0/1}$-SVM 的最优性理论,提出一种快速的交替方向乘子法,通过数值实验验证了该算法求解 $L_{0/1}$-SVM 模型更高效。

在支持向量机中,常用到的正则项有 l_0,l_1,l_p $(0<p<1)$,l_2 范数等,其中 l_2 范数可以改善模型的过拟合,为大多数模型所使用的,但是不具备稀疏性,l_0 范数指向量非零元素个数,具有很好的稀疏性,但由于其非凸比较难求解,而 l_1 范数是 l_0 范数的最优凸近似[14],可以实现特征的自动选择,通过学习过滤掉冗余特征,因此很多学者致力于研究 l_1 范数为正则项的稀疏模型。Tanveer 和 Sharma[15]等人提出了具有弹球损失的稀疏支持向量机(Pin-SSVM),所提出的模型使用 l_1 范数,保证了模型的稀疏性、鲁棒性和不敏感性;Zheng[16]等人提出稀疏支持矩阵机,新的正则化项是 l_1 范数和核范数的线性组合,同时考虑稀疏性和低秩性,使用广义前向后向算法进行求解,实验结果也证明了该模型具有优异的性能。

因此受其启发,本文在 $L_{0/1}$ 损失函数的基础上,使用 l_1 范数作为模型的正则项,建立一个新的稀疏支持向量机模型,同时建立其最优性理论,并提出相应的 ADMM 算法求解,证明该算法是收敛的。

1 预备知识

1.1 符号与定义

本文向量默认为列向量,$\mathbb{R}^n$表示 n 维向量空间,$x\in\mathbb{R}^n$是一个 n 维向量,其中 x_i 表示向量的第 n 行。

定义 1[17] 设 $f:R^n\to R$ 是下半连续函数,则 f 在给定的 $x\in R$ 处关于参数 $\alpha>0$ 的临近点算子定义为

$$prox_f(x):=\underset{v\in R}{\operatorname{argmin}} f(v)+\frac{1}{2\alpha}(v-x)^2. \quad (1)$$

定义 2[18] 对给定参数 $\eta>0$,l_1 范数的临近点算子为

$$Prox_{\frac{1}{\eta}\|\cdot\|_1}(x)=sign(x_i)\max\left\{|x_i|-\frac{1}{\eta}\right\}$$
$$=\begin{cases} x_i-\dfrac{1}{\eta}, & x_i>\dfrac{1}{\eta} \\ 0, & |x_i|\leqslant\dfrac{1}{\eta} \\ x_i+\dfrac{1}{\eta}, & x_i<-\dfrac{1}{\eta} \end{cases}. \quad (2)$$

定义 3[13] 对给定参数 $\gamma>0$,$C>0$,$L_{0/1}$ 范数的临近点算子为

$$Prox_{\gamma C\|(\cdot)_+\|_0}(x)=\begin{cases} 0, & 0<x<\sqrt{2\gamma C} \\ 0 \text{ or } x, & x=\sqrt{2\gamma C} \\ x, & x>\sqrt{2\gamma C} \text{ or } x\leqslant 0 \end{cases}. \quad (3)$$

定义 4[19] 设 $f:R^n\to R$ 在 $x\in R$ 是局部 Lipschitz 的,那么 f 在 x 处的正则次微分是

$$\hat{\partial} f(x):=\left\{\begin{array}{l} v\in R \mid \liminf\limits_{y\to x,\ y\neq x} \\ \dfrac{f(y)-f(x)-\langle v,y-x\rangle}{\|y-x\|}\geqslant 0 \end{array}\right\}. \quad (4)$$

定义 5[20] 设 $\Omega\subseteq\mathbb{R}^n$是任意非空闭集,$x^*\in\Omega$,则 Ω 在 $x^*\in\Omega$ 处的 Bouligand 切锥 $T_\Omega(x^*)$定义为

$$T_\Omega(x^*):=\left\{\begin{array}{l} d\in\mathbb{R}^n:\exists\{x^k\}\to x^*, \\ \{t_k\}\downarrow 0 \text{ s.t. } \lim\limits_{k\to\infty}\dfrac{x^k-x^*}{t_k}=d \end{array}\right\}. \quad (5)$$

则 Ω 在 $x^*\in\Omega$ 处的 Clarke 切锥 $T_\Omega^C(x^*)$定义为

$$T_\Omega^C(x^*):=\left\{\begin{array}{l} d\in\mathbb{R}^n:\exists\{x^k\}\to x^*, \\ \quad\{t_k\}\downarrow 0,\exists\{y^k\}\to x^* \\ \text{s.t. } \lim\limits_{k\to\infty}\dfrac{x^k-y^k}{t_k}=d \end{array}\right\}. \quad (6)$$

2 模型

2.1 模型发展

支持向量机的基本思想是寻找一个最优超平面将样本尽可能分开,并且使样本到超平面的距离尽可能的远。给定一个训练集$\{(\boldsymbol{x}_i,y_i\mid i=1,\cdots,m)\}$,其中 $x_i\in\mathbb{R}^n$是输入向量,$y_i\in\{-1,1\}$是输出标签,$y_i=+1$ 和 $y_i=-1$ 分别对应正类和负类。训练一个超平面$\langle\boldsymbol{w},\boldsymbol{x}\rangle+b=w_1x_1+\cdots+w_nx_n+$

$b=0$,使得对任意的输入向量 x,都可以正确预测 x 的标签 y,并且使输入向量距超平面的距离 $\frac{2}{\|\boldsymbol{w}\|_2}$ 尽可能的远,将上述思想应用到二分类问题中可以得到如下模型:

$$\begin{aligned}&\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R}}\ \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}\xi_i\\&\text{s.t.}\ \ y_i(\langle\boldsymbol{w},\boldsymbol{x}\rangle+b)\geqslant 1-\xi_i,\end{aligned}\tag{7}$$

其中 $C>0$ 是惩罚参数,ξ_i 是松弛变量,若样本可以完全分开,则不需要惩罚。或者是引入损失函数 $L(\cdot)$对分类错误的点进行惩罚,模型如下:

$$\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R}}\ \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}L[1-y_i(\langle\boldsymbol{w},\boldsymbol{x}\rangle+b)],\tag{8}$$

其中前一项是正则项,衡量模型的复杂度,控制模型的稀疏;后一项损失函数控制模型的误差。理想的损失函数是 $L_{0/1}$ 损失函数,2020 年修乃华团队[13]提出 $L_{0/1}$ 软间隔损失 SVM,模型如下:

$$\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R}}\ \frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{m}L_{0/1}(1-y_i\langle\boldsymbol{w},\boldsymbol{x}\rangle+b),\tag{9}$$

其中 $L_{0/1}$ 损失函数的表达式为:

$$L_{0/1}(t)=\begin{cases}1, & t>0\\0, & t\leqslant 0\end{cases}.\tag{10}$$

该模型很好的捕捉到二分类的本质并且保证模型使用的支持向量更少,该模型的解决是支持向量机领域向前迈进了一大步,对于模型、理论和算法的更深入研究成为了有意义的方向。

2.1 $L_{0/1}$-SSVM 模型

由文献[14]可知,l_1 范数具有良好的稀疏性,为了减少冗余特征和提高模型的稀疏性,在 $L_{0/1}$ 范数软间隔 SVM 模型的基础上,使用 l_1 范数作为正则项,提出如下新模型:

$$\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R}}\ \|\boldsymbol{w}\|_1+C\sum_{i=1}^{m}L_{0/1}[1-y_i(\boldsymbol{w}^\top x_i+b)],\tag{11}$$

其中 $\boldsymbol{w}\in\mathbb{R}^n$ 是向量,$b\in\mathbb{R}$是一个实数,$\|\cdot\|_1$ 是向量中所有元素绝对值之和,$C>0$ 是惩罚参数,$L_{0/1}$ 是损失函数。

为了便于分析,设 $\boldsymbol{u}:=\boldsymbol{e}-A\boldsymbol{w}-b\boldsymbol{y}\in\mathbb{R}^m$,其中 $\boldsymbol{e}:=(1,1,\cdots 1)^\top\in\mathbb{R}^m$,$A:=[y_1\boldsymbol{x}_1\cdots y_m\boldsymbol{x}_m]^\top\in\mathbb{R}^{m\times n}$,$\boldsymbol{y}:=(y_1,y_2,\cdots,y_m)^\top\in\mathbb{R}^m$,则模型有如下转化:

$$\begin{aligned}&\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R},\boldsymbol{u}\in\mathbb{R}^m}\ \|\boldsymbol{w}\|_1+C\|\boldsymbol{u}_+\|_0\\&s.t.\quad \boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}=\boldsymbol{e},\end{aligned}\tag{12}$$

其中 $\boldsymbol{u}_+:=((u_1)_+,\cdots,(u_m)_+)^\top\in\mathbb{R}^m=\max\{\boldsymbol{u},\boldsymbol{0}\}$,则 $\sum_{i=1}^{m}L_{0/1}[1-y_i(\boldsymbol{w}^\top\boldsymbol{x}_i+b)]=L_{0/1}(\boldsymbol{u})=\|\boldsymbol{u}_+\|_0$,因此模型是本文提出的具有 l_1 正则项的稀疏支持向量机模型,简称 $L_{0/1}$-SSVM,其中 $\|\cdot\|_0$ 是非凸非连续的,$\|\cdot\|_1$ 是不可微的,因此上述问题是 NP-难的,下文对其最优性条件进行分析。

2.2 最优性理论

通过引入变量 $\boldsymbol{s}\in\mathbb{R}^m$,模型有如下等价转化:

$$\begin{aligned}&\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R},\boldsymbol{u}\in\mathbb{R}^m,\boldsymbol{s}\in\mathbb{R}^m}\ \|\boldsymbol{w}\|_1+C\|\boldsymbol{u}_+\|_0\\&s.t.\qquad A\boldsymbol{w}+b\boldsymbol{y}=\boldsymbol{s},\\&\qquad\quad\ \boldsymbol{e}-\boldsymbol{s}=\boldsymbol{u}.\end{aligned}\tag{13}$$

对应的拉格朗日函数为:

$$\begin{aligned}&L(\boldsymbol{w},b,\boldsymbol{u},\boldsymbol{s},\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)\\&=\|\boldsymbol{w}\|_1+C\|\boldsymbol{u}_+\|_0+\langle\boldsymbol{\rho}_1,\boldsymbol{s}-A\boldsymbol{w}-b\boldsymbol{y}\rangle+\\&\quad\langle\boldsymbol{\rho}_2,\boldsymbol{u}-\boldsymbol{e}+\boldsymbol{s}\rangle,\end{aligned}$$

其中$\boldsymbol{\rho}_1\in\mathbb{R}^m$,$\boldsymbol{\rho}_2\in\mathbb{R}^m$是拉格朗日乘子。

首先证明模型的全局最优解存在并且全局最优解解集是有界的。

定理 1 若给定实数 $C\in[-K,K]$,其中 K 是正数,则问题(12)的全局最优解存在且解集是有界的。

证明:由问题可知

$$\min_{\boldsymbol{w}\in\mathbb{R}^m,b\in\mathbb{R}}\ f(\boldsymbol{w},b)\leqslant f(\boldsymbol{e},b)<n+Cm<+\infty.$$

对于 f,由水平集

$$S:=\{(\boldsymbol{w},b)\mid f(\boldsymbol{w},b)<n+Cm\}.$$

由于$(\boldsymbol{e};b)\in S$,所以 $S\neq\varnothing$,因此全局最优解存在。对于$\forall(\boldsymbol{w},b)\in S$,有

$$\|\boldsymbol{w}\|_1<f(\boldsymbol{w},b)\leqslant n+Cm.$$

实数 $C\in[-K,K]$,则 S 是有界的。由于 $L_{0/1}$ 损失函数和范数均为下半连续函数,所以函数 f 也是下半连续函数,即问题的全局最优解存在且解集是有界的。

接下来给出模型的 KKT 点,并证明全局最优解与 KKT 点的关系。

定义 6 对于模型(13),给定 $C>0$,称$(\boldsymbol{w}^*,b^*,\boldsymbol{u}^*,\boldsymbol{s}^*)$是问题的 KKT 点,即如果存在$\boldsymbol{\rho}_1^*\in\mathbb{R}^m$,$\boldsymbol{\rho}_2^*\in\mathbb{R}^m$满足下式

$$\begin{cases}0\in\partial\|\boldsymbol{w}\|_1-\boldsymbol{\rho}_1^*A\\ \boldsymbol{y}\boldsymbol{\rho}_1^*=\boldsymbol{0}\\ \boldsymbol{0}\in C\partial\|(\boldsymbol{u}^*)_+\|_0+\boldsymbol{\rho}_2^*\\ \boldsymbol{\rho}_1^*+\boldsymbol{\rho}_2^*=\boldsymbol{0}\\ A\boldsymbol{w}^*+b^*\boldsymbol{y}=\boldsymbol{s}^*\\ \boldsymbol{e}-\boldsymbol{s}^*=\boldsymbol{u}^*\end{cases}. \tag{14}$$

定理 2 若$(\boldsymbol{w}^*,b^*,\boldsymbol{u}^*,\boldsymbol{s}^*)$是问题(13)的全局最优解,如果存在 $D=\{(\boldsymbol{w},b,\boldsymbol{s})\mid\boldsymbol{s}-A\boldsymbol{w}-b\boldsymbol{y}=\boldsymbol{0}\}$,$\Pi_d=\{d\mid d=\|\boldsymbol{w}\|_1\}$,使得 $N_D(\boldsymbol{w})\cap N_{\Pi_d}(\boldsymbol{w})=\{\boldsymbol{0}\}$,则点$(\boldsymbol{w}^*,b^*,\boldsymbol{u}^*,\boldsymbol{s}^*)$也是问题(13)的 KKT 点。

证明:若$(\boldsymbol{w}^*,b^*,\boldsymbol{u}^*,\boldsymbol{s}^*)$是问题(13)的全局最优解,则也是可行解。即

$$\begin{cases}(\boldsymbol{w}^*,b^*,\boldsymbol{s}^*)\in D_1=\{(\boldsymbol{w},b,\boldsymbol{s})\mid\boldsymbol{s}-A\boldsymbol{w}-b\boldsymbol{y}=0\}\\ (\boldsymbol{u}^*,\boldsymbol{s}^*)\in D_2=\{(\boldsymbol{u},\boldsymbol{s})\mid\boldsymbol{u}-\boldsymbol{e}+\boldsymbol{s}=0\}\end{cases}.$$

要证明$(\boldsymbol{w}^*,b^*,\boldsymbol{u}^*,\boldsymbol{s}^*)$也是 KKT 点,则需要证明

$$\begin{aligned}&\hat{\partial}\{\|\boldsymbol{w}\|_1+C\|\boldsymbol{u}_+\|_0+\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})\}\\&=(\hat{\partial}\|\boldsymbol{w}\|_1,0,0,0)+(0,0,C\hat{\partial}\|\boldsymbol{u}_+\|_0,0)+\\&\quad N_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+N_{D_2}(\boldsymbol{u},\boldsymbol{s}).\end{aligned} \tag{15}$$

(1)对于$\forall(\boldsymbol{w},b,\boldsymbol{s})\in D_2$,$\forall(\boldsymbol{u},\boldsymbol{s})\in D_2$,根据定理显然有

$$\begin{aligned}&(\hat{\partial}\|\boldsymbol{w}\|_1,0,0,0)+(0,0,C\hat{\partial}\|\boldsymbol{u}_+\|_0,0)+\\&N_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+N_{D_2}(\boldsymbol{u},\boldsymbol{s})\subseteq\hat{\partial}\{\|\boldsymbol{w}\|_1+\\&C\|\boldsymbol{u}_+\|_0+\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})\},\end{aligned}$$

其中 $N_{D_1}(\boldsymbol{w},b,\boldsymbol{s})$,$N_{D_2}(\boldsymbol{u},\boldsymbol{s})$各为 D_1,D_2 在$(\boldsymbol{w},b,\boldsymbol{s})$,$(\boldsymbol{u},\boldsymbol{s})$处的法锥。

(2)给定$(\boldsymbol{r},r_1,\boldsymbol{r}_2)\in\hat{\partial}_{(\boldsymbol{w},b,\boldsymbol{s})}\{\|\boldsymbol{w}\|_1+\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s})\}$,根据定义 4 正则次微分,对$\forall(\boldsymbol{w},b,\boldsymbol{s})\to(\boldsymbol{w}^*,b^*,\boldsymbol{s}^*)$,有

$$\begin{aligned}&o\{\|(\boldsymbol{w},b,\boldsymbol{s})-(\boldsymbol{w}^*,b^*,\boldsymbol{s}^*)\|\}\\&\leqslant\|\boldsymbol{w}\|_1+\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s})-\|\boldsymbol{w}^*\|_1-\\&\quad\delta_{D_1}(\boldsymbol{w}^*,b^*,\boldsymbol{s}^*)-\langle\boldsymbol{r},(\boldsymbol{w}-\boldsymbol{w}^*)\rangle-\\&\quad r_1(b-b^*)-\langle\boldsymbol{r}_2,(\boldsymbol{s}-\boldsymbol{s}^*)\rangle.\end{aligned}$$

当$(\tilde{\boldsymbol{w}},\tilde{b},\tilde{\boldsymbol{s}})\to(\boldsymbol{w}^*,b^*,\boldsymbol{s}^*)$且$\tilde{\boldsymbol{w}}\in\delta_{D_1}(\boldsymbol{w},b)\cap\Pi_d$,有$\|\tilde{\boldsymbol{w}}\|_1=\|\boldsymbol{w}\|_1$,则

$$\begin{aligned}&o\{\|(\boldsymbol{w},b,\boldsymbol{s})-(\boldsymbol{w}^*,b^*,\boldsymbol{s}^*)\|\}\\&\leqslant-\langle\boldsymbol{r},(\boldsymbol{w}-\boldsymbol{w}^*)\rangle-r_1(b-b^*).\end{aligned}$$

则$(\boldsymbol{r},r_1,\boldsymbol{r}_2)\in N_{D_1\cap(\Pi_d\times\mathbb{R}\times\mathbb{R}^m)}$,由于 Π_d,D_1 均为凸集,且 $N_D(\boldsymbol{w})\cap N_{\Pi_d}(\boldsymbol{w})=\{\boldsymbol{0}\}$,可得

$$\begin{aligned}&N_{D_1\cap(\Pi_d\times\mathbb{R}\times\mathbb{R}^m)}(\boldsymbol{w},b,\boldsymbol{s})\\&=N_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+N_{(\Pi_d\times\mathbb{R}\times\mathbb{R}^m)}(\boldsymbol{w},b,\boldsymbol{s}).\end{aligned}$$

因此

$$\begin{aligned}&\hat{\partial}_{(\boldsymbol{w},b,\boldsymbol{s})}(\|\boldsymbol{w}\|_1+\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s}))\\&=\hat{\partial}\|\boldsymbol{w}\|_1+\hat{\partial}_{(\boldsymbol{w},b,\boldsymbol{s})}\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s}).\end{aligned}$$

同理给定$(\boldsymbol{k}_1,\boldsymbol{k}_2)\in\hat{\partial}_{(\boldsymbol{u},\boldsymbol{s})}\{C\|\boldsymbol{u}_+\|_0+\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})\}$,根据定义 4 正则次微分的定义,对于$\forall(\boldsymbol{u},\boldsymbol{s})\to(\boldsymbol{u}^*,\boldsymbol{s}^*)$,有

$$\begin{aligned}&o\{\|(\boldsymbol{u},\boldsymbol{s})-(\boldsymbol{u}^*,\boldsymbol{s}^*)\|\}\\&\leqslant C\|\boldsymbol{u}_+\|_0+\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})-C\|\boldsymbol{u}^*{}_+\|_0-\\&\quad\delta_{D_2}(\boldsymbol{u}^*,\boldsymbol{s}^*)-\langle\boldsymbol{k}_1,\boldsymbol{u}-\boldsymbol{u}^*\rangle-\langle\boldsymbol{k}_2,\boldsymbol{s}-\boldsymbol{s}^*\rangle.\end{aligned}$$

当$(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{s}})\to(\boldsymbol{u}^*,\boldsymbol{s}^*)$,且$\tilde{\boldsymbol{u}}\in\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})$,$\tilde{\boldsymbol{u}}\in\mathbb{R}_{I_x}$,其中 $I_x=\{i\mid\boldsymbol{u}_i\neq0\}$且$\mathbb{R}_{I_y}=\{\boldsymbol{u}\in\mathbb{R}^m\mid\boldsymbol{u}_i\leqslant0,i\notin I_x\}$,则$\|\tilde{\boldsymbol{u}}_+\|_0=\|\boldsymbol{u}_+\|_0$,则

$$\begin{aligned}&o\{\|(\boldsymbol{u},\boldsymbol{s})-(\boldsymbol{u}^*,\boldsymbol{s}^*)\|\}\\&\leqslant-\langle\boldsymbol{k}_1,\boldsymbol{u}-\boldsymbol{u}^*\rangle-\langle\boldsymbol{k}_2,\boldsymbol{s}-\boldsymbol{s}^*\rangle.\end{aligned}$$

则$(\boldsymbol{u},\boldsymbol{s})\in N_{(\mathbb{R}_{I_x}\times\mathbb{R}^m)\cap D_2}$,由于$\mathbb{R}_{I_x}$,$D_2$ 均为凸集,可得

$$N_{(\mathbb{R}_{I_x}\times\mathbb{R}^m)\cap D_2}(\boldsymbol{u},\boldsymbol{s})=N_{\mathbb{R}_{I_x}}(\boldsymbol{u},\boldsymbol{s})+N_{D_2}(\boldsymbol{u},\boldsymbol{s}).$$

因此

$$\begin{aligned}&\hat{\partial}_{(\boldsymbol{u},\boldsymbol{s})}\{C\|\boldsymbol{u}_+\|_0+\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})\}\\&=C\hat{\partial}_{(\boldsymbol{u},\boldsymbol{s})}\|\boldsymbol{u}_+\|_0+N_{D_2}(\boldsymbol{u},\boldsymbol{s}).\end{aligned}$$

即

$$\begin{aligned}&\hat{\partial}\{\|\boldsymbol{w}\|_1+C\|\boldsymbol{u}_+\|_0+\delta_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+\delta_{D_2}(\boldsymbol{u},\boldsymbol{s})\}\\&\subseteq(\hat{\partial}\|\boldsymbol{w}\|_1,0,0,0)+(0,0,C\hat{\partial}\|\boldsymbol{u}_+\|_0,0)+\\&N_{D_1}(\boldsymbol{w},b,\boldsymbol{s})+N_{D_2}(\boldsymbol{u},\boldsymbol{s}).\end{aligned}$$

综上所述,对于全局最优解$(\boldsymbol{w}^*,b^*,\boldsymbol{u}^*,\boldsymbol{s}^*)$有式(15)成立,定理得证。

最后给出模型的 P-稳定点。

通过引入变量 $\boldsymbol{z}=\boldsymbol{w}\in\mathbb{R}^n$,将模型转化为

$$\begin{aligned}&\min_{\boldsymbol{w}\in\mathbb{R}^n,b\in\mathbb{R},\boldsymbol{z}\in\mathbb{R}^n,\boldsymbol{u}\in\mathbb{R}^m}\|\boldsymbol{z}\|_1+C\|\boldsymbol{u}_+\|_0\\&\text{s.t.}\quad\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}=\boldsymbol{s},\\&\qquad\boldsymbol{z}=\boldsymbol{w}.\end{aligned} \tag{16}$$

定义 7 对于模型(16),给定 $C>0$,我们称 $(\boldsymbol{w}^*,b^*,\boldsymbol{z}^*,\boldsymbol{u}^*)$ 是临近点算子稳定点,也称 P-稳定点,即如果存在拉格朗日乘子 $\boldsymbol{\lambda}^*\in\mathbb{R}^m$,$\boldsymbol{\lambda}_1^*\in\mathbb{R}^n$ 和常数 $\gamma_1,\gamma_2>0$,满足

$$\begin{cases} A^\top\boldsymbol{\lambda}^*-\boldsymbol{\lambda}_1^*=0 \\ \langle \boldsymbol{y},\boldsymbol{\lambda}^*\rangle=0 \\ Prox_{\gamma_1\|\cdot\|_1}(\boldsymbol{z}^*-\gamma_1\boldsymbol{\lambda}_1^*)=\boldsymbol{z}^* \\ Prox_{\gamma_2 C\|(\cdot)_+\|_0}(\boldsymbol{u}^*-\gamma_2\boldsymbol{\lambda}^*)=\boldsymbol{u}^* \\ \boldsymbol{u}^*+A\boldsymbol{w}^*+b^*\boldsymbol{y}-\boldsymbol{e}=0 \\ \boldsymbol{z}^*-\boldsymbol{w}^*=0 \end{cases}. \tag{17}$$

这里的 $[Prox_{\gamma_1\|\cdot\|_1}(\boldsymbol{g}^*)]_i=\begin{cases} g_i^*-\gamma_1, & g_i^*>\gamma_1 \\ 0, & |g_i^*|\leqslant\gamma_1, \\ g_i^*+\gamma_1, & g_i^*<-\gamma_1 \end{cases}$

其中 $\boldsymbol{g}^*:=z^*-\gamma_1\boldsymbol{\lambda}_1^*$,$[Prox_{\gamma_2 C\|(\cdot)_+\|_0}(\boldsymbol{h}^*)]_i=\begin{cases} 0, & 0<h_i^*\leqslant\sqrt{2\gamma_2 C} \\ h_i^*, & h_i^*>\sqrt{2\gamma_2 C}\ or\ h_i^*\leqslant 0 \end{cases}$,其中 $\boldsymbol{h}^*:=\boldsymbol{u}^*-\gamma_2\boldsymbol{\lambda}^*$。

3 算法及收敛性分析

本节设计相应的算法($L_{0/1}$-SSVM)求解该模型,并对算法进行收敛性分析

3.1 算法

对于模型(16),相应的增广拉格朗日函数为

$$\begin{aligned} &L_\sigma(\boldsymbol{w},b,\boldsymbol{z},\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{\lambda}_1) \\ &=\|\boldsymbol{z}\|_1+C\|\boldsymbol{u}_+\|_0+ \\ &\quad\langle\boldsymbol{\lambda},\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}\rangle+ \\ &\quad\langle\boldsymbol{\lambda}_1,\boldsymbol{z}-\boldsymbol{w}\rangle+\frac{\sigma}{2}\|\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}\|^2+ \\ &\quad\frac{\sigma}{2}\|\boldsymbol{z}-\boldsymbol{w}\|^2, \end{aligned}$$

其中 $\boldsymbol{\lambda}\in\mathbb{R}^m$,$\boldsymbol{\lambda}_1\in\mathbb{R}^n$ 是拉格朗日乘子,$\sigma>0$ 是罚参数。

为简化求解,令 $\boldsymbol{v}=\frac{1}{\sigma}\boldsymbol{\lambda}_1$ 为缩放对偶变量,则缩放后的 $ADMM$ 的增广拉格朗日函数为

$$\begin{aligned} &L_\sigma(\boldsymbol{w},b,\boldsymbol{z},\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{v}) \\ &=\|\boldsymbol{z}\|_1+C\|\boldsymbol{u}_+\|_0+\langle\boldsymbol{\lambda},\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}\rangle+ \\ &\quad\frac{\sigma}{2}\|\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}\|^2+\frac{\sigma}{2}\|\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v}\|^2. \end{aligned}$$

对应的 ADMM 迭代框架如下

$$\begin{aligned} \boldsymbol{u}^{k+1}&=\arg\min_{\boldsymbol{u}\in\mathbb{R}^m}L_\sigma(\boldsymbol{w}^k,b^k,\boldsymbol{z}^k,\boldsymbol{u},\boldsymbol{\lambda}^k,\boldsymbol{v}^k), \\ \boldsymbol{z}^{k+1}&=\arg\min_{\boldsymbol{z}\in\mathbb{R}^n}L_\sigma(\boldsymbol{w}^k,b^k,\boldsymbol{z},\boldsymbol{u}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{v}^k), \\ \boldsymbol{w}^{k+1}&=\arg\min_{\boldsymbol{w}\in\mathbb{R}^n}L_\sigma(\boldsymbol{w},b^k,\boldsymbol{z}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{v}^k)+ \\ &\quad\frac{\sigma}{2}\|\boldsymbol{w}-\boldsymbol{w}^k\|_{V_k}^2, \\ b^{k+1}&=\arg\min_{b\in\mathbb{R}}L_\sigma(\boldsymbol{w}^{k+1},b,\boldsymbol{z}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{v}^k), \\ \boldsymbol{\lambda}^{k+1}&=\boldsymbol{\lambda}^k+\eta\sigma(\boldsymbol{u}^{k+1}-\boldsymbol{e}+A\boldsymbol{w}^{k+1}+b^{k+1}\boldsymbol{y}), \\ \boldsymbol{v}^{k+1}&=\boldsymbol{v}^k+\eta(\boldsymbol{w}^{k+1}-\boldsymbol{z}^{k+1}). \end{aligned} \tag{18}$$

其中 $\eta>0$ 是对偶步长,近端项 $\|\boldsymbol{w}-\boldsymbol{w}^k\|_{V_k}^2=\langle\boldsymbol{w}-\boldsymbol{w}^k,V_k(\boldsymbol{w}-\boldsymbol{w}^k)\rangle$,由于 $\boldsymbol{w}$-子问题是强凸的,因此 V_k 为一个不定矩阵。

每个子问题的具体迭代过程如下:

(1)$\boldsymbol{u}$-子问题的迭代

$\boldsymbol{u}$-子问题可以表示为

$$\begin{aligned} \boldsymbol{u}^{k+1}=&\arg\min_{\boldsymbol{u}\in\mathbb{R}^m}C\|\boldsymbol{u}_+\|_0+ \\ &\frac{\sigma}{2}\left\|\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}+\frac{\boldsymbol{\lambda}}{\sigma}\right\|^2. \end{aligned}$$

利用定义,得到 $\boldsymbol{u}$ 的迭代如下:

$$\boldsymbol{u}^{k+1}=Prox_{\frac{C}{\sigma}\|(\cdot)_+\|_0}(\boldsymbol{e}-A\boldsymbol{w}^k+b^k\boldsymbol{y}-\frac{\boldsymbol{\lambda}_k}{\sigma}). \tag{19}$$

(2)$\boldsymbol{z}$-子问题的迭代

$\boldsymbol{z}$-子问题可以表示为

$$\boldsymbol{z}^{k+1}=\arg\min_{\boldsymbol{z}\in\mathbb{R}^n}\|\boldsymbol{z}\|_1+\frac{\sigma}{2}\|\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v}\|^2.$$

$0\in\partial\left(\|\boldsymbol{z}\|_1+\frac{\sigma}{2}\|\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v}\|^2\right)=\sigma(\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v})+\partial\|\boldsymbol{z}\|_1$ 即 $-\sigma(\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v})=\partial\|\boldsymbol{z}\|_1=t_i$,其中

$$t_i=\begin{cases} \{1\}, & \boldsymbol{z}_i>0 \\ \{-1\}, & \boldsymbol{z}_i<0,i=1,\cdots,m. \\ [-1,1], & \boldsymbol{z}_i=0 \end{cases}$$

结合之后进行化简归类,可以得到

$$\begin{cases} \boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v}=-\frac{1}{\sigma}sign(\boldsymbol{z}_i), & \boldsymbol{z}_i\neq 0 \\ |\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v}|=\frac{1}{\sigma}, & \boldsymbol{z}_i=0 \end{cases}.$$

令 $\boldsymbol{r}^k=\boldsymbol{w}^k-\boldsymbol{v}^k$,得到 $\boldsymbol{z}$ 的迭代如下:

$$\boldsymbol{z}^{k+1}=S_{\frac{1}{\sigma}}(r_i^k)=\begin{cases} r_i^k-\frac{1}{\sigma}, & r_i^k>\frac{1}{\sigma} \\ 0, & |r_i^k|\leqslant\frac{1}{\sigma}. \\ r_i^k+\frac{1}{\sigma}, & r_i^k<\frac{1}{\sigma} \end{cases} \tag{20}$$

(3)$\boldsymbol{w}$-子问题的迭代

$\boldsymbol{w}$-子问题可以表示为

$$\boldsymbol{w}^{k+1}=\underset{\boldsymbol{w}\in\mathbb{R}^n}{\operatorname{argmin}}\langle\boldsymbol{\lambda},A\boldsymbol{w}\rangle+\frac{\sigma}{2}\|\boldsymbol{w}-\boldsymbol{w}^k\|^2_{-A^\top_{\bar V_k}A_{\bar V_k}}+\frac{\sigma}{2}\|\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}\|^2+\frac{\sigma}{2}\|\boldsymbol{z}-\boldsymbol{w}+\boldsymbol{v}\|^2.$$

$\boldsymbol{w}$-子问题是一个关于 $\boldsymbol{w}$ 的凸规划问题,即有

$$\boldsymbol{0}=A^\top\boldsymbol{\lambda}^k-\sigma A^\top_{\bar V_k}A_{\bar V_k}(\boldsymbol{w}-\boldsymbol{w}^k)+\sigma A^\top(\boldsymbol{u}^{k+1}+A\boldsymbol{w}+b^k\boldsymbol{y}-\boldsymbol{e})-\sigma(\boldsymbol{z}^k-\boldsymbol{w}+\boldsymbol{v}^k).$$

得到 $\boldsymbol{w}$ 的迭代如下:

$$(I+A^\top_{\bar V_k}A_{\bar V_k})\boldsymbol{w}^{k+1}=A^\top\left[\boldsymbol{e}-\boldsymbol{u}^{k+1}-b^k\boldsymbol{y}-\frac{\boldsymbol{\lambda}^k}{\sigma}+\frac{1}{A^\top}(\boldsymbol{z}^{k+1}-\boldsymbol{v}^k)\right].\tag{21}$$

(4)b-子问题的迭代

b-子问题可以表示为

$$b^{k+1}=\underset{b\in\mathbb{R}}{\operatorname{argmin}}\langle\boldsymbol{\lambda},b\boldsymbol{y}\rangle+\frac{\sigma}{2}\|\boldsymbol{u}+A\boldsymbol{w}+b\boldsymbol{y}-\boldsymbol{e}\|^2.$$

得到 b 的迭代如下:

$$b^{k+1}=\frac{-\langle\boldsymbol{y},A\boldsymbol{w}^{k+1}-\boldsymbol{e}+\boldsymbol{u}^{k+1}+\frac{\boldsymbol{\lambda}^k}{\sigma}\rangle}{\boldsymbol{y}^\top\boldsymbol{y}}.\tag{22}$$

(5)更新 $\boldsymbol{\lambda}$

$\boldsymbol{\lambda}$ 的更新如下

$$\boldsymbol{\lambda}^{k+1}=\boldsymbol{\lambda}^k+\eta\sigma(\boldsymbol{u}^{k+1}+A\boldsymbol{w}^{k+1}+b^{k+1}\boldsymbol{y}-\boldsymbol{e}).\tag{23}$$

(6)更新 $\boldsymbol{v}$

$\boldsymbol{v}$ 的更新如下

$$\boldsymbol{v}^{k+1}=\boldsymbol{v}^k+\eta(\boldsymbol{w}^{k+1}+\boldsymbol{z}^{k+1}).\tag{24}$$

根据以上迭代规则,给出 $L_{0/1}$ 损失稀疏支持向量机的 ADMM 算法迭代框架:

算法($L_{0/1}$-SSVM)

初始化 $\boldsymbol{w}^0\in\mathbb{R}^n,\boldsymbol{z}^0\in\mathbb{R}^n,\boldsymbol{u}^0\in\mathbb{R}^m,b^0\in\mathbb{R},\boldsymbol{\lambda}^0\in\mathbb{R}^m,\boldsymbol{v}^0\in\mathbb{R}^m$;

循环 for k =0:MaxIter do

步 1 通过公式(19)计算$\boldsymbol{u}^{k+1}$;

步 2 通过公式(20)计算$\boldsymbol{z}^{k+1}$;

步 3 通过公式(21)计算$\boldsymbol{w}^{k+1}$;

步 4 通过公式(22)计算 b^{k+1};

步 5 通过公式(23)更新$\boldsymbol{\lambda}^{k+1}$;

步 6 通过公式(24)更新$\boldsymbol{v}^{k+1}$;

步 7 当满足终止条件(17)时迭代停止,否则置 $k=k+1$,转步 1。

输出 $(\boldsymbol{w},b)$

3.2 收敛性分析

定理 3 假设$(\boldsymbol{w}^*,b^*,\boldsymbol{z}^*,\boldsymbol{u}^*,\boldsymbol{\lambda}^*,\boldsymbol{v}^*)$为序列$\{(\boldsymbol{w}^k,b^k,\boldsymbol{z}^k,\boldsymbol{u}^k,\boldsymbol{\lambda}^k,\boldsymbol{v}^k)\}$通过算法 ADMM 生成的极限点,那么$(\boldsymbol{w}^*,b^*,\boldsymbol{z}^*,\boldsymbol{u}^*)$是一个 P-稳定点,而且是问题(16)的局部最优解。

证明:设$\boldsymbol{h}^k:=\boldsymbol{e}-A\boldsymbol{w}^k-b^k\boldsymbol{y}-\frac{\boldsymbol{\lambda}^k}{\sigma}$,由于 $V_k=\left\{\boldsymbol{z}^k\mid z_i^k\in\left(0,\sqrt{\frac{2C}{\sigma}}\right],i\in\mathbb{N}_m\right\}$以及其补集 $\overline{V_k}:=\mathbb{N}_m\setminus V_k$,因此 $V_k\subseteq\mathbb{N}_m$有有限多个元素,对于足够大的 k,有子集 $J\subseteq\{1,2,3,\cdots\}$使

$$V_j\equiv:V\quad\forall j\in J.$$

为了表示简单,设 $\Lambda^k:=(\boldsymbol{w}^k,b^k,\boldsymbol{z}^k,\boldsymbol{u}^k,\boldsymbol{\lambda}^k,\boldsymbol{v}^k)$及其极限点 $\Lambda^*:=(\boldsymbol{w}^*,b^*,\boldsymbol{z}^*,\boldsymbol{u}^*,\boldsymbol{\lambda}^*,\boldsymbol{v}^*)$,即$\{\Lambda^k\}\to\Lambda^*$,这也表示$\{\Lambda^j\}_{j\in J}\to\Lambda^*$ 和$\{\Lambda^{j+1}\}_{j+1\in J}\to\Lambda^*$。

(1)将$\boldsymbol{\lambda}^k$ 在 J 处取极限,即 $k\in J,k\to\infty$,有

$$\begin{cases}\boldsymbol{\lambda}^*_V=\boldsymbol{\lambda}^*_V-\eta\sigma(\boldsymbol{u}^*_V+A\boldsymbol{w}^*_V+b^*_V\boldsymbol{y}-\boldsymbol{e})\\ \boldsymbol{\lambda}^*_{\bar V}=\boldsymbol{0}\end{cases}.$$

由上式可知$\boldsymbol{u}^*_V+A\boldsymbol{w}^*_V+b^*_V\boldsymbol{y}-\boldsymbol{e}=\boldsymbol{0}$

(2)将$\boldsymbol{u}^k$ 在 J 处取极限,即 $k\in J,k\to\infty$,有

$$\boldsymbol{h}^*=\boldsymbol{e}-A\boldsymbol{w}^*-b^*\boldsymbol{y}-\frac{\boldsymbol{\lambda}^*}{\sigma}=[\boldsymbol{e}-A\boldsymbol{w}^*-b^*\boldsymbol{y}-\boldsymbol{u}^*]+\boldsymbol{u}^*-\frac{\boldsymbol{\lambda}^*}{\sigma}.$$

又由 $L_{0/1}$ 近端算子的定义有

$$\begin{cases}\boldsymbol{u}^*_V=\boldsymbol{0} & \text{(a)}\\ \boldsymbol{u}^*_{\bar V}=\boldsymbol{h}^*_{\bar V} & \text{(b)}\end{cases}$$

因此由(b)和$\boldsymbol{\lambda}^*_{\bar V}=\boldsymbol{0}$ 有

$$\boldsymbol{u}^*_{\bar V}=\boldsymbol{h}^*_{\bar V}=[\boldsymbol{e}-A\boldsymbol{w}^*_{\bar V}-b^*_{\bar V}\boldsymbol{y}-\boldsymbol{u}^*_{\bar V}]+\boldsymbol{u}^*_{\bar V}-\frac{\boldsymbol{\lambda}^*_{\bar V}}{\sigma}=[\boldsymbol{e}-A\boldsymbol{w}^*_{\bar V}-b^*_{\bar V}\boldsymbol{y}-\boldsymbol{u}^*_{\bar V}]+\boldsymbol{u}^*_{\bar V}.$$

可以得出 $\boldsymbol{e}-A\boldsymbol{w}^*_{\bar V}-b^*_{\bar V}\boldsymbol{y}-\boldsymbol{u}^*_{\bar V}=\boldsymbol{0}$,

又因为$\boldsymbol{u}^*_V+A\boldsymbol{w}^*_V+b^*_V\boldsymbol{y}-\boldsymbol{e}=\boldsymbol{0}$,

得到$\boldsymbol{u}^*+A\boldsymbol{w}^*+b^*\boldsymbol{y}-\boldsymbol{e}=\boldsymbol{0}$,即$\boldsymbol{h}^*=\boldsymbol{u}^*-\frac{\boldsymbol{\lambda}^*}{\sigma}$,

因此由 $L_{0/1}$ 近端算子的定义有

$$\boldsymbol{u}^*=Prox_{\frac{C}{\sigma}\|(\cdot)_+\|_0}(h^*)=Prox_{\frac{C}{\sigma}\|(\cdot)_+\|_0}\left(\boldsymbol{u}^*-\frac{\boldsymbol{\lambda}^*}{\sigma}\right).$$

(3)将$\boldsymbol{v}^k$ 在 J 处取极限,即 $k\in J,k\to\infty$,有

$$\boldsymbol{v}^*=\boldsymbol{v}^*+\sigma(\boldsymbol{z}^*-\boldsymbol{w}^*).$$

得到$\boldsymbol{z}^*-\boldsymbol{w}^*=0$。

(4)将$\boldsymbol{z}^k$ 在 J 处取极限,即 $k\in J$,$k\to\infty$,有

$$\boldsymbol{z}^*=Prox_{\frac{1}{\sigma}\|\cdot\|_1}(\boldsymbol{w}^*-\boldsymbol{v}^*)=Prox_{\frac{1}{\sigma}\|\cdot\|_1}(\boldsymbol{z}^*-\boldsymbol{v}^*).$$

(5)将$\boldsymbol{w}^k$ 在 J 处取极限,即 $k\in J$,$k\to\infty$,有

$$\begin{aligned}&(I+A_V^\top A_V)\boldsymbol{w}^*\\&=A_V^\top\left[\boldsymbol{e}-\boldsymbol{u}_V^*-b^*\boldsymbol{y}_V-\frac{\boldsymbol{\lambda}_V^*}{\sigma}+\frac{1}{A_V^\top}(\boldsymbol{z}^*+\boldsymbol{v}^*)\right]\\&=A_V^\top A_V\boldsymbol{w}^*-\frac{A_V^\top\boldsymbol{\lambda}_V^*}{\sigma}+\boldsymbol{z}^*+\boldsymbol{v}^*.\end{aligned}$$

即$\boldsymbol{w}^*+\boldsymbol{w}^*A_V^\top A_V=A_V^\top A_V\boldsymbol{w}^*-\frac{A_V^\top\boldsymbol{\lambda}_V^*}{\sigma}+\boldsymbol{z}^*+\boldsymbol{v}^*$,

有$\boldsymbol{0}=-\frac{A_V^\top\boldsymbol{\lambda}_V^*}{\sigma}+\boldsymbol{v}^*$,由于$\boldsymbol{v}^*=\frac{\boldsymbol{\lambda}_1}{\sigma}$,得到

$$A_V^\top\boldsymbol{\lambda}_V^*-\boldsymbol{\lambda}_1^*=\boldsymbol{0}.$$

(6)将$\boldsymbol{b}^k$ 在 J 处取极限,即 $k\in J$,$k\to\infty$,有

$$\begin{aligned}b^*&=\frac{-\langle\boldsymbol{y},A\boldsymbol{w}^*-\boldsymbol{e}+\boldsymbol{u}^*+\frac{\boldsymbol{\lambda}^*}{\sigma}\rangle}{\boldsymbol{y}^\top\boldsymbol{y}}\\&=b^*-\frac{\langle\boldsymbol{y},\boldsymbol{\lambda}^*\rangle}{\sigma\boldsymbol{y}^\top\boldsymbol{y}}.\end{aligned}$$

得到$\langle\boldsymbol{y},\boldsymbol{\lambda}^*\rangle=\boldsymbol{0}$。

综上所述$(\boldsymbol{w}^*,b^*,\boldsymbol{z}^*,\boldsymbol{u}^*)$是一个 P-稳定点,而且是问题的局部最优解,证明完成。

4 总结与展望

4.1 总结

本文建立了一个基于 $L_{0/1}$ 损失函数,使用 l_1 范数作为正则项的稀疏支持向量机($L_{0/1}$-SSVM),给出了模型的一阶最优性条件,包括解的性质,模型的 KKT 点、P-稳定点以及与全局最优解的关系。提出相应的 ADMM 算法,通过 l_1 范数的临近点算子推导出子问题的迭代表达式,并证明算法是收敛的。

4.2 展望

本文对 $L_{0/1}$ 稀疏支持向量机进行了一定的研究,但仍不太完善,在模型方可以使用更稀疏的 l_0 范数作为正则项,还可以拓展到 n 维空间,建立基于 $L_{0/1}$ 损失的稀疏支持张量机等;对于模型的二阶最优性条件也可以作深入研究;本文在数值实验方面的研究有所欠缺,可以将其应用于实际问题,验证算法的有效性。

利益冲突:作者声明无利益冲突。

参考文献(References)

[1] Boyd S,Parikh N,Chu E,et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends® in Mac-hine learning,2011,3(1):1-122.
https://doi.org/10.1561/2200000016.

[2] Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization[J]. IEEE Signal Processing L-etters,2007,14(10):707-710.
https://doi.org/10.1109/LSP.2007.898300.

[3] Wright J,Ma Y,Mairal J,et al. Sparse representation for computer vision and pattern recognition[J]. Proceedings of the IEEE,2010,98(6):1031-1044.

[4] Liu J,Chen J,Ye J. Large-scale sparse logistic regression [C]. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009:547-556.
https://doi.org/10.1145/1557019.1557082.

[5] Bienstock D. Computational study of a famliy of mixed-integer quadratic programming problems[J]. Mathematical programming,1996,74:121-140.
https://doi.org/10.1007/BF02592208.

[6] Zou H,Hastie T,Tibshirani R. Sparse principal component analysis[J]. Journal of computational and graphical statistics,2006,15(2):265-286.
https://doi.org/10.1198/106186006X113430.

[7] Cortes C,Vapnik V. Support-vector networks[J]. Machine Learning,1995,20:273-297.
https://doi.org/10.1007/BF00994018.

[8] Schölkopf B,Smola A J. Learning with kernels:support vector machines,regularization,optimization,and beyond [M]. MIT press,2002.
https://doi.org/10.1198/jasa.2003.s269.

[9] Elad M,Figueiredo M A T,Ma Y. On the role of sparse and redundant representations in image processing[J]. Proceedings of the IEEE,2010,98(6):972-982.
https://doi.org/10.1109/jproc.2009.2037655.

[10] Bartlett P L,Wegkamp M H. Classification with a Reject Option using a Hinge Loss. Journal of Machine

Learning Research,2008,9(8).

[11] Jumutc V,Huang X,Suykens J A K. Fixed-size Pegasos f-or hinge and pinball loss SVM[C]. The 2013 International Joint Conference on Neural Networks(IJCNN). IEEE,2013:1-7.
https://doi.org/10.1109/ijcnn.2013.6706864.

[12] Shen X,Tseng G C,Zhang X,et al. On psi-Learning [J]. Journal of the American Statistical Association, 2003,98(1):724-734.

[13] Wang H,Shao Y,Zhou S,et al. Support Vector Machine Classifier via $L_{0/1}$ Soft-Margin Loss [J]. IEEE transactions on pattern analysis and machine intelligence,2021,44(10):7253-7265.
https://doi.org/10.1109/TPAMI.2021.3092177.

[14] Zhu J,Rosset S,Tibshirani R,et al. 1-norm support vector machines[J]. Advances in neural information processing systems,2003,16.

[15] Tanveer M,Sharma S,Rastogi R,et al. Sparse support vector machine with pinball loss[J]. Transactions on Emerging Telecommunications Technologies, 2021, 32 (2):e3820.
https://doi.org/10.1002/ett.3820.

[16] Zheng Q,Zhu F,Qin J,et al. Sparse support matrix machine[J]. Pattern Recognition,2018,76:715-726.
https://doi.org/10.1016/j.patcog.2017.10.003.

[17] Wen F,Adhikari L,Pei L,et al. Nonconvex regularization-based sparse recovery and demixing with application to color image inpainting[J]. IEEE Access,2017, 5:11513-11527.
https://doi.org/10.1109/access.2017.2705646.

[18] Shao Y H,Li C N,Huang L W,et al. Joint sample and feature selection via sparse primal and dual LSSVM [J]. Knowledge-Based Systems,2019,185:104915.
https://doi.org/10.1016/j.knosys.2019.104915.

[19] Makela M M,Neittaanmaki P. Nonsmooth optimization:analysis and algorithms with applications to optimal control[M]. 5th Edition. World Scientific,1992.

[20] 赵晨,罗自炎,修乃华.稀疏优化理论与算法若干新进展[J].运筹学学报,2020,24(4):1-24.
https://doi.org/10.15960/j.cnki.issn.1007-6093.2020.04.001.

Optimality Conditions and Algorithms for Sparse Support Vector Machines with l_1 Regular Terms

YANG Xiaohui, GAO Caixia*

(School of Mathematical Sciences, Inner Mongolia University, Hohhot 010021, China)

Abstract: Support Vector Machine (SVM), as one of the main methods of machine learning, is a popular learning tool used to solve classification and regression tasks, and has attracted much attention in the fields of image classification, pattern recognition and disease diagnosis. In the context of the support vector machine model (SVM), the loss function is considered optimal, with most existing loss functions acting as proxies. Since l_1 norm has good sparsity, redundant features can be removed through feature selection. In this paper, a loss-based norm sparse support vector machine (called $L_{0/1}$-SSVM) is proposed based on the $L_{0/1}$ soft margin loss model. The existence of the model solution is proved, the KKT points and P-stable points of the model are given, and the relationship between the global optimal solution and KKT points is proved. The iterative framework of ADMM algorithm is designed by using the proximal operator of l_1 norm, and the convergence analysis is carried out to prove that the algorithm converges to the P-stable point.

Keywords: l_1 norm; $L_{0/1}$ loss function; sparse support vector machine; optimality conditions; ADMM algorithm

DOI: 10.48014/fcpm.20240411001

Citation: YANG Xiaohui, GAO Caixia. Optimality conditions and algorithms for sparse support vector machines with l_1 regular terms[J]. Frontiers of Chinese Pure Mathematics, 2024, 2(3): 16-24.

Copyright © 2024 by author(s) and Science Footprint Press Co., Limited. This article is open accessed under the CC-BY License (http://creativecommons.org/licenses/by/4.0/).