L1-TWSVM

闫茹钰, 喻望, 张芙作, 郑洁, 朱榕平

自动化系

2018 年 12 月 27 日

目录

1 TWSVM

2 L1-TWSVM

3 L1-TWSVM 收敛性证明

4 数值实验

5 参考文献

目录

1. TWSVM
2. L1-TWSVM
3. L1-TWSVM 收敛性证明
4. 数值实验
5. 参考文献

TWSVM

TWSVM (孪生支持向量机) 是 Jayadeva 等人于 2007 年提出的一种改进的双分界面支持向量机，用于解决二分类问题。

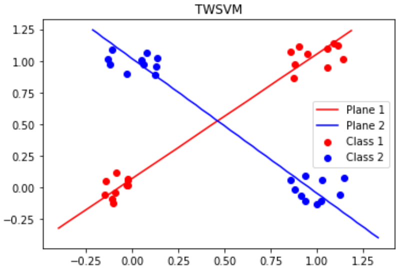
$$
\begin{aligned}
\min_{\mathbf{w}_1, b_1} \ & \frac{1}{2}\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_2^2 + c_1 \mathbf{e}_2^T \mathbf{q}_1 \\
s.t. \ & -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 \geq \mathbf{e}_2 \\
& \mathbf{q}_1 \geq 0
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\min_{\mathbf{w}_2, b_2} \ & \frac{1}{2}\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|_2^2 + c_2 \mathbf{e}_1^T \mathbf{q}_2 \\
s.t. \ & (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \mathbf{q}_2 \geq \mathbf{e}_1 \\
& \mathbf{q}_2 \geq 0
\end{aligned} \tag{2}
$$

TWSVM (孪生支持向量机) 是 Jayadeva 等人于 2007 年提出的一种改进的双分界面支持向量机，用于解决二分类问题。

$$
\begin{aligned}
\min_{\mathbf{w}_1, b_1} \quad & \frac{1}{2}\|\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1\|_2^2 + c_1 \mathbf{e}_2^T \mathbf{q}_1 \\
s.t. \quad & -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 \geq \mathbf{e}_2 \\
& \mathbf{q}_1 \geq 0
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
\min_{\mathbf{w}_2, b_2} \quad & \frac{1}{2}\|\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2\|_2^2 + c_2 \mathbf{e}_1^T \mathbf{q}_2 \\
s.t. \quad & (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \mathbf{q}_2 \geq \mathbf{e}_1 \\
& \mathbf{q}_2 \geq 0
\end{aligned}
\tag{2}
$$



TWSVM 对偶问题

$$\max_{\boldsymbol{\alpha}} \mathbf{e}_2^T\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^T\mathbf{G}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{G}^T\boldsymbol{\alpha} \quad (3)$$
$$s.t.\ 0 \leq \boldsymbol{\alpha} \leq c_1\mathbf{e}_2$$

$$\max_{\boldsymbol{\beta}} \mathbf{e}_1^T\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}^T\mathbf{H}(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{H}^T\boldsymbol{\beta} \quad (4)$$
$$s.t.\ 0 \leq \boldsymbol{\beta} \leq c_2\mathbf{e}_1$$

目录

1 TWSVM

2 L1-TWSVM

3 L1-TWSVM 收敛性证明

4 数值实验

5 参考文献

L1-TWSVM 优化问题

TWSVM 有良好的分类性能，已成为数据分类研究的热点。但 TWSVM 使用对离群值较为敏感的 L2 范数来度量距离，导致异常观测点可能会对其结果有较大影响。由于 L1 范数是 L2 范数距离的鲁棒替代，(Yan, Ye, and Yu, 2018) 提出基于 L1 范数的鲁棒分类器。

$$\begin{aligned} \min_{\mathbf{w}_1, b_1} \; & \frac{1}{2}||\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1||_1 + c_1 \mathbf{e}_2^T \mathbf{q}_1 \\ s.t. \; & -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 \geq \mathbf{e}_2, \mathbf{q}_1 \geq 0 \end{aligned} \tag{5}$$

$$\begin{aligned} \min_{\mathbf{w}_2, b_2} \; & \frac{1}{2}||\mathbf{B}\mathbf{w}_2 + \mathbf{e}_2 b_2||_1 + c_2 \mathbf{e}_1^T \mathbf{q}_2 \\ s.t. \; & (\mathbf{A}\mathbf{w}_2 + \mathbf{e}_1 b_2) + \mathbf{q}_2 \geq \mathbf{e}_1, \mathbf{q}_2 \geq 0 \end{aligned} \tag{6}$$

原问题可改写为

$$\min_{\mathbf{w}_1, b_1} \frac{1}{2}\left(\sum_{i=1}^{m_1} \frac{(\mathbf{a}_i^T \mathbf{w}_1 + e_1^i b_1)^2}{d_i}\right) + c_1 \mathbf{e}_2^T \mathbf{q}_1$$
$$s.t. \ -(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 \geq \mathbf{e}_2, \mathbf{q}_1 \geq 0 \tag{7}$$

其中 $d_i = |\mathbf{a}_i^T \mathbf{w}_1 + e_1^i b_1| \neq 0$。令 $\mathbf{D}_1 = diag(1/d_{11}, 1/d_{12}, \ldots, 1/d_{1m_1})$，则问题即为

$$\min_{\mathbf{z}_1} \frac{1}{2}\mathbf{z}_1^T \mathbf{H}^T \mathbf{D}_1 \mathbf{H} \mathbf{z}_1 + c_1 \mathbf{e}_2^T \mathbf{q}_1$$
$$s.t. \ -(\mathbf{G}\mathbf{z}_1 + \mathbf{q}_1) \geq \mathbf{e}_2, \mathbf{q}_1 \geq 0 \tag{8}$$

Input: $A \in \mathbb{R}^{m_1 \times n}$ and $B \in \mathbb{R}^{m_2 \times n}$;
Construct the matrices $H = (A\ e_1)$ and $G = (B\ e_2)$;
Set $p = 0$. Initialize z^p, a standard solution of TWSVM;
while *not converge* **do**
Compute D_1^p;
Compute $z_1^{(p+1)}$ by solving

$$z_1^{(p+1)} = \arg\min_{z_1} \frac{1}{2} z_1^T H^T D_1^p H z_1 + c_1 e_2^T q_1, \ s.t. -Gz_1 + q_1 \geq e_2, \ q_1 \geq 0 \tag{9}$$

p=p+1;
end
Output: The learned solution of z_1.

目录

1 TWSVM

2 L1-TWSVM

3 L1-TWSVM 收敛性证明

4 数值实验

5 参考文献

L1-TWSVM 收敛性证明

引理

对任意非零向量 $\mathbf{u}, \mathbf{u}^p \in \mathbb{R}^1$, 有以下不等式:

$$\|\mathbf{u}\|_1 - \frac{\|\mathbf{u}\|_1^2}{2\|\mathbf{u}^p\|_1} \leq \|\mathbf{u}^p\|_1 - \frac{\|\mathbf{u}^p\|_1^2}{2\|\mathbf{u}^p\|_1} \tag{10}$$

证明.

$$\begin{aligned} &(\sqrt{\mathbf{v}} - \sqrt{\mathbf{vp}})^2 \geq 0 \Rightarrow \mathbf{v} - 2\sqrt{\mathbf{v}\mathbf{v}^p} + \mathbf{v}^p \geq 0 \\ &\Rightarrow \sqrt{\mathbf{v}} - \frac{\mathbf{v}}{2\sqrt{\mathbf{v}^p}} \leq \frac{\sqrt{\mathbf{v}^p}}{2} \Rightarrow \sqrt{\mathbf{v}} - \frac{\mathbf{v}}{2\sqrt{\mathbf{v}^p}} \leq \sqrt{\mathbf{v}^p} - \frac{\mathbf{v}^p}{2\sqrt{\mathbf{v}^p}} \end{aligned} \tag{11}$$

□

定理

算法 *1* 在每步迭代中都使问题 (8) 的目标值单调递减。

【证明】 首先，用以下等式重写 (9) 中的问题：

$$\mathbf{z}_1^{(p+1)} = \arg\min_{\mathbf{z}_1} \frac{1}{2}\mathbf{z}_1^T\mathbf{H}^T\mathbf{D}_1^T\mathbf{H}\mathbf{z}_1 + c_1\mathbf{e}_2^T\max(0, \mathbf{e}_2 + \mathbf{G}\mathbf{z}_1) \tag{12}$$

即：

$$\mathbf{z}_1^{(p+1)} = \arg\min_{\mathbf{z}_1} \frac{1}{2}(\mathbf{H}\mathbf{z}_1)^T\mathbf{D}_1^p\mathbf{H}\mathbf{z}_1 + c_1\mathbf{e}_2^T\max(0, \mathbf{e}_2 + \mathbf{G}\mathbf{z}_1) \tag{13}$$

因此，在第 $(p+1)$ 步迭代中，有

$$\begin{aligned}&\frac{1}{2}(\mathbf{H}\mathbf{z}_1^{(p+1)})^T\mathbf{D}_1^p(\mathbf{H}\mathbf{z}_1^{(p+1)}) + c_1\mathbf{e}_2^T\max(0, \mathbf{e}_2 + \mathbf{G}\mathbf{z}_1^{(p+1)})\\ &\leq \frac{1}{2}(\mathbf{H}\mathbf{z}_1^p)^T\mathbf{D}_1^p(\mathbf{H}\mathbf{z}_1^p) + c_1\mathbf{e}_2^T\max(0, \mathbf{e}_2 + \mathbf{G}\mathbf{z}_1^p)\end{aligned} \tag{14}$$

L1-TWSVM 收敛性证明 II

将式 (10) 中的 $\mathbf{u}$ 和 $\mathbf{u}^p$ 替换为 $\mathbf{H}\mathbf{z}_1^{(p+1)}$ 和 $\mathbf{H}\mathbf{z}_1^p$，可以得到：

$$\|\mathbf{H}\mathbf{z}_1^{(p+1)}\|_1 - \frac{\|\mathbf{H}\mathbf{z}_1^{(p+1)}\|_1^2}{2\|\mathbf{H}\mathbf{z}_1^p\|_1} \leq \|\mathbf{H}\mathbf{z}_1^p\|_1 - \frac{\|\mathbf{H}\mathbf{z}_1^p\|_1^2}{2\|\mathbf{H}\mathbf{z}_1^p\|_1} \tag{15}$$

因此可得如下不等式：

$$\sum_{i=1}^{m_1}\left(\left|\mathbf{h}_i^T\mathbf{z}_1^{(p+1)}\right| - \frac{(\mathbf{h}_i^T\mathbf{z}_1^{(p+1)})^2}{2\left|\mathbf{h}_i^T\mathbf{z}_1^p\right|}\right) \leq \sum_{i=1}^{m_1}\left(\left|\mathbf{h}_i^T\mathbf{z}_1^p\right| - \frac{(\mathbf{h}_i^T\mathbf{z}_1^p)^2}{2\left|\mathbf{h}_i^T\mathbf{z}_1^p\right|}\right) \tag{16}$$

该式可被简化为：

$$\begin{aligned} &\|\mathbf{H}\mathbf{z}_1^{(p+1)}\|_1 - \frac{1}{2}(\mathbf{H}\mathbf{z}_1^{(p+1)})^T\mathbf{D}_1^p(\mathbf{H}\mathbf{z}_1^{(p+1)}) \\ \leq\ &\|\mathbf{H}\mathbf{z}_1^p\|_1 - \frac{1}{2}(\mathbf{H}\mathbf{z}_1^p)^T\mathbf{D}_1^p(\mathbf{H}\mathbf{z}_1^p) \end{aligned} \tag{17}$$

综合式 (14) 和 (17)，可得：

$$
\begin{aligned}
&\|\mathbf{H}\mathbf{z}_1^{(p+1)}\|_1 + c_1\mathbf{e}_2^T \max(0, \mathbf{e}_2 + \mathbf{G}\mathbf{z}_1^{(p+1)}) \\
&\leq \|\mathbf{H}\mathbf{z}_1^{p}\|_1 + c_1\mathbf{e}_2^T \max(0, \mathbf{e}_2 + \mathbf{G}\mathbf{z}_1^{p})
\end{aligned} \tag{18}
$$

因为 (8) 中的问题恒小于零，因此算法 1 收敛，(18) 中的不等式成立。

定理

算法 *1* 收敛至问题 (8) 的一个局部最优解。

【证明】问题 (8) 的拉格朗日函数如下：

$$L_2(\mathbf{z}_1, \mathbf{q}_1) = \frac{1}{2}\|\mathbf{H}\mathbf{z}_1\|_1 + c_1\mathbf{e}_2^T\mathbf{q}_1 - \alpha^T(-\mathbf{G}\mathbf{z}_1 + \mathbf{q}_1 - \mathbf{e}_2) - \beta^{\mathbf{T}}\mathbf{q}_1 \quad (19)$$

其中，α 和 β 是拉格朗日乘子向量。通过对其求导并取零，可以得到问题 (8) 的 KKT 条件：

$$\mathbf{H}^T\mathbf{D}_1\mathbf{H}\mathbf{z}_1 + \mathbf{G}\alpha = 0, c_1\mathbf{e}_2 - \alpha - \beta = 0 \quad (20)$$

在算法 1 的每步迭代中，寻找问题 (9) 中的最优 $\mathbf{z}_1^{(p+1)}$ 。因此，算法 1 的收敛解满足问题的 KKT 条件。接下来，定义算法 1 中问题 (9) 的拉格朗日函数如下：

$$L_3(\mathbf{z}_1,\mathbf{q}_1)=\frac{1}{2}\mathbf{H}^T\mathbf{D}_1\mathbf{H}\mathbf{z}_1+c_1\mathbf{e}_2^T\mathbf{q}_1-\alpha^T(-\mathbf{G}\mathbf{z}_1+\mathbf{q}_1-\mathbf{e}_2)-\beta^{\mathbf{T}}\mathbf{q}_1 \quad (21)$$

同样对其求导并取零，得到：

$$\mathbf{H}^T\mathbf{D}_1\mathbf{H}\mathbf{z}_1+\mathbf{G}\alpha=0,c_1\mathbf{e}_2-\alpha-\beta=0 \quad (22)$$

根据算法 1 中 $\mathbf{D}_1$ 的定义，等式 (20) 和 (22) 在算法 1 收敛时成立。这说明算法 1 的收敛解 $\mathbf{z}_1^{(p+1)}$ 满足问题 (8) 的 KKT 条件，是问题 (8) 的一个局部最优解。

目录

1 TWSVM

2 L1-TWSVM

3 L1-TWSVM 收敛性证明

4 数值实验

5 参考文献

实验代码采用 Python 编写，调用 cvxpy (Diamond and Boyd, 2016) 求解凸优化问题。使用异或问题测试 L1-TWSVM 算法的正确性。

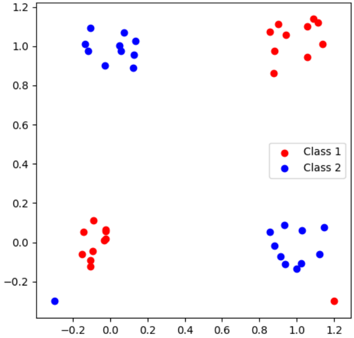


图: 数据散点图

数值实验

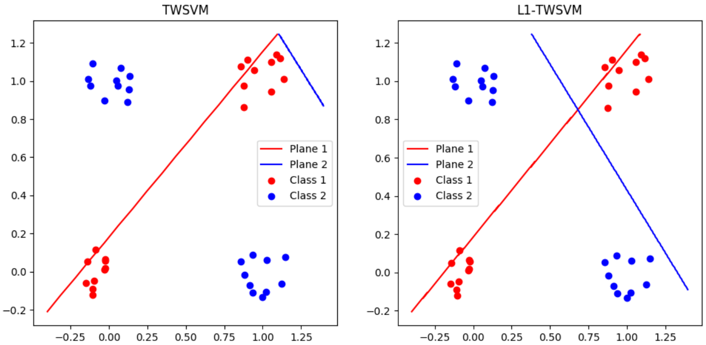


图: 不带离群点，无参数规范化

数值实验

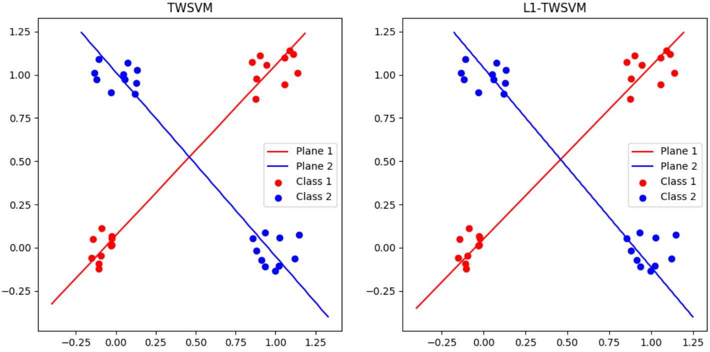


图: 不带离群点，参数规范化

数值实验

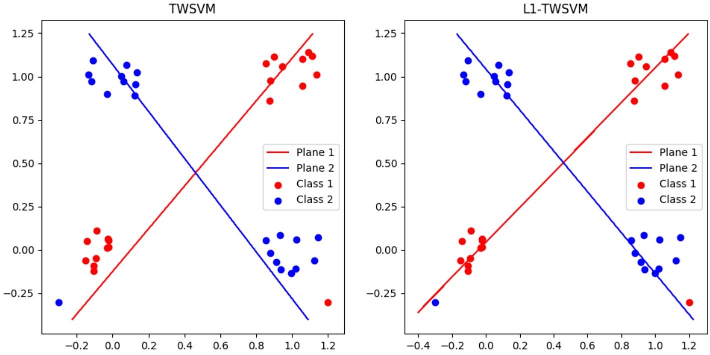


图: 带离群点，参数规范化

目录

1. TWSVM
2. L1-TWSVM
3. L1-TWSVM 收敛性证明
4. 数值实验
5. 参考文献

Diamond, Steven and Stephen Boyd (2016). "CVXPY: A Python-Embedded Modeling Language for Convex Optimization". In: *Journal of Machine Learning Research* 17.83, pp. 1–5 (cit. on p. 19).

Yan, He, Qiao-Lin Ye, and Dong-Jun Yu (2018). "Efficient and robust TWSVM classification via a minimum L1-norm distance metric criterion". In: *Machine Learning*, pp. 1–26 (cit. on p. 8).

谢谢