THU-70250403, Convex Optimization (Fall 2018) **Homework: 2**

L1-TWSVM

Lecturer: Shuning Wang, Li Li swang@tsinghua.edu.cn li-li@tsinghua.edu.cn

Student: 郑洁 j-zheng18@mails.tsinghua.edu.cn

3 L1-TWSVM

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 (Ding et al. 2006; Gao 2008; Kwak 2008; Li et al. 2015a; Nie et al. 2015; Wright et al. 2009)，本文提出基于 L1 范数的鲁棒分类器。优化问题如下：

$$\begin{aligned}
&\min_{\mathbf{w}_1,b_1} \frac{1}{2}||\mathbf{A}\mathbf{w}_1+\mathbf{e}_1b_1||_1+c_1\mathbf{e}_2^T\mathbf{q}_1\\
&s.t. -(\mathbf{B}\mathbf{w}_1+\mathbf{e}_2b_1)+\mathbf{q}_1\geq \mathbf{e}_2, \mathbf{q}_1\geq 0 \qquad (1)\\
&\min_{\mathbf{w}_2,b_2} \frac{1}{2}||\mathbf{B}\mathbf{w}_2+\mathbf{e}_2b_2||_1+c_2\mathbf{e}_1^T\mathbf{q}_2\\
&s.t. (\mathbf{A}\mathbf{w}_2+\mathbf{e}_1b_2)+\mathbf{q}_2\geq \mathbf{e}_1, \mathbf{q}_2\geq 0 \qquad (2)
\end{aligned}$$

其中 $||\cdot||_1$ 表示 L1 范数。在最小化目标函数时，每个平面要尽可能靠近两个分类中的一类，并尽可能远离另一类。由于公式 (这里记得改编号!!!!!!!!!) 中不等式为非凸约束，具有局部最优解，可以求解得到两个不平行的超平面：

$$\mathbf{x}^T\mathbf{w}_1+b_1=0, \mathbf{x}^T\mathbf{w}_2+b_2=0 \qquad (3)$$

则原问题可优化为：

$$\begin{aligned}
&\min_{\mathbf{w}_1,b_1} \frac{1}{2}(\sum_{i=1}^{m_1}\frac{(\mathbf{a}_i^T\mathbf{w}_1+e_1^ib_1)^2}{d_i})+c_1\mathbf{e}_2^T\mathbf{q}_1\\
&s.t. -(\mathbf{B}\mathbf{w}_1+\mathbf{e}_2b_1)+\mathbf{q}_1\geq \mathbf{e}_2, \mathbf{q}_1\geq 0 \qquad (4)\\
&\min_{\mathbf{w}_2,b_2} \frac{1}{2}(\sum_{j=1}^{m_2}\frac{(\mathbf{b}_j^T\mathbf{w}_2+e_2^jb_2)^2}{d_j})+c_2\mathbf{e}_1^T\mathbf{q}_2\\
&s.t. (\mathbf{A}\mathbf{w}_2+\mathbf{e}_1b_2)+\mathbf{q}_2\geq \mathbf{e}_1, \mathbf{q}_2\geq 0 \qquad (5)
\end{aligned}$$

其中 $d_i=|\mathbf{a}_i^T\mathbf{w}_1+e_1^ib_1|\neq 0$，$d_j=|\mathbf{b}_j^T\mathbf{w}_2+e_2^jb_2|\neq 0$，$e_1^i, e_2^j$ 分别表示 e_1 的第 i 个元素和 e_2 的第 j 个元素。由于上述两个式子都包含绝对值运算，难以直接求解，本文提出了一种迭代凸优化策略，基本思想为迭代更新

增广向量 z_1 直到连续两次迭代式 (这里记得改编号!!!!!!!!!) 的目标值小于一个固定值 (如 0.001)，则 z_1 为局部最优解。记 z_1^p 为第 p 次迭代结果，则第 $p+1$ 次迭代结果 $z_1^{(p+1)}$ 可等价为下述问题的解：

$$\begin{aligned}
&\min_{\mathbf{z}_1}\ \frac{1}{2}(\sum_{i=1}^{m_1}\frac{(\mathbf{h}_i^T\mathbf{z}_1)^2}{d_{1i}})+c_1\mathbf{e}_2^T\mathbf{q}_1\\
&s.t.\ -(\mathbf{G}\mathbf{z}_1+\mathbf{q}_1)\geq\mathbf{e}_2,\mathbf{q}_1\geq 0
\end{aligned}\tag{6}$$

$$\begin{aligned}
&\min_{\mathbf{z}_2}\ \frac{1}{2}(\sum_{j=1}^{m_2}\frac{(\mathbf{g}_j^T\mathbf{z}_2)^2}{d_{2j}})+c_2\mathbf{e}_1^T\mathbf{q}_2\\
&s.t.\ (\mathbf{H}\mathbf{z}_2+\mathbf{q}_2)\geq\mathbf{e}_1,\mathbf{q}_2\geq 0
\end{aligned}\tag{7}$$

其中 $d_{1i}=|\mathbf{h}_i^T\mathbf{z}_1^p|$，$d_{2j}=|\mathbf{g}_j^T\mathbf{z}_2^p|$，$\mathbf{g}_j^T=(\mathbf{b}_j^Te_2^j)$，则公式 (19)(20 这里记得改编号!!!!!!!!!) 可改写为

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

$$\begin{aligned}
&\min_{\mathbf{z}_2}\ \frac{1}{2}\mathbf{z}_2^T\mathbf{H}^T\mathbf{D}_2\mathbf{G}\mathbf{z}_2+c_2\mathbf{e}_1^T\mathbf{q}_2\\
&s.t.\ (\mathbf{H}\mathbf{z}_2+\mathbf{q}_2)\geq\mathbf{e}_1,\mathbf{q}_2\geq 0
\end{aligned}\tag{9}$$

其中 $\mathbf{D}_1=diag(1/d_{11},1/d_{12},...,1/d_{1m_1})$ $\mathbf{D}_2=diag(1/d_{21},1/d_{22},...,1/d_{2m_2})$ 为对角矩阵。则问题 (21)(22 这里记得改编号!!!!!!!!!) 等价于：

$$\begin{aligned}
&\min_{\mathbf{z}_1}\ \frac{1}{2}||\mathbf{H}\mathbf{z}_1||_1+c_1\mathbf{e}_2^T\mathbf{q}_1\\
&s.t.\ -(\mathbf{G}\mathbf{z}_1+\mathbf{q}_1)\geq\mathbf{e}_2,\mathbf{q}_1\geq 0
\end{aligned}\tag{10}$$

$$\begin{aligned}
&\min_{\mathbf{z}_2}\ \frac{1}{2}||\mathbf{G}\mathbf{z}_2||_1+c_2\mathbf{e}_1^T\mathbf{q}_2\\
&s.t.\ (\mathbf{H}\mathbf{z}_2+\mathbf{q}_2)\geq\mathbf{e}_1,\mathbf{q}_2\geq 0
\end{aligned}\tag{11}$$

公式 (14 这里记得改编号!!!!!!!!!) 是不等式约束 (非凸) 的凸优化问题，因此它存在解析解。其拉格朗日函数为：

$$\begin{aligned}
\boldsymbol{L}_1(\mathbf{w}_1,b_1,\mathbf{q}_1,\boldsymbol{\alpha},\boldsymbol{\beta})=&\frac{1}{2}(\mathbf{A}\mathbf{w}_1+\mathbf{e}_1b_1)^T\mathbf{D}_1(\mathbf{A}\mathbf{w}_1+\mathbf{e}_1b_1)\\
&+c_1\mathbf{e}_2^T\mathbf{q}_1-\boldsymbol{\alpha}^T(-(\mathbf{B}\mathbf{w}_1+\mathbf{e}_2b_1)+\mathbf{q}_1-\mathbf{e}_2)-\boldsymbol{\beta}^T\mathbf{q}_1
\end{aligned}\tag{12}$$

其中 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, ..., \alpha_{m_2})^T, \boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, ..., \beta_{m_1})^T$ 为拉格朗日乘子，$\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta} \geq 0$，令 $\boldsymbol{L}_1$ 对 $\mathbf{w}_1, b_1, \mathbf{q}_1$ 的偏导分别为 0，可得 Karush-Kuhn-Tucker (KKT) 条件为：

$$\frac{\partial L}{\partial \mathbf{w}_1} = \mathbf{A}^T \mathbf{D}_1 (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + \mathbf{B}^T \boldsymbol{\alpha} = 0 \tag{13}$$

$$\frac{\partial L}{\partial b_1} = \mathbf{e}_1^T \mathbf{D}_1 (\mathbf{A}\mathbf{w}_1 + \mathbf{e}_1 b_1) + \mathbf{e}_2 \boldsymbol{\alpha} = 0 \tag{14}$$

$$\frac{\partial L}{\partial \mathbf{q}_1} = c_1 \mathbf{e}_2 - \boldsymbol{\alpha} - \boldsymbol{\beta} = 0 \tag{15}$$

$$-(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 \geq \mathbf{e}_2, \mathbf{q}_1 \geq 0 \tag{16}$$

$$\boldsymbol{\alpha}^T(-(\mathbf{B}\mathbf{w}_1 + \mathbf{e}_2 b_1) + \mathbf{q}_1 - \mathbf{e}_2) = 0, \boldsymbol{\beta}^T \mathbf{q}_1 = 0 \tag{17}$$

从式 (28 这里记得改编号!!!!!!!!!) 可推出 $0 \leq \boldsymbol{\alpha} \leq c_1 \mathbf{e}_2$，结合公式 (26)(27) 可得：

$$\mathbf{A}^T \mathbf{e}_1^T \mathbf{D}_1 (\mathbf{A}\mathbf{e}_1)(\mathbf{w}_1 b_1)^T + (\mathbf{B}^T \mathbf{e}_2^T)\boldsymbol{\alpha} = 0 \tag{18}$$

结合之前定义的矩阵 $(\mathbf{H}, \mathbf{G})$ 及增广向量 $(\mathbf{z}_1, \mathbf{z}_2)$，可以得到

$$\mathbf{H}^T \mathbf{D}_1^p \mathbf{H} \mathbf{z}_1^{(p+1)} + \mathbf{G}^T \boldsymbol{\alpha} = 0 \tag{19}$$

即

$$\mathbf{z}_1^{(p+1)} = -(\mathbf{H}^T \mathbf{D}_1^p \mathbf{H})^{-1} \mathbf{G}^T \boldsymbol{\alpha} \tag{20}$$

由于 $(\mathbf{H}^T \mathbf{D}_1^p \mathbf{H})^{-1}$ 为半正定矩阵，因此可能得到不稳定或不准确的解，在实际应用中，本文使用正则化方法 (Jayadeva and Chandra (2007), Mangasarian andWild (2006)) 解决这个问题。$(\mathbf{H}^T \mathbf{D}_1^p \mathbf{H} + \varepsilon \mathbf{I})$ 为正定矩阵 (其中 ε 为一个小扰动)，不受奇点的影响。则逆矩阵 $(\mathbf{H}^T \mathbf{D}_1^p \mathbf{H})^{-1}$ 可由 $(\mathbf{H}^T \mathbf{D}_1^p \mathbf{H} + \varepsilon \mathbf{I})$ 代替，因此 $\mathbf{z}_1^{(p+1)}$ 可推导为：

$$\mathbf{z}_1^{(p+1)} = -(\mathbf{H}^T \mathbf{D}_1^p \mathbf{H} + \varepsilon \mathbf{I})^{-1} \mathbf{G}^T \boldsymbol{\alpha} \tag{21}$$

同样的

$$\mathbf{z}_2^{(p+1)} = -(\mathbf{G}^T \mathbf{D}_2^p \mathbf{G} + \varepsilon \mathbf{I})^{-1} \mathbf{H}^T \boldsymbol{\beta} \tag{22}$$

将增广向量 $\mathbf{z}_1^{(p+1)}, \mathbf{z}_2^{(p+1)}$ 分别代入拉格朗日函数中。在 KKT 条件下，原问题 (14)(15 这里记得改编号!!!!!!!!!) 转变为 Wolfe 对偶问题：

$$\begin{aligned} &\max_{\boldsymbol{\alpha}} \ \mathbf{e}_2^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G} (\mathbf{H}^T \mathbf{D}_1 \mathbf{H})^{-1} \mathbf{G}^T \boldsymbol{\alpha} \\ &s.t. 0 \leq \boldsymbol{\alpha} \leq c_1 \mathbf{e}_2 \end{aligned} \tag{23}$$

$$\begin{aligned} &\max_{\boldsymbol{\beta}} \ \mathbf{e}_1^T \boldsymbol{\beta} - \frac{1}{2} \boldsymbol{\beta}^T \mathbf{H} (\mathbf{G}^T \mathbf{D}_2 \mathbf{G})^{-1} \mathbf{H}^T \boldsymbol{\beta} \\ &s.t. 0 \leq \boldsymbol{\beta} \leq c_2 \mathbf{e}_1 \end{aligned} \tag{24}$$

通过求解对偶问题可得拉格朗日乘子 $\boldsymbol{\alpha} \in \boldsymbol{R}^{m_2\times 1}, \boldsymbol{\beta} \in \boldsymbol{R}^{m_1\times 1}$ 以及权向量 $\mathbf{w}_1, \mathbf{w}_2$、偏差 b_1, b_2，即获得两个不平行的超平面。对于新加入的点 $\mathbf{x} \in \mathbf{R}^n$，根据决策方程 $f(\mathbf{x})$(选择最近的超平面) 将其分配到对应类别中

$$f(\mathbf{x}) = arg \min_{i=1,2} (|\mathbf{x}^T\mathbf{w}_i + b_i|/|||\mathbf{w}_i|) \tag{25}$$

其中 $|\cdot|$ 表示取绝对值。式 (14 这里记得改编号!!!!!!!!!) 中的目标函数为非凸约束的凸问题，因此 $\mathbf{z}_1^{(p+1)}$ 为此问题的局部最优解。而在式 (33 这里记得改编号!!!!!!!!!) 中，$\mathbf{D}_1^p$ 依赖于 $\mathbf{z}_1^{(p+1)}$，因此它是一个未知变量，可看作 (14) 中目标的隐变量，可以同相同的迭代算法交替优化求解。我们根据前一次迭代结果 $\mathbf{z}_1^{(p+1)}$ 来更新 $\mathbf{D}_1^p$，又通过 $\mathbf{D}_1^p$ 来改变 $\mathbf{z}_1^{(p+1)}$，增加 p 直到连续两次迭代结果小于一个固定值。此外，适当的初始化可有效加快算法收敛速度。本文通过求解公式 (8)(9 这里记得改编号!!!!!!!!) 得到初始解，仿真结果较优。算法 1 总结了 L1-TWSVM 的迭代过程。