

L1 范数最大间隔分类器设计

寇振宇¹, 杨绪兵¹, 张福全¹, 杨红鑫¹, 许等平²

(1. 南京林业大学信息科学技术学院, 南京 210037)

(2. 国家林业局调查规划设计院, 北京 100714)

[摘要] 以 L1 范数为例, 设计了一个 L1 范数的大间隔分类器 L1MMC (L1-norm Maximum Margin Classifier), 主要特点如下: (1) 间隔由 L1 范数的点到平面距离解析表示; (2) 该分类器与 SVM 一样, 通过最大化 L1 间隔, 达到同时最小化经验风险和结构风险的目的; (3) 只需要通过线性规划进行求解, 避免了 SVM 的二次规划问题; (4) 分类精度达到甚至超过 SVM. 最后, 在人工数据和国际标准 UCI 数据集上, 验证了该方法的有效性.

[关键词] L1 范数, 支持向量机, 间隔, 线性规划

[中图分类号] TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2018)04-0059-06

Design of L1 Norm Maximum Margin Classifier

Kou Zhenyu¹, Yang Xubing¹, Zhang Fuquan¹, Yang Hongxin¹, Xu Dengping²

(1. College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China)

(2. Survey & Planning Institute of State Forestry Administration, Beijing 100714, China)

Abstract: L1 norm is taken as an example to design an L1 norm L1MMC (L1-norm Maximum Margin Classifier). The main features are as follows: (1) The interval is represented by the point-to-plane distance analysis of the L1 norm; (2) This classifier, like SVM, maximizes the L1 interval to minimize the risk of both empirical and structural risks; (3) Only need to be solved through linear programming to avoid the quadratic programming problem of SVM; (4) Classification accuracy reaches or even exceeds SVM. Finally, on the artificial data and the international standard UCI data set, verify the effectiveness of the method.

Key words: L1 norm, support vector machine, margin, linear programming

在机器学习和模式识别领域中, 模式之间的相似性常采用距离来度量. 距离度量^[1]是一项基础性工作, 对模型设计及其几何解释均是至关重要的. 如 SVM^[2], 其泛化能力是由间隔决定的, 该间隔是由 L2 范数的距离解析表示, 几何上可解释为两类样本最大程度上分离, 同时, 在模型约束条件中通过不等式控制经验误差, 故 SVM 模型兼顾了经验误差和泛化能力, 加之几何意义明确并且求解方便, 诞生了很多优秀的学习算法, 如最小二乘支持向量机 (LSSVM: Least Square SVM)^[3]、近似支持向量机 (PSVM: Proximal SVM)^[4]、广义特征值最接近支持向量机 (GEPSVM: PSVM via Generalized Eigenvalue)^[5]等. 然而, 随着大数据时代的来临, L2 范数支持向量机由于其模型设计机制的制约, 在应用中受到很多局限^[6-8].

近年, 在大数据^[9]背景下, 海量数据中隐藏着多种有价值信息, 而在处理信息时, 人们面临着存储和计算两大问题. 稀疏学习^[10]无疑是解决此类问题的有效方法之一. 稀疏学习模型常采用 L1 范数^[11-12]进行设计, 导出的虽然是凸优化问题, 甚至是严格凸, 理论上存在最优解, 但由于 L1 范数的不可导性, 目前只能采用迭代方式求解. 许多基于 L1 范数的学习算法被提出^[13], 如 Zhu^[14]等提出了 L1 范数支持向量机, 采用 L1 范数的正则化项替代 SVM 的间隔, 使得学习结果满足稀疏化, 有利于提取特征. Wang^[15]等提出了用 L1 范数作为损失函数的支持向量机, 在理论上使得求解速度和稀疏性得到了保证. Kujala^[16]等研究了一种基于 L1 范数的支持向量机, 可显著减少 L1 范数最小化问题中的求解迭代次数, 使得算法更加高效.

收稿日期: 2018-08-16.

基金项目: 国家自然科学基金 (31670554、61871444)、江苏省自然科学基金 (BK20161527、BK20171453).

通讯联系人: 杨绪兵, 博士, 副教授, 研究方向: 模式识别, 机器学习. Email: xbyang@njfu.edu.cn

稳定. 然而, 由于缺少 L1 范数下点到平面距离的解析表达式, 上述的 L1 范数的 SVM 在设计模型时, 要么仍采用了 L2 范数间隔, 要么直接牺牲间隔思想, 故不能称为真正意义上的 L1 范数学习机. 2017 年, 杨^[17-18]等给出了 L1 范数下点到平面距离及其投影的解析表达式. 在此基础上, 本文提出了一种真正意义上的 L1 范数最大间隔分类器 L1MMC, 与 SVM 相比, 该方法不仅几何意义明确, 还具有如下特点: (1) 间隔由 L1 范数的点到平面距离解析表示; (2) 该分类器与 SVM 一样, 通过最大化 L1 间隔, 达到兼顾最小化经验风险的目的; (3) L1MMC 只需要通过线性规划进行求解, 避免了 SVM 的二次规划问题; (4) 分类精度达到甚至超过 SVM.

本文内容安排如下: 第 1 节是相关工作简介. 第 2 节中, 详述本文提出的 L1 范数最大间隔分类器, 从模型设计、求解方法、几何解释等方面展开. 在第 3 节, 给出了验证分类器正确性的可视化例子以及对比实验. 第 4 节总结全文.

1 相关工作

1.1 支持向量机

基于小样本理论和结构化风险最小原理基础, Vapnik 等人提出的著名的机器学习方法: SVM, 其根据有限样本信息来寻找模型的复杂性和学习能力之间的最佳策略, 以此获得较高的推广性能. 二分类学习的目标是: 基于训练集在样本空间中寻找一个分类超平面, 将不同类别的样本尽可能分开. 设给定的样本集: $\{(\mathbf{x}_i, y_i)\}, i=1, 2, \dots, n, \mathbf{x}_i \in R^d, y_i \in \{\pm 1\}$ 在样本空间中, 记分类超平面为:

$$\mathbf{w}^T \mathbf{x} + b = 0,$$

式中, $\mathbf{w}(\in R^d)$ 为法向量, b 为阈值. 在 L2 范数的欧氏距离度量下, 根据点到平面距离公式, 易知 SVM 的两个支撑超平面 (supporting planes) 间的距离 (即间隔) 为 $2/\|\mathbf{w}\|_2$. 由 PAC (Probably Approximately Correct) 理论知, SVM 的最小错分上界可通过间隔来度量, 即通过最大化间隔达到提升推广能力的目的. 结合经验风险最小化原理, SVM 的优化目标如下:

$$\begin{aligned} \min_{\mathbf{w}, b, \delta_i} \quad & \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^n \delta_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \delta_i, \\ & \delta_i \geq 0, i = 1, 2, \dots, n, \end{aligned} \quad (1)$$

式中, C 是平衡参数, 用来调整模型在经验风险和泛化能力之间的折衷, δ_i 为松弛变量. 该问题可通过二次规划 (QP: quadratic programming) 求解.

1.2 L1-norm 支持向量机

Zhu 等为追求解的稀疏性, 采用 L1 范数正则化项取代了 (1) 中的间隔项, 并称之为 L1 范数支持向量机 (L1SVM: L1 norm SVM), 并理论上证明了 L1SVM 的时间复杂度为 $O(nq * \min(n, q))$, n 为训练样本个数, q 为解的非 0 元素个数. 极端情况, 当解只保留一个非 0 分量, 复杂度可退化为 $O(n)$. 同时在实验上也验证了该方法具有一定抗噪声能力, 在基因数据集上与 SVM 具有相当的分类能力. 其优化目标可由下式表达:

$$\begin{aligned} \min_{\mathbf{w}, b, \delta_i} \quad & \sum_{i=1}^n \delta_i + \lambda \|\mathbf{w}\|_1 \\ \text{s.t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \delta_i, \\ & \delta_i \geq 0, i = 1, 2, \dots, n. \end{aligned} \quad (2)$$

L1SVM 上述优缺点通过对比模型 (1) 和 (2) 易知, 此处不再细述.

2 L1 范数最大间隔分类器

如前所述, L1SVM 采用 L1 正则化项代替了间隔项, 以达到稀疏解的目的. 本节拟采用 L1 范数来度量间隔, 设计真正意义上的 L1 范数 SVM. 下文先从 L1 范数的点到平面距离出发, 导出基于 L1 范数的最大间隔分类器 L1MMC, 再考虑问题求解.

2.1 模型设计

据文献[17],点 v 到超平面 $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0$ 的L1距离存在解析表示,为方便阅读,本文以引理1描述如下:

引理1 空间 R^d 中任一点 v 到给定超平面 $g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0$ 的L1距离,可解析表示为

$$s = \frac{|\mathbf{w}^T \mathbf{v} + b|}{\|\mathbf{w}\|_{\infty}}, \quad (3)$$

式中, $\|\mathbf{w}\|_{\infty}$ 表示向量 \mathbf{w} 的无穷范数,定义为 $\|\mathbf{w}\|_{\infty} = \max_{1 \leq i \leq d} |w_i|$.

证明参见文献[17].

借鉴SVM的间隔的思想,两个支撑超平面间的距离(间隔),可由落在支撑平面上的两类支持向量联合导出,以定理1描述之.

定理1 L1范数的间隔为 $r = \frac{2}{\|\mathbf{w}\|_{\infty}}$.

证明 为方便描述,记 $\mathbf{w}^T \mathbf{x} + b = 0$ 为分类面,其两个支撑平面为 $\mathbf{w}^T \mathbf{x} + b = \pm 1$. 设 $\mathbf{v}_1, \mathbf{v}_2$ 分别是位于不同支撑平面上的两个支持向量,满足 $\mathbf{w}^T \mathbf{v}_1 + b = 1, \mathbf{w}^T \mathbf{v}_2 + b = -1$,由引理1得 $r = \frac{1}{\|\mathbf{w}\|_{\infty}} + \frac{1}{\|\mathbf{w}\|_{\infty}} = \frac{2}{\|\mathbf{w}\|_{\infty}}$.

按SVM设计思想,同样通过引入松弛变量,设计软间隔的L1MMC,优化目标如下:

$$\begin{aligned} \min_{(\mathbf{w}, b)} & \frac{\|\mathbf{w}\|_{\infty}}{2} + C \sum_{i=1}^n \delta_i \\ \text{s.t. } & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \delta_i, \\ & \delta_i \geq 0, i = 1, 2, \dots, n, \end{aligned} \quad (4)$$

式中, \mathbf{w} 和 b 为分类超平面的法向量和阈值, y_i 为类别标记, \mathbf{x}_i 为训练样本, C 为平衡因子,控制模型在经验风险和泛化能力之间折衷. δ_i 为松弛变量.该模型与SVM的设计思想完全一致,即:在最大化L1间隔的同时亦最小化经验风险.

2.2 问题求解

形如(4)的模型,可转化为线性规划问题求解,以定理2描述之.

定理2 模型(4)等价于如下线性规划问题:

$$\begin{aligned} \min_{\boldsymbol{\eta}} & \mathbf{f}^T \boldsymbol{\eta} \\ \text{s.t. } & \mathbf{A} \boldsymbol{\eta} \leq \mathbf{h}, \end{aligned} \quad (5)$$

式中, $\mathbf{f} = [1 \quad \mathbf{0}_{1 \times d} \quad C \mathbf{e}_{1 \times n} \quad \mathbf{0}]^T$, $\mathbf{h} = [-\mathbf{e}_{1 \times n} \quad \mathbf{0}_{1 \times d} \quad \mathbf{0}_{1 \times d} \quad \mathbf{0}_{1 \times n}]^T$, $\mathbf{0}_{m \times n}$ 是 $m \times n$ 阶的零矩阵, $\mathbf{e}_{m \times n}$ 为元素全为1的 $m \times n$ 阶矩阵, $\mathbf{D} = \text{diag}(y_1, y_2, \dots, y_n)$,

$$\mathbf{A} = \begin{bmatrix} -\mathbf{0}_{n \times 1} & -\mathbf{D}\mathbf{X} & -\mathbf{I}_{n \times n} & -\mathbf{D}\mathbf{e}_{n \times 1} \\ -\mathbf{e}_{d \times 1} & \mathbf{I}_{d \times d} & \mathbf{0}_{d \times n} & \mathbf{0}_{d \times 1} \\ -\mathbf{e}_{d \times 1} & -\mathbf{I}_{d \times d} & \mathbf{0}_{d \times n} & \mathbf{0}_{d \times 1} \\ \mathbf{0}_{n \times 1} & \mathbf{0}_{n \times d} & -\mathbf{I}_{n \times n} & \mathbf{0}_{n \times 1} \end{bmatrix}.$$

证明 式(4)中,可令 $t = \|\mathbf{w}\|_{\infty}$,由无穷范数定义 $\|\mathbf{w}\|_{\infty} = \max |w_i|$,知 $-t \leq w_i \leq t (1 \leq i \leq d)$.

式(4)转化为如下优化问题:

$$\begin{aligned} \min_{t, \mathbf{w}, \delta_i, b} & t + C \sum_{i=1}^n \delta_i \\ \text{s.t. } & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \delta_i, \\ & w_i - t \leq 0, \\ & -w_i - t \leq 0, \\ & \delta_i \geq 0. \end{aligned} \quad (6)$$

记 $\boldsymbol{\eta} = [t, \mathbf{w}^T, \boldsymbol{\delta}^T, b]^T$,将式(6)转化为线性规划标准模型:

$$\begin{aligned} & \min [1 \quad \mathbf{0}_{1 \times d} \quad C\mathbf{e}_{1 \times n} \quad \mathbf{0}] \times [t, \mathbf{w}^T, \boldsymbol{\delta}^T, b]^T \\ & \text{s.t.} \begin{bmatrix} -\mathbf{0}_{n \times 1} & -D\mathbf{X} & -I_{n \times n} & -D\mathbf{e}_{n \times 1} \\ -\mathbf{e}_{d \times 1} & I_{d \times d} & \mathbf{0}_{d \times n} & \mathbf{0}_{d \times 1} \\ -\mathbf{e}_{d \times 1} & -I_{d \times d} & \mathbf{0}_{d \times n} & \mathbf{0}_{d \times 1} \\ \mathbf{0}_{n \times 1} & \mathbf{0}_{n \times d} & -I_{n \times n} & \mathbf{0}_{n \times 1} \end{bmatrix} \begin{bmatrix} t \\ \mathbf{w}_{d \times 1} \\ \boldsymbol{\delta}_{n \times 1} \\ b \end{bmatrix} \leq \begin{bmatrix} -\mathbf{e}_{n \times 1} \\ \mathbf{0}_{d \times 1} \\ \mathbf{0}_{d \times 1} \\ \mathbf{0}_{n \times 1} \end{bmatrix}. \end{aligned} \quad (7)$$

式(7)按定理 2 约定的符号整理,可得式(5). 其中,按 $\boldsymbol{\eta}$ 定义, $\mathbf{w} = \boldsymbol{\eta}(2:d+1, 1)$, $b = \boldsymbol{\eta}(\text{end}, 1)$.

至此,由以上分析可见,本文提出的 LIMMC 方法也可通过线性规划问题求解. 由 1.2 节知, L1SVM 的时间复杂度为 $O(nq * \min(n, q))$. 而本文所提出的方法,按文献[19],其时间复杂度为 $O(n * (n+2d)(d+1))$,其中 n 为训练样本个数, q 为 L1SVM 解中非 0 元素个数, $(n+2d)(d+1)$ 为式(5)中约束矩阵的非零元素个数. 故时间复杂度略高于 L1SVM.

3 实验验证

为验证本文方法的有效性,本文选择标准 SVM 及 L1SVM 为比较对象. 标准 SVM 是基于最大间隔思想提出的,同时兼顾了经验风险最小化,其模型的设计目标实现了经验风险和结构风险的高度统一. 而 L1SVM 主要是考虑解的稀疏性,即采用 L1 范数正则化项取代了标准 SVM 的间隔,相对于标准 SVM 的二次规划求解, L1SVM 只需要求解一个线性规划问题. 本文提出的采用 L1 范数度量的 LIMMC 方法,既考虑了间隔思想,同时也兼顾了经验风险问题,且亦可通过线性规划问题求解.

实验分为两个部分:一是本文方法的两个可视化例子,二是在 UCI 数据集下验证本文方法的分类效果.

3.1 可视化实验

为验证本文提出的 L1 范数最大间隔分类器的有效性,实验分别在 Iris 数据集中对线性可分和线性不可分的两类样本进行了分类.

第一部分可视化实验在 Iris 数据集中线性可分的样本上进行,选取了 100 个 2 维样本进行分类. 分类结果如图 1 所示.

图 1 中,“+”号为正类样本,“○”为负类样本. 实线为本文分类器求得的最大间隔分类面及对应的支撑平面. 虚线为 SVM 求得的分类面及对应的支撑平面. 点线为 L1 范数支持向量机求得的分类面. 由图可见本文方法在此类线性可分数据集上可取得很好的分类效果,效果上与 L2 范数 SVM 相当. 同时可以看到, L1 范数 SVM 为了使所求解满足稀疏性,所求解的其中一个分量约等于 0,导致分类面的法向量平行于 x 轴,追求稀疏性而牺牲了最大间隔. 本文方法相较于 L1 范数 SVM,在最大间隔上具有优势.

第二部分可视化实验在 Iris 数据集中线性不可分的样本上进行,同样选取了 100 个 2 维样本进行分类. 分类结果由图 2 所示.

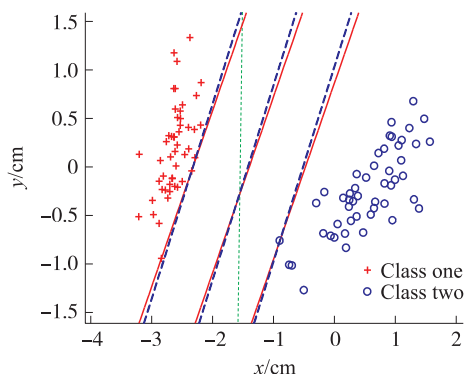


图 1 线性可分样本分类

Fig. 1 Linearly separable sample classification

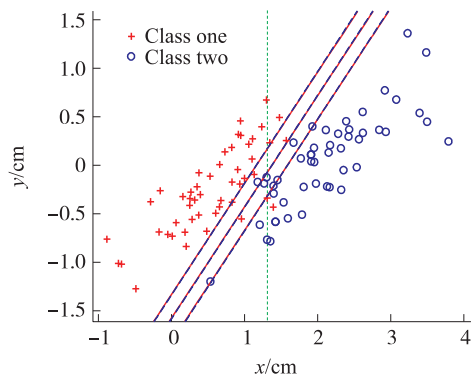


图 2 线性不可分样本分类

Fig. 2 Linearly inseparable sample classification

图 2 中,“+”号为正类样本,“○”为负类样本. 实线为本文分类器求得的最大间隔分类面及对应的支撑平面. 虚线为 SVM 求得的分类面及对应的支撑平面. 点线为 L1 范数支持向量机求得的分类面. 由图可见本文方法在此类线性不可分数据集上同样可取得很好的分类效果.

3.2 UCI 数据集试验

为了检验本文分类器对真实数据集的分类能力,本组实验选取了 10 个 UCI 数据集进行实验,数据集的基本信息如表 1 所示.

表 1 UCI 数据集基本信息

Table 1 Main information about UCI data sets

数据集	维数	样本数	类别信息	数据集	维数	样本数	类别信息
monk1	6	432	2[216,216]	monk3	6	432	2[204,228]
monk2	6	432	2[290,142]	cmc	8	1 473	2[109,1364]
sonar	60	208	2[97,111]	glass	9	214	2[51,163]
Iris23	4	100	2[50,50]	spambase	57	4 601	2[2788,1813]
ionosphere	34	351	2[126,225]	mushroom	22	8 124	2[4208,3916]

在 10 个 UCI 数据集上的实验结果如表 2 所示. 从表 2 可以看到,本文提出的方法对大部分数据集的训练精度和测试精度均有优势.

表 2 UCI 数据集上的实验结果

Table 2 Experimental results about UCI data sets

数据集	方法	训练精度/%	测试精度/%	训练时间/s	数据集	方法	训练精度/%	测试精度/%	训练时间/s
monk1	L1MMC	66.67	66.67	0.021 3	monk3	L1MMC	80.56	80.57	0.022 7
	SVM	65.97	68.83	0.147 4		SVM	80.56	80.57	0.082 0
	L1-SVM	66.67	66.67	0.014 8		L1-SVM	80.56	80.57	0.016 3
monk2	L1MMC	67.13	67.12	0.012 5	cmc	L1MMC	92.60	92.59	0.088 3
	SVM	67.13	67.12	0.090 4		SVM	92.60	92.59	5.965 4
	L1-SVM	67.13	67.12	0.012 6		L1-SVM	92.60	92.59	0.045 1
sonar	L1MMC	100.00	91.33	0.123 0	glass	L1MMC	95.74	93.89	0.019 1
	SVM	97.71	82.07	0.012 5		SVM	95.53	93.41	0.058 4
	L1-SVM	82.90	91.83	0.053 9		L1-SVM	94.50	92.93	0.025 6
Iris23	L1MMC	98.11	96.00	0.007 3	spambase	L1MMC	93.42	92.98	1.301 0
	SVM	97.44	94.00	0.000 8		SVM	99.91	83.18	83.272 2
	L1-SVM	96.89	94.00	0.005 7		L1-SVM	93.51	93.07	2.765 5
ionosphere	L1MMC	95.57	88.88	0.225 8	mushroom	L1MMC	0.815 7	0.815 5	0.631 3
	SVM	95.38	89.45	0.549 4		SVM	0.895 1	0.891 9	84.035 5
	L1-SVM	93.26	88.60	0.169 6		L1-SVM	0.815 5	0.815 5	0.471 0

4 结语

在 L1 投影的解析计算方法的基础上,本文将该方法应用到实际的分类问题中,提出了一种 L1 范数度量的最大间隔分类器. 该分类器有着明确的几何意义,能够使用线性规划进行求解. 相较传统的 L2 范数支持向量机,当样本数据量较大时,大大缩短了求解时间. 相较此前人们提出的 L1 范数支持向量机(使用 L1 范数进行惩罚),是真正意义上使用 L1 范数进行度量的分类器.

[参考文献]

- [1] HUANG K, YANG H, KING I, et al. Maxi-min margin machine: learning large margin classifiers locally and globally[J]. IEEE transactions on neural networks, 2008, 19(2): 260-272.
- [2] VAPNIK V. The nature of statistical learning theory[C]//Conference on artificial intelligence. Montreal, Canada: Springer-Verlag, 1995: 988-999.
- [3] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3): 293-300.
- [4] FUNG G. Proximal support vector machine classifier[J]. Machine learning, 2005, 59(1/2): 77-97.
- [5] MANGASARIAN O L, WILD E W. Multisurface proximal support vector machine classification via generalized eigenvalues[J]. IEEE transactions on pattern analysis & machine intelligence, 2006, 28(1): 69-74.

- [6] HUANG K,ZHENG D,KING I,et al. Arbitrary norm support vector machines[J]. Neural computation,2014,21(2):560-582.
- [7] GU B,SHENG V S. A robust regularization path algorithm for γ -support vector classification[J]. IEEE transactions on neural networks and learning systems,2017,28(5):1241-1248.
- [8] RASTOGI R,SAIGAL P,CHANDRA S. Angle-based twin parametric-margin support vector machine for pattern classification[J]. Knowledge-based systems,2018,139:64-77.
- [9] NAN S,SUN L,CHEN B,et al. Density-dependent quantized least squares support vector machine for large data sets[J]. IEEE transactions on neural networks and learning systems,2017,28(1):94-106.
- [10] TIPPING M E. Sparse Bayesian learning and the relevance vector machine[J]. Journal of machine learning research,2001,1(3):211-244.
- [11] ZHENG W,LIN Z,WANG H. L1-norm kernel discriminant analysis via Bayes error bound optimization for robust feature extraction[J]. IEEE transactions on neural networks and learning systems,2014,25(4):793-805.
- [12] WANG L,SHEN X. On L1-norm multiclass support vector machines:methodology and theory[J]. Journal of the American statistical association,2007,102(478):583-594.
- [13] YAN H,YE Q,YU D J,et al. Least squares twin bounded support vector machines based on L1-norm distance metric for classification[J]. Pattern recognition,2018,74:434-447.
- [14] ZHU J,ROSSET S,TIBSHIRANI R,et al. 1-norm support vector machines[C]//Advances in neural information processing systems. Vancouver,Canada:MIT Press,2003:49-56.
- [15] WANG L,SHEN X,ZHENG Y F. On L1-norm multi-class support vector machines[C]//International conference on machine learning and applications. Orlando,USA:IEEE Computer Society,2006:83-88.
- [16] KUJALA J,AHO T,ELOMAA T. A walk from 2-norm SVM to 1-norm SVM[C]//Ninth IEEE international conference on data mining. Miami,USA:IEEE Computer Society,2009:836-841.
- [17] 杨绪兵,顾一凡,陈松灿,等. L1 投影的解析计算方法[J]. 南京大学学报(自然科学版),2017,53(3):476-482.
- [18] 杨绪兵,王一雄,陈斌. 马氏度量学习中的几个关键问题研究及几何解释[J]. 南京大学学报(自然科学版),2013,49(2):133-141.
- [19] YEN E H,ZHONG K,HSIEH C J,et al. Sparse linear programming via primal and dual augmented coordinate descent[C]//International conference on neural information processing systems. Montreal,Canada:MIT Press,2015:2368-2376.

[责任编辑:丁 蓉]