

## 基于DC规划的 $L_1$ 范数稀疏线性判别分析\*

翁婷, 李国权, 张家豪

(重庆师范大学 数学科学学院, 重庆 401331)

**摘要:**针对传统线性判别分析抗噪性能不足,以及在高维小样本数据集上(即样本特征数量远远大于样本个数)泛化性能不佳等问题,现有文献提出了鲁棒线性判别分析(robust linear discriminant analysis, RLDA)和鲁棒稀疏线性判别分析(robust sparse linear discriminant analysis, RSLDA)模型,并提出基于交替方向乘子法的优化算法求解RLDA和RSLDA模型。但上述算法不具备收敛性,在有些数据集上算法不收敛。因此,根据RLDA、RSLDA的模型结构特点,提出一种基于DC函数规划的优化算法DC\_SLDA,该算法通过将原问题的目标函数进行DC分解,转换为DC规划问题,进而利用DC算法进行迭代求解。所提出的算法不仅具有收敛性保证而且子问题具有解析解,使得模型的训练效率得到明显提升。

**关键词:**线性判别分析;稀疏;DC规划

**中图分类号:** O221.2

**文献标志码:** A

**文章编号:** 1672-6693(2025)06-0022-08

线性判别分析(linear discriminant analysis, LDA)是一种监督降维和分类方法,旨在通过线性投影将高维数据映射到低维空间,使得不同类别的样本在该空间中尽可能分离,同类样本尽可能聚集。LDA广泛应用于文本分类<sup>[1]</sup>、信息检索<sup>[2]</sup>、微阵列数据分析<sup>[3]</sup>和人脸识别<sup>[4-5]</sup>等领域。经典的LDA模型由Fisher于1936年首次提出<sup>[6]</sup>,它通过最大化类间散度与类内散度的比值,寻找最优投影方向,并转换为广义特征值问题求解。但经典的LDA存在一些局限性:1)对异常值或噪声敏感;2)类间散度矩阵需可逆;3)最多提取 $K-1$ 个判别特征( $K$ 为类别数)。

1989年,Friedman<sup>[7]</sup>提出了正则化判别分析(regularized discriminant analysis, RDA),通过引入双重正则化机制,结合交叉验证策略优化参数组合,有效解决了高维小样本数据下LDA和二次判别分析可能带来的协方差矩阵奇异性和估计偏差问题,明显提高了模型分类的准确性和稳定性。但该模型未考虑稀疏性。2005年,Ye<sup>[8]</sup>提出了一种广义判别分析框架,通过新优化准则 $\text{trace}((S_L^T)^+ S_b^T)$ 采用同时对角化技术解决小样本问题中的矩阵奇异问题,并首次引入正交线性判别分析,通过QR分解避免了矩阵的直接求逆运算,保证投影矩阵的正交性。但该模型同样未引入稀疏性,导致模型解释性较弱。2014年,Wang等人<sup>[9]</sup>提出了基于 $L_1$ 范数的LDA模型LDA-L1,通过使用 $L_1$ 范数替代 $L_2$ 范数计算类间和类内散度,解决了传统LDA对异常值敏感及样本限制问题。同年,Zheng等人<sup>[10]</sup>提出了 $L_1$ -LDA,基于 $L_1$ 范数类间和类内散度,通过Bayes误差界优化目标函数,使用迭代算法求解,每次迭代中只需求解1个凸规划问题,且该问题有闭式解。2016年,Zhang等人<sup>[11]</sup>在不相关线性判别分析中引入稀疏性约束,提出了稀疏不相关线性判别分析模型,通过将 $L_1$ 范数最小化和正交约束优化相结合,在保证特征不相关性的同时实现稀疏性,有效解决了高维小样本问题中的特征选择与分类任务,并使用线性化交替方向乘子法(alternating direction method of multipliers, ADMM)或加速线性化Bregman算法求解,但未考虑鲁棒性,仅关注稀疏性。2017年,Li等人<sup>[12]</sup>提出基于绝对值不等式损失和 $L_1$ 正则化的绝对值不等式判别分析方法,通过重构优化问题实现了鲁棒且稀疏的监督降维,有效克服了传统LDA的局限性,通过连续线性算法求解,但需要求解一系列线性规划问题,计算成本较高。2020年,Li等人<sup>[13]</sup>提出基于 $L_1$ 范数和Bhattacharyya误差界优化的鲁棒稀疏线性判别分析(robust sparse linear discriminant analysis, RSLDA),通过ADMM求解该非凸问题,在保持抗噪性、处理小样本优势的同时,RSLDA增强了对异常值的鲁棒性,提升了模

\* 收稿日期:2025-06-17 修回日期:2025-10-09 网络出版时间:2026-01-08T16:58

资助项目:国家自然科学基金面上项目(No. 11871128, No. 12171063);重庆市教育委员会科学技术研究计划重点项目(No. KJZD-K202300509)

第一作者简介:翁婷,女,研究方向为最优化理论与算法,E-mail: wengting\_xs@163.com;通信作者:李国权,男,教授,博士,E-mail: ligq@cqnu.edu.cn

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20260108.1547.012

型的特征选择能力。由于RSLDA的目标函数是非凸且非光滑的,采用ADMM算法求解该问题无法保证算法的收敛性,文献[13]中的实验结果也表明在某些数据集上ADMM算法是不收敛的。此外,使用ADMM的求解过程中需额外引入参数,参数的取值会对模型的泛化性能有影响。因此,本文从文献[13]提出的鲁棒性判别分析(robust linear discriminant analysis,RLDA)和RSLDA模型出发,基于差分凸(difference of convex,DC)规划方法对模型进行重构,利用DC算法(difference of convex functions algorithm,DCA)进行求解,并给出了该算法的收敛性,且DC子问题具有解析解。从而本文所提的新算法不仅保证了收敛性,而且更加高效,使得模型的训练效率得到明显提升。

## 1 LDA 简介

LDA凭借简单性和较高的预测准确性,成为监督分类任务中的重要工具。LDA源于Fisher判别问题,主要通过以下2个关键矩阵定义进行判别:

$$\mathbf{W} = \frac{1}{n} \sum_{k=1}^K \sum_{j \in C_k} (\mathbf{x}_j - \boldsymbol{\mu}_k)(\mathbf{x}_j - \boldsymbol{\mu}_k)^T, \mathbf{B} = \frac{1}{n} \sum_{k=1}^K n_k \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T.$$

其中: $n$ 为样本总数; $n_k$ 为第 $k$ 类的样本数; $K$ 为类别数; $C_k$ 为第 $k$ 类的样本索引集合; $\mathbf{x}_j$ 为样本索引 $j$ 对应的样本特征向量; $\boldsymbol{\mu}_k = \frac{1}{n_k} \sum_{j \in C_k} \mathbf{x}_j$ 是第 $k$ 类样本的均值向量; $\mathbf{W}$ 表示类内散度矩阵,用于衡量同一类别内部的紧凑性; $\mathbf{B}$ 表示类间散度矩阵,用于衡量不同类别之间的差异。

LDA旨在寻找一组正交的判别方向 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{K-1}$ ,这些判别方向可通过依次求解下列问题获得:

$$\begin{aligned} \max_{\mathbf{w} \in \mathbb{R}^p} \quad & \frac{\mathbf{w}^T \mathbf{B} \mathbf{w}}{\mathbf{w}^T \mathbf{W} \mathbf{w}}, \\ \text{s. t.} \quad & \mathbf{w}_h^T \mathbf{W} \mathbf{w}_l = 0, 1 \leq l < h < K. \end{aligned} \quad (1)$$

利用Lagrange乘法,可将问题(1)转换为广义特征值问题:

$$\mathbf{B} \mathbf{w}_h = \lambda \mathbf{W} \mathbf{w}_h.$$

从而问题简化为求解 $\mathbf{W}$ 的最大特征所对应的特征向量,即最优投影向量对应于最大特征值的特征向量。当 $\mathbf{W}$ 非奇异时,解由 $\mathbf{W}^{-1} \mathbf{B}$ 的前 $K$ 个最大非零特征值对应的特征向量给出,投影向量需满足正交性,以保证各维度间的独立性。

然而,LDA存在一些局限:1)对异常值敏感,类间散度矩阵 $\mathbf{B}$ 是基于 $L_2$ 范数计算的,类间距离的估计易受异常值的干扰;2)对于小样本问题,即当样本维度远大于样本数量时,类内散度矩阵 $\mathbf{W}$ 可能不可逆,导致广义特征值问题无法求解;3)所能提取的特征数量受限,受限于 $\mathbf{B}$ 的秩,LDA至多可提取 $K-1$ 个判别特征,不适用于高维数据的处理。

## 2 RLDA 和 RSLDA

### 2.1 RLDA 及 DC 规划方法

文献[13]提出的RLDA模型使用作差代替了经典LDA中比值的形式,并引入 $L_1$ 范数替代 $L_2$ 范数计算类间差异,以降低对异常值的敏感性,使得模型在存在噪声或异常值时仍能保持稳定性能;并突破特征数限制,可生成任意数量的判别方向。当只求解1个判别向量时,模型为:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \lambda \mathbf{w}^T \mathbf{W} \mathbf{w}. \quad (2)$$

其中: $\bar{\mathbf{x}}_i$ 表示第 $i$ 类样本的均值, $\bar{\mathbf{x}}$ 表示所有样本的均值, $\lambda > 0$ 为正则化参数。上式中前一项,最大化 $L_1$ 范数类间散度,减少异常值影响。后一项,最小化 $L_2$ 范数类内散度,保持同类样本聚集。为求解式(2),文献[13]提出了一个基于ADMM框架的求解算法,通过交替优化变量得到最优解。这样的变化虽增强了LDA的鲁棒性,但算法可能不收敛。因此,本文提出一种新的优化算法,将问题进行变形并采用DCA对问题进行求解,其中生成的子问题利用软阈值算子可求得解析解。

令 $\mathbf{W} = \mu \mathbf{I} - (\mu \mathbf{I} - \mathbf{W})$ , $\mathbf{I}$ 为单位矩阵, $\mu$ 是一个大于 $\mathbf{W}$ 最大特征值的常数,即满足 $\mu > \lambda_{\max}(\mathbf{W})$ 。因式(2)中参数 $\lambda > 0$ ,故可将 $\lambda$ 换至第1项,而不改变模型的结构,得到:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} -\lambda \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \mathbf{w}^T \mathbf{W} \mathbf{w}. \quad (3)$$

将  $\mathbf{W} = \mu \mathbf{I} - (\mu \mathbf{I} - \mathbf{W})$  代入式(3)中,问题变为:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} -\lambda \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \mathbf{w}^T (\mu \mathbf{I} - (\mu \mathbf{I} - \mathbf{W})) \mathbf{w},$$

整理后得到:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \mu \mathbf{w}^T \mathbf{w} - \left[ \lambda \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \mathbf{w}^T (\mu \mathbf{I} - \mathbf{W}) \mathbf{w} \right]. \quad (4)$$

观察所得式子,减号前一项  $\mu \mathbf{w}^T \mathbf{w}$  显然为关于变量  $\mathbf{w}$  的凸函数。减号后由 2 部分组成,由于  $\mu > \lambda_{\max}(\mathbf{W})$ ,因此  $\mu \mathbf{I} - \mathbf{W}$  为正定矩阵。由此可知,  $\mathbf{w}^T (\mu \mathbf{I} - \mathbf{W}) \mathbf{w}$  关于变量  $\mathbf{w}$  是凸函数。令:

$$g(\mathbf{w}) = \mu \mathbf{w}^T \mathbf{w}, h(\mathbf{w}) = \lambda \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \mathbf{w}^T (\mu \mathbf{I} - \mathbf{W}) \mathbf{w},$$

则问题(4)可写成如下形式:

$$\min_{\mathbf{w}} g(\mathbf{w}) - h(\mathbf{w}). \quad (5)$$

因  $g(\mathbf{w}), h(\mathbf{w})$  都是连续凸函数,故问题(5)是一个标准的 DC 规划问题,可采用 DCA 进行求解。在每一次迭代中,需求解子问题:

$$\min_{\mathbf{w}} g(\mathbf{w}) - \langle \mathbf{d}^t, \mathbf{w} \rangle, \quad (6)$$

其中:  $t$  为迭代次数,  $\mathbf{d}^t$  为  $h(\mathbf{w})$  在  $\mathbf{w}^t$  处的次梯度,即  $\mathbf{d}^t \in \partial h(\mathbf{w}^t)$ 。由无约束优化最优性充要条件可知子问题

(6)的解为  $\mathbf{w}^{t+1} = \frac{\mathbf{d}^t}{2\mu}$ 。

下面给出求解问题(5)的迭代算法。

**算法 1** 问题(5)的 DCA。

步 0,初始化:令  $t=0$ ,选择初始点  $\mathbf{w}^0 \in \mathbf{R}^p$ ;

步 1,计算  $\mathbf{d}^t \in \partial h(\mathbf{w}^t)$ ;

步 2,计算  $\mathbf{w}^{t+1}$ ,通过  $\mathbf{w}^{t+1} = \frac{\mathbf{d}^t}{2\mu}$ ;

步 3,若满足终止条件,算法终止;否则,令  $t=t+1$ ,转步 1。

算法 1 中,  $h(\mathbf{w})$  在  $\mathbf{w}^t$  处的次梯度可由下式计算:

$$\mathbf{d}^t = \lambda \sum_{i=1}^K (n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \text{sgn} \langle \mathbf{w}^t, (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \rangle) + 2(\mu \mathbf{I} - \mathbf{W}) \mathbf{w}^t. \quad (7)$$

其中:  $\text{sgn}(\cdot)$  为符号函数。

算法 1 是标准的 DCA,收敛性结果见定理 1。

**定理 1**<sup>[14]</sup> 给定初始点  $\mathbf{w}^0$ ,若算法 1 产生的点列  $\{\mathbf{w}^k\}$  是有界的,则该点列的任一极限点都是问题(5)的稳定点。

## 2.2 RSLDA 及 DC 规划方法

文献[13]为了提升模型的特征选择能力,在问题(2)的目标函数中添加 1 个  $L_1$  范数正则化项,得到以下模型:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} - \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \lambda \mathbf{w}^T \mathbf{W} \mathbf{w} + \rho \|\mathbf{w}\|_1. \quad (8)$$

其中:  $\rho > 0$  为正则化参数。

与 2.1 节类似,模型可改写为:

$$\min_{\mathbf{w}} \mu \mathbf{w}^T \mathbf{w} - h(\mathbf{w}) + \rho \|\mathbf{w}\|_1,$$

其中:  $h(\mathbf{w})$  与 2.1 节一致,即  $h(\mathbf{w}) = \lambda \sum_{i=1}^K (n_i | \mathbf{w}^T \bar{\mathbf{x}}_i - \mathbf{w}^T \bar{\mathbf{x}} |) + \mathbf{w}^T (\mu \mathbf{I} - \mathbf{W}) \mathbf{w}$ 。可以看到添加  $L_1$  正则项后的问题仍可看作 2 个下半连续凸函数之差,因此对于稀疏后的模型仍是一个标准的 DC 规划问题。在每一次迭代中,需要计算  $h(\mathbf{w})$  在  $\mathbf{w}^t$  处的次梯度  $\mathbf{d}^t$  并求解子问题:

$$\min_{\mathbf{w}} \mu \mathbf{w}^T \mathbf{w} + \rho \|\mathbf{w}\|_1 - \langle \mathbf{d}^t, \mathbf{w} \rangle, \quad (9)$$

子问题(9)等价于:

$$\min_{\mathbf{w}} \frac{1}{2} \left\| \mathbf{w} - \frac{\mathbf{d}^t}{2\mu} \right\|_2^2 + \frac{\rho}{2\mu} \|\mathbf{w}\|_1. \quad (10)$$

由软阈值算子可直接求得问题(10)的解析解:

$$\mathbf{w}_i^{t+1} = \begin{cases} \frac{d_i^t + \rho}{2\mu}, & d_i^t < -\rho; \\ 0, & |d_i^t| < \rho; \\ \frac{d_i^t - \rho}{2\mu}, & d_i^t > \rho. \end{cases} \quad (11)$$

下面给出求解问题(8)的算法。

**算法 2** 问题(8)的 DCA。

步 0, 初始化: 令  $t=0$ , 选择初始点  $\mathbf{w}^0 \in \mathbf{R}^p$ ;

步 1, 根据式(7)计算  $\mathbf{d}^t \in \partial h(\mathbf{w}^t)$ ;

步 2, 根据式(11)计算  $\mathbf{w}^{t+1}$ ;

步 3, 若满足终止条件, 算法停止; 否则, 令  $t=t+1$ , 转步 1。

然而, 对于一个真实的数据集, 1 个判别方向通常是不够的, 它应该被投影到一个更高维的空间, 这更有利于对数据的不同特征进行分析, 而不仅仅是通过 1 个判别方向得到 1 个一维的投影数据集。因此, 为了获得更多的判别方向, 采用文献[15]中的方法, 利用已经求得的判别方向对数据集进行处理, 得到一个新的数据集, 再利用该数据集按照相同方法求得新的判别方向, 并不断重复这个步骤, 具体而言: 将原本计算得到的第 1 个判别方向表示为  $\mathbf{w}_1$ , 将后续需要计算的第  $r$  个判别方向记为  $\mathbf{w}_r$ , 其中  $r > 1$ 。则  $\mathbf{w}_l$  由以下样本集求得:

$$\mathbf{x}^{\text{new}} = \mathbf{x} - \sum_{l=1}^{r-1} \mathbf{w}_l (\mathbf{w}_l^T \mathbf{x}), \quad (12)$$

其中:  $\mathbf{w}_l$  是第  $l$  个判别方向,  $\mathbf{x}$  表示数据集更新  $r-1$  次后任意一个样本点,  $\mathbf{x}^{\text{new}}$  表示该样本点对应生成的新的样本点。因此, 式(12)确保能不断获取新的样本集, 并且使得包含在先前获得的判别方向中的信息被扣除。通过这个更新样本点的方法, 就可以得到不受样本类别数  $K$  所限制的多个判别方向。

基于上述讨论, 总结优化问题(8)的 DC 规划求解流程如下, 称为算法 3。

**算法 3** 基于 DC 规划的  $L_1$  稀疏线性判别分析(DC\_SLDA)。

步 0, 初始化: 令  $k=1$ , 选择需要的判别向量个数  $d, d > 1$ 。

步 1, 对数据集  $\mathbf{X}$  使用算法 2 得到判别方向  $\mathbf{w}_k$ ;

步 2, 通过式(12)更新  $\mathbf{X}$ ;

步 3, 若  $k=d$ , 则算法终止; 否则  $k=k+1$ , 转步 1。

### 3 数值实验

为了说明本文所提出的算法 DC\_SLDA 的有效性, 将 DC\_SLDA 与几种经典的方法进行比较, 包括 LDA、RDA<sup>[16]</sup>、惩罚线性判别分析 (penalized linear discriminant analysis, PLDA)<sup>[17]</sup>、稀疏 Fisher 线性判别分析 (sparse Fisher linear discriminant analysis, SFLDA)<sup>[18]</sup> 和 RSLDA<sup>[13]</sup>。所有数值实验环境均为笔记本电脑 Intel (R) Core(TM) i5-7300HQ 处理器和 8 GB 内存, 所有算法均通过 Matlab 2016a 编码实现。对于所有算法所设置最大迭代次数均为 200, 并采用  $\|\mathbf{w}^{t+1} - \mathbf{w}^t\| \leq 10^{-3}$  作为算法终止准则。RDA 的正则化参数和其他 2 个参数均从集合  $\{10^{-4}, 10^{-3}, \dots, 10^{-1}, 1, 10\}$  中选出。RSLDA 中的参数  $\rho$  和  $\lambda$  分别从  $\{0.1, 0.5, 1, 5, 10, 50, 100, 500\}$  和  $\{0.1, 0.5, 1, 5, 10, 50, 100\}$  中选取,  $\delta$  从  $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$  中选取。PLDA 中罚参数从  $\{0.1, 0.5, 1, 5, 10, 50, 100\}$  中选取。SFLDA 中对于  $L_0$  范数的近似函数的近似参数  $\alpha=5$ 。所有实验均利用五折交叉验证的方法对超参数进行优化。

#### 3.1 仿真数据集

下面将本文算法 DC\_SLDA 与 LDA、RDA、PLDA 和 SFLDA 进行对比。首先, 研究算法的鲁棒性。需要构

造一个仿真数据集。数据分为 2 类,每类 50 个样本。第 1 类样本满足均值为  $(\sqrt{3}, 1)$  和协方差为  $\begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$  的高斯分布,而第 2 类样本点满足均值为  $(4\sqrt{3}, 3)$  和协方差为  $\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$  的高斯分布。仿真数据集样本点分布如图 1 所示。

显然,当不添加异常值点时,上述数据最理想的判别方向与  $x$  轴之间的夹角应为  $30^\circ$ 。将 LDA 运用到上述仿真数据集,并在图 2 中绘制出它所得的判别向量。

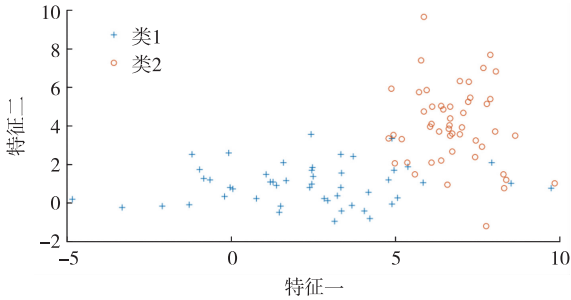


图 1 仿真数据集

Fig. 1 Artificial dataset

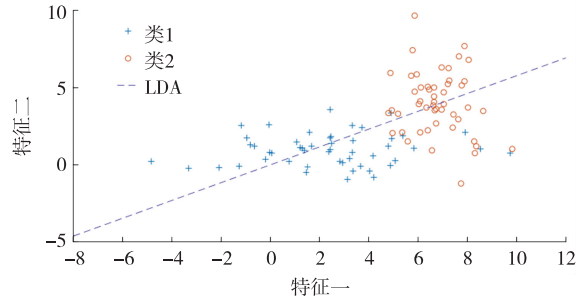


图 2 LDA 在仿真数据集上的表现

Fig. 2 The performance of LDA on artificial datasets

现对 2 个类别的样本点均添加若干游离于类别之外的异常点,即在仿真数据集当中添加噪声。再应用 LDA、RDA、PLDA、SFLDA 以及本文提出的算法 DC\_SLDA 到上述仿真数据集,并在图中绘制这些算法的判别向量,见图 3。

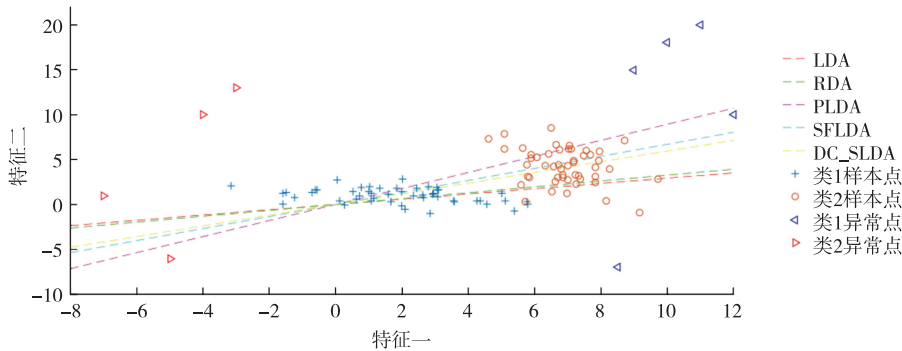


图 3 添加噪声后各算法表现

Fig. 3 The performance of each algorithm after adding noise

通过比较加入噪声后各种方法得到的判别向量与  $x$  轴的夹角,可以分析出各类方法的鲁棒性。若所得夹角越趋于  $30^\circ$ ,则该类方法有更好的鲁棒性。反之,越偏离  $30^\circ$ ,此类方法在面对拥有异常值的数据集时表现越糟糕。各算法所得判别方向与  $x$  轴的夹角分别为  $\theta_{LDA} = 15.16^\circ, \theta_{RDA} = 17.33^\circ, \theta_{PLDA} = 38.63^\circ, \theta_{DC\_SLDA} = 28.46^\circ$ 。

结果表明,在将样本投影点间的距离以  $L_1$  范数来度量后,算法 DC\_SLDA 具有良好的鲁棒性。

为了体现本文提出的算法在 RSLDA 基础上改进后算法变得收敛,建立一个新的仿真数据集,并记录每次迭代后的目标函数值,以此观察 RSLDA 和 DC\_SLDA 这 2 种算法的收敛情况。数据集包含 2 类样本点,每个类别包含 100 个样本点。第 1 类样本点为满足  $X \sim N\left(\frac{1}{3}, 1\right)$  的高斯分布,第 2 类样点为满足  $X \sim N\left(\frac{2}{3}, 2\right)$  的高斯分布。目标函数值随着迭代次数的变化,如图 4 和图 5 所示,结果表明改进后的算法 DC\_SLDA 在仿真数据集上是收敛的,但 RSLDA 却不收敛。

### 3.2 真实数据集

将本文算法与 LDA、PLDA、SFLDA、RSLDA 进行比较,实验选取了 UCR 数据库的 8 个基准数据集。表 4 详细列出了这些数据集的信息。此次实验旨在通过这些多样化的数据集,全面评估所提出方法的性能。具体实验结果见表 5。

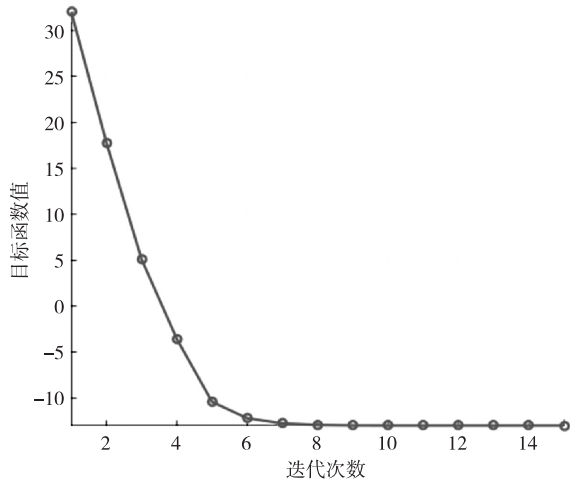


图 4 DC-SLDA 在仿真数据集上表现

Fig. 4 DC-SLDA performs on artificial datasets

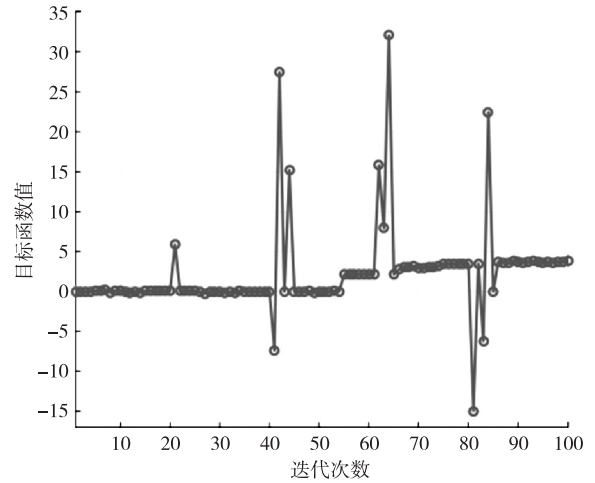


图 5 RSLDA 在仿真数据集上表现

Fig. 5 RSLDA performs on artificial datasets

表 4 实验中使用的基准数据集

Tab. 4 The benchmark dataset used in the experiment

数据集	样本大小/个	特征数个数	数据集	样本大小/个	特征数个数
BirdChicken	40	512	ToeSegmentation2	36	343
ECGFiveDays	23	136	ShapeletSim	20	500
Gun	50	150	Herring	64	512
MoteStrain	20	84	TwoLeadECG	20	512

表 5 实验结果

Tab. 5 Experimental results

数据集		算法				
		LDA	PLDA	RSLDA	SFLDA	DC_SLDA
BirdChicken	分类准确率	(67.03±14.31)%	(76.73±21.12)%	(78.43±14.88)%	(74.62±8.61)%	<b>(82.47±7.48)%</b>
	特征数个数	512	131	262	107	234
	训练耗时/s	0.003 2	0.002 8	0.036	0.062 6	0.018
ECGFiveDays	分类准确率	(75.37±7.46)%	(78.26±18.07)%	<b>(88.72±13.53)%</b>	(72.56±17.96)%	(87.20±10.39)%
	特征数个数	136	42	91	50	80
	训练耗时/s	0.004 6	0.000 7	0.003 1	0.000 7	0.001 2
Gun	分类准确率	(62.41±12.18)%	67.13±3.97)%	(74.02±16.23)%	<b>(77.14±6.39)%</b>	(72.47±18.83)%
	特征数个数	150	24	97	54	113
	训练耗时/s	0.003 2	0.000 6	0.011 2	0.002 1	0.000 8
MoteStrain	分类准确率	(75.29±19.32)%	(74.53±18.68)%	(84.27±29.11)%	(87.28±24.63)%	<b>(89.33±6.21)%</b>
	特征数个数	84	42	57	50	53
	训练耗时/s	0.000 5	0.264 1	0.006 9	0.000 8	0.000 8
ToeSegmentation2	分类准确率	(62.36±23.18)%	(58.24±13.31)%	(64.27±25.37)%	(66.67±7.33)%	<b>(70.41±6.34)%</b>
	特征数个数	343	120	63	74	56
	训练耗时/s	0.003	0.001	0.033 1	0.052 5	0.004 9
ShapeletSim	分类准确率	(73.50±23.97)%	<b>(95.89±2.92)%</b>	(74.62±8.49)%	(78.32±23.83)%	(74.31±14.23)%
	特征数个数	500	66	23	50	30
	训练耗时/s	0.001 6	0.002 5	0.049 7	0.012 4	0.006 6

续表 5

数据集	算法					
	LDA	PLDA	RSLDA	SFLDA	DC_SLDA	
Herring	分类准确率	(60.72±17.04)%	(53.98±4.12)%	(60.94±12.21)%	<b>(63.98±21.42)%</b>	(61.29±11.49)%
	特征数个数	512	143	255	179	227
	训练耗时/s	0.003	0.002 6	0.036 2	0.001 5	0.004 2
TwoLeadECG	分类准确率	(56.43±20.84)%	(91.23±5.32)%	(67.42±31.43)%	(65.00±23.48)%	<b>(73.27±20.38)%</b>
	特征数个数	512	459	503	421	485
	训练耗时/s	0.001 4	0.002 4	0.009 4	0.014 4	0.003 6
平均准确率		66.64%	74.5%	74.09%	73.19%	<b>76.34%</b>

注:表中加黑的数据表示最优结果。

表 5 给出了分类器在基准数据集上的数值结果,不难观察到:在特征选择方面,LDA 的解几乎没有稀疏性,与其余 4 个具有一定特征选择能力的分类器相比,LDA 的泛化表现不佳,这也说明了提高 LDA 的特征选择能力是很有必要的。从实验结果可以发现,DC\_SLDA 有着较高的分类准确率,且具备较好的特征选择能力,能有效解决 LDA 面临的两大问题中的过拟合问题。与 RSLDA 相比,DC\_SLDA 的运行时间得到了明显的降低,并且解的稀疏性却没有降低。

DC\_SLDA 是一个能减少维度的方法,数据点可经判别向量投影到一个低维空间当中,图 6 生动地说明了数据集 Gun 投影到 2 个获得的判别向量上的情况。可以看到,不同类数据集在投影后尽可能地分散开。总的来说,本文算法与 RSLDA 相比是收敛的,这是一大突破。

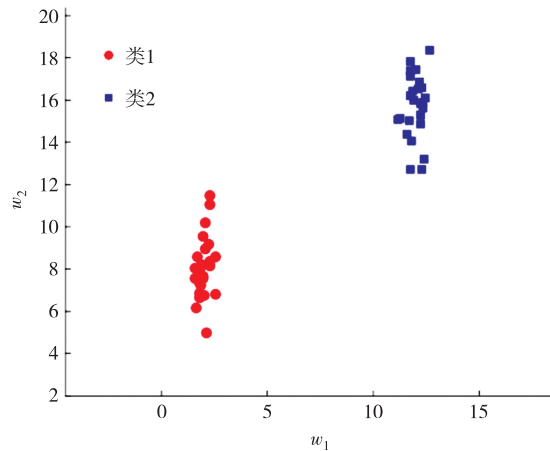


图 6 数据集 Gun 的投影

Fig. 6 The projection of the dataset Gun

## 4 结束语

本文研究了  $L_1$  范数正则化项的稀疏线性判别分析,将原问题通过简单的变形转化为 DC 规划问题,最后使用 DCA 进行高效求解,与修改前的方法相比,保证了算法的收敛性并不失解的稀疏性,且新算法中的子问题具有解析解。实验表明本文所提出的方法具有较好的特征选择能力和鲁棒性,与 RSLDA 收敛性的对比实验也表明所提算法是收敛的。真实数据集上的对比实验也能直观看到本文所提算法在运行时间上比 RSLDA 更短。

### 参考文献:

- [1] JAIN A K, DUBES R C. Algorithms for clustering data[M]. Englewood Cliffs: Prentice Hall, 1988.
- [2] KOWALSKI G J. Information retrieval systems: theory and implementation[M]. New York: Springer, 2007.
- [3] DUDOIT S, FRIDLAND J, SPEED T P. Comparison of discrimination methods for the classification of tumors using gene expression data[J]. Journal of the American Statistical Association, 2002, 97(457): 77-87.

- [4] JIN Z, YANG J Y, HU Z S, et al. Face recognition based on the uncorrelated discriminant transformation[J]. Pattern Recognition, 2001, 34(7): 1405-1416.
- [5] SWETS D L, WENG J J. Using discriminant eigenfeatures for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 831-836.
- [6] FISHER R A. The use of multiple measurements in taxonomic problems[J]. Annals of Eugenics, 1936, 7(2): 179-188.
- [7] FRIEDMAN J H. Regularized discriminant analysis[J]. Journal of the American Statistical Association, 1989, 84(405): 165-175.
- [8] YE J P. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems[J]. The Journal of Machine Learning Research, 2005, 6: 483-502.
- [9] WANG H X, LU X S, HU Z L, et al. Fisher discriminant analysis with  $L_1$ -norm[J]. IEEE Transactions on Cybernetics, 2014, 44(6): 828-842.
- [10] ZHENG W M, LIN Z C, WANG H X.  $L_1$ -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.
- [11] ZHANG X W, CHU D L, TAN R C E. Sparse uncorrelated linear discriminant analysis for undersampled problems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(7): 1469-1485.
- [12] LI C N, ZHENG Z R, LIU M Z, et al. Robust recursive absolute value inequalities discriminant analysis with sparseness[J]. Neural Networks, 2017, 93: 205-218.
- [13] LI C N, SHAO Y H, YIN W T, et al. Robust and sparse linear discriminant analysis via an alternating direction method of multipliers[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(3): 915-926.
- [14] LE THI H A, PHAM DINH T. DC programming and DCA: thirty years of developments[J]. Mathematical Programming, 2018, 169(1): 5-68.
- [15] TAO Q, CHU D J, WANG J. Recursive support vector machines for dimensionality reduction[J]. IEEE Transactions on Neural Networks, 2008, 19(1): 189-193.
- [16] GUO Y Q, HASTIE T, TIBSHIRANI R. Regularized linear discriminant analysis and its application in microarrays[J]. Biostatistics, 2007, 8(1): 86-100.
- [17] WITTEN D M, TIBSHIRANI R. Penalized classification using Fisher's linear discriminant[J]. Journal of the Royal Statistical Society Series B, Statistical Methodology, 2011, 73(5): 753-772.
- [18] LE THI H A, PHAN D N. DC programming and DCA for sparse Fisher linear discriminant analysis[J]. Neural Computing and Applications, 2017, 28(9): 2809-2822.

## Operations Research and Cybernetics

### $L_1$ Norm Sparse Linear Discriminant Analysis Based on DC Programming

WENG Ting, LI Guoquan, ZHANG Jiahao

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

**Abstract:** To address traditional linear discriminant analysis's insufficient robustness to noise and poor generalization performance on high-dimensional datasets with small sample sizes (i. e., where the number of features far exceeds the number of samples), literature proposed robust linear discriminant analysis (RLDA) and robust sparse linear discriminant analysis (RSLDA), along with an ADMM-based optimization algorithm for solving them. However, this method lacks convergence guarantees and fails to converge on some datasets. Based on the structural features of RLDA and RSLDA, an improved optimization algorithm DC\_SLDA based on differences of convex functions (DC) programming is proposed. The algorithm transforms the original problem into a DC programming problem through DC decomposition, and then applies the difference of convex functions algorithm for iterative solution. The algorithm proposed not only has convergence guarantees but also features closed-form solutions for its subproblems, which significantly improves the training efficiency of the model.

**Keywords:** linear discriminant analysis; sparsity; DC programming

(责任编辑 黄 颖)