

一种无参数的局部线性判别分析方法

黄礼泊, 凌永权

广东工业大学信息工程学院, 广东 广州

Email: www.huanglibo@gmail.com, yongquanling@gdut.edu.cn

收稿日期: 2021年3月23日; 录用日期: 2021年4月19日; 发布日期: 2021年4月26日

摘要

为了解决因引入局部化思想的线性判别分析(Linear Discriminant Analysis, LDA)方法需要人工设置邻居个数而无法以自适应的方式挖掘数据的局部结构问题, 提出了一种无参数的局部线性判别分析(Parameter-free Local Linear Discriminant Analysis, Pf-LLDA)方法。该方法首先建立了一个关于权重矩阵和变换矩阵的统一优化模型。然后, 通过使用交替方向的方法迭代求解出了与数据局部结构相关的权重矩阵和与判别分析相关的变换矩阵。从而使得Pf-LLDA在无需人为设定邻居个数的情况下, 自适应地挖掘出了数据的局部结构并最终实现了局部线性判别分析的能力。在仿真数据集和手写体真实数据集上的实验结果表明, Pf-LLDA挖掘出数据局部结构的同时实现了更优的判别分析结果。

关键词

无参, 近邻, 线性判别分析, 局部结构, 降维

A Parameter-Free Local Linear Discriminant Analysis Method

Libo Huang, Wing-Kuen Ling

School of Information Engineering, Guangdong University of Technology, Guangzhou Guangdong

Email: www.huanglibo@gmail.com, yongquanling@gdut.edu.cn

Received: Mar. 23rd, 2021; accepted: Apr. 19th, 2021; published: Apr. 26th, 2021

Abstract

For the Linear Discriminant Analysis (LDA) method which utilized the idea of localization and setting the number of neighbors by hand, it cannot mine the local structure embedded in the data adaptively. In order to solve this problem, this paper proposes a Parameter-free Local Linear Discriminant Analysis (Pf-LLDA) method. This method first establishes a unified optimization model about the weight matrix and the transformation matrix. Then, by using the alternating direction

method, the solution of the weight matrix, which is associated with the local structure embedded in the data, and the transformation matrix, which is associated with the discriminant analysis are iteratively well found. Thus, Pf-LLDA adaptively mines the local structure of the data and finally achieves the capability of local linear discriminant analysis without artificially setting the number of neighbors. Experimental results on both synthetic and handwritten real datasets show that Pf-LLDA achieves better discriminant analysis results while mining the local structure of the data.

Keywords

Parameter-Free, Nearest Neighbor, Linear Discriminant Analysis, Local Structure, Dimensionality Reduction

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

维度约简在解决维度灾难问题上具有很重要的作用。降维作为维度约简中最为直观的一种方法在模式识别和机器学习领域具有很重要的研究意义[1]。

通常来讲, 根据是否使用了数据标签信息, 降维方法可以被简单的分为有监督的降维方法和无监督的降维方法[2]。相对于无监督的降维方法, 有监督的降维方法因为利用了已知的数据标签信息而被更广泛的采用。线性判别分析(Linear Discriminant Analysis, LDA) [1] [2]便是一个典例。LDA 通过最大化类间距离的同时并最小化类内距离的方式找到了具有判别性的线性变换。但由于 LDA 用样本均值代表整体样本的方式构造类内散度矩阵和整体散度矩阵, 当同类样本不满足高斯分布假设(即多模态现象)时[3], LDA 将失去判别能力。为了使 LDA 能有效地应对多模态现象, 国内外众多学者在传统 LDA 中引入了数据的局部结构。比较典型的有: 非参数¹判别分析(Nonparametric Discriminant Analysis, NDA) [4], 局部 Fisher 判别分析(Local Fisher Discriminant Analysis, LFDA) [3]和局部敏感判别分析[5] (Locality Sensitive Discriminant Analysis, LSDA)等。更为具体地, NDA 通过将每个样本替换表示成其近邻样本加权的形式引入了数据的局部结构。LFDA 首先将 LDA 中的类内和整体散度矩阵重新表示成样本间距离平方和的形式, 然后再引入局部保留投影的思想使样本在降维前后保持着一致的邻居关系, 从而达到局部结构的引入。不同于 LFDA, LSDA 从谱图理论出发, 通过将图表示成拉普拉斯矩阵的形式引入了数据的局部结构。此外, 还有许多在 LFDA 和 LSDA 模型基础上的变体[6] [7] [8], 这些变体本质上都是通过改变样本间相似关系的权重值或者图节点间边值的大小得到对数据结构更准确的近邻表达。不幸的是, 这些方法在引入数据局部结构时都必须人为设定样本的近邻个数。该近邻个数参数决定性地影响着数据降维后的可分性, 且其具体值的确定需要科研人员对数据的先验知识具有很全面的认识, 这很大程度上限制了局部线性判别分析方法的广泛应用[9]。因此, 仅根据已有的数据样本, 自动对数据进行局部线性判别分析是一个非常值得研究的问题[10], 同时在图像检索[4]和人脸识别[11]等方面具有很重要的应用价值。

受自权重自适应局部判别分析(Self-weighted Adaptive Locality Discriminant Analysis, SALDA) [9]的启发, 本文提出了一种无参数的局部判别分析(Parameter-free Local Linear Discriminant Analysis, Pf-LLDA)方法。Pf-LLDA 首先建立一个与 LDA 类似的模型。但不同于 LDA 模型只需要求解与降维相关的变换矩阵, Pf-LLDA 不仅要求解与降维相关的变换矩阵还要求解与数据局部结构相关的权重矩阵。通过运¹需要注意的是, 这里的“非参数”意指无高斯假设, 直译于 Nonparametric, 与本文标题“无参数”区分开。

用交替方向的优化求解思路[12], Pf-LLDA 模型在无需设置近邻个数的前提下得到了反映数据结构图的权重矩阵和考虑了数据局部结构性质的变换矩阵。最后, 根据变换矩阵得到了测试样本的低维表示, 并将其输入到 k 近邻分类器中进行验证。在仿真和真实数据集上的实验结果表明, Pf-LLDA 不仅在多模态问题上实现了无参数的局部判别分析性能, 而且在手写字体识别问题上相比其他方法得到了更优的结果。

2. 算法描述

本章从已有的工作出发, 引出了无参数的局部判别分析(Parameter-free Local Linear Discriminant Analysis, Pf-LLDA)模型, 并给出了具体的求解方法与伪代码。更具体地, 2.1 节、2.2 节和 2.3 节分别概述了线性判别分析(Linear Discriminant Analysis, LDA)、局部 Fisher 判别分析(Local Fisher Discriminant Analysis, LFDA)和自权重自适应局部判别分析(Self-weighted Adaptive Locality Discriminant Analysis, SALDA)模型; 2.4 节详细阐述了 Pf-LLDA 模型与求解方案。

2.1. 线性判别分析(LDA)

给定 n 个 m 维的数据样本 $X = [x_1, \dots, x_n] \in \mathbb{R}^{m \times n}$, 并且已知这些样本来自 c 个不同的类别, LDA 可以通过以下固定整体分散度的同时最小化类内分散度的方式得到具有判别性的变换矩阵 $A \in \mathbb{R}^{m \times d}$ 。

$$\begin{aligned} \min_A \operatorname{Tr}(A^T S_w A) \\ \text{s.t. } A^T S_t A = I_d \end{aligned} \quad (1)$$

这里, d 表示低维目标空间的维数; $\operatorname{Tr}(X)$ 表示对矩阵 X 的求迹运算; X^T 表示对矩阵 X 转置; I_d 表示 $d \times d$ 维的单位阵。 S_w 和 S_t 分别表示数据的类内散度矩阵和整体散度矩阵, 并分别由以下公式计算[3][7]:

$$\begin{aligned} S_w &= \sum_{i=1}^c \sum_{j=1}^{n_i} (x_j^i - \mu^i)(x_j^i - \mu^i)^T \\ &= \frac{1}{2} \sum_{i=1}^c \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} (x_j^i - x_k^i)(x_j^i - x_k^i)^T, \end{aligned} \quad (2)$$

$$\begin{aligned} S_t &= \sum_{j=1}^n (x_j - \mu)(x_j - \mu)^T \\ &= \frac{1}{2n} \sum_{j=1}^n \sum_{k=1}^n (x_j - x_k)(x_j - x_k)^T. \end{aligned} \quad (3)$$

其中, x_j^i 表示属于类别 i 的样本 x_j ; n_i 表示属于类别 i 的样本个数; μ 和 μ^i 分别表示所有样本的均值和属于类别 i 的样本均值。最终, LDA 的变换矩阵 A 可以通过求解如下广义特征值问题的最小的特征值所对应的特征向量得到[1][2]:

$$S_w A = \lambda S_t A. \quad (4)$$

2.2. 局部 Fisher 判别分析(LFDA)

局部 Fisher 判别分析通过将样本的近邻结构引入到构建类内散度矩阵和整体散度矩阵中, 使得模型具有局部判别能力并能很好地对多模态分布下的数据进行有效降维。具体地, LFDA 首先以如下方式重新定义了 LDA 的类内散度矩阵(2)和整体散度矩阵(3):

$$\begin{aligned} \hat{S}_w &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{jk}^{(w)} (x_j - x_k)(x_j - x_k)^T \\ &= \sum_{j=1}^n \left(\sum_{k=1}^n W_{jk}^{(w)} \right) x_j x_k^T - \sum_{j=1}^n \sum_{k=1}^n W_{jk}^{(w)} x_j x_k^T \\ &= X D^{(w)} X^T - X W^{(w)} X^T = X L^{(w)} X^T, \end{aligned} \quad (5)$$

$$\begin{aligned}\hat{S}_t &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n W_{jk}^{(w)} (x_j - x_k)(x_j - x_k)^T \\ &= XD^{(t)}X^T - XW^{(t)}X^T = XL^{(t)}X^T.\end{aligned}\quad (6)$$

其中, $D^{(i)} \in \mathbb{R}^{n \times n}, i = \{w, t\}$ 表示以对称矩阵 $W^{(i)} \in \mathbb{R}^{n \times n}, i = \{w, t\}$ 的行和为对角元素构造的对角矩阵; 而权重矩阵 $W^{(i)} \in \mathbb{R}^{n \times n}, i = \{w, t\}$ 以如下形式定义:

$$W_{jk}^{(w)} = \begin{cases} W_{ij}/n_i, & \text{if } y_j = y_k = i \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$W_{jk}^{(t)} = \begin{cases} W_{ij}/n, & \text{if } y_j = y_k \\ 1/n, & \text{otherwise} \end{cases} \quad (8)$$

这里, y_j 表示样本 x_j 的类别标签; W 表示与样本近邻相关的仿射矩阵。一种简单的定义方式是: 当样本 x_j 与样本 x_k 互为近邻时, $W_{ij} = 1$, 否则 $W_{ij} = 0$ 。更多关于 W 的定义方式可以参考文献[3], 需要强调的是, 无论哪种定义方式, 样本近邻点的个数都是需要人为指定的。最后 LFDA 使用改进的类内散度矩阵 \hat{S}_w 和整体散度矩阵 \hat{S}_t 对应地替换 LDA 中的 S_w 和 S_t , 得到与 LDA 类似的模型:

$$\begin{aligned}\min_A \text{Tr}(A^T \hat{S}_w A) \\ \text{s.t. } A^T \hat{S}_t A = I_d\end{aligned}\quad (9)$$

从而以类似于公式(4)的方式求得最终的变换矩阵。

2.3. 自权重自适应局部判别分析(SALDA)

不同于 LFDA 使用人为指定样本近邻个数的方式来挖掘数据的局部结构, SALDA 尝试根据数据自身信息自适应地挖掘数据的局部流型结构来达到局部判别分析的目的。具体而言, 首先根据矩阵的迹运算和矩阵元素形式的 F-范数等价的性质, 即对于方阵 $X \in \mathbb{R}^{m \times n}$, $\text{Tr}(X^T X) = \|X\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n X_{ij}^2$, 以及公式(2), 易得:

$$\text{Tr}(A^T S_w A) = \frac{1}{2} \sum_{i=1}^c \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} \|A^T (x_j^i - x_k^i)\|_F^2, \quad (10)$$

这里 $\|X\|_F$ 表示矩阵 X 元素形式的 F-范数。因此, 单独将控制样本间差异的权重 W 提出来并控制其为正数, 则由公式(10)与传统 LDA 模型(1)便可得到 SALDA 模型[9]:

$$\begin{aligned}\min_A \sum_{i=1}^c \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} W_{jk}^i \|A^T (x_j^i - x_k^i)\|_F^2 \\ \text{s.t. } A^T S_t A = I_d, W_{jk}^i \geq 0,\end{aligned}\quad (11)$$

需要注意的是, 这里的权重 W 只是作为常数值单独进行表示而非模型的求解变量。参考 LFDA 中定义近邻结构的仿射矩阵发现, 其内部的每一个元素表示样本相关性的大小。换句话说, 当样本距离越近时, 对应在仿射矩阵中的值就越大。因此, 根据 F-范数与欧拉距离的相似性定义, SALDA 通过以下对同类别 i 内的样本对 x_j 和 x_k 的求欧拉距离的倒数形式定义 W 中的每一个元素值[13] [14]:

$$W_{jk}^i = \frac{1}{2 \|A^T (x_j^i - x_k^i)\|_F}, \quad (12)$$

当得到 W 时, 模型就可以重新表示成:

$$\begin{aligned}\min_A \text{Tr}(A^T XLX^T A) \\ \text{s.t. } A^T S_t A = I_d\end{aligned}\quad (13)$$

这里, 拉普拉斯矩阵 $L = D - W$, D 是由权重矩阵 W 得到的对角矩阵, 其对角线上的元素为 W 对应行中所有元素值之和, 即 $D_{ii} = \sum_{j=1}^n W_{ij}$ 。显然, 若记 $\tilde{S}_w = XLX^T$, 模型(13)与模型(1)具有类似的结构, 所以同样可以通过广义特征值分解的方法进行求解。

值得强调的是, SALDA 为了运用近邻思想并避开近邻个数的设置来挖掘数据的局部结构, 采用构建具有仿射矩阵性质的方式固定计算权重矩阵 W 。这种方式虽然在构建仿射矩阵方面似乎是有效的, 但是在整体模型上却不是很合理。一方面, 模型(11)中的待求解变量只有变换矩阵 A , 但模型却是以迭代计算(12)和求解子模型(13)的方式进行求解; 另一方面, W 并不是模型要求解的参数, 但又参与模型的更新求解, 这不仅模糊了模型的可解释性也导致模型的求解并不一定收敛。

2.4. 无参数的局部判别分析(Pf-LLDA)

受 SALDA 启发, 我们将权重矩阵 W 假定为一个需要求解的未知量, 得到如下 Pf-LLDA 模型:

$$\begin{aligned} \min_{(A,W)} \sum_{i=1}^c n_i \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} W_{ij}^2 \left\| A^T (x_j^i - x_k^i) \right\|_F^2 \\ \text{s.t. } A^T S_i A = I_d, W_{jk}^i \geq 0, \sum_{k=1}^{n_i} W_{jk}^i = \frac{n_i}{n}. \end{aligned} \quad (14)$$

相比于 SALDA 模型(11), Pf-LLDA 主要有以下几方面的不同:

- Pf-LLDA 模型将变换矩阵 A 与权重矩阵 W 共同作为模型待求解的未知量进行建模, 而 SALDA 只对变换矩阵 A 进行了建模;
- Pf-LLDA 通过控制同类别样本的权重和为定值, 理论地给出了 W 的自适应更新公式。相比于 SALDA 直接设定而无法保证收敛的方式, Pf-LLDA 关于 W 的更新不仅具有可解释性而且保证模型收敛;
- Pf-LLDA 通过设定不同类样本的权重和为 n^i/n 的方式对不同类别进行了区分使得模型适用于类别失衡问题从而更好地提升降维性能。

模型(14)可以通过交替方向的优化求解思路[12]迭代进行求解。下面具体给出了针对 W 和 A 的求解方式。

2.4.1. 固定 W , 更新 A

当权重矩阵固定, 即 W 为已知数时, 模型(14)转换成:

$$\begin{aligned} \min_A \sum_{i=1}^c n_i \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} W_{ij}^2 \left\| A^T (x_j^i - x_k^i) \right\|_F^2 \\ \text{s.t. } A^T S_i A = I_d \end{aligned} \quad (15)$$

与公式(10)的推导类似, 模型(15)可进一步转换成模型(13)的形式。因此以上问题(15)可以通过广义特征值的方法进行求解。

2.4.2. 固定 A , 更新 W

当变换矩阵固定, 即 A 为已知数时, 模型(14)转换成:

$$\begin{aligned} \min_W \sum_{i=1}^c n_i \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} W_{ij}^2 \left\| A^T (x_j^i - x_k^i) \right\|_F^2 \\ \text{s.t. } W_{jk}^i \geq 0, \sum_{k=1}^{n_i} W_{jk}^i = \frac{n_i}{n}. \end{aligned} \quad (16)$$

模型(16)可以通过分别求解它的一般化子模型来得到最终的 W 。具体而言, 对于任意确定的 i 和 j , 令向量 $u = W_j^i \in \mathbb{R}^{n_i \times 1}$, 其中 $W_j^i = [W_{j1}^i, \dots, W_{jn_i}^i]^T$ 。同时令对角矩阵 $V = \text{diag} \left(\left\| A^T (x_j^i - x_k^i) \right\|_F^2 \right), k = 1, \dots, n_i$,

即 V 的对角线元素为 $V_{kk} = \left\| A^T (x_j^i - x_k^i) \right\|_F^2$ 。所以模型(16)在固定 i 和 j 时, 可得到通用子模型:

$$\begin{aligned} & \min_u u^T V u \\ & \text{s.t. } u \geq 0, u^T \mathbf{1} = \frac{n_i}{n}. \end{aligned} \quad (17)$$

这里 $\mathbf{1} \in \mathbb{R}^{n_i \times 1}$ 表示长度为 n_i 的全 1 列向量。模型(17)可以通过拉格朗日乘子法[1]计算得出 u 的每一个元素, 具体如下:

$$u_k = \frac{n_i}{n \cdot V_{kk}} \left(\sum_{t=1}^{n_i} \frac{1}{V_{tt}} \right)^{-1}, k = 1, \dots, n_i. \quad (18)$$

所以, 模型(16)的最优解 W^* 满足:

$$W_{jk}^{i*} = \frac{n_i}{n \cdot \left\| A^T (x_j^i - x_k^i) \right\|_F^2} \left(\sum_{t=1}^{n_i} \frac{1}{\left\| A^T (x_j^i - x_k^i) \right\|_F^2} \right)^{-1}, i = 1, \dots, c; j = 1, \dots, n_i; k = 1, \dots, n_i. \quad (19)$$

有趣的是, Pf-LLDA 通过设置 W 为模型变量的方式求解出来的结果与 SALDA 分析仿射矩阵性质的方式固定 W 的结果具有类似的结构。但由于 SALDA 缺乏理论的推导, 同时更新公式(12)与(19)并非完全一样, 导致 SALDA 模型求解时并不一定收敛。最终求解模型(14)的算法流程如算法 1 所示, 其中初始化目标函数值 $o_0 = +\infty, o_1 = 1$ 保证了算法能正常进行循环迭代求解过程。当目标函数值在迭代更新前后的变化量小于精度误差 $\varepsilon = 10^{-6}$ 时, 循环终止, 并返回最终的变换矩阵 A 。

Algorithm 1. The algorithm of parameter-free local linear discriminant analysis

算法 1. 无参数的局部判别分析算法

输入:	来自 c 个类别的数据矩阵 $X = [x_1, \dots, x_n] \in \mathbb{R}^{m \times n}$, 低维空间的维度 d
输出:	最优的变换矩阵 $A^* \in \mathbb{R}^{m \times d}$
1.	初始化 $t=1$, 目标函数值 $o_0 = +\infty, o_1 = 1$; 初始化权重矩阵 $W_{jk}^{i(0)} = \frac{1}{n_i}, i = 1, \dots, c; j = 1, \dots, n_i; k = 1, \dots, n_i$; 以及精度误差 $\varepsilon = 10^{-6}$
2.	while $ o_t - o_{t-1} > \varepsilon$
3.	通过计算模型(15)更新变换矩阵 $A^{(t)}$
4.	通过公式(19)更新权重矩阵 $W^{(t)}$
5.	将 $A^{(t)}$ 和 $W^{(t)}$ 代入模型(14)的目标函数更新 o_t
6.	$t := t + 1$
7.	end while
8.	返回 $A^* = A^{(t)}$

3. 实验与分析

为了系统检验算法的有效性, 我们分别对比了本文提出的算法与其它相关算法在两个仿真数据集与一个真实数据集上的表现。

3.1. 仿真数据集上的实验

如图 1 所示, 每个数据集中有两个类, 每类包含 200 个二维的数据点。图 1a 中两个类别的数据由均

值为 $[-1, 0]$ 和 $[1, 0]$, 方差同为 $[0.1, 0; 0, 1]$ 的二维高斯分布函数生成。图 1b 中类别 I 的数据由均值为 $[-3, 0]$ 和 $[3, 0]$, 方差同为 $[0.5, 0; 0, 0.5]$ 的两个二维高斯分布函数生成; 类别 II 的数据由均值为 $[0, 0]$, 方差为 $[0.1, 0; 0, 1]$ 的二维高斯分布函数生成。显然, 图 1a 中的不同类别的数据点只由一个高斯分布产生, 即只存在单模态现象; 而图 1b 中的类别 I 的数据集是由两个高斯分布的数据点构成, 所以存在多模态现象。

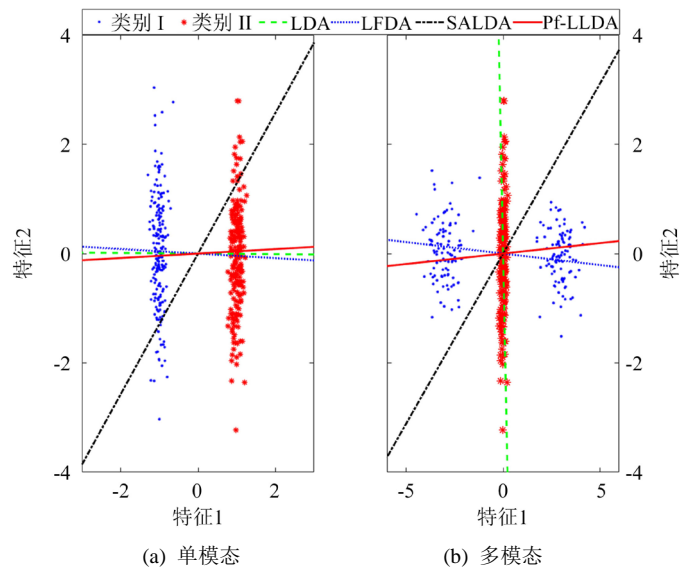


Figure 1. The projection of different method on simulation datasets
图 1. 不同方法在仿真数据集的投影方向

我们在这两个仿真数据上对比了线性判别分析(Linear Discriminant Analysis, LDA) [4]、局部 Fisher 判别分析(Local Fisher Discriminant Analysis, LFDA) [3]、自权重自适应局部判别分析(Self-weighted Adaptive Locality Discriminant Analysis, SALDA) [9]和我们提出的无参数的局部判别分析(Parameter-free Local Linear Discriminant Analysis, Pf-LLDA)方法。如图 1 所示, 每种方法的一维投影线已经被描绘出来。不难发现, 在单模态数据集上(图 1a), LDA、LFDA 和 Pf-LLDA 都能找到合理的投影方向。但在多模态数据集上(图 1b), LDA 却失败了。另一方面, SALDA 无论在何种模态的数据集上似乎都不具备判别能力。其主要的原因有: 1) 由 2.4 节的分析和 2.3 节中权重更新公式(12)可知, SALDA 通过简单对样本距离求倒数的方式很难保证模型收敛; 2) 分析 SALDA 的代码²实现发现, SALDA 主要是因为迭代了超过所设置的最大次数(1000 次)而跳出模型求解。这就导致了 SALDA 得到的投影结果更倾向于陷入随机的一个局部最优值。相比之下, 我们不仅通过理论证明了 Pf-LLDA 权重的更新方式实现了如 LFDA 的局部判别性分析, 同时在这两个仿真数据集上以算法 1 的实现方式都能保证在 100 次迭代内完成收敛。

3.2. 真实数据集上的实验

USUP³ 是一个手写数字图像数据集。它是模式识别领域研究所广泛使用的数据集之一, 扫描于美国邮局的邮件中, 具有很好的普适性。我们随机取了其中 5500 张手写数字图像作为我们的实验数据, 每张图像是由 16×16 像素点的灰度图组成。图 2 显示了该数据集中的部分样本。

需要说明的是, 为了避免降维算法中存在的过度降维[11]和小样本[15]问题, 与文献[13]类似, 我们首先对所有数据进行主成分分析[1] [2]处理, 将每个样本表示成 79 维的向量。然后, 将训练样本输入到

²<https://github.com/guomuhan/salda2/blob/master/ADLDA.m>

³<https://www.kaggle.com/bistaumanga/usps-dataset>

LDA 相关的算法中得到变换矩阵和低维表示的训练样本。接着, 根据变换矩阵对测试样本进行降维, 得到低维表示的测试样本。最后, 将低维表示的测试样本和训练样本作为 1 近邻算法[1] [9]的输入, 得到关于测试样本的分类准确率。需要说明的是, 为了结果的准确性, 每组实验都会独立重复 30 次并将最终准确率的平均值记录下来。



Figure 2. Part of samples in USUP dataset

图 2. USUP 数据集中的部分样本

我们在真实数据集对比了 LDA、NDA、FLDA、SALDA 和局部敏感判别分析[5] (Locality Sensitive Discriminant Analysis, LSDA)方法。并用控制变量法从训练样本个数、所降的维度、近邻个数和类别个数四个角度进行实验验证。每个实验角度的参数由表 1 给出。具体而言, 表 1 中关于“训练样本个数”角度的实验参数由第二列给出。此时, 样本最终的维度为 60; 通过引入近邻个数实现的局部线性判别分析算法的近邻个数参数为 5; 数据的类别个数为 5; 而来自每类的训练样本个数为*: 表示它是这组实验所要研究的变量。其他三个角度的实验参数设置同理。

Table 1. The detailed parameters for each experiment

表 1. 各个实验的具体参数

实验的角度 固定的参数	训练样本个数	所降的维度	近邻个数	类别个数
每类训练样本个数	*	20	20	20
所降的维度	60	*	60	60
近邻个数	5	5	*	5
类别个数	5	5	5	*

表 2 对比了当每类训练样本为 5、10、20、40、80、160 和 320 时, 各种方法在 USUP 数据集上的分类准确率。不难发现, 当所降维度、近邻个数和类别个数相同时, 所有方法的分类准确率会随着训练样本的增加而有所提升。相比之下, Pf-LLDA 总体上得到了更好的分类效果。特别是在训练样本数较少的情况下, Pf-LLDA 的优势更为明显。

Table 2. The classification accuracy with different number of training samples (%)

表 2. 不同训练样本个数下的分类准确率(%)

方法 每类的 训练样本个数	LDA	NDA	FLDA	LSDA	SALDA	Pf-LLDA
5	27.36	59.55	22.62	54.82	51.11	61.73
10	29.85	64.82	21.96	53.30	70.60	76.21

Continued

20	62.20	63.45	61.97	47.48	64.38	88.98
40	85.40	86.73	86.28	73.89	63.56	90.52
80	90.29	91.94	90.90	87.65	73.94	92.74
160	91.91	94.66	93.52	93.95	81.58	94.27
320	92.89	95.67	95.52	95.58	86.69	95.80

表 3 对比了当训练样本数、近邻个数和类别个数相同时, 不同方法将样本从高维降到不同低维度时的准确率。需要注意的是, 每个样本经过主成分分析表示成 79 维的向量, 所以所降的维度最高被设置为 78。对比发现, 在所降的维度较低时, Pf-LLDA 相比于其他方法还是可以接受的。并且随着所降的维度的升高, Pf-LLDA 实现了更高的准确率上限。

Table 3. The classification accuracy with different reduced dimensions (%)

表 3. 不同所降维度下的分类准确率(%)

方法 所降的维度	LDA	NDA	FLDA	LSDA	SALDA	Pf-LLDA
4	85.62	77.28	84.26	32.63	67.96	71.60
8	85.57	84.06	87.39	50.00	62.53	77.05
16	85.07	86.59	86.86	74.82	66.94	83.45
32	85.69	87.76	87.27	77.92	70.81	88.95
64	85.92	86.93	86.17	72.80	64.40	90.47
70	85.47	86.43	86.06	70.36	63.03	90.28
78	85.46	85.94	85.47	69.28	66.02	90.40

图 3 描绘了对于不同的局部判别分析方法, 近邻个数对分类准确率的影响。这里我们只展示了近邻个数从 1 到 39 时的结果。可以看到, 不同的近邻个数对 NDA、LFDA 和 LSDA 的分类效果影响很大。特别地, LSDA 的分类准确率由近邻个数为 3 时的 73.73% 降到近邻个数为 37 时的 60.07%。相比之下, SALDA 通过自权重的方式避开对近邻个数的依赖从而实现了稳定的结果输出。但与 2.4 节中的分析一致, SALDA 陷入随机的一个局部最优值导致其分类准确率并不理想。Pf-LLDA 克服了 SALDA 的缺点, 不仅实现了稳定的结果, 同时也得到了较高的准确率。

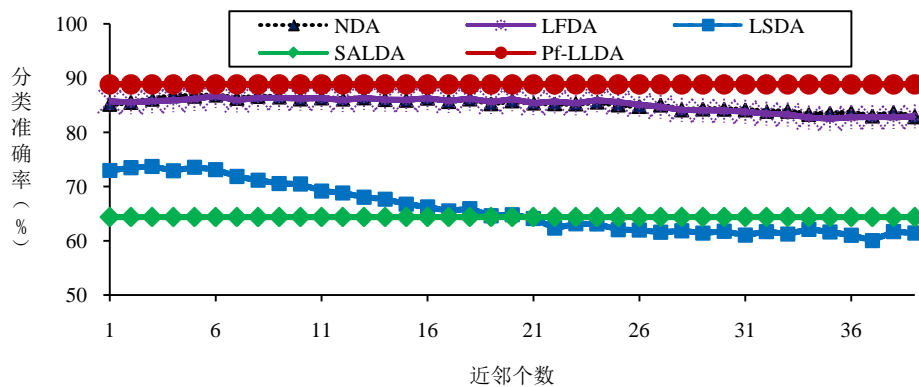


Figure 3. The influence of the number of nearest neighbors on classification accuracy

图 3. 近邻个数对分类准确率的影响

最后, 我们验证了当类别个数从 4 增加到 10 时, 不同方法的分类结果。如表 4 所示, 相比于其他四种局部判别分析方法和作为基线的 LDA 方法, Pf-LLDA 得到了显著的提升。这不仅依赖于 Pf-LLDA 利用了数据的局部结构, 也依赖于我们所引入的无参方式自适应地挖掘出数据局部几何结构的差异性。

Table 4. The classification accuracy with different number of classes (%)

表 4. 不同类别个数下的分类准确率(%)

方法 类别个数	LDA	NDA	FLDA	LSDA	SALDA	Pf-LLDA
4	43.24	53.39	43.17	34.82	75.43	89.45
5	60.86	64.38	61.08	47.87	65.99	88.86
6	68.86	72.90	70.44	53.47	65.01	85.81
7	74.21	76.92	75.44	56.13	64.76	85.95
8	72.85	75.13	74.27	54.90	60.09	83.15
9	72.94	73.88	73.84	54.23	58.81	80.93
10	75.33	75.21	75.89	56.95	60.52	81.07

4. 总结

本文提出了一种无参数的局部判别分析方法。该方法通过利用已知的数据信息, 自适应地学习到了数据的低维局部结构。系统的实验结果表明, 该方法在无需人为设置近邻个数的情况下, 不仅挖掘出了数据的局部结构还有效的提高了分类准确率。未来将在模型中引入非线性或张量的思想, 使模型适用于挖掘更复杂数据的数据结构。

参考文献

- [1] Fukunaga, K. (2013) Introduction to Statistical Pattern Recognition. Elsevier, New York.
- [2] Farzana, A., Sadaoui, S. and Selim, B. (2021) Conceptual and Empirical Comparison of Dimensionality Reduction Algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, **40**, Article ID: 100378. <https://doi.org/10.1016/j.cosrev.2021.100378>
- [3] Sugiyama, M. (2007) Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis. *Journal of Machine Learning Research*, **8**, 1027-1061.
- [4] Cao, G.Q., Iosifidis, A. and Gabbouj, M. (2017) Multi-View Nonparametric Discriminant Analysis for Image Retrieval and Recognition. *IEEE Signal Processing Letters*, **24**, 1537-1541. <https://doi.org/10.1109/LSP.2017.2748392>
- [5] Yu, H.Y., Gao, L.R., Li, W., et al. (2017) Locality Sensitive Discriminant Analysis for Group Sparse Representation-Based Hyperspectral Imagery Classification. *IEEE Geoscience and Remote Sensing Letters*, **14**, 1358-1362. <https://doi.org/10.1109/LGRS.2017.2712200>
- [6] Yan, S.C., Xu, D., Zhang, B.Y., et al. (2006) Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 40-51. <https://doi.org/10.1109/TPAMI.2007.250598>
- [7] 谢钧, 刘剑. 一种新的局部判别投影方法[J]. 计算机学报, 2011, 34(11): 2243-2250.
- [8] Nie, F.P., Xiang, S.M. and Zhang, C.S. (2007) Neighborhood Minmax Projections. In: *International Joint Conferences on Artificial Intelligence*, AAAI Press, Hyderabad, 993-998.
- [9] Guo, M.H., Nie, F.P. and Li, X.L. (2018) Self-Weighted Adaptive Locality Discriminant Analysis. In: *International Conference on Image Processing*, IEEE Press, Athens, 3378-3382. <https://doi.org/10.1109/ICIP.2018.8451023>
- [10] Shi, Z.H., Wu, D.R., Huang, J., et al. (2020) Supervised Discriminative Sparse PCA with Adaptive Neighbors for Dimensionality Reduction. In: *International Joint Conference on Neural Networks*, IEEE Press, Glasgow, 1-8. <https://doi.org/10.1109/IJCNN48605.2020.9206927>

- [11] Wan, H., Guo, G.D., Wang, H., *et al.* (2015) A New Linear Discriminant Analysis Method to Address the Over-Reducing Problem. In: *International Conference on Pattern Recognition and Machine Intelligence*, Springer Press, Warsaw, 65-72. https://doi.org/10.1007/978-3-319-19941-2_7
- [12] He, B.S. and Yuan, X.M. (2012) On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method. *SIAM Journal on Numerical Analysis*, **50**, 700-709. <https://doi.org/10.1137/110836936>
- [13] Nie, F.P., Yuan, J.J. and Huang, H. (2014) Optimal Mean Robust Principal Component Analysis. In: *International Conference on Machine Learning*, ACM Press, Beijing, 1062-1070.
- [14] Jiang, W.H., Nie, F.P. and Huang, H. (2015) Robust Dictionary Learning with Capped l_1 -Norm. In: *International Joint Conference on Artificial Intelligence*, AAAI Press, Buenos Aires, 3590-3596.
- [15] Alok, S. and Kuldip, P.K. (2015) Linear Discriminant Analysis for the Small Sample Size Problem: An Overview. *International Journal of Machine Learning and Cybernetics*, **6**, 443-454. <https://doi.org/10.1007/s13042-013-0226-9>