

利用潜在稀疏表示学习的增强局部保持投影方法^①



彭 帅^{1,2}, 胡良臣^{1,2}

¹(安徽师范大学 计算机与信息学院, 芜湖 241003)

²(工业智能数据安全安徽省重点实验室, 芜湖 241002)

通信作者: 胡良臣, E-mail: csls.hu@ahnu.edu.cn

摘 要: 降维在机器学习和模式识别领域中起着至关重要的作用. 目前, 现有的基于投影的方法往往只单一地利用了数据之间的距离信息或表示关系来保持数据的结构, 难以有效捕捉高维空间中数据流形的非线性特征和复杂相关性. 为了解决这个问题, 本文提出了一种利用潜在稀疏表示学习的增强局部保持投影 (enhanced locality preserving projection with latent sparse representation learning, LPP_SRL) 方法. 所提出方法不仅利用距离信息以保留数据的局部结构, 而且利用多重局部线性表示来揭示数据的全局非线性结构. 此外, 为了在投影学习和稀疏自表示之间建立联系, 本文采用了一种新策略, 将稀疏自表示中的字典替换为低维表示的重构样本. 通过这种方法, 能够有效地过滤掉不相关的特征和噪声, 从而更好地保留原始特征空间中的主要成分. 在多个公开可用的基准数据集上进行的大量实验证明了所提出方法的有效性和优越性.

关键词: 降维; 投影学习; 稀疏表示; 主成分; 图像分类

引用格式: 彭帅, 胡良臣. 利用潜在稀疏表示学习的增强局部保持投影方法. 计算机系统应用, 2024, 33(9): 14-27. <http://www.c-s-a.org.cn/1003-3254/9609.html>

Enhanced Locality Preserving Projection with Latent Sparse Representation Learning

PENG Shuai^{1,2}, HU Liang-Chen^{1,2}

¹(School of Computer and Information, Anhui Normal University, Wuhu 241003, China)

²(Key Laboratory of Industrial Intelligent Data Security of Anhui Province, Wuhu 241002, China)

Abstract: Dimensionality reduction plays a crucial role in machine learning and pattern recognition. The existing projection-based methods tend to solely utilize distance information or representation relationships among data points to maintain the data structure, which makes it difficult to effectively capture the nonlinear features and complex correlations of data manifolds in high-dimensional space. To address this issue, this study proposes a method: enhanced locality preserving projection with latent sparse representation learning (LPP_SRL). The method not only utilizes distance information to preserve the local structure of the data but also leverages multiple local linear representations to unveil the global nonlinear structure of the data. Moreover, to establish a connection between projection learning and sparse self-representation, this study employs a novel strategy by replacing the dictionary in sparse self-representation with reconstructed samples from the low-dimensional representation. This approach effectively filters out irrelevant features and noise, thereby better preserving the principal components in the original feature space. Extensive experiments conducted on multiple publicly available benchmark datasets have demonstrated the effectiveness and superiority of the proposed method.

Key words: dimensionality reduction; projection learning; sparse representation; principal component; image classification

① 基金项目: 安徽省高校自然科学基金项目 (2022AH040028); 国家自然科学基金青年项目 (62306010); 安徽师范大学高峰和奖补学科建设项目 (2023 GFKK171)

收稿时间: 2024-03-02; 修改时间: 2024-04-01; 采用时间: 2024-04-19; csa 在线出版时间: 2024-07-24

CNKI 网络首发时间: 2024-07-25

1 引言

高维数据由于高维空间中样本之间距离的稀疏性,导致数据稀疏和冗余,增加了计算和存储成本,因此在处理和分析时面临挑战.作为常用的降维算法之一,投影学习在高维数据处理和分析中具有重要意义^[1-3].通过保留数据的基本信息并将其映射到低维空间,投影学习可以显著减少数据的冗余和复杂性,同时保持数据点之间的关键关系.这简化了问题求解,提高了计算效率,并促进了数据可视化和理解^[4].

降维算法的演进经历了从传统线性方法,如主成分分析 (principal component analysis, PCA)^[5] 和线性判别分析 (linear discriminant analysis, LDA)^[6], 到基于流形学习的非线性方法的过程.在这个过程中,出现了一系列基于流形学习的降维算法,包括等距映射 (isometric mapping, ISOMAP)^[7]、局部线性嵌入 (locally linear embedding, LLE)^[8] 和拉普拉斯特征映射 (Laplacian eigenmap, LE)^[9] 等.这些方法致力于将高维数据映射到低维流形空间,以便更好地捕捉数据中的非线性特征和结构.然而,尽管非线性流形学习方法在处理复杂数据结构方面表现出色,但在样本外推和泛化方面仍然存在挑战.在一些应用中,线性投影技术表现突出成为最常用和稳定的方法.因此,经典的局部保持投影 (locality preserving projection, LPP)^[10] 作为一种线性降维方法,利用 LE 在原始和嵌入空间之间建立线性映射.同时,通过保持样本之间的局部邻域关系, LPP 可以更好地揭示数据的局部结构.随着研究的深入,逐渐出现了许多将其他优秀的数据挖掘工具 (如低秩或稀疏属性) 与投影学习相结合的投影学习方法^[11-14].

众所周知,类似于 LPP 的方法在执行之前需要预定义一个图来描述数据样本之间的局部关联.显然,在稀疏高维空间中,随着欧氏度量逐渐失效,基于任何假设的任何预定义图都将是徒劳的.近年来,提出了许多同时优化图和投影矩阵的方法^[15-18].在类内样本之间的亲和度满足二次条件的前提下, Li 等人^[19] 研究了全连接的类内图,并提出了局部自适应判别分析 (locality adaptive discriminant analysis, LADA), 可以很好地揭示潜在的数据流形结构. Pang 等人^[20] 提出了同时学习邻居和投影矩阵的方法 (simultaneously learning neighborhood and projection matrix, SLNP), 使相似度和邻居成为低维子空间中的变量.然后通过最小化统一的目标函

数获得最佳相似度和投影矩阵. Nie 等人^[21] 提出了自适应局部线性判别分析 (adaptive local linear discriminant analysis, ALLDA), 它从数据本身自适应地学习了一个 k 近邻图, 以提取数据的局部连通性.但在特定情况下, 伴随投影学习的持续优化图可能不会超越预定义图的性能, 主要原因有 3 点: (1) 特定假设可能导致学习到的亲和度分布与真实数据分布之间存在差异; (2) 当数据在高维空间中稀疏分布时, 仅依赖数据距离进行相似性估计可能无法准确表示数据点之间的真实局部关系; (3) 亲和度的初始估计误差可能导致投影中的方向偏差, 随后在学习过程中引起累积误差.

在流形学习的背景下, 基本假设是数据分布在一个低维流形上, 这意味着该流形的局部区域可以同胚于某个欧几里得空间 (通常是低维空间). 在这些局部区域中, 数据在流形上呈近似线性, 允许在欧几里得空间中使用线性关系对其进行近似. LLE 是一个经典的为此目的设计的流形学习算法, 而邻域保持嵌入 (neighborhood preserving embedding, NPE)^[22] 是一个具有线性投影学习的扩展版本. LLE 和 NPE 使每个样本可以通过其相邻样本的线性组合在低维空间中近似重构, 同时保持数据点之间的局部关系. 然而, 在环境空间 (ambient space) 中选择 k 个最近邻可能并不总是理想的. 为此, 稀疏保持投影 (sparsity preserving projection, SPP)^[23] 利用稀疏编码学习原始数据的稀疏表示, 而表示系数直接描述了数据点之间的局部关系, 实现了有效的特征提取和冗余信息的减少. 考虑到 NPE 和 SPP 在显式判别时的局限性和对损坏样本的脆弱性, 低秩保持嵌入 (low-rank preserving embedding, LRPE)^[24] 联合获得所有数据的自表示, 从而能够提取数据集的整体全局结构. 一般来说, 表示学习和投影学习通常被认为是两个不相关的学习过程, 这限制了实现最佳性能的潜力. 因此, 低秩嵌入 (low-rank embedding, LRE)^[25] 通过使用鲁棒的低秩表示 (low-rank representation, LRR) 来解决对损坏数据的无监督鲁棒线性降维问题, 同时寻找最佳的 LRR 和最佳子空间. 此外, 扩展近似低秩投影学习 (extended approximate low-rank projection learning, EALPL)^[26] 和潜在低秩和稀疏嵌入 (latent low-rank and sparse embedding, LLRSE)^[27] 与 LRPE 和 LRE 把原始数据用作字典不同, 而是使用投影样本作为字典. 更准确地说, 对于低秩表示学习, 字典是通过利用投影后的样本重构形成的, 从而使噪声效应得以

抑制. 近年来, 人们对稀疏和低秩表示揭示数据固有局部结构的能力越来越感兴趣^[28-30].

以上的介绍涵盖了2种代表性的技术. 其中一种源于LPP方法, 从预定义的亲和度图演变为动态亲和度图. 另一种是基于NPE方法构建的, 从 k 个最近邻线性表示演化为稀疏或低秩表示. 现有的方法往往只沿着其中一条思路开展研究, 难以有效捕捉高维空间中数据的复杂结构. 在本文的工作中, 我们结合了以上2种思路, 创新地提出了一种监督降维方法, 即LPP_SRL, 它不仅考虑了在欧几里得度量下保持邻域关系, 还结合了稀疏表示, 以促进主成分的提取, 并在环境空间中过滤噪声和冗余特征. 总之, 本文的研究具有以下主要贡献: (1) 本文将基于欧几里得距离的邻域保持和局部稀疏自表示两个模块自适应地纳入一个统一的框架, 丰富了对数据局部拓扑结构的描述, 并克服了传统方法的潜在缺点; (2) 在稀疏自表示模块中, 本文选择重构样本作为字典, 并建立了与投影学习的联合学习模式, 使模型能够在投影学习过程中捕获数据的主要成分, 消除冗余特征, 并消除噪声; (3) 为了优化所提出模型, 我们开发了一个高效的迭代算法, 并在不同场景下对各种公开可用数据集进行了实验, 证明了所提出的算法在降维方面表现出色.

本文第2节介绍本文写作中的一些符号, 并简要概述了该领域的相关方法. 接下来, 在第3节中, 本文提出一种创新的模型, 讨论了优化过程, 并分析了其计算复杂性. 实验设置和结果在第4节详细说明. 最后, 在第5节中, 我们总结了全文.

2 相关工作

2.1 符号与定义

假设有一个数据矩阵 $X = [X_1, X_2, \dots, X_c] = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{d \times n}$, 由 d 维空间中的 n 个样本组成, 其中每个 X_i 表示属于第 i 个类别的子矩阵. 在降维任务中, 我们的目标是学习一个投影矩阵 $Q \in \mathbb{R}^{m \times d}$, 有效地将原始样本 $x_i \in \mathbb{R}^{d \times 1}$ 投影到一个 m 维子空间中, 旨在提取最显著的特征, 如下所示:

$$y_i = Qx_i \quad (1)$$

其中, $y_i \in \mathbb{R}^{m \times 1}$ 表示与 x_i 对应的低维向量表示. 此外, 对于向量 $x = [x_1, x_2, \dots, x_n]$, ℓ_2 范数被定义为 $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$. 对于矩阵 $X \in \mathbb{R}^{d \times n}$, 我们将 ℓ_1 范数、 $\ell_{2,1}$ 范

数和Frobenius范数分别定义为 $\|X\|_1 = \sum_{j=1}^n \sum_{i=1}^d |x_{ij}|$ 、 $\|X\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^d x_{ij}^2}$ 和 $\|X\|_F^2 = \sum_{j=1}^n \sum_{i=1}^d x_{ij}^2$.

2.2 局部保持投影

LPP是一种线性降维方法, 旨在保持高维数据的局部邻域, 并确保低维空间中的数据样本投影与其在高维空间中的原始局部关系紧密相符. 一般来说, 其目标函数可以表示为:

$$\min_Q \frac{1}{2} \sum_{i,j=1}^n \|Qx_i - Qx_j\|_2^2 w_{ij} \quad (2)$$

其中, $w_{ij} = \exp(-\|x_i - x_j\|_2^2 / t)$ 表示样本 x_i 和 x_j 之间的亲和度, 它们彼此是 k -最近邻的(否则, $w_{ij} = 0$), t 是一个可调节参数. 通过简单的数学推导, 可以得到以下等式:

$$\frac{1}{2} \sum_{i,j=1}^n \|Qx_i - Qx_j\|_2^2 w_{ij} = \text{tr}(QXLX^T Q^T) \quad (3)$$

其中, $L = D - W$ 是拉普拉斯矩阵, D 是度矩阵, 其对角线元素为 $d_{ii} = \sum_{j=1}^n w_{ij}$, W 是亲和矩阵, 其第 (i, j) 个元素表示为 w_{ij} . 因此, 通过对式(3)加上约束, 可以将其转化为式(4):

$$\min_Q \text{tr}(QXLX^T Q^T) \quad \text{s.t. } QXD^T Q^T = I \quad (4)$$

利用拉格朗日乘法, 约束问题(4)等价于以下广义特征值分解问题:

$$XLX^T q^T = \alpha XDX^T q^T \quad (5)$$

其中, q 是对应于特征值 α 的特征向量, 并且作为行向量组成投影矩阵 Q . LPP中的最优投影由与 m 个最小广义特征值对应的 m 个特征向量组成.

2.3 稀疏表示学习

稀疏表示^[31-33]是信号处理和机器学习领域常用的技术之一, 旨在将信号或数据表示为尽可能少的非零元素的线性组合, 有助于从信号中提取关键特征, 降低数据维度, 并展现出了优秀的噪声鲁棒性. 假设要表示的信号是一个向量 $x \in \mathbb{R}^d$, 字典由矩阵 $D \in \mathbb{R}^{d \times n}$ 表示, 其中 d 是信号维度, n 是字典中原子(基)的数量. 则稀疏表示的数学模型可以描述为:

$$\min_w \|w\|_0 \quad \text{s.t. } x = Dw \quad (6)$$

其中, $w \in \mathbb{R}^n$ 是稀疏表示系数向量, $\|w\|_0$ 表示 w 的 ℓ_0 范数, 即非零元素的数量. 目标是找到一个稀疏系数向

量 w ,使得信号 x 可以表示为字典 D 中原子的线性组合,同时最小化 w 的 ℓ_0 范数,即尽可能少地使用基来表示信号。

然而,问题(6)是一个NP-hard问题,因为计算 ℓ_0 范数是一个非凸优化问题.为了解决这个问题,通常使用 ℓ_1 范数作为 ℓ_0 范数的近似,得到以下近似问题:

$$\min_w \|w\|_1 \quad \text{s.t. } x = Dw \quad (7)$$

其中, $\|w\|_1$ 表示 w 的 ℓ_1 范数,即其元素的绝对值之和。 ℓ_1 范数问题是一个凸优化问题,可以用各种优化方法求解.在实践中, ℓ_1 范数提供了良好的稀疏性,同时计算效率更高.此外,在实际场景中,我们要考虑到真实数据通常存在噪声.因此,模型(7)中的条件可以放宽为 $x = Dw + e$,其中 e 表示噪声向量.然后,模型(7)可以进一步写为:

$$\min_w \|w\|_1 \quad \text{s.t. } \|x - Dw\| \leq \epsilon \quad (8)$$

其中, ϵ 是误差容限或误差阈值,表示原始信号和利用稀疏系数向量 w 和字典 D 复原的信号之间的最大允许误差或偏差。

2.4 稀疏保持投影

SPP是LLE和NPE的一种变体,专注于在将高维数据投影到低维空间时保持数据的稀疏性.LLE、NPE和SPP使用一个共同目标函数,可以表达为:

$$\min_Y \sum_{i=1}^n \|y_i - \sum_{j=1}^n w_{ij} y_j\|_2^2 \quad \text{s.t. } YY^T = I \quad (9)$$

其中, $Y = [y_1, y_2, \dots, y_n] \in \mathbb{R}^{m \times n}$ 是嵌入数据矩阵,第 i 列表示嵌入数据样本 y_i .对于每个数据样本 x_i ,LLE和NPE计算其邻域中每个样本的局部重构权重 w_{ij} ,使得以下重构误差最小化:

$$\min_W \sum_{i=1}^n \|x_i - \sum_{j=1}^n w_{ij} x_j\|_2^2 \quad \text{s.t. } \sum_j w_{ij} = 1 \quad (10)$$

其中, $W = \{w_{ij}\}_{i,j=1}^n \in \mathbb{R}^{n \times n}$ 是描述原始数据的邻接属性的权重矩阵.与LLE和NPE不同,SPP的权重矩阵 W 是通过考虑每个样本作为其他样本的稀疏线性组合而构建的,并且可以一次完成,即:

$$\min_{w_i} \|w_i\|_1 \quad \text{s.t. } \|x_i - Xw_i\|_2^2 \leq \epsilon, \sum_j w_{ij} = 1 \quad (11)$$

其中,字典选择为 X 自身.在使用SPP的情况下,无需预定义每个样本的邻域,而是由稀疏表示学习引导的

自适应机制,确定哪些样本是样本 x_i 的邻居。

3 利用潜在稀疏表示学习的增强局部保持投影方法

考虑到现有方法的优点和局限性,在本节中,我们将逐步完善所提出的模型LPP_SRL,并设计了一个高效的迭代算法来优化它.此外,我们还分析了所提出模型的求解算法的计算复杂度。

3.1 模型建立

如第1节所述,在投影学习过程中学习到的图在许多情景下表现出了良好的性能,但原始空间中基于欧氏距离定义的图也确实包含了对后续分类任务有益的某些特征信息.因此,所提出的模型继承了原始LPP如下:

$$\min_Q \frac{1}{2} \sum_{i,j=1}^n \|Qx_i - Qx_j\|_2^2 w_{ij} \quad (12)$$

此外,考虑到特定亲度假设的适用性可能不够灵活,我们只利用环境空间中样本的 k 近邻连接信息.具体来说,亲和度矩阵 W 定义如下:

$$w_{ij} = \begin{cases} 1/k, & x_i \in N_k(x_j) \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

其中, $x_i \in N_k(x_j)$ 表示样本 x_i 是 x_j 同一类中的 k 个最近邻之一。

此外,我们认为稀疏局部表示能揭示数据样本之间的内在连接.因此,得到以下目标函数:

$$\min_{z_i} \|z_i\|_1 + \lambda \|e_i\|_1 \quad \text{s.t. } x_i = Xz_i + e_i \quad (14)$$

其中, z_i 表示第 i 个样本的表示系数,而 e_i 可能包含一些错误或噪声.令 $Z = [z_1, z_2, \dots, z_n]$ 和 $E = [e_1, e_2, \dots, e_n]$ 可以以矩阵形式表示为:

$$\min_Z \|Z\|_1 + \lambda \|E\|_1 \quad \text{s.t. } X = XZ + E \quad (15)$$

事实上,式(15)不能直接与投影学习过程集成,除非利用表示系数以图的形式描述数据样本之间的连接.此外,从现实场景中收集到的数据 X 不可避免地包含噪声,这意味着直接使用 X 作为字典来学习表示系数可能会受到负面影响.为了解决这些问题,我们假设投影空间具有一组正交基向量 $\{p_1, p_2, \dots, p_m\}$ (也可以称为原始数据的主成分),使用这些基向量上的投影坐标 $y_i = [y_{1i}, y_{2i}, \dots, y_{mi}]^T$ 重构样本 x_i ,如下所示:

$$x_i = p_1 y_{1i} + p_2 y_{2i} + \cdots + p_m y_{mi} + r_i \quad (16)$$

其中, 残差向量 r_i 可能包含一些低级特征或噪声. 令 $y_i = Qx_i$, $P = [p_1, p_2, \cdots, p_m]$, $R = [r_1, r_2, \cdots, r_m]$, 式 (16) 可以以矩阵形式表示为:

$$X = PQX + R \quad (17)$$

然后, 将式 (17) 替换进方程 $X = XZ + E$ 的右侧, 得到:

$$X = PQXZ + RZ + E \quad (18)$$

其中, RZ 作为重新组装的残差向量, 可以与 E 合并成为一个矩阵. 结合式 (15) 和式 (18), 得到新的稀疏表示模型如下:

$$\min_Z \|Z\|_1 + \lambda \|E\|_{2,1} \quad \text{s.t. } X = PQXZ + E \quad (19)$$

其中, 用 PQX 替换 X 作为字典可以揭示数据样本之间更本质的连接, 残差矩阵 E 合并了 RZ 和原始的 E , 范数 $\|\cdot\|_{2,1}$ 用于过滤掉特定样本的损坏和离群值.

综合考虑式 (12) 和式 (19) 可以实现相互促进学习的目的, 导出最终目标函数如下:

$$\begin{cases} \min_{P,Q,Z,E} \frac{1}{4} \sum_{i,j=1}^n \|Qx_i - Qx_j\|_2^2 w_{ij} + \lambda_1 \|Z\|_1 \\ + \lambda_2 \|E\|_{2,1} + \frac{\lambda_3}{2} \|Q\|_F^2 \\ \text{s.t. } X = PQXZ + E, P^T P = I \end{cases} \quad (20)$$

其中, λ_1 、 λ_2 和 λ_3 是 3 个正则化参数, 约束 $P^T P = I$ 使模型在求解过程中避免平凡解.

3.2 模型优化

在本节中, 我们将通过采用交替方向乘子法 (alternating direction method of multipliers, ADMM) 框架^[34] 解决式 (20) 中的优化问题. 关于 ADMM 在非凸问题上的详细收敛分析可以在文献 [35] 中找到. 为了简化优化过程, 我们首先引入一个辅助变量 B 和一个约束 $Z = B$, 使式 (20) 可分离. 因此, 通过放松原始问题, 问题 (20) 可以重新表述如下:

$$\begin{cases} \min_{P,Q,Z,E,B} \frac{1}{2} \text{tr}(QXL_W X^T Q^T) + \lambda_1 \|B\|_1 + \lambda_2 \|E\|_{2,1} + \frac{\lambda_3}{2} \|Q\|_F^2 \\ \text{s.t. } X = PQXZ - E, P^T P = I, Z = B \end{cases} \quad (21)$$

其中, 拉普拉斯矩阵 L_W 定义为 $L_W = D - (W + W^T)/2$, $D \in \mathbb{R}^{n \times n}$ 是对角矩阵, 其第 ii 项为 $\sum_j (w_{ij} + w_{ji})/2$.

紧接着, 问题 (21) 的增广拉格朗日函数定义如下:

$$\begin{cases} \mathcal{L}(P, Q, Z, E, B) \\ = \frac{1}{2} \text{tr}(QXL_W X^T Q^T) + \lambda_1 \|B\|_1 + \lambda_2 \|E\|_{2,1} \\ + \frac{\lambda_3}{2} \|Q\|_F^2 + \langle C_1, X - PQXZ + E \rangle + \langle C_2, Z - B \rangle \\ + \frac{\mu}{2} (\|X - PQXZ - E\|_F^2 + \|Z - B\|_F^2) \\ \text{s.t. } P^T P = I \end{cases} \quad (22)$$

其中, $\langle A, B \rangle = \text{tr}(A^T B)$, μ 是惩罚参数, C_1 和 C_2 是两个拉格朗日乘子.

通过使用一些数学技巧, 问题 (22) 可以重新表述为:

$$\begin{cases} \mathcal{L}(P, Q, Z, E, B) \\ = \frac{1}{2} \text{tr}(QXL_W X^T Q^T) + \lambda_1 \|B\|_1 + \lambda_2 \|E\|_{2,1} + \frac{\lambda_3}{2} \|Q\|_F^2 \\ + \frac{\mu}{2} \left(\|X - PQXZ - E + \frac{C_1}{\mu}\|_F^2 + \|Z - B + \frac{C_2}{\mu}\|_F^2 \right) \\ - \frac{1}{2\mu} (\|C_1\|_F^2 + \|C_2\|_F^2) \\ \text{s.t. } P^T P = I \end{cases} \quad (23)$$

为了解决问题 (23), 需要交替更新变量 P 、 Q 、 Z 、 E 和 B .

• 更新 Z : 给定 B 、 Q 、 P 和 E 值, 关于 Z 的子问题如下:

$$\min_Z \|X - PQXZ - E + \frac{C_1}{\mu}\|_F^2 + \|Z - B + \frac{C_2}{\mu}\|_F^2 \quad (24)$$

定义 $F = X - E + C_1/\mu$, $T = B - C_2/\mu$ 并且让 $\partial \mathcal{L}(Z)/\partial Z = 0$, 我们有:

$$(X^T Q^T QX + I)Z = X^T Q^T P^T F + T \quad (25)$$

然后, 从式 (25) 计算出:

$$Z = (X^T Q^T QX + I)^{-1} (X^T Q^T P^T F + T) \quad (26)$$

• 更新 B : 给定固定的 Z , 关于 B 的子问题如下:

$$\min_B \lambda_1 \|B\|_1 + \frac{\mu}{2} \|B - J\|_F^2 \quad (27)$$

其中, $J = Z + C_2/\mu$.

B 可以通过以下方式获得:

$$B = \Omega_{\frac{\lambda_1}{\mu}}(J) \quad (28)$$

其中, Ω 是标量缩放算子^[36]. 具体来说, B 中的每个元素应为:

$$B_{ij} = \text{sgn}(J_{ij}) \cdot \max\left(|J_{ij}| - \frac{\lambda_1}{\mu}, 0\right) \quad (29)$$

其中, sgn 是用于确定实数的符号 (正或负) 的逻辑函数, $\max(\alpha, \beta)$ 是一个取 α 和 β 之间最大值的函数。

- 更新 Q : 给定 P 、 Z 和 E , 关于 Q 的问题变为:

$$\min_Q \frac{1}{2} \text{tr}(QXLWX^TQ^T) + \frac{\lambda_3}{2} \|Q\|_F^2 + \frac{\mu}{2} \|X - PQXZ - E + \frac{C_1}{\mu}\|_F^2 \quad (30)$$

对 $\mathcal{L}(Q)$ 关于 Q 的导函数取零, 即 $\partial\mathcal{L}(Q)/\partial Q = 0$, 我们有:

$$Q(\lambda_3 I + XLWX^T + \mu XZZ^T X^T) = \mu P^T FZ^T X^T \quad (31)$$

其中, $F = X - E + C_1/\mu$, 然后, 由式 (31) 可计算得到 Q :

$$Q = \mu P^T FZ^T X^T (\lambda_3 I + XLWX^T + \mu XZZ^T X^T)^{-1} \quad (32)$$

- 更新 P : 给定 Q 、 Z 和 E , 并删除与 P 无关的项, 关于 P 的子问题如下:

$$\min_P \|F - PQXZ\|_F^2 \quad \text{s.t. } P^T P = I \quad (33)$$

问题 (33) 是一个正交普鲁克问题 (orthogonal Procrustes problem, OPP)^[37], 其解可以直接从 thin SVD 中获得, 即:

$$FZ^T X^T Q^T = U\Sigma V^T \quad (34)$$

然后, 我们得到:

$$P = UV^T \quad (35)$$

其中, U 和 V 分别是 $FZ^T X^T Q^T$ 的左奇异向量和右奇异向量。

- 更新 E : 给定 P 、 Q 和 Z , 更新 E 的子问题为:

$$\min_E \lambda_2 \|E\|_{2,1} + \frac{\mu}{2} \|X - PQXZ - E + \frac{C_1}{\mu}\|_F^2 \quad (36)$$

令 $K = X - PQXZ + C_1/\mu$, 问题 (36) 等价于:

$$\min_E \frac{\lambda_2}{\mu} \|E\|_{2,1} + \frac{1}{2} \|E - K\|_F^2 \quad (37)$$

可以分解为:

$$\min_{e_i} \frac{\lambda_2}{\mu} \|e_i\|_2 + \frac{1}{2} \|e_i - k_i\|_2^2 \quad (38)$$

其中, e_i 和 k_i 分别是矩阵 E 和 K 的第 i 列. 问题 (38) 的解^[38]如下:

$$e_i = \begin{cases} \left(1 - \frac{\lambda_2/\mu}{\|k_i\|_2}\right) k_i, & \|k_i\|_2 > \frac{\lambda_2}{\mu} \\ 0, & \|k_i\|_2 \leq \frac{\lambda_2}{\mu} \end{cases} \quad (39)$$

- 更新 C_1 、 C_2 和 μ : C_1 、 C_2 和 μ 分别通过以下方式

更新:

$$\begin{cases} C_1 = C_1 + \mu(X - PQXZ - E) \\ C_2 = C_2 + \mu(Z - B) \\ \mu = \min(\rho\mu, \mu_{\max}) \end{cases} \quad (40)$$

模型 (20) 的详细优化过程见算法 1.

算法 1. 通过 ADMM 优化模型 (20) 的算法

输入: 训练样本矩阵 $X \in \mathbb{R}^{d \times n}$, 超参数 $\lambda_1, \lambda_2, \lambda_3$, 邻域数量 k 和投影维度 m .

初始化: $P = \arg \max_{P^T P = I} \text{tr}(P^T \Sigma P)$, 其中 Σ 是训练样本协方差矩阵; $Q = P^T, Z = B = 0, E = 0$, 并且 $C_1 = C_2 = 0, \mu = 0.1, \rho = 1.1, \mu_{\max} = 10^8$ 且 $\epsilon = 10^{-6}$.

重复:

1. 使用式 (26) 更新 Z .

2. 使用式 (29) 更新 B .

3. 使用式 (32) 更新 Q .

4. 使用式 (35) 更新 P .

5. 使用式 (39) 更新 E .

6. 使用式 (40) 更新 C_1, C_2 和 μ .

直到: 以下约束条件被满足

$$\begin{cases} \|X - PQXZ - E\|_{\infty} \leq \epsilon \\ \|Z - B\|_{\infty} \leq \epsilon \end{cases}$$

输出: 判别投影矩阵 Q .

3.3 复杂度分析

在本节中, 对解决所提出模型 LPP_SRL 所涉及的计算复杂性进行了全面分析. 算法 1 涉及几个关键步骤, 其中矩阵求逆是解决 Q 和 Z 所需的主要耗时操作. 对于计算 E 的 $\ell_{2,1}$ 范数的优化以及使用 ℓ_1 范数解决 B , 相对而言计算成本较小, 可以忽略不计. 考虑到矩阵求逆, 求解 Q 的计算复杂度为 $O(d^3)$, 而求解 Z 的计算复杂度为 $O(n^3)$. 对于 P 的更新主要由奇异值分解 (singular value decomposition, SVD) 计算所决定, 其复杂度为 $O(m^3)$ ^[27]. 总的来说, 算法 1 的计算复杂度为 $O(\tau(n^3 + d^3 + m^3))$, 其中 τ 表示迭代次数. 在实际情况中, 通常有 $d \leq n$ 且 m 较小, 因此可以将总计算复杂度的近似估计为 $O(\tau n^3)$.

4 实验

在本节中, 我们评估了所提出方法在 6 个广泛使用的公共数据集上的有效性, 这些数据集收集自不同的场景, 包括: (1) EYaleB^[39]; (2) YTC^[40]; (3) Binalpha (<https://cs.nyu.edu/~roweis/data.html>); (4) USPS^[41]; (5) Caltech101 Silhouettes^[42]; (6) ETH80^[43]. 为了进行有说服力的比较, 我们将所提方法与一些先进的方法进行比较, 包括 LDA^[6], Supervised LPP (SLPP)^[44], LSDA^[45], OLSDA^[46], FOLPP^[47], LADA^[19], RSLDA^[48], SLNP^[20],

ALLDA^[21], SN-TSL^[49], LRPER^[3]. 在本文的实验中, 使用最近邻 (1-NN) 分类器利用欧氏距离对样本进行分类, 并将上述方法的参数调试到最佳范围之内或直接采用作者推荐的值. 此外, 我们针对每个数据集固定了所有参数, 并同时改变训练样本的数量. 为了公平起见, 我们对所有算法在比较实验中进行了 10 次实验, 并计算出平均识别率. 在每一次实验中, 我们随机选择每个类别的一些样本作为训练数据 (#Tr), 同时使用剩余的样本作为测试数据. 为了提高计算效率, 我们首先对样本进行归一化, 即 $x_i = x_i / \|x_i\|_2$, 然后使用 PCA, 在所有数据集中通过保留 98% 能量来减少数据的维度. 保留 98% 能量指的是仅选择满足 $\sum_{i=1}^k \lambda_i / \sum_{i=1}^n \lambda_i = 0.98$ 的最大的 k 个特征值对应的特征向量, 其中 λ_i 表示原始数据的协方差矩阵的第 i 大特征值, n 是总特征值的数量. 这些选定的向量用于组成 PCA 的投影矩阵进行降维.

4.1 在 Extended Yale B 数据集上的实验

Extended Yale B (EYaleB) 数据集包括从 38 名志愿者那里获取的共 2414 张人脸图像, 每位志愿者提供了 59–64 张具有不同光照条件的图像, 此外, 每个人的一半图像包含不同程度的阴影. 在本文的实验中, 每个样本被调整大小为 32×32 像素, 图 1 展示了 EYaleB 数据集的一些典型样本. 对于所有的比较算法, 我们随机选择每个人的 10、15、20 和 25 个样本作为训练数据, 剩余的样本用于测试, 从表 1 可以明显看出, 所提出的方法与其他监督学习方法相比具有优势.

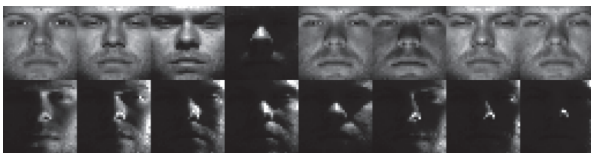


图 1 EYaleB 人脸数据集的一些样本

表 1 在 EYaleB 数据集上不同方法的平均识别准确率 (%) 和标准差

#Tr	10	15	20	25
LDA	83.13 (1.02)	91.38 (1.02)	94.07 (0.49)	96.34 (0.24)
SLPP	86.26 (0.96)	92.81 (0.94)	94.72 (0.22)	96.39 (0.43)
LSDA	85.41 (1.06)	92.54 (0.70)	94.76 (0.41)	96.24 (0.44)
OLSDA	88.30 (1.28)	91.62 (0.83)	94.12 (0.37)	95.29 (0.58)
FOLPP	88.89 (1.32)	93.34 (0.81)	94.51 (0.65)	95.86 (0.75)
LADA	88.36 (1.33)	93.77 (0.88)	95.49 (0.68)	96.76 (0.63)
RSLDA	88.86 (1.00)	93.16 (0.80)	94.66 (0.42)	96.17 (0.48)
SLNP	85.95 (0.99)	92.71 (0.86)	94.66 (0.29)	96.76 (0.38)
ALLDA	82.77 (1.22)	90.04 (0.91)	92.90 (0.58)	95.11 (0.77)
SN-TSL	88.62 (1.02)	93.64 (0.78)	95.39 (0.56)	96.83 (0.54)
LRPER	81.59 (1.09)	89.62 (0.85)	92.55 (0.53)	95.00 (0.54)
LPP_SRL	90.52 (1.10)	94.40 (0.45)	95.68 (0.44)	96.92 (0.34)

4.2 在 YTC 数据集上的实验

YouTube-Celebrities (YTC) 数据集包括来自 YouTube 的 47 位名人的 1910 个视频片段. 这些视频来自真实场景, 展示了在面部表情、外貌和姿势上的显著变化, 同时还包含噪音、错位、低画质和遮挡等问题, 每个视频片段包含 8–400 帧, 每个名人由 1000 多张图像表示. 对于本文的实验, 每位名人的 200 张图像被随机选取为实验数据 (即总共 9400 张图像), 这些人脸图像被统一调整大小为 30×30 像素, 图 2 展示了不同个体的一些人脸图像. 我们随机选择每位名人的 15、20、25 和 30 个样本作为训练数据, 剩余的图像作为测试数据. 表 2 所示的实验结果表明, 所提出的方法在所有比较算法中表现最佳.

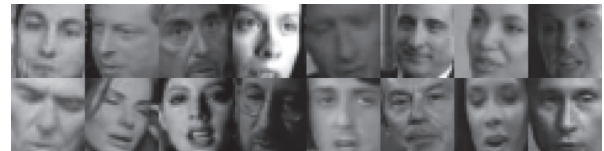


图 2 YTC 人脸数据集的一些样本

表 2 在 YTC 数据集上不同方法的平均识别准确率 (%) 和标准差

#Tr	15	20	25	30
LDA	66.44 (0.86)	72.84 (0.41)	76.28 (0.59)	78.90 (0.51)
SLPP	66.40 (0.97)	73.60 (0.55)	77.31 (0.64)	79.93 (0.66)
LSDA	65.45 (1.38)	77.69 (0.65)	84.08 (0.62)	87.72 (0.44)
OLSDA	32.67 (0.85)	33.77 (1.00)	34.44 (1.04)	36.44 (0.87)
FOLPP	69.15 (0.80)	75.49 (0.70)	78.85 (0.49)	81.21 (0.45)
LADA	58.07 (6.78)	39.29 (6.64)	39.70 (7.58)	40.57 (6.63)
RSLDA	78.59 (0.76)	82.73 (0.79)	84.96 (0.51)	86.79 (0.44)
SLNP	73.01 (0.96)	80.53 (0.70)	84.49 (0.57)	87.52 (0.53)
ALLDA	63.37 (1.02)	72.62 (0.96)	77.71 (0.35)	81.77 (0.59)
SN-TSL	78.54 (0.89)	82.46 (0.63)	84.59 (0.43)	86.24 (0.60)
LRPER	64.76 (0.81)	70.99 (0.73)	75.13 (0.61)	78.61 (0.55)
LPP_SRL	80.39 (0.92)	84.20 (0.81)	86.42 (0.48)	88.33 (0.61)

4.3 在 Binalpha 数据集上的实验

该数据集包含分布在 36 个类别中的 1404 个样本, 每个样本由一个大小为 20×16 像素的二进制图像表示. 值得注意的是, 该数据集不仅包括从“0”到“9”的数字, 还包括大写字母从“A”到“Z”, 因此对分类任务构成了挑战, 图 3 展示了一些样本图像. 我们随机选择每个类别的 10、15 和 20 个样本作为训练数据, 剩余的作为测试数据, 此数据集上的平均识别结果如表 3 所示, 我们可以发现, 与其他监督学习方法相比, 所提出的方法具有优势.



图3 Binalpha数据集的一些样本

表3 在 Binalpha 数据集上不同方法的平均识别准确率 (%) 和标准差

#Tr	10	15	20
LDA	27.27 (1.41)	41.86 (1.70)	50.06 (1.18)
SLPP	26.38 (1.97)	43.13 (1.88)	53.90 (1.67)
LSDA	25.31 (1.48)	41.41 (2.12)	50.37 (1.06)
OLSDA	17.92 (3.54)	12.64 (0.79)	10.56 (0.96)
FOLPP	42.43 (2.12)	53.87 (1.50)	60.80 (2.08)
LADA	46.10 (2.01)	40.76 (3.33)	23.48 (5.79)
RSLDA	65.28 (1.68)	67.71 (0.88)	69.94 (1.29)
SLNP	26.21 (2.33)	46.32 (1.41)	57.70 (2.24)
ALLDA	15.30 (1.69)	26.85 (1.20)	34.85 (1.88)
SN-TSL	49.17 (1.54)	52.62 (0.90)	56.30 (0.99)
LRPER	21.62 (1.33)	37.12 (1.12)	45.19 (1.31)
LPP_SRL	66.03 (1.69)	68.92 (0.84)	71.51 (1.27)

4.4 在 USPS 数据集上的实验

USPS 数据集是一个手写数字图像集合, 包括从“0”至“9”的 10 类, 该数据集包含共 9289 个图像, 每个类别包含不同数量的样本, 从 708 到 1553 个不等. 为了方便分析, 所有图像都已调整大小为 16×16 像素的灰度图像, 因此样本的原始维度为 256, 图 4 展示了数据集中的一些样本, 代表了不同的数字. 在我们的实验中, 我们随机选择每个类别的 10、20、30 和 40 个图像作为训练集, 剩余的作为测试集, 表 4 表明, 所提出的方法在比较的算法中获得了最高的准确率.

4.5 在 Caltech101 Silhouettes 数据集上的实验

Caltech101 数据集包含 101 个不同类别的物体图像, 每个类别包含 31–798 张图像, 大多数类别大约有 50 张图像, 该数据集中的每个图像都包含一个高质量的多边形轮廓, 勾勒出场景中的主要对象. 在这里, 我们利用了 Caltech101 Silhouettes 数据集, 它专门关注物体轮廓, 以研究每种方法在形状识别中的性能, 在该数据集中, 轮廓被呈现为填充的黑色多边形, 背景为白色. 该数据集包含来自原始数据集的 8641 个样本, 其中每个样本都是一个大小为 16×16 像素的轮廓图像, 图 5 提供了原始 Caltech101 数据集中飞机和雷龙的一些样本. 在本文的实验中, 我们从每个类别中随机提取了 10、15 和 20 个样本, 形成 3 组训练集, 剩余的样本被作为测试集, 从表 5 可以看出, 只有 RSLDA 表现略优于所提出的方法.



图4 USPS数据集的一些样本

表4 在 USPS 数据集上不同方法的平均识别准确率 (%) 和标准差

#Tr	10	20	30	40
LDA	72.85 (1.90)	84.57 (0.76)	86.91 (0.55)	88.02 (0.59)
SLPP	74.69 (1.88)	85.83 (0.63)	88.74 (0.53)	89.90 (0.66)
LSDA	70.05 (2.40)	84.95 (1.13)	88.96 (0.62)	90.47 (0.54)
OLSDA	69.14 (2.52)	68.52 (1.35)	68.21 (1.81)	68.78 (1.90)
FOLPP	80.12 (1.42)	86.62 (0.97)	88.81 (0.84)	89.82 (0.61)
LADA	77.78 (1.97)	81.53 (1.45)	80.81 (1.44)	82.04 (2.05)
RSLDA	83.45 (1.53)	88.58 (0.74)	90.00 (0.34)	90.72 (0.49)
SLNP	73.98 (2.36)	86.25 (0.56)	88.66 (0.44)	90.00 (0.76)
ALLDA	65.18 (3.50)	78.32 (1.57)	83.30 (0.56)	85.46 (1.04)
SN-TSL	82.84 (1.25)	86.67 (0.66)	88.16 (0.33)	88.73 (0.65)
LRPER	31.19 (1.44)	42.33 (0.69)	43.69 (0.81)	47.01 (0.83)
LPP_SRL	84.97 (1.73)	89.12 (0.54)	90.89 (0.23)	91.64 (0.49)



图5 Caltech101数据集的一些样本

表5 在 Caltech101 Silhouettes 数据集上不同方法的平均识别准确率 (%) 和标准差

#Tr	10	15	20
LDA	35.86 (1.41)	40.63 (0.99)	43.55 (0.87)
SLPP	37.82 (1.59)	44.15 (1.14)	48.47 (0.91)
LSDA	36.85 (1.55)	43.24 (1.14)	47.56 (1.00)
OLSDA	13.64 (1.11)	14.56 (0.62)	15.80 (0.99)
FOLPP	40.66 (1.71)	46.77 (1.09)	50.50 (0.77)
LADA	15.30 (1.31)	15.49 (0.68)	17.94 (1.57)
RSLDA	50.94 (1.77)	55.02 (0.81)	57.09 (0.67)
SLNP	39.68 (1.53)	46.56 (1.20)	49.88 (0.57)
ALLDA	31.93 (0.90)	39.04 (1.05)	43.35 (1.20)
SN-TSL	40.50 (1.51)	45.22 (1.09)	48.60 (0.79)
LRPER	36.43 (1.32)	42.22 (0.99)	46.68 (0.74)
LPP_SRL	50.85 (1.73)	54.87 (0.88)	56.89 (0.75)

4.6 在 ETH80 数据集上的实验

ETH80 数据集包括属于 8 个不同类别的物体图像, 包括苹果、汽车、奶牛、杯子、狗、马、梨和番茄, 每个类别有 10 个对象实例, 并且每个对象实例有来自 41 个不同视角的图片, 图 6 展示了 ETH80 数据集中的一些样本. 在本文的实验中, 我们将灰度图像调

整大小为 20×20 像素, 并将其转换为 400 维向量, 我们随机选择每个类别的 10、20、30 和 40 个图像作为训练集, 剩余的作为测试集. 从表 6 呈现的结果中, 可以明显看出所提出方法的优越性能.

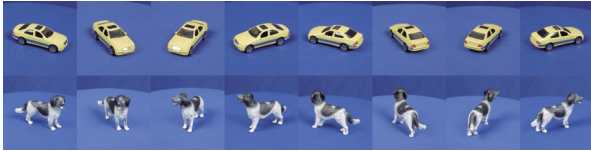


图 6 ETH80 数据集的一些样本

表 6 在 ETH80 数据集上不同方法的平均识别准确率 (%) 和标准差

#Tr	10	20	30	40
LDA	25.62 (3.31)	33.04 (2.28)	43.90 (2.06)	51.89 (1.50)
SLPP	47.42 (1.83)	34.20 (2.75)	47.60 (1.26)	56.71 (0.99)
LSDA	44.81 (2.42)	31.84 (2.49)	43.93 (1.64)	52.36 (1.21)
OLSDA	39.84 (1.84)	34.36 (1.80)	32.74 (2.34)	31.10 (1.35)
FOLPP	52.10 (1.58)	46.95 (1.43)	54.02 (1.72)	60.53 (1.17)
LADA	50.92 (2.50)	54.83 (1.92)	55.64 (1.88)	54.69 (1.39)
RSLDA	51.74 (1.90)	60.24 (1.69)	64.78 (1.48)	68.41 (1.42)
SLNP	24.63 (2.02)	23.19 (1.68)	43.41 (1.85)	51.93 (1.44)
ALLDA	44.45 (2.27)	47.41 (1.69)	46.35 (1.20)	46.51 (1.27)
SN-TSL	53.62 (2.19)	59.70 (1.69)	60.35 (1.15)	63.41 (1.31)
LRPER	23.11 (2.55)	32.84 (2.00)	39.91 (1.89)	48.89 (1.44)
LPP_SRL	54.98 (1.85)	62.79 (1.38)	66.77 (1.32)	70.03 (1.07)

4.7 实验结果与分析

根据表 1-表 6 中呈现的实验结果, 所提出的方法在包括人脸识别、手写识别、形状识别和物体识别在内的所有 6 个图像数据集上表现出优越的性能. 这说明了所提出方法的普适性, 并且我们从观察中发现了以下有趣的现象.

(1) 保留高维数据的局部结构通常会带来更好的性能. 例如, LPP 胜过 LDA, 因为 LPP 有效地保留了数据的局部结构, 而 LDA 则忽略了这一点. 此外, FOLPP 比 LPP 具有更强的局部保持能力^[47], 从而得到更高的识别率. 另外, 正如本文前面所讨论的一样, 预定义图方法和自适应优化图方法的性能没有显著差异.

(2) 本文提出的方法和 RSLDA 在所有数据集上都能稳定表现出较高的识别率, 而其他方法则无法保持稳定. 例如, SLNP 在 Binalpha 和 ETH80 数据集中, 以及 SN-TSL 在 Binalpha 数据集中, 表现出较差的识别率. 原因在于所提出的方法和 RSLDA 通过在投影学习过程中考虑数据重建来有效地保留数据的主要

成分.

(3) 在大多数情况下, 所提出的方法优于 RSLDA. 这是因为我们的方法不仅保留了基于距离的邻域, 而且自适应地学习了局部稀疏自表示, 丰富了数据的局部拓扑结构的描述. 相比之下, RSLDA 忽略了对于有效分类至关重要的结构. 这表明我们方法中的局部保持增强策略的确有助于保持数据的局部内在结构.

(4) 在 ETH80 数据集上使用 LADA、SLNP 和 ALLDA 时, 需要预先消除全局散布矩阵 S_l 的零空间, 或者对矩阵 S_l 的逆矩阵施加正则化扰动. 相比之下, 所提出的方法绕过了这些要求, 并且不需要任何特定的数据预处理步骤, 这说明了所提出方法没有小样本 (small sample size, SSS) 问题.

总之, 实验结果表明了我们所提出的方法在各种图像识别任务中的有效性和泛用性.

4.8 参数敏感性研究

目标模型 (20) 的分类性能可能会受到 4 个可调整参数的影响, 即 λ_1 、 λ_2 、 λ_3 和 k . 然而, 自适应地选择合适的参数仍然是一个开放问题, 到目前为止尚未提出有效的解决方案.

在本节中, 我们测试不同参数组合下的分类准确率, 来分析这 4 个参数的影响. 通过实验, 我们观察到 λ_3 和 k 对其他参数的影响较小. 因此, 我们独立确定它们的值. 然而 λ_1 和 λ_2 表现出一定的互相关性, 因此我们进行了网格搜索, 以确定这两个参数的最佳值. 具体来说, λ_1 和 λ_2 从 $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1, 10, 10^2, 10^3, 10^4\}$ 中选择, λ_3 从 $\{10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1, 10, 10^2\}$ 中选择, 而 k 从 $\{1, 2, 3, \dots, 8, 9\}$ 和 $n_c - 1$ 中选择, 其中 n_c 表示训练集中每个类别的样本数. 图 7 展示了在 4 个数据集上不同参数组合下所提出算法的性能. 对于 YTC 数据集, 在 $\lambda_1 \in \{10^{-5}, 10^{-4}\}$, $\lambda_2 \in \{10^{-2}, 0.1\}$ 和 $\lambda_3 \in \{10^{-2}, 0.1\}$ 时获得最佳性能. 对于 Binalpha 数据集, 最佳值为 $\lambda_1 \in \{10^{-4}, 10^{-3}\}$, $\lambda_2 \in \{10^{-2}, 0.1\}$ 和 $\lambda_3 = 1$. 对于 USPS 数据集, 最高识别率在 $\lambda_1 \in \{10^{-5}, 10^{-4}\}$, $\lambda_2 \in \{10^{-2}, 0.1\}$ 和 $\lambda_3 \in \{0.1, 1\}$ 时获得. 最后, 对于 ETH80 数据集, 当 $\lambda_1 \in \{10^{-5}, 10^{-4}\}$, $\lambda_2 \in \{10^{-2}, 0.1\}$ 和 $\lambda_3 = 0.1$ 时, 获得最佳性能. 图 8 显示了在 YTC、Binalpha、USPS 和 ETH80 数据集上选择的邻域样本数与识别率之间的关系. 在大多数情况下, k 的最优值相对较小.

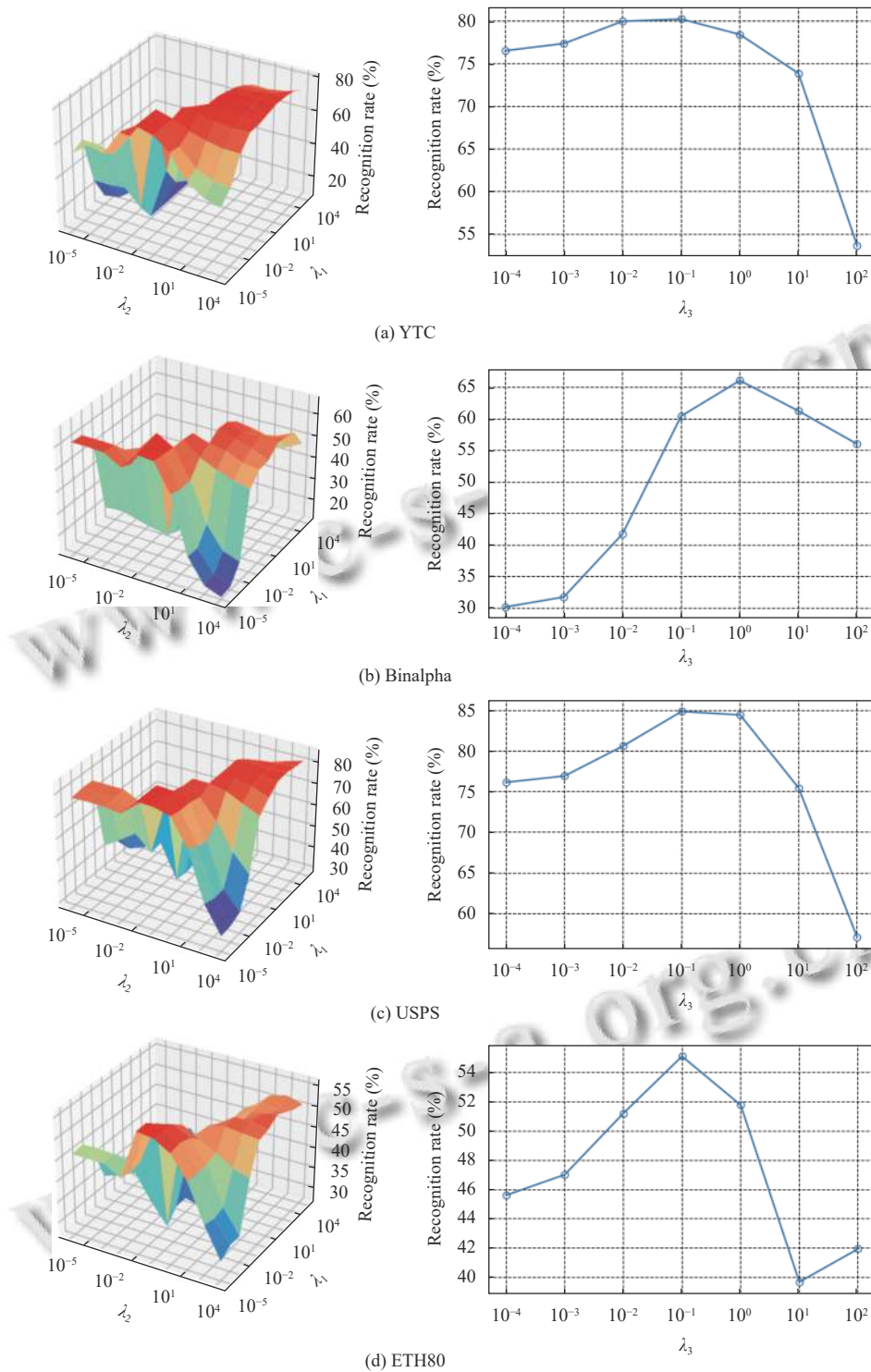


图 7 识别准确率和参数 λ_1 , λ_2 和 λ_3 的关系, 在 4 个数据集上每个类别的训练样本个数分别为 15、10、10 和 10

4.9 对维度敏感性的研究

为了进行特征提取, 我们在 4 个不同的数据集上分别评估了所提出方法在各种维度下的性能. 图 9 展示了识别率与降维后维度之间的关系. 从图 9 中可以

观察到, 只要不设定过低的维度, LPP_SRL 在各种特征维度上都能表现出一致且可观的性能. 在表 1-表 6 的实验结果中, 我们对所有除了 SN-TSL 外的方法设定了统一的子空间维度, 在 EYaleB、YTC、Binalpha、

USPS、Caltech101 Silhouettes 和 ETH80 数据集上分别为 80、40、50、25、50 和 20. 对于 SN-TSL, 降维后

的维度则始终设置为 c , 其中 c 对应于训练样本中的类别数.

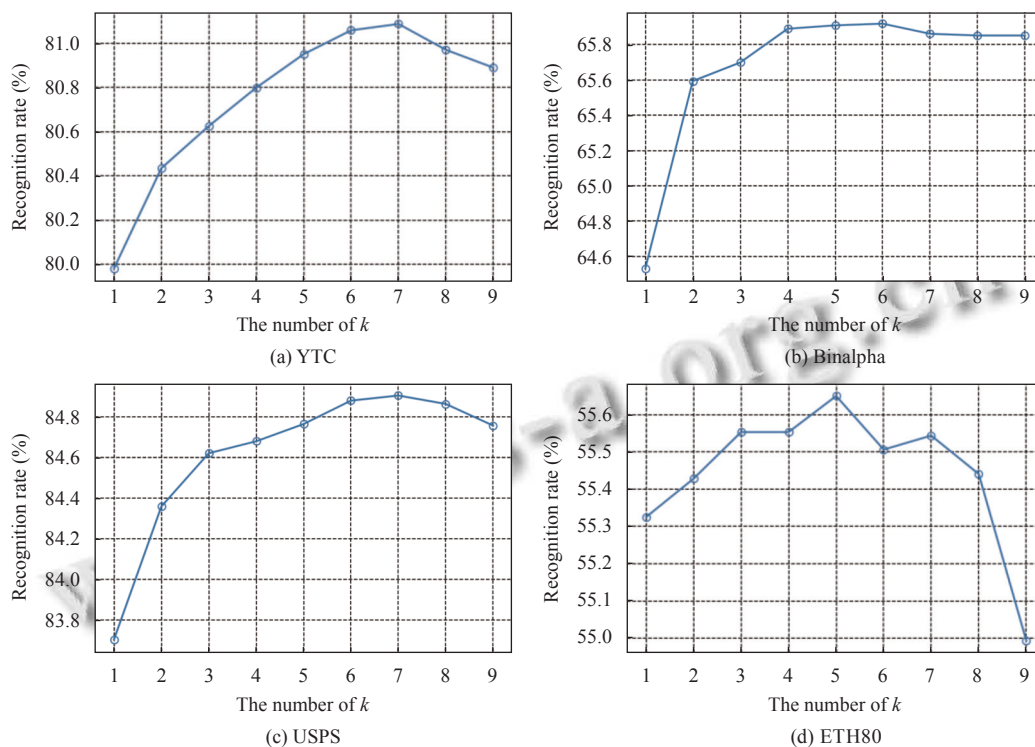


图 8 识别准确率和邻域个数之间的关系, 在 4 个数据集上每个类别的训练样本个数分别为 15、10、10 和 10

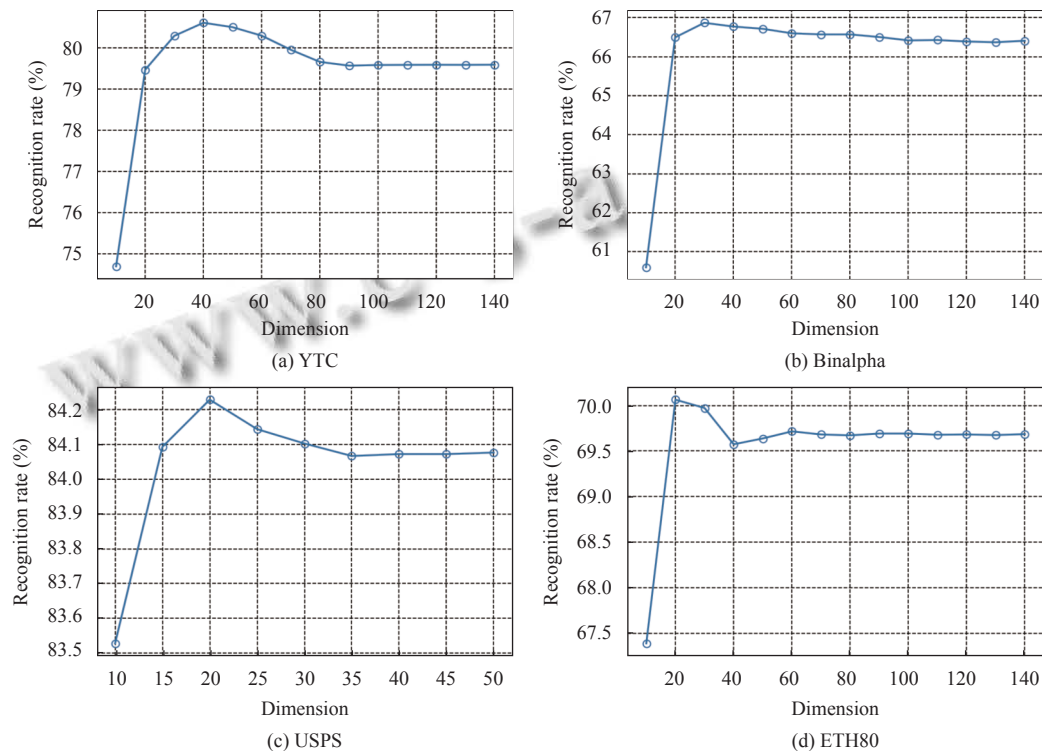


图 9 识别准确率和降维维度之间的关系, 在 4 个数据集上每个类别的训练样本个数分别为 15、10、10 和 40

4.10 收敛性研究

在本节中,我们从实验的角度展示了所提出方法的收敛性,在 YTC、Binalpha、USPS 和 ETH80 这 4 个数据集上进行了实验.图 10 显示了模型在 4 个数

据集的不同迭代次数下的目标函数值和约束误差.从中可以观察到,所提出的学习模型 (20) 的目标函数损失和约束误差稳步减小,并最终趋于稳定.这一观察结果表明我们设计的优化算法表现出了良好的收敛特性.

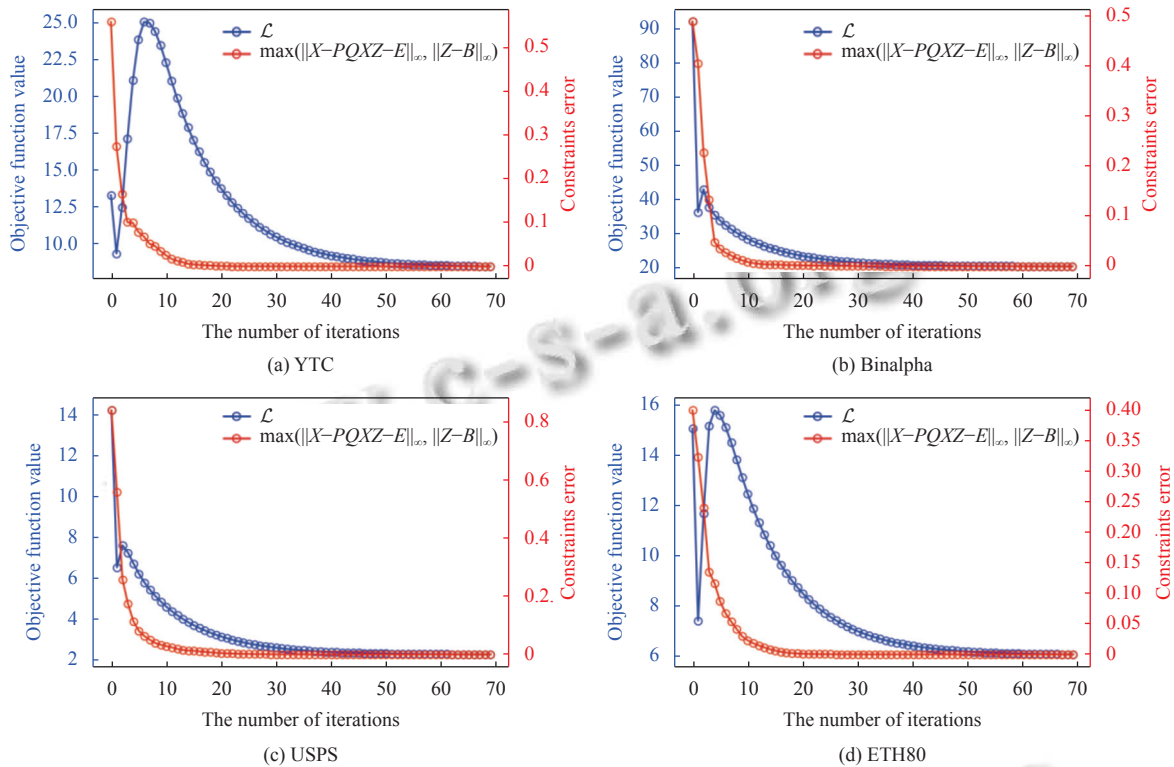


图 10 所提出方法的收敛曲线,在 4 个数据集上每个类别的训练样本个数分别为 15、10、10 和 40

5 结论

本文提出了一种监督降维方法,通过潜在稀疏表示学习增强了局部保持投影的性能.本文将保持邻域结构和局部稀疏自表示集成到统一框架中,解决了传统方法的局限性,并丰富了数据的局部拓扑结构描述.

在稀疏自表示模块中,本文使用重建样本作为字典,并将其与投影学习相结合,以有效捕获数据的主要成分、消除冗余特征和去除噪声.通过高效的迭代算法和在多个公开可用数据集上的实验,所提出的方法在降维方面表现出了出色的性能.对于未来的工作,我们计划设计一种不基于任何假设的图学习方法,并进一步将其整合到我们的方法中,以增强其在不同场景下的泛化能力.

参考文献

1 Meenakshi, Srirangarajan S. Locality-aware discriminative

- subspace learning for image classification. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 5015414. [doi: 10.1109/TIM.2022.3187735]
- 2 Zhao SP, Wu JG, Zhang B, *et al.* Adaptive graph embedded preserving projection learning for feature extraction and selection. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(2): 1060–1073. [doi: 10.1109/TSMC.2022.3193131]
- 3 Zhang T, Long CF, Deng YJ, *et al.* Low-rank preserving embedding regression for robust image feature extraction. IET Computer Vision, 2024, 18(1): 124–140. [doi: 10.1049/cvi2.12228]
- 4 Lu XH, Long J, Wen J, *et al.* Locality preserving projection with symmetric graph embedding for unsupervised dimensionality reduction. Pattern Recognition, 2022, 131: 108844. [doi: 10.1016/j.patcog.2022.108844]
- 5 Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems,

- 1987, 2(1–3): 37–52. [doi: [10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)]
- 6 Martinez AM, Kak AC. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(2): 228–233. [doi: [10.1109/34.908974](https://doi.org/10.1109/34.908974)]
- 7 Tenenbaum JB, Silva VD, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, 290(5500): 2319–2323. [doi: [10.1126/science.290.5500.2319](https://doi.org/10.1126/science.290.5500.2319)]
- 8 Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(5500): 2323–2326. [doi: [10.1126/science.290.5500.2323](https://doi.org/10.1126/science.290.5500.2323)]
- 9 Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. Vancouver: MIT Press, 2001. 585–591.
- 10 He XF, Niyogi P. Locality preserving projections. *Proceedings of the 16th International Conference on Neural Information Processing Systems (NIPS)*. Whistler: MIT Press, 2003. 153–160.
- 11 Wen J, Han N, Fang XZ, *et al.* Low-rank preserving projection via graph regularized reconstruction. *IEEE Transactions on Cybernetics*, 2019, 49(4): 1279–1291. [doi: [10.1109/TCYB.2018.2799862](https://doi.org/10.1109/TCYB.2018.2799862)]
- 12 Liu ZH, Lu Y, Lai ZH, *et al.* Robust sparse low-rank embedding for image dimension reduction. *Applied Soft Computing*, 2021, 113: 107907. [doi: [10.1016/j.asoc.2021.107907](https://doi.org/10.1016/j.asoc.2021.107907)]
- 13 Wan MH, Yao Y, Zhan TM, *et al.* Supervised low-rank embedded regression (SLRER) for robust subspace learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(4): 1917–1927. [doi: [10.1109/TCSVT.2021.3090420](https://doi.org/10.1109/TCSVT.2021.3090420)]
- 14 Jiang L, Fang XZ, Sun WJ, *et al.* Low-rank constraint based dual projections learning for dimensionality reduction. *Signal Processing*, 2023, 204: 108817. [doi: [10.1016/j.sigpro.2022.108817](https://doi.org/10.1016/j.sigpro.2022.108817)]
- 15 Lu JL, Wang HL, Zhou J, *et al.* Low-rank adaptive graph embedding for unsupervised feature extraction. *Pattern Recognition*, 2021, 113: 107758. [doi: [10.1016/j.patcog.2020.107758](https://doi.org/10.1016/j.patcog.2020.107758)]
- 16 Hu LC, Zhang WS, Dai ZL. Joint sparse locality-aware regression for robust discriminative learning. *IEEE Transactions on Cybernetics*, 2022, 52(11): 12245–12258. [doi: [10.1109/TCYB.2021.3080128](https://doi.org/10.1109/TCYB.2021.3080128)]
- 17 Hu LC, Dai ZL, Tian L, *et al.* Class-oriented self-learning graph embedding for image compact representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(1): 74–87. [doi: [10.1109/TCSVT.2022.3197746](https://doi.org/10.1109/TCSVT.2022.3197746)]
- 18 Wen J, Zhong ZF, Zhang Z, *et al.* Adaptive locality preserving regression. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(1): 75–88. [doi: [10.1109/TCSVT.2018.2889727](https://doi.org/10.1109/TCSVT.2018.2889727)]
- 19 Li XL, Wang Q, Nie FP, *et al.* Locality adaptive discriminant analysis framework. *IEEE Transactions on Cybernetics*, 2022, 52(8): 7291–7302. [doi: [10.1109/TCYB.2021.3049684](https://doi.org/10.1109/TCYB.2021.3049684)]
- 20 Pang YW, Zhou B, Nie FP. Simultaneously learning neighborhood and projection matrix for supervised dimensionality reduction. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(9): 2779–2793. [doi: [10.1109/TNNLS.2018.2886317](https://doi.org/10.1109/TNNLS.2018.2886317)]
- 21 Nie FP, Wang Z, Wang R, *et al.* Adaptive local linear discriminant analysis. *ACM Transactions on Knowledge Discovery from Data*, 2020, 14(1): 9. [doi: [10.1145/3369870](https://doi.org/10.1145/3369870)]
- 22 He XF, Cai D, Yan SC, *et al.* Neighborhood preserving embedding. *Proceedings of the 10th IEEE International Conference on Computer Vision*, Vol. 1. Beijing: IEEE, 2005. 1208–1213. [doi: [10.1109/ICCV.2005.167](https://doi.org/10.1109/ICCV.2005.167)]
- 23 Qiao LS, Chen SC, Tan XY. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 2010, 43(1): 331–341. [doi: [10.1016/j.patcog.2009.05.005](https://doi.org/10.1016/j.patcog.2009.05.005)]
- 24 Zhang YP, Xiang M, Yang B. Low-rank preserving embedding. *Pattern Recognition*, 2017, 70: 112–125. [doi: [10.1016/j.patcog.2017.05.003](https://doi.org/10.1016/j.patcog.2017.05.003)]
- 25 Wong WK, Lai ZH, Wen JJ, *et al.* Low-rank embedding for robust image feature extraction. *IEEE Transactions on Image Processing*, 2017, 26(6): 2905–2917. [doi: [10.1109/TIP.2017.2691543](https://doi.org/10.1109/TIP.2017.2691543)]
- 26 Fang XZ, Han N, Wu JG, *et al.* Approximate low-rank projection learning for feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(11): 5228–5241. [doi: [10.1109/TNNLS.2018.2796133](https://doi.org/10.1109/TNNLS.2018.2796133)]
- 27 Ren ZW, Sun QS, Wu B, *et al.* Learning latent low-rank and sparse embedding for robust image feature extraction. *IEEE Transactions on Image Processing*, 2020, 29: 2094–2107. [doi: [10.1109/TIP.2019.2938859](https://doi.org/10.1109/TIP.2019.2938859)]
- 28 Zhou JH, Zhang B, Zeng SN, *et al.* Joint discriminative latent subspace learning for image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4653–4666. [doi: [10.1109/TCSVT.2021.3135316](https://doi.org/10.1109/TCSVT.2021.3135316)]
- 29 Zhang XQ, Tan Z, Sun HJ, *et al.* Orthogonal low-rank projection learning for robust image feature extraction. *IEEE*

- Transactions on Multimedia, 2022, 24: 3882–3895. [doi: [10.1109/TMM.2021.3109442](https://doi.org/10.1109/TMM.2021.3109442)]
- 30 Hui KF, Shen XJ, Abhadiomhen SE, *et al.* Robust low-rank representation via residual projection for image classification. Knowledge-based Systems, 2022, 241: 108230. [doi: [10.1016/j.knosys.2022.108230](https://doi.org/10.1016/j.knosys.2022.108230)]
- 31 Donoho DL, Elad M. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. Proceedings of the National Academy of Sciences of the United States of America, 2003, 100(5): 2197–2202. [doi: [10.1073/pnas.0437847100](https://doi.org/10.1073/pnas.0437847100)]
- 32 Donoho DL. For most large under determined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics, 2006, 59(6): 797–829. [doi: [10.1002/cpa.20132](https://doi.org/10.1002/cpa.20132)]
- 33 Candes EJ, Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory, 2006, 52(12): 5406–5425. [doi: [10.1109/TIT.2006.885507](https://doi.org/10.1109/TIT.2006.885507)]
- 34 Boyd S, Parikh N, Chu E, *et al.* Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning, 2011, 3(1): 1–112. [doi: [10.1561/22000000016](https://doi.org/10.1561/22000000016)]
- 35 Hong MY, Luo ZQ, Razaviyayn M. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM Journal on Optimization, 2016, 26(1): 337–364. [doi: [10.1137/140990309](https://doi.org/10.1137/140990309)]
- 36 Candès EJ, Li XD, Ma Y, *et al.* Robust principal component analysis? Journal of the ACM, 2011, 58(3): 11. [doi: [10.1145/1970392.1970395](https://doi.org/10.1145/1970392.1970395)]
- 37 Zou H, Hastie T, Tibshirani R. Sparse principal component analysis. Journal of Computational and Graphical Statistics, 2006, 15(2): 265–286. [doi: [10.1198/106186006X113430](https://doi.org/10.1198/106186006X113430)]
- 38 Liu J, Ji SW, Ye JP. Multi-task feature learning via efficient $\ell_{2,1}$ -norm minimization. Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal: AUAI Press, 2009. 339–348.
- 39 Georgiades AS, Belhumeur PN, Kriegman DJ. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 643–660. [doi: [10.1109/34.927464](https://doi.org/10.1109/34.927464)]
- 40 Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. Proceedings of CVPR 2011. Colorado: IEEE, 2011. 529–534. [doi: [10.1109/CVPR.2011.5995566](https://doi.org/10.1109/CVPR.2011.5995566)]
- 41 Hull JJ. A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(5): 550–554. [doi: [10.1109/34.291440](https://doi.org/10.1109/34.291440)]
- 42 Hu LC, Xu JK, Tian L, *et al.* Self-centralized jointly sparse maximum margin criterion for robust dimensionality reduction. Knowledge-based Systems, 2020, 206: 106343. [doi: [10.1016/j.knosys.2020.106343](https://doi.org/10.1016/j.knosys.2020.106343)]
- 43 Leibe B, Schiele B. Analyzing appearance and contour based methods for object categorization. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison: IEEE, 2003. II–409. [doi: [10.1109/CVPR.2003.1211497](https://doi.org/10.1109/CVPR.2003.1211497)]
- 44 He XF, Yan SC, Hu YX, *et al.* Face recognition using Laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328–340. [doi: [10.1109/TPAMI.2005.55](https://doi.org/10.1109/TPAMI.2005.55)]
- 45 Cai D, He XF, Zhou K, *et al.* Locality sensitive discriminant analysis. Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad: Morgan Kaufmann Publishers Inc., 2007. 708–713.
- 46 Nie FP, Xiang SM, Liu Y, *et al.* Orthogonal vs. uncorrelated least squares discriminant analysis for feature extraction. Pattern Recognition Letters, 2012, 33(5): 485–491. [doi: [10.1016/j.patrec.2011.11.028](https://doi.org/10.1016/j.patrec.2011.11.028)]
- 47 Wang R, Nie FP, Hong RC, *et al.* Fast and orthogonal locality preserving projections for dimensionality reduction. IEEE Transactions on Image Processing, 2017, 26(10): 5019–5030. [doi: [10.1109/TIP.2017.2726188](https://doi.org/10.1109/TIP.2017.2726188)]
- 48 Wen J, Fang XZ, Cui JR, *et al.* Robust sparse linear discriminant analysis. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(2): 390–403. [doi: [10.1109/TCSVT.2018.2799214](https://doi.org/10.1109/TCSVT.2018.2799214)]
- 49 Chen Z, Wu XJ, Cai YH, *et al.* Sparse non-negative transition subspace learning for image classification. Signal Processing, 2021, 183: 107988. [doi: [10.1016/j.sigpro.2021.107988](https://doi.org/10.1016/j.sigpro.2021.107988)]

(校对责编: 张重毅)