

中图分类号: TP181 文献标志码: A 文章编号: 1006-8961(2011)10-1802-10

论文索引信息:陶剑文,王士同.  $L_1$  范局部线性嵌入 [J]. 中国图象图形学报,2011,16(10):1802-1811

# $L_1$ 范局部线性嵌入

陶剑文<sup>1),2)</sup>,王士同<sup>2)</sup>

<sup>1)</sup>(江南大学信息工程学院,无锡 214122) <sup>2)</sup>(浙江工商职业技术学院工学院,宁波 315012)

**摘要:** 数据降维问题存在于包括机器学习、模式识别、数据挖掘等多个信息处理领域。局部线性嵌入(LLE)是一种用于数据降维的无监督非线性流行学习算法,因其优良的性能,LLE得以广泛应用。针对传统的LLE对离群(或噪声)敏感的问题,提出一种鲁棒的基于 $L_1$ 范数最小化的LLE算法( $L_1$ -LLE)。通过 $L_1$ 范数最小化来求取局部重构矩阵,减小了重构矩阵能量,能有效克服离群(或噪声)干扰。利用现有优化技术, $L_1$ -LLE算法简单且易实现。证明了 $L_1$ -LLE算法的收敛性。分别对人造和实际数据集进行应用测试,通过与传统LLE方法进行性能比较,结果显示 $L_1$ -LLE方法是稳定、有效的。

**关键词:** 降维;  $L_1$ -范数; 流形学习; 局部线性嵌入; 鲁棒性

## $L_1$ -norm locally linear embedding

Tao Jianwen<sup>1),2)</sup>, Wang Shitong<sup>2)</sup>

<sup>1)</sup>(School of Information Engineering, Southern Yangtze University, Wuxi 214122 China)

<sup>2)</sup>(School of Information Engineering, Zhejiang Business Technology Institute, Ningbo 315012 China)

**Abstract:** The problem of dimensionality reduction arises in many fields of information processing, including machine learning, pattern recognition, data mining etc. Locally linear embedding (LLE) is an unsupervised and nonlinear learning algorithm for dimensionality reduction, well-known for its outperformance. Unlike classical LLE, which is based on the  $L_2$ -norm, a novel  $L_1$ -norm based LLE ( $L_1$ -LLE) algorithm is proposed in this article, which is robust to outliers because it utilizes the  $L_1$ -norm, which is less sensitive to outliers. The proposed  $L_1$ -norm optimization technique is intuitive, simple, and easy to implement. It is also proven to find a globally minimal solution. The proposed method is applied to several data sets and the performance is compared to those of other conventional methods.

**Keywords:** dimensionality reduction;  $L_1$ -norm; manifold learning; locally linear embedding; robust

## 0 引言

高维数据处理的研究,因其在Web文本数据、图像数据、生物数据、网络数据,以及金融市场交易数据等领域的广泛应用,而受到普遍重视。由于“维数灾祸”(curses of dimensionality)问题,通常

对于高维数据的处理(如Web文本分类、图像分类、模式识别等),存在着种种困难。一种常用的做法是在保持数据所含感兴趣信息的前提下,尽可能降低数据的维数,即降维。降维是将高维数据向低维表达空间变换的过程。低维表达对应于数据的本质维,数据的本质维是能客观描述数据特征的最小参数个数。降维技术除了能简化数据

收稿日期:2010-09-25;修回日期:2010-11-15

基金项目:国家自然科学基金项目(60975027,60903100),宁波市自然科学基金项目(2009A610080)。

第一作者简介:陶剑文(1973—),男,副教授。江南大学轻工信息技术专业博士研究生,主要研究方向为模式识别、Web挖掘等。E-mail:jianwen\_tao@yahoo.com.cn。

使数据集得以有效处理外,还能应用于数据可视化。另外,数据降维过程还有助于揭示某些嵌入在高维空间的数据集的潜在低维流形结构,或称流形学习,其中,低维空间反映潜在的控制参数,高维空间为数据特征空间。

很多数据常常是嵌入在高维空间的一个低维流形, ISOMAP (isometric map)<sup>[1]</sup> 和 LLE (locally linear embedding)<sup>[2]</sup> 是寻找描述这样低维流形子空间的最有代表性的方法。ISOMAP 在降维过程中通过计算点对之间的测地距离,采用 MDS (multi-dimensional scale)<sup>[1]</sup> 方法来获取全局最优的几何结构,获得了较好的效果,目前已经发展了很多改进算法。LLE<sup>[2]</sup> 方法是 2000 年由 Roweis 和 Saul 提出的一种新的非线性降维方法。LLE 在降维过程中能保持数据的局部几何结构不变,并能够避免局部极小,从而可获得全局的低维嵌入<sup>[3]</sup>。目前,已有多种 LLE 的变体提出<sup>[4-8]</sup>。

虽然上述数据降维方法对于不同性状的数据均具有良好的数据降维能力,但其对离群数据均具有较强的敏感性,即在噪声环境下不具备一定的鲁棒性。为了缓解传统的基于  $L_2$  范数的主成分分析 ( $L_2$ -PCA)<sup>[9]</sup> 技术对离群敏感的问题,提出许多基于  $L_1$  范数的 PCA ( $L_1$ -PCA 或 PCA- $L_1$ ) 技术<sup>[9-10]</sup>, 其明显提升了传统 PCA 的抗噪能力。针对 LLE 对噪声敏感问题,目前已有许多鲁棒的 LLE 的变体先后被提出<sup>[3,11-12]</sup>。最近, Chang 等人提出一种鲁棒的 LLE 方法<sup>[3]</sup>,其基于一种鲁棒的 PCA 技术,实现低维数据的加权嵌入;徐雪松等人提出一种基于核函数的稳健 LLE 方法<sup>[11-12]</sup>。

传统的基于  $L_2$  范数的 LLE ( $L_2$ -LLE) 方法因在求取重构矩阵时会放大噪声 (或离群) 影响因子,导致  $L_2$ -LLE 输出因噪声 (或离群) 影响而偏移目标。为了提升在噪声环境下  $L_2$ -LLE 算法的鲁棒性,受 PCA- $L_1$ <sup>[9-10]</sup> 方法的启发,提出一种鲁棒的基于  $L_1$  范数的 LLE 算法 ( $L_1$ -LLE),采用  $L_1$  范数来实现局部误差最小化重构和嵌入,从而获得更好的抗噪性。另外, $L_1$ -LLE 优化算法简单且易实现,同样能获得全局最优解。

## 1 LLE 算法

假设输入数据为采样于流形空间的  $N$  个  $D$  维实向量样本点  $\mathbf{x}_i, \mathbf{x}_i \in \mathbf{R}^D, i = 1, 2, \dots, N$ , 输出为

降维后的  $N$  个  $d$  维向量  $\mathbf{y}_i, \mathbf{y}_i \in \mathbf{R}^d, i = 1, 2, \dots, N$  ( $d \ll D$ )。LLE 算法可以归结为 3 步。

- 1) 寻找每个样本点  $\mathbf{x}_i \in \mathbf{R}^D$  的  $k$  近邻集  $N_i \in \mathbf{R}^D$ ;
- 2) 基于局部数据点的最小化误差线性重构, 计算  $k$  近邻集的局部重构权值矩阵  $\mathbf{W}$ ;
- 3) 为了保持最好的局部几何拓扑结构, 根据局部重构权值矩阵  $\mathbf{W}$ , 计算出高维数据点  $\mathbf{x}_i \in \mathbf{R}^D$  的低维嵌入坐标  $\mathbf{y}_i \in \mathbf{R}^d$ 。

在步骤 1) 中,假定高维空间中的数据为非线性分布,采用 diijstra 距离计算样本点的  $k$  近邻集。在步骤 2) 中,计算样本点的局部线性重构权值矩阵,为使样本点的线性重构误差最小化,定义基于  $L_2$  范数的局部重构残差函数

$$\min \varepsilon_2(\mathbf{W}) = \sum_{i=1}^N \left\| \mathbf{x}_i - \sum_{j=1}^k \mathbf{w}_j^i \mathbf{x}_j^i \right\|_2^2 = \sum_{i=1}^N \left\| \sum_{j=1}^k \mathbf{w}_j^i (\mathbf{x}_i - \mathbf{x}_j^i) \right\|_2^2 \quad (1)$$

式中,  $\mathbf{x}_j^i (j = 1, 2, \dots, k)$  为  $\mathbf{x}_i$  的  $k$  个近邻点,  $\mathbf{w}_j^i$  是  $\mathbf{x}_i$  与  $\mathbf{x}_j^i$  之间的权值,且满足

$$\sum_{j=1}^k \mathbf{w}_j^i = 1 \quad (2)$$

$\|\cdot\|_2$  指矩阵或向量的  $L_2$  范数。这里求取  $\mathbf{W}$  矩阵,需要构造一个局部协方差矩阵  $\mathbf{Q}^i$ 。

$$\mathbf{Q}_{j_m}^i = (\mathbf{x}_i - \mathbf{x}_j^i)^T (\mathbf{x}_i - \mathbf{x}_m^i) \quad (3)$$

将式(3)与式(2)相结合,并采用拉格朗日乘子法,即可求出局部最优化重建权值矩阵

$$\mathbf{w}_j^i = \frac{\sum_{m=1}^k (\mathbf{Q}^i)_{j_m}^{-1}}{\sum_{p=1}^k \sum_{q=1}^k (\mathbf{Q}^i)_{pq}^{-1}} \quad (4)$$

在实际运算中,  $\mathbf{Q}^i$  可能是一个奇异矩阵,此时必须正则化  $\mathbf{Q}^i$ , 即

$$\mathbf{Q}^i = \mathbf{Q}^i + r\mathbf{I} \quad (5)$$

式中,  $r$  是正则化参数,  $\mathbf{I}$  是一个  $k \times k$  的单位矩阵。

在步骤 3) 中, LLE 将所有的样本点映射到低维空间中。映射条件满足下式

$$\min \varepsilon_2(\mathbf{Y}) = \sum_{i=1}^N \left| \mathbf{y}_i - \sum_{j=1}^k \mathbf{w}_j^i \mathbf{y}_j^i \right|^2 \quad (6)$$

式中,  $\varepsilon_2(\mathbf{Y})$  为损失函数值,  $\mathbf{y}_i$  是  $\mathbf{x}_i$  的输出向量,  $\mathbf{y}_j^i (j = 1, 2, \dots, k)$  是  $\mathbf{y}_i$  的  $k$  个近邻点,且要满足两个条件,即

$$\sum_{i=1}^N \mathbf{y}_i = \mathbf{0}, \quad \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^T = \mathbf{I} \quad (7)$$

式中,  $\mathbf{I}$  是  $N \times N$  的单位矩阵。这里的  $\mathbf{w}_j^i (i = 1, 2, \dots, N)$  可以存储在  $N \times N$  的稀疏矩阵  $\mathbf{W}$  中, 当  $\mathbf{x}_j$  是  $\mathbf{x}_i$  的近邻点时,  $W_{i,j} = \mathbf{w}_j^i$ , 否则,  $W_{i,j} = 0$ ,  $W_{ij}$  为  $\mathbf{w}$  的第  $i$  行第  $j$  列元素。则损失函数可重写为

$$\min \varepsilon_2(\mathbf{Y}) = \sum_{i=1}^N \sum_{j=1}^N M_{i,j} \mathbf{y}_i^T \mathbf{y}_j \quad (8)$$

式中,  $\mathbf{M}$  是一个  $N \times N$  的对称矩阵, 其表达式为

$$\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W}) \quad (9)$$

要使损失函数值达到最小, 则取  $\mathbf{Y}$  为  $\mathbf{M}$  的最小  $m$  个非零特征值所对应的特征向量。在处理过程中, 将  $\mathbf{M}$  的特征值从小到大排列, 第 1 个特征值接近于零, 那么舍去第 1 个特征值。通常取第 2 ~  $m + 1$  间的特征值所对应的特征向量作为输出结果。

$L_2$ -LLE 降维算法中重构矩阵  $\mathbf{W}$  是否准确将直接影响低维嵌入的最终效果。由于  $L_2$ -LLE 中局部线性重构矩阵  $\mathbf{W}$  是由数据集  $\mathbf{X}$  解出的, 故当数据集  $\mathbf{X}$  受噪声污染时, 计算对噪声非常敏感, 特别是当  $\mathbf{X}^T \mathbf{X}$  的特征值较小时, 甚至得不到需要的结果, 受 PCA- $L_1$  思想的启发, 本文提出一种基于  $L_1$  范数的局部重构误差最小化的 LLE 算法 ( $L_1$ -LLE), 重点关注  $L_1$ -LLE 算法的重构矩阵的构建方法。

## 2 $L_1$ -LLE 算法

假设输入数据为采样于具有噪声污染的流形空间的  $N$  个  $D$  维实向量样本点  $\mathbf{x}_i + \varepsilon_i, \mathbf{x}_i \in \mathbf{R}^D (i = 1, 2, \dots, N), \varepsilon_i (i = 1, 2, \dots, N)$  为各点相应的污染噪声, 输出为降维后的  $N$  个  $d$  维向量  $\mathbf{y}_i, \mathbf{y}_i \in \mathbf{R}^d, i = 1, 2, \dots, N (d \ll D)$ 。  $L_1$ -LLE 算法步骤描述如下:

- 1) 对于各数据点  $\mathbf{x}_i \in \mathbf{R}^D$ , 寻找其  $k$  近邻集  $\mathbf{X}_N$ ;
- 2) 基于  $L_1$  范数的邻域重构矩阵  $\mathbf{W}$  的构建;
- 3) 由局部重建权值矩阵  $\mathbf{W}$  及其近邻集  $\mathbf{X}_N$ , 计算各样本点的低维嵌入坐标  $\mathbf{y}_i \in \mathbf{R}^d$ 。

步骤 1) 实现方法和  $L_2$ -LLE 中计算  $k$  近邻集方法相同。下面重点研究步骤 2) 的实现算法。

根据式(1), 定义基于  $L_1$  范数的局部重构残差函数

$$\min \varepsilon_1(\mathbf{w}) = \sum_{i=1}^N \left\| \sum_{j=1}^K \mathbf{w}_j^i (\mathbf{x}_i - \mathbf{x}_j^i) \right\|_1 \quad (10)$$

s. t.

$$\|\mathbf{w}^i\|_2^2 = 1, \quad \mathbf{w}_j^i > 0 \quad (11)$$

式中,  $\|\cdot\|_1$  指矩阵或向量的  $L_1$  范数。对式(10)最小化问题的求解等价于如下优化问题

$$\begin{aligned} \min \mathbf{w}_j^* &= \arg \min_w \sum_{i=1}^N \left| \sum_{j=1}^K \mathbf{w}_j^i (\mathbf{x}_i - \mathbf{x}_j^i) \right| = \\ & \arg \min_w \sum_{i=1}^N \sum_{p=1}^D \left| \sum_{j=1}^K \mathbf{w}_j^i (x_{pi} - x_{pj}^i) \right| \end{aligned} \quad (12)$$

s. t.  $\|\mathbf{w}^i\|_2^2 = 1, \mathbf{w}_j^i > 0$

式中,  $x_{pi}$  指  $\mathbf{x}_i$  的第  $p$  个元素,  $\mathbf{w}^i$  指与  $\mathbf{x}_i$  对应的权值向量,  $x_{pj}^i$  指邻居向量  $\mathbf{x}_j^i$  的第  $p$  个元素。

令

$$f(\mathbf{w}) = \sum_{i=1}^N \sum_{p=1}^D \left| \sum_{j=1}^K \mathbf{w}_j^i (x_{pi} - x_{pj}^i) \right| \quad (13)$$

定义符号函数  $P_p^i(t)$  为

$$P_p^i(t) = \begin{cases} 1 & \sum_{j=1}^K \mathbf{w}_j^i (x_{pi} - x_{pj}^i) > 0 \\ -1 & \sum_{j=1}^K \mathbf{w}_j^i (x_{pi} - x_{pj}^i) \leq 0 \end{cases} \quad (14)$$

将式(14)代入式(13)得

$$f(\mathbf{w}(t)) = \sum_{i=1}^N \sum_{p=1}^D P_p^i(t) \left( \sum_{j=1}^K \mathbf{w}_j^i(t) (x_{pi} - x_{pj}^i) \right) \quad (15)$$

为了最小化式(15)且满足

$$\|\mathbf{w}^i(t)\|_2^2 = 1, \quad \mathbf{w}_j^i(t) > 0 \quad (16)$$

引入拉格朗日函数

$$\begin{aligned} L(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\beta}) &= \sum_{i=1}^N \sum_{p=1}^D P_p^i(t) \sum_{j=1}^K \mathbf{w}_j^i(t) (x_{pi} - x_{pj}^i) + \\ & \sum_{i=1}^N \lambda_i (1 - \|\mathbf{w}^i(t)\|_2^2) - \sum_{j=1}^K \beta_j \mathbf{w}_j^i(t) \end{aligned} \quad (17)$$

式中,  $\lambda_i \geq 0, \beta_j \geq 0$  为拉格朗日乘子变量, 式(17)对  $\mathbf{w}_j^i(t)$  求偏导数得

$$\frac{\partial L}{\partial \mathbf{w}_j^i(t)} = \sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - 2\lambda_i \mathbf{w}_j^i(t) - \beta_j = 0$$

从而

$$\mathbf{w}_j^i(t) = \frac{\sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - \beta_j}{2\lambda_i} \quad (18)$$

根据式(16)可知  $\sum_{j=1}^K (\mathbf{w}_j^i(t))^2 = 1$ , 将式(18)代入可得

$$\lambda_i = \frac{1}{2} \sqrt{\sum_{j=1}^K \left( \sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - \beta_j \right)^2} \quad (19)$$

将式(19)代入式(18)可得

$$w_j^i(t) = \frac{\sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - \beta_j}{\sqrt{\sum_{j=1}^K \left( \sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - \beta_j \right)^2}} \quad (20)$$

将式(20)代入式(17)可得原始问题的对偶问题为

$$\max_w F(\mathbf{w}(t)) = \sum_{i=1}^N \sum_{p=1}^D P_p^i(t) \sum_{j=1}^K w_j^i(t) (x_{pi} - x_{pj}^i) - \sum_{j=1}^K \beta_j w_j^i(t) \quad (21)$$

为了使得式(21)最大化,采用如下迭代公式

$$w_j^i(t+1) = \frac{\sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - \beta_j}{\sqrt{\sum_{j=1}^K \left( \sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) - \beta_j \right)^2}} \quad (22)$$

每一次迭代后,  $P_p^i(t+1)$  根据式(14)进行更新。为了确保每次迭代后  $w_j^i(t) > 0$ , 其中变量  $\beta_j \geq 0$  可由下式替代

$$\beta_j = \min \left\{ \sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i) \right\} - \Delta_j$$

式中,  $\Delta_j$  为一可调的极小正参数。根据 KKT 条件, 当  $w_j^i(t) > 0$  时,  $\beta_j = 0$ , 从而式(21)变为

$$\max_w F(\mathbf{w}(t)) = \sum_{i=1}^N \sum_{p=1}^D P_p^i(t) \sum_{j=1}^K w_j^i(t) (x_{pi} - x_{pj}^i) \quad (23)$$

局部邻域重构权值矩阵构建算法描述如下:

输入  $N$  个输入样本数据  $\{\mathbf{x}_i\}_{i=1}^N$ , 参数初始化。任意取邻域重构权值矩阵  $\mathbf{w}(0)$ , 令  $\mathbf{w}(0) \leftarrow \mathbf{w}(0) / \|\mathbf{w}(0)\|_2$ , 迭代次数  $t = 0, i = 0$ 。

输出 局部邻域重构权值矩阵  $\mathbf{w}^*$ 。

步骤:

1) 计算样本数据  $\{\mathbf{x}_i\}_{i=1}^N$  的 KNN ( $k$ -最近邻集) 邻域矩阵  $\mathbf{X}_N = \{\mathbf{x}_j^i\}_{j=1}^K$ 。

2) 符号检查 根据式(14)进行符号判别, 得到初始化符号矩阵  $\mathbf{P}(0)$ 。

3) 重构矩阵迭代 根据式(22)进行迭代计算

局部重构矩阵  $\mathbf{w}(t+1)$ , 令  $\mathbf{w}(t+1) \leftarrow \mathbf{w}(t+1) / \|\mathbf{w}(t+1)\|_2$ , 并根据式(14)进行符号矩阵更新  $\mathbf{P}(t+1)$ 。

4) 收敛检查

(1) 如果  $F(\mathbf{w}(t+1)) \neq F(\mathbf{w}(t))$ , 返回到 3) 继续进行迭代;

(2) 如果存在  $i$  使得  $\sum_{j=1}^K w_j^i(t) (x_{pi} - x_{pj}^i) = 0$ , 则令  $\mathbf{w}(t+1) \leftarrow (\mathbf{w}(t+1) + \Delta\mathbf{w}) / \|\mathbf{w}(t+1) + \Delta\mathbf{w}\|_2$ , 然后转到步骤 3) 继续进行迭代;

(3) 否则, 设  $\mathbf{w}^* = \mathbf{w}(t+1)$ , 结束。

为了证明算法 1 的收敛性, 引出如下  $L_1$ -LLE 收敛性定理。

**定理 1** ( $L_1$ -LLE 收敛性) 算法 1 能够收敛到一个优化权值矩阵  $\mathbf{w}^*(t)$ , 且该优化矩阵为对偶问题式(23)中目标函数  $F(\mathbf{w}(t))$  的局部最大点。

**证明** 首先, 证明  $F(\mathbf{w}(t))$  为一个关于参数  $t$  的非减函数

$$F(\mathbf{w}(t)) = \sum_{i=1}^N \sum_{p=1}^D P_p^i(t) \sum_{j=1}^K w_j^i(t) (x_{pi} - x_{pj}^i) \geq \sum_{i=1}^N \sum_{p=1}^D \sum_{j=1}^K P_p^i(t-1) w_j^i(t) (x_{pi} - x_{pj}^i) \geq \sum_{i=1}^N \sum_{p=1}^D \sum_{j=1}^K P_p^i(t-1) w_j^i(t-1) (x_{pi} - x_{pj}^i) = F(\mathbf{w}(t-1)) \quad (24)$$

故目标函数  $F(\mathbf{w}(t))$  为关于  $t$  的非减函数, 且训练样本数有限, 故算法 1 能收敛到优化值  $\mathbf{w}^*(t)$ 。

其次, 证明  $\mathbf{w}^*(t)$  为目标函数  $F(\mathbf{w}(t))$  的局部极大点:

因为  $\mathbf{w}(t)$  能根据算法 1 收敛到  $\mathbf{w}^*(t)$ , 故对于所有  $p (p = 1, \dots, D)$ ,  $P_p^i(t) \sum_{j=1}^K (w_j^i(t))^* (x_{pi} - x_{pj}^i) > 0$ , 又因样本点数有限, 且由算法 1 中(2)保证使得  $\sum_{j=1}^K w_j^i(t) (x_{pi} - x_{pj}^i) \neq 0$ , 则存在  $\mathbf{w}^*(t)$  的一个较小邻域  $N(\mathbf{w}^*(t))$ , 使得  $P_p^i(t) \sum_{j=1}^K w_j^i(t) (x_{pi} - x_{pj}^i) \geq 0$  成立。而又  $\mathbf{w}^*(t)$  平行于  $\sum_{p=1}^D P_p^i(t) (x_{pi} - x_{pj}^i)$ , 则不等式  $F(\mathbf{w}^*(t)) \geq F(\mathbf{w}(t))$ 。从而对偶问题式(23)中目标函数  $F(\mathbf{w}(t))$  能够取得局部最大点  $\mathbf{w}^*(t)$ 。

证毕

在计算出局部重构矩阵  $\mathbf{W}$  后, 通过

式(6)–(9)计算原始高维空间数据点的低维嵌入坐标  $y_i \in R^d$ 。

### 3 $L_1$ -LLE 抗噪性分析

为简化标识,令  $x_0$  代表样本点  $x_i$ ,  $U(x_0)$  代表  $x_0$  的邻域, 设  $x_1, x_2, \dots, x_k \in U(x_0)$ , 则  $x_0 = \sum_{i=1}^k w^i x_i$ ,  $\sum_{i=1}^k w^i = 1$ 。令  $x'_i = x_i + \varepsilon_i$  ( $i = 0, 1, 2, \dots, k$ ) 代表相应的受噪声污染的数据点,  $\varepsilon_i$  ( $i = 0, 1, 2, \dots, k$ ) 为各点相应的污染噪声, 则  $x'_0 = \sum_{i=1}^k w'^i x'_i$ ,  $\sum_{i=1}^k w'^i = 1$ , 其中,  $x'_i \in U(x'_0)$ ,  $w'^i$  代表受噪声污染后的数据点局部重构矩阵元素。再令  $X^0 = (x_1, x_2, \dots, x_k)$   $X^{0'} = (x'_1, x'_2, \dots, x'_k)$   $W = (w^1, w^2, \dots, w^k)^T$   $W' = (w'^1, w'^2, \dots, w'^k)^T$  则有

$$x_0 = X^0 W \quad x'_0 = X^{0'} W'$$

从而,有如下定理。

**定理 2** (重构矩阵误差界) 若数据源各噪声间、不同维数间以及  $W' = (w'^1, w'^2, \dots, w'^k)^T$  与噪声之间是相互独立的, 各数据点噪声均值为 0 且同方差, 则在噪声污染环境下, 对于数据点局部线性重构矩阵误差  $\Delta W = W' - W$  的期望界存在如下估计

$$E \|\Delta W\|^2 \leq \frac{k(k+1)\sigma^2}{\lambda_{\min} l} E \|W'\|^2 \quad (25)$$

式中,  $\|\cdot\|$  代表欧几里得范数,  $\sigma^2 = \sum_{i=1}^D \sigma_i^2$ ,  $\sigma_i^2 = \text{Var}(\varepsilon_i)$  ( $i = 1, 2, \dots, D$ ),  $l = \text{rank}(X^0)$ ,  $\lambda_{\min}$  为  $X^{0T} X^0$  的最小非零特征值。

**证明** 详细证明请参考文献[11]。

从定理 2 可知, 影响 LLE 算法鲁棒性的因素有 3 个方面<sup>[2]</sup>: 1) 噪声源分布 ( $\sigma^2$ ); 2) 近邻集基数  $k$ ; 3) 局部重构矩阵能量  $\|W'\|$ 。由此可知, 在噪声环境下, 基于  $L_2$  范数的  $L_2$ -LLE 算法使局部重构矩阵的范数因受噪声干扰而被放大, 使得重构矩阵计算误差上升, 导致降维性能下降。而在噪声环境下, 基于  $L_1$  范数的  $L_1$ -LLE 算法明显缩小了重构矩阵的能量范数, 进而减小了重构矩阵的误差, 增强了 LLE 算法的抗噪能力, 这从第 4 部分的实验结果可明显看出。

## 4 实验结果

### 4.1 人造数据

采用 3 个通用的人造数据集 (Swiss roll、Scurve、Helix)<sup>[3]</sup> 来验证  $L_1$ -LLE 算法的抗噪能力。实验参数设置如表 1 所示。图 1—3 分别显示  $L_2$ -LLE 和  $L_1$ -LLE 算法对 3 个人造数据集的低维嵌入效果。从图 1—3 中 (c) 的低维嵌入效果可看出,  $L_1$ -LLE 方法在上述 3 种人造数据集上均具有较  $L_2$ -LLE 明显增强的流形保持性能, 在噪声干扰情况下,  $L_1$ -LLE 的低维嵌入结果较  $L_2$ -LLE 更平滑, 从而使得数据流形的局部几何结构保持更完美。

表 1 人造数据实验参数设置

Tab. 1 Parameter settings of the LLE/ $L_1$ -LLE experiments reported in Fig. 1-3

参数	Swiss roll	S-curve	Helix
$D$	3	3	2
$d$	2	2	1
$k$	12	12	10
正常数据点数	1 500	1 500	500
噪声百分比/%	5	10	15

### 4.2 实际数据

#### 4.2.1 UCI 数据集

实验采用 10 个 UCI 数据集<sup>[9]</sup>, 分别评价  $L_1$ -LLE 和 LLE 在加入 5% 噪声数据情况下的分类性能, 数据集分类器采用经典的支持向量机 (SVM) 分类器<sup>[4]</sup>。UCI 数据集详细信息如表 2 所示。

表 2 中 10 个数据集作为实验数据, 从中分别随机抽取 5% 的数据作为噪声数据, 各类数据集随机分割成训练集和测试集, 分别采用 LLE 和  $L_1$ -LLE 算法对数据进行降维后通过 SVM 分类器进行二元数据分类。实验中, 通过 10 重交叉验证进行参数选择, 每个数据集分别实验 10 次, 取其平均值作为各自实验结果。图 4 分别显示在不同本质嵌入维下 10 个数据集的平均分类正确率。从分类结果看出, 在有噪声的情况下,  $L_1$ -LLE 对 UCI 数据的低维嵌入 (低维空间) 性能在绝大多数情况下均优于 LLE 方法。

另外, 从 UCI 数据实验结果可看出, 在不同的嵌入维度下, 两种方法的模式识别率有明显变化, 这也说明高维数据本质维的估计对分类精度率的影响较大。本文后续实验均在相同或各自优化的本质维

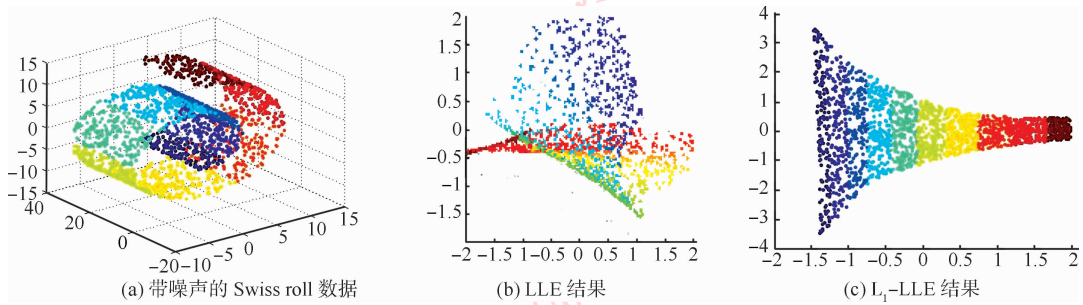


图 1 LLE/  $L_1$ -LLE 对有噪声的 Swiss roll 数据结果  
Fig. 1 LLE/  $L_1$ -LLE applied to the noisy Swiss roll data set

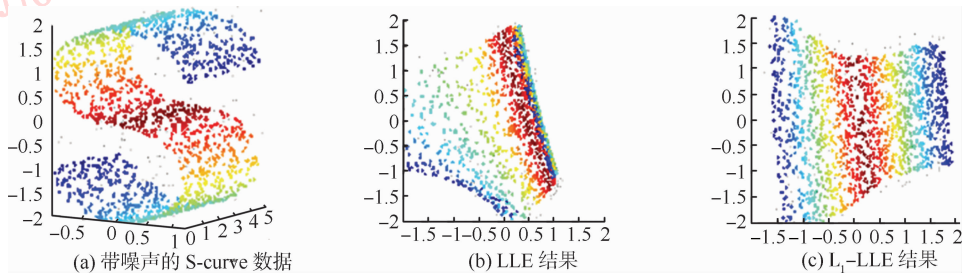


图 2 LLE/  $L_1$ -LLE 对有噪声的 S-curve 数据结果  
Fig. 2 LLE/  $L_1$ -LLE applied to the noisy S-curve data set

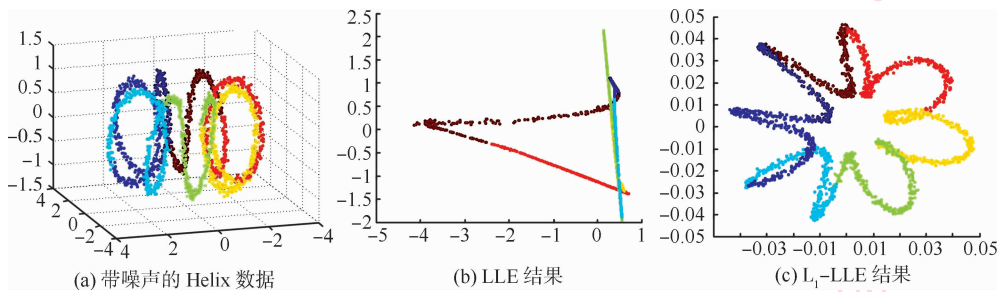


图 3 LLE/  $L_1$ -LLE 对有噪声的 Helix 数据结果  
Fig. 3 LLE/  $L_1$ -LLE applied to the noisy Helix data set

表 2 实验中采用的 UCI 数据集

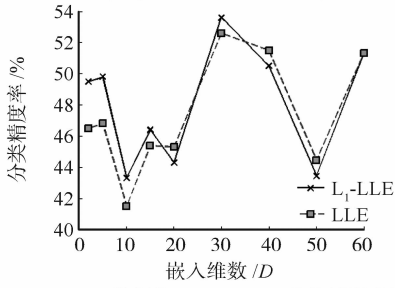
Tab. 2 UCI datasets used in experiments

数据集	正类数	负类数	维数
Sonar-mines	111	97	60
Cancer-Wpbc	151	47	33
Biomed	127	67	5
Liver	145	200	6
Heart	120	150	13
Balance-scale left	288	337	4
Ionosphere-good	225	126	34
Waveform 0	300	600	21
Hepatitis	123	32	19
Diabetes	500	268	8

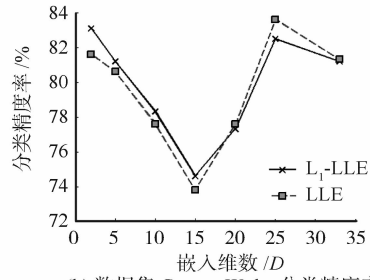
下进行比较。

#### 4.2.2 MNIST 手写数字数据集

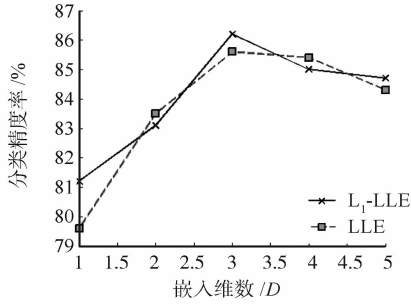
为了显示  $L_1$ -LLE 在高维实际数据上的有效性,将  $L_1$ -LLE 作用在著名的 MNIST 手写数字数据库上进行实验<sup>[3]</sup>,库中手写数字图像大小被规范化为  $28 \times 28$  灰度级图像,这样整个数字空间维数为 784。实验中,从 MNIST 数字库中随机选择 1 000 张数字“8”图像作为实验样本集,图 5 显示 LLE 对该数据集的 2 维嵌入结果。在嵌入流形中贯穿两条线路,一条线路为水平线,另一条为垂直线。从图 5 可看出,第 1 维刻画手写数字的斜度变化,第 2 维刻画手写数字的宽度变化,沿上述两条贯穿路线的斜度和宽度变化是比较平滑的。由此可见,当将手写数字投射到 2 维空间时,LLE 能较好地保持流形的局部几何结构。



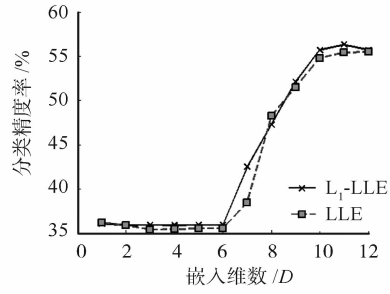
(a) 数据集 Sonar-mines 分类精度率比较



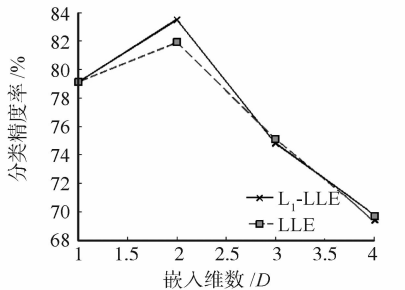
(b) 数据集 Cancer-Wpbc 分类精度率比较



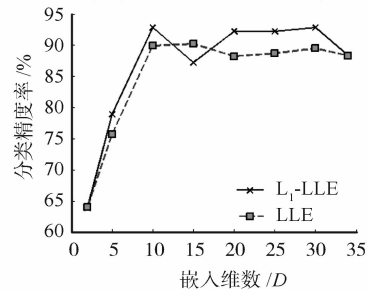
(c) 数据集 Cancer-Wpbc 分类精度率比较



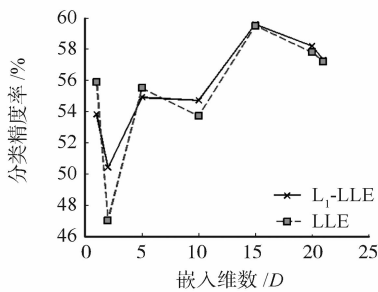
(d) 数据集 Heart 分类精度率比较



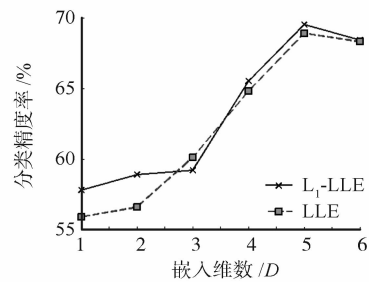
(e) 数据集 Balance-scale 分类精度率比较



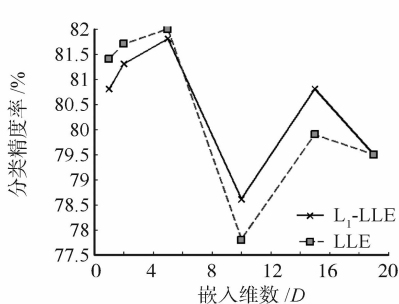
(f) 数据集 Ionosphere-good 分类精度率比较



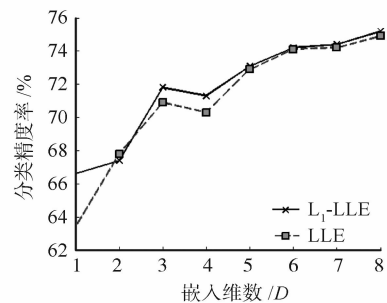
(g) 数据集 Waveform 0 分类精度率比较



(h) 数据集 Liver 分类精度率比较



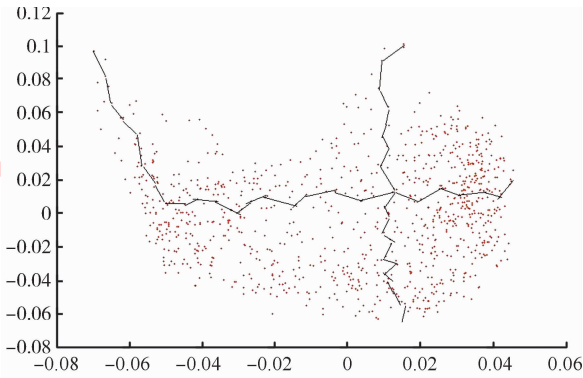
(i) 数据集 Hepatitis 分类精度率比较



(j) 数据集 Diabetes 分类精度率比较

图 4 LLE 和 L<sub>1</sub>-LLE 在 UCI 数据集上的分类正确率

Fig. 4 Correct classification rates of LLE and L<sub>1</sub>-LLE on UCI datasets



(a) LLE 将数字“8”投射到 2 维空间的嵌入结果



(b) 与图 (a) 中贯穿线路相应的数字图像

图 5 LLE 对无噪声手写数字“8”2 维嵌入结果

Fig. 5 LLE on handwritten digit “8”

在向原始数据集中加入部分噪声数据后(如图 6 所示),通过  $L_1$ -LLE 和 LLE 分别将带有噪声的手写数字数据集投射到 2 维嵌入空间,两种算法中最近邻大小设置为  $k = 12$ ,图 7—8 分别显示 LLE 和  $L_1$ -LLE 的实验结果。通过两个图示的比较分析可知,在低维( $R^2$ )嵌入空间, $L_1$ -LLE 在保持手写数字的斜度和宽度变化的连续性上优于 LLE,从而验证了  $L_1$ -LLE 能有效削减噪声数据影响。

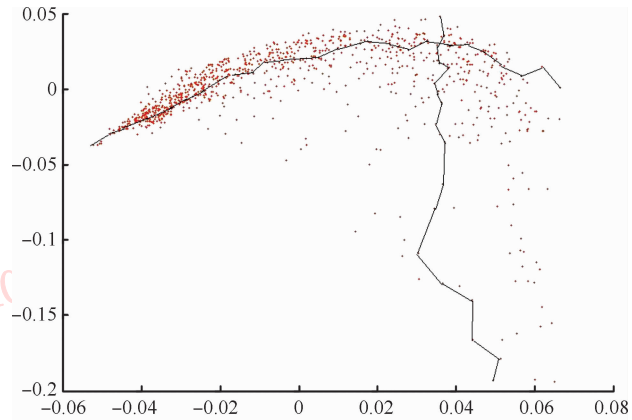


图 6 手写数字噪声数据样例

Fig. 6 Some noisy images generated to serve as outliers

### 4.2.3 人脸数据集

为了进一步说明  $L_1$ -LLE 算法的抗噪性能,本节将  $L_1$ -LLE 算法应用于人脸重构问题,并与 LLE 算法进行性能比较。实验数据采用著名的 Yale<sup>[13-14]</sup>人脸数据集,该数据集由 15 个对象共 165 张灰度级图像组成,每个对象包括 11 幅表情不同的图像,每幅图像大小为  $64 \times 64$ ,每一像素作为一个输入数据维,这样整个 Yale 数据库的输入样本空间为 4 096 维。从 165 幅图像中随机选出 20% 的图像,分别被一个大小至少为  $15 \times 10$  的随机黑白点组成的噪声矩阵块随机覆盖,如图 9 所示。对于这些受随机噪声污染的图像,利用 LLE、 $L_1$ -LLE 对其进行不同维



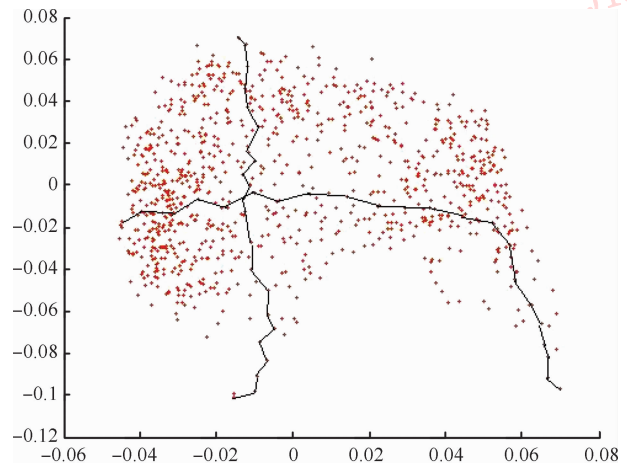
(a) LLE 对带噪声的手写数字的 2 维嵌入结果



(b) 与图 (a) 中贯穿线路相应的数字图像

图 7 LLE 对有噪声手写数字嵌入结果

Fig. 7 LLE on handwritten digit “8” with generated outliers



(a)  $L_1$ -LLE 对带噪声的手写数字的 2 维嵌入结果



(b) 与图 (a) 中贯穿线路相应的数字图像

图 8  $L_1$ -LLE 对有噪声手写数字嵌入结果

Fig. 8  $L_1$ -LLE on handwritten digit “8” with generated outliers

度的降维,然后在低维空间对图像分别进行重构,还原出原图像如图 9(c)(d)所示。从图 9 的实验结果可看出, $L_1$ -LLE 对噪声人脸图像的重构效果明显优于 LLE 方法,进一步证明, $L_1$ -LLE 算法的抗噪能力明显强于 LLE 算法。

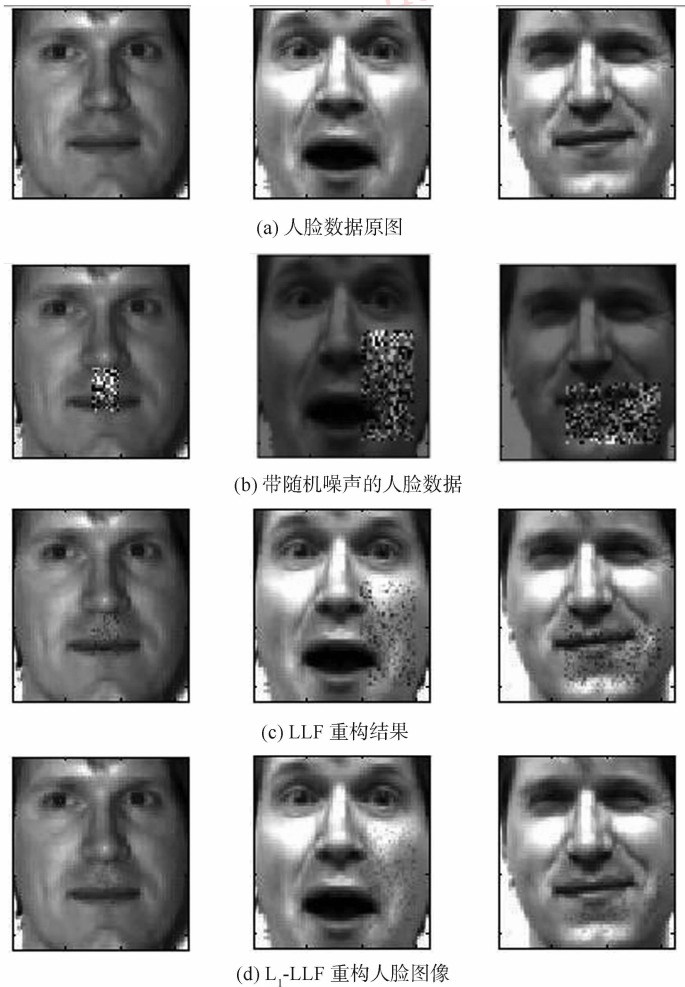
图9 噪声情况下 LLE/ $L_1$ -LLE 人脸重构效果

Fig.9 Face images with occlusion and the reconstructed faces

## 5 结 论

提出一种基于  $L_1$  范数最小化的 LLE 算法。通过大量实验证明,该方法应用于有噪声的数据流形嵌入中具有优于传统 LLE 方法的抗噪性能。所提方法对重构矩阵的求取是通过迭代过程来完成的,理论证明了迭代过程的收敛性。尽管实验证明了  $L_1$ -LLE 方法的有效性和较强的鲁棒性,但是所提算法的迭代过程耗时明显多于传统 LLE 方法,关键瓶颈在于重构矩阵的迭代次数较大。通过一定研究发现,重构矩阵的初值选取比较关键,这也是本方法有待进一步研究的工作。

## 参考文献 (References)

[ 1 ] Tenenbaum J B, de Silva V, Langford J C. A global geometric

framework for nonlinear dimensionality reduction [J]. Science, 2000, 290 (5500):2319-2323.

[ 2 ] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290 (5500): 2323-2326.

[ 3 ] Chang H, Yeung D Y. Robust locally linear embedding [J]. Pattern Recognition, 2006, 39(6):1053-1065.

[ 4 ] Yin J S, Hu D W, Zhou Z T. Manifold learning using growing locally linear embedding [C]// Proc. of the 2007 IEEE Symp. on Computational Intelligence and Data Mining. Honolulu: IEEE Press, 2007: 73-80.

[ 5 ] Varini C, Degenhard A, Nattkemper T W. Isolle: LLE with geodesic distance [J]. Neuro Computing, 2006, 69:1768-1771.

[ 6 ] Wen Guihua, Lu Tinghui, Jiang Lijun, et al. Locally linear embedding based on relative manifold [J]. Journal of Software, 2009, 20(9): 2376-2386. [文贵华,陆庭辉,江丽君,等.基于相对流形的局部线性嵌入[J].软件学报,2009,20(9):2376-2386.]

[ 7 ] Wen G H. Relative transformation-based neighborhood optimization for isometric embedding [J]. Neuro Computing,

- 2009, 72(4-6): 1205-1213.
- [ 8 ] Zhu Minghan, Luo Dayong, Yi Liqun, et al. Incremental locally linear embedding algorithm based on orthogonal iteration method [J]. *Acta Electronica Sinica*, 2009, 37(1): 132-136. [朱明早, 罗大庸, 易励群, 等. 基于正交迭代的增量LLE算法[J]. 电子学报, 2009, 37(1): 132-136.]
- [ 9 ] Kwak N. Principal component analysis based on  $L_1$ -norm maximization [J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008, 30(9): 1672-1680.
- [10] Li Xuelong, Pang Yanwei, Yuan Yuan.  $L_1$ -norm-based 2DPCA [J]. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 2010, 40(4): 1170-1175.
- [11] Xu Xuesong, Song Dongming, Zhang Xu, et al. Robust linear embedding based on a kernel function [J]. *Journal of Image and Graphics*, 2009, 14(6): 1141-1147. [徐雪松, 宋东明, 张谱, 等. 基于核函数的稳健线性嵌入方法[J]. 中国图象图形学报, 2009, 14(6): 1141-1147.]
- [12] Xu Xuesong, Song Dongming, Zhang Xu, et al. The research of detection of outliers based on manifold learning [J]. *Engineering Science*, 2009, 11(2): 82-87. [徐雪松, 宋东明, 张谱, 等. 基于流形学习的离群点检测方法[J]. 中国工程科学, 2009, 11(2): 82-87.]
- [13] Eimad E. Abusham, Wong E K. Locally linear discriminate embedding for face recognition [J]. *Discrete Dynamics in Nature and Society*, 2009, 2009(1): 1-8.
- [14] Goldberg Y, Ritov Y. LDR-LLE: LLE with low-dimensional neighborhood representation [J]. *Advances in Visual Computing, Lecture Notes in Computer Science*, 2008, 5359(2008): 43-54.