Image dimension-reducing based on improved LLE algorithm

Fang Li and Xiang Gao

School of Mathematical Sciences, Ocean University of China, Lane 238, Songling Road, Laoshan District, Qingdao City, Shandong Province, 266100, People's Republic of China

Email: gaoxiangshuli@126.com

Abstract. The locally linear embedding (LLE) is one of the most promising algorithm among nonlinear dimensionality reduction technique. When the number of the nearest neighbor is larger than the input dimension, LLE will make the local covariance matrix of the weights vectors singular in the process of calculating local reconstruction weights vectors. In this case, regularization is needed, but at the same time, which leads to not reflect optimally intrinsic structure of neighbours. To solve the problem, we propose NWLLE (New Weighted LLE) to improve LLE and test the new algorithm from embedded performance and quantitative error. The numerical experiment results show that the NWLLE yields less error than LLE and can produce a good embedding effect.

1. Introduction

Nonlinear dimensionality reduction technology is applied to discover effective and compact representation of complex high-dimensional data. In recent years, diverse algorithms have been researched. Isometric Mapping (ISOMAP) [1], The locally linear embedding (LLE) [2], Laplacian Eigenmaps (LE) [3], Local Tangent Space Alignment (LTSA) [4], Maximum Variance Unfolding (MVU) [5], diffusion maps [6,7], t-Distributed Stochastic Neighbor Embedding (t-SNE) [8], and deep learning such as Autoencoder (AE) [9], Restricted Boltzmann Machines (RBM) [10,11]. This paper mainly studies the LLE algorithm.

LLE preserves local structure and learns low dimensional embeddings of high dimensional inputs, which is extensively applied in many areas, such as feature extraction [12,13], fault diagnosis [14], pattern recognition [15] and image processing [16,17]. Though its computational simplicity and intuitive approach, there are many related works aimed at problems of LLE. In 2002, selection of the optimal parameter value for the LLE was discussed [18]. In 2003, Hessian Eigenmaps (HE) based on the hessian expanded LLE [19], and efficient LLE were proposed for imperfect manifolds [20]. In 2006, robust LLE addressed the problem of outliers by pretraining the input [21]. Subsequently, various improved LLE produced through new distance [22-24]. In 2011, SLLEP applied local linear regression to find embedding for fault diagnosis [25]. In 2014, an improved LLE proved to be effective elimination of redundant information [26]. In 2017, sparse LLE can be applied to larger data sets [27-29]. Although above improved algorithms, some details bring to our notice. In this paper, we still discuss and modify the problem of LLE.

This paper is structured as follows. Section 2 describes LLE algorithm. Section 3 discusses the problem of LLE, and improved algorithm is depicted in section 4. Section 5 is numerical experiments and results.
2. LLE algorithm

(1) Finding \( k \) nearest neighbours for each \( x_i (i = 1, 2, \cdots, N) \) as measured by Euclidean distance in the high-dimensional space. Where \( k \) is a fixed value.

(2) For each \( x_i (i = 1, 2, \cdots, N) \), LLE compute local reconstruction weights vectors \( W_i \) by its \( k \) nearest neighbours. Constraint \( \sum_{j=1}^{k} w_{ij} = 1 \) ought to be satisfied:

\[
W_i = \operatorname{argmin} \epsilon(W_i) = \left| x_i - \sum_{j=1}^{k} w_{ij} x_{ij} \right|^2
\]

(3) For each \( x_i (i = 1, 2, \cdots, N) \), LLE finds the low-dimensional embedding \( Y \), which satisfies two constraints: \( \sum_{i=1}^{N} y_i = 0 \) and \( \frac{1}{N} \sum_{i=1}^{N} y_i y_i^T = I \). Minimize following cost function:

\[
\min \Phi(Y) = \sum_{i=1}^{N} \left| y_i - \sum_{j=1}^{k} w_{ij} y_{ij} \right|^2
\]

3. Problem of LLE

In this section, we firstly demonstrate that calculation of \( W_i \) in the second step of LLE. Minimize reconstruction error:

\[
\epsilon(W_i) = \left| x_i - \sum_{j=1}^{k} w_{ij} x_{ij} \right|^2 = \left| \sum_{j=1}^{k} w_{ij} (x_i - x_{ij}) \right|^2 = \left( (x_i - x_{ij})W_i^T \right)^T \left( (x_i - x_{ij})W_i \right)
\]

\[
= W_i^T (x_i - x_{ij})^T (x_i - x_{ij})W_i = W_i^T C_i W_i
\]

To our knowledge, we can combine \( \sum_{j=1}^{k} w_{ij} = 1 \) and solve above problem by Lagrange multiplier method. But when the number of nearest neighbour \( k \) is larger than dimension of the input space input \( D \), LLE makes \( C_i \) singular and regularization is needed [30]. Then

\[
\epsilon(W_i) = \left| x_i - \sum_{j=1}^{k} w_{ij} x_{ij} \right|^2 + \delta |W_i|^2
\]

\[
= W_i^T C_i W_i + \Delta \text{trace}(C_i) |W_i|^2 = W_i^T \left( C_i + \Delta \text{trace}(C_i) I \right) W_i
\]

Where \( \delta = \Delta \text{trace}(C_i) \) is a constant and \( \Delta \leq 1 \) for LLE. Noted that regularization constant \( \Delta \) must be tuned carefully. In theory, \( \Delta \) is too small for LLE to approach singular of \( C_i \) properly. On the contrary, a relatively large \( \Delta \) can improve embedding, but average local information of \( C_i \) to some extent, which further affect \( W_i \). Next, we take swissrol dataset as an example and show that LLE can yield completely different embeddings for different \( \Delta \), as shown in figure 1.
4. Improved LLE

Based on the above idea, we always use large \( \Delta \) for better embedding when face practical problems, which will lead \( W_i \) can’t reflect optimally local information of neighbours to some extent. Thus, we employ a new computation for \( W_i \) and propose NWLLE to improve LLE algorithm \([31-34]\).

Given the symmetric positive semidefinite matrix \( C_i(k \times k) \), then by spectral decomposition:

\[
P^T C_i P = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)
\]

Where \( P \) is an orthogonal matrix. When \( k > D \), the rank \( r \) of \( C_i \) satisfies \( r \leq D < k \). suppose that \( \lambda_1 = \cdots = \lambda_{k-r} = 0 < \lambda_{k-r+1} \leq \cdots \leq \lambda_k \) are \( k \) real eigenvalues of \( C_i \). Further, \( \xi_1, \xi_2, \ldots, \xi_k \) are orthogonal eigenvectors corresponding to \( k \) eigenvalues of \( C_i \), where \( \xi_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{ik})^T \). Let

\[
P = [\xi_1, \xi_2, \ldots, \xi_k],
\]

then

\[
C_i = P \Lambda P^T = [P_1 \quad P_2]\begin{bmatrix} \Lambda_1 & \end{bmatrix} \begin{bmatrix} P_1^T \\ P_2^T \end{bmatrix} = P_1 \Lambda_1 P_1^T + P_2 \Lambda_2 P_2^T
\]

Where

\[
P = [P_1 \quad P_2]; \quad \Lambda = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix};
\]

\[
P_1 = [\xi_1, \xi_2, \ldots, \xi_{k-r}]_{k \times (k-r)}; \quad P_2 = [\xi_{k-r+1}, \ldots, \xi_k]_{k \times r};
\]

\[
\Lambda_1 = \begin{bmatrix} \lambda_{11} & 0 & \ldots & 0 \\ 0 & \ddots & \vdots & \vdots \\ 0 & \ldots & \lambda_{k-r+1-k \times (k-r)} & \xi_{k-r+1} \\ 0 & \ldots & 0 & \lambda_k \end{bmatrix}_{(k-r) \times (k-r)}; \quad \Lambda_2 = \begin{bmatrix} \lambda_{k-r+1-k \times (k-r)} & \ldots & 0 \\ 0 & \ldots & \lambda_k \end{bmatrix}_{r \times \text{tr}}.
\]

Since \( \Lambda_1 = 0 \), \( P_1 \Lambda_1 P_1^T = 0 \). Thus \( C_i = P_2 \Lambda_2 P_2^T \). According to the upper form, we know that \( \xi_{k-r+1}, \ldots, \xi_k \) corresponding to non-zero eigenvalues \( \lambda_{k-r+1}, \ldots, \lambda_k \) can account for \( C_i \). Then we take \( r \) eigenvectors as the starting point and look for optimal representation of \( W_i \). We suppose

\[
W_i = \varphi_{k-r+1} \xi_{k-r+1} + \varphi_{k-r+2} \xi_{k-r+2} + \cdots + \varphi_k \xi_k = \sum_{i=k-r+1}^{k} \varphi_i \xi_i = P_2 \varphi
\]

Then the solution of \( W_i \) is converted to
\[ \varepsilon(W_t) = W_t^T C_t W_t = W_t^T P_2 \Lambda_2 P_2^T W_t \approx |W_t^T W_t| = |\varphi^T P_2^T P_2 \varphi| = \left| \sum_{i=k-r+1}^{k} \phi_i \xi_i \right|^2 \approx \frac{1}{2} \left| \sum_{i=k-r+1}^{k} \phi_i \xi_i \right|^2 \]

By Lagrange multiplier technique and combining constraint. Hence, the solution for the minimization problem is

\[ W_t = \frac{\sum_{i=k-r+1}^{k} \xi_i \left( \sum_{j=1}^{k} \xi_ij \right)}{\sum_{i=k-r+1}^{k} \left( \sum_{j=1}^{k} \xi_ij \right)^2} \]

It should be noted NWLLE algorithm take \( r \) eigenvectors as the starting point and look for representation of \( W_t \) for preserving optimally local information of \( k \) nearest neighbours.

5. Numerical experiments

Selecting \( N = 2000 \), \( k = 12 \) according to intrinsic dimension \( d = 2 \) of swissroll data, we apply NWLLE to obtain 2-D embedding space of swissroll, as shown in figure 2.

![Figure 2. 2-D embedding space of swissroll for NWLLE algorithm.](image-url)

We assume that the following two principles ought to be satisfied for the optimal embedding of LLE. Firstly, if two data points \( \{x_1, x_2\} \) in the high dimensional space are different, their corresponding data points \( \{y_1, y_2\} \) must be different in a lower dimensional space. Moreover, if \( \{x_{i1}, x_{i2}, \cdots, x_{ik}\} \) are \( k \) nearest neighbours of data point \( x_i \) in the high dimensional space, then \( \{y_{i1}, y_{i2}, \cdots, y_{ik}\} \) must be \( k \) nearest neighbours of data point \( y_i \) in a lower dimensional space [35].

As has been noted, we need a quantitative measure that is necessary to estimate the mapping quality between the high dimensional input and output in the embedded space. The residual variance appears to be appropriate for this measure:

\[ \text{error} = 1 - \rho(D_X D_Y) \]

Where \( D_X \) and \( D_Y \) are the distance matrix of between paired data points in the high dimensional space \( X \in R^D \) and the corresponding embedding space \( Y \in R^d \), respectively, \( \rho \) is the standard linear correlation coefficient [18]. Please note that the smaller the error is, the better high dimensional input is approximated in the embedding space [36-38].
Figure 3. Error comparison between LLE and NWLLE.

As shown in figure 2 and figure 3 show that, the swissroll numerical experiment results show that the NWLLE can produce a good dimension reduction effect and yields less error than LLE.

For persuasiveness and credibility, we present more empirical results to evaluate NWLLE algorithm. The relevant dataset are as follows. The MNIST database [39] is comprised of 60000 train images and 10000 test images, which each image is 28×28 pixels gray-scale handwritten digits image from 1 to 9. We randomly select 2000 images from train set and experiment. figure 4 illustrates real handwritten digits samples in the MNIST database. The hand database [40] consisted of 481 images with various motions. each image is 512×480 pixels gray-scale image. Motions images samples in the hand database are shown in figure 5.

Figure 4. Handwritten digits samples in the MNIST database.

Figure 5. Motions samples in the hand database.

We apply NWLLE algorithm to above datasets. Note that all motions images are normalized to 10 ×10 for saving the computing resource. The visualizations of low dimensional is available.

2-D embedding space of the handwritten digits and the motions are presented in figure 6 and figure 7. Note the coordinates of embedding spaces relate to meaningful attributes [41], such as digital styles and hand's positions. Obviously, the experimental results demonstrate that NWLLE algorithm succeeds in finding the intrinsic information and internal laws hidden in high-dimensional data to achieve dimensionality reduction. The improved algorithm makes the application of LLE more extensive in dimensionality reduction.
Figure 6. 2-D embedding space of real handwritten digits.

Figure 7. 2-D embedding space of motions. $N = 481$.

References
[1] Tenenbaum J B, Silva V D and Langford J C 2000 *Science* **290** (5500) 2319
[2] Roweis S T, Saul L K 2000 *Science* **290** (5500) 2323-2326
[3] Belkin M, Niyogi P 2002 *International Conference on Neural Information Processing Systems: Natural & Synthetic* **14**(6) 585-591
[4] Zhang Z, Zha H 2005 *Society for Industrial and Applied Mathematics* **26**(4) 406-424
[5] Weinberger K Q, Saul L K 2006 *International Journal of Computer Vision* **70**(1) 77-90
[6] Nadler B, Lafon S, Coifman R R and Kevrekidis I G 2006 *Applied & Computational Harmonic Analysis* **21**(1) 113-127
[7] Coifman R R, Lafon S 2006 *Applied & Computational Harmonic Analysis* **21**(1) 5-30
[8] Laurens V D M, Hinton G 2008 *Journal of Machine Learning Research* **9**(2605) 2579-2605
[9] Hinton G E, Salakhutdinov R R 2006 *Science* **313**(5786) 504-507
[10] Salakhutdinov R, Mnih A and Hinton G 2007 *International Conference on Machine Learning* **227** 791-798
[11] Tang T B, Murray A F 2007 *Neurocomputing* **70**(7) 1198-1206
[12] Fu M, Xu Q, Kong M and Luo B 2010 *International Conference on Computer Science & Education* 557-560
[13] Xu J, Mu H, Wang Y and Huang F 2018 *Computational & Mathematical Methods in Medicine*
[14] Wang X, Zheng Y, Zhao Z and Wang J 2015 Sensors 15(7) 16225-16247
[15] Wang L 2014 Robust face recognition based on spectral regression optimized by local linear embedding Computer Engineering & Applications
[16] Shen H, Tao D and Ma D 2013 Plos One 8(12) e82409
[17] Sha L, Dan S and Wang J 2017 IEEE International Conference on Acoustics
[18] Olga Kouropteva, Oleg Okun and Matti Pietikäinen 2002 Scandinavian Conference on Image Analysis 3540(10) 359-363
[19] Donoho D L, Grimes C 2003 Proceedings of the National Academy of Sciences of the United States of America 100(10) 5591-5596
[20] Hadid A, Pietikäinen M 2003 International Conference on Machine Learning & Data Mining in Pattern Recognition 2734 188—201
[21] Chang H, Yeung D Y 2006 Pattern Recognition 39(6) 1053-1065
[22] Varini C, Degenhard A and Nattkemper T W 2006 Neurocomputing 69(13-15) 1768-1771
[23] Wang H, Zheng J, Yao Z and Li L 2006 International Symposium on Advances in Neural Networks-issn 39711326-1333
[24] Genaro D S, German C D and Jose C 2011 Neurocomputing 80(2) 19-30
[25] Li B, Zhang Y 2011 Mechanical Systems & Signal Processing 25(8) 3125-3134
[26] Liu S M , Deng Y N, Lv Y X 2014 Applied Mechanics & Materials 644-650 2160-2163
[27] Ziegelmeier L, Kirby M, Peterson C 2017 S Procedia Computer Science 108 635-644
[28] Chen B, Wang B 2017 American Journal of Neural Networks and Applications 3(4) 40-48
[29] Wang L, Xu X, Liu G, Chen B and Chen Z 2017 Chinese Journal of Oceanology and Limnology 35(5) 1002-1009
[30] Saul L K, Roweis S T 2003 Journal of Machine Learning Research 4(2) 119-155
[31] Chen B, Escalante S, Guyon I, Ponce-López V, Shah N and Liu M O 2016 Overcoming calibration problems in pattern labeling with pairwise ratings: application to personality traits. Computer Vision-ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016 Proceedings Part III 419-432
[32] Ponce-López V, Chen B, Places A , liu M O, Corneanu C, Baro X, Escalante H J, Escalante, Guyon I and Guyon S 2016 Escalera ChaLearn LaP 2016: first round challenge on first impressions - dataset and results. Computer Vision -- ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016 Proceedings Part III 400-418
[33] Ponce-López V, Ponce-López, Chen B, liu M O and liu C 2016 Corneanu, A. Clapés, I. Guyon, X. Baró, H. J. Escalante and S. Escalera Chalearn lap 2016: First round challenge on first impressions-dataset and results, Computer Vision -- ECCV 2016 Workshops. Springer
[34] Chen B, Chen Z, Yang S, Huang X, Du Z, Cui J and Bhimani Xie X 2017 36th IEEE International Performance Computing and Communications Conference, Special Session on Cyber Physical Systems: Security, Computing, and Performance (IPCCC-XPS 2017)
[35] Wu F, Hu Z 2006 The LLE and a linear mapping Pattern Recognition 39(9) 1799-1804
[36] Wang L, Chen B, Chen C, Chen Z and Liu G 2016 China Ocean Engineering 30(1) 149-160
[37] Wang L, Chen B , Zhang J and Chen Z 2013 Natural Hazards 67(2) 129-143
[38] Chen B, Liu G and Wang L 2017 International Journal of Energy and Environmental Engineering 2(6) 117-126
[39] Yann Le Cun, Corinna Cortes, Christopher J C Burges THE MNIST DATABASE of handwritten digits. http://yann.lecun.com/exdb/mnist/
[40] Zhao Liang The Robotics Institute, Carnegie Mellon University GMU Image Data: hand. http://vasc.ri.cmu.edu/idb/html/motion/hand/index.html
[41] Escalante H J, Ponce-Lopez V, Wan J, Riegler M A, Chen B, Clape’s A, Escalera S, Guyon I, Baro X, Halvorsen P and et al 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun 67-73