Improved Locally Linear Embedding by Using Adaptive Neighborhood Selection Techniques

Zhiyou Zhang[1]

Nanjing College of Information Technology, Nanjing, Jiangsu, 210023, China
E-mail: `zhangzy@njcit.cn`

Jiayan Zhou[2]

JiangSu Province Nanjing Engineering Vocational College, Nanjing, Jiangsu, 210023, China
E-mail: `ivy2zhou@126.com`

Haijian Shao

School of CSE, Jiangsu University of Science and Technology, Jiangsu, 212000, Zhenjiang, China
E-mail: `jsj_shj@just.edu.cn`

Anping Bao

Nanjing College of Information Technology, Nanjing, Jiangsu, 210023, China
E-mail: `baoap@njcit.cn`

Unsupervised learning algorithm locally linear embedding (LLE) is a typical technique which applies the preserving embedding method of high dimensional data to low dimension. The number of neighborhood nodes of LLE is a decisive parameter because the improper value will affect the manifold structure in the local neighborhood and lead to the lower computational efficiency. Based on the fact that the shortest path in low-density can be established easily, this paper proposes an improved LLE method by using the sparse matrix in combination with the weights related to each point used for the linear combination in local neighborhood. The correlation dimension between high and low dimension is used to estimate the proper number of the reduced dimension, thereby selecting the best upper bound for the non-uniform manifold. Finally, we provide the experimental evaluation to verify the effectiveness of the proposed algorithm.

---

1. Introduction

The data collection and the storage techniques have been enhanced over the past few decades as the development of modern industry and sensing technology. The number of available data used for observation or measurement about the engineering, economics and population is on the rise. This increases the difficulty of the theoretical analysis related to the multi-dimensional data such as signals, two-dimensional or three-dimensional images and stereo space etc in the statistical pattern [1].

Typically, it is not all of the data in the given samples are "useful" and "necessary" in the theoretical development. The data collected from the real world contains many noises, corrupt and missing value, so the data preprocessing such as data cleaning and filter is necessary before the model analysis. Specially, the data with high-dimension presents many computational efficiency and theoretical analysis challenges. The dimensionality reduction method is an effectiveness approach to extract the feature. This is because, on the one hand, the computational complexity will increase if all the available data are utilized for model analysis and sample investigation. On the other hand, the sample only contains the limited information because some data collected from one resource usually provide the redundant information. The data dimensionality reduction can capture the data characteristic as well as reduce the data dimensionality. Many methods had been developed for this issue in recent years although these methods came from many different disciplines [2]. Many famous techniques, such as principal components analysis (PCA) [3], multidimensional scaling (MDS) [4], independent component analysis (ICA) [5] and Factor analysis [6], can model the linear subspace (manifolds) for a given high dimensional observation. However, these methods have one common drawback: They only focus on the data characteristics in linear subspace. Therefore, many excellent methods, such as kernel PCA [7, 8], locally linear embedding (LLE) [9, 10], Laplacian eigenmaps (LEM) [11], Hessian eigenmaps (Hessian LLE) [12], locality preserving projections (LPP) [13] and Landmark isomap[14] can be used to solve the high dimensionality problem in nonlinear space.

In the real application, the structure of data used for the analysis usually is mainly nonlinear. Locally linear embedding (LLE) can preserve embedding via the low-dimension calculation, which is a typically unsupervised learning algorithm through finding a point set in the nearest neighbors. The main advantage of LLE is that its topology of the original high-dimension data used for observation is same as the low-dimensional one. However, its performance depends on the number of neighborhood nodes k, which is a decisive parameter because the improper k will affect the manifold structure in the local neighborhood, and even lower the computational efficiency.

Based on outlined discussion, this paper proposes an adaptive neighborhood selection based on the sparse matrix and the weights about each point used for the linear combination according to fundamental fact that the shortest path in low-density can be established easily. This paper is organized as follows. In section 2, the formalization for manifold algorithm, dimensionality reduction representation is given and the locally linear embedding method is illustrated and the number of neighborhood nodes k selection method is presented. The correlation dimension measurement is given to calculate the proper number of the low-dimension in Section 3. The numerical simulation is applied to demonstrate the performance of the improved LLE in Section 4. In Section 5, the derived results and prospective research issues is summarized and discussed.

2. Theoretical Analysis

In this section, the manifold algorithm principle and the mathematical representation related to dimensionality reduction are given.

2.1 Dimensionality Reduction and LLE

The basic processing steps of the dimensionality reduction can be indicated as: $s = (s_1, s_2, ..., s_k)^T$ is a proper dimension with lower size $k \leq p$, which is used to represent the variable $x = (x_1, x_2, ..., x_p)^T$, with the mean $E(x) = \mu = (\mu_1, \mu_2, ..., \mu_p)^T$ and variance $\sigma\{(x - \mu)(x - \mu)^T\} = \sum_{p \times p}(x, \mu)$. Note that, "variable" can be appropriately treated as the "features" in machine learning. The dimensionality reduction is defined by

$$s = wx, s_i = \sum_{j=1,...,n} w_{i,j} x_i = w_{i,1} x_1 + ... + w_{i,n} x_n, i = 1, ..., p \tag{2.1}$$

where $\hat{\mu}_i = \frac{1}{n} \times \sum_{j=1,...,n} x_{ij}$ and $\hat{\sigma}_i = \frac{1}{n} \times \sum_{j=1,...,n} (x_{ij} - \hat{\mu}_i)^2$ are the mean and variance with respect to $x = \{x_{ij}, 1 \leq i \leq p, 1 \leq j \leq n\}$, respectively. w is the corresponding transformation weight matrix. In real world applications, many data have nonlinear structure compared to the linear data.

Locally linear embedding (LLE) is designed almost the same time about the isomap, which is a typically nonlinear dimensionality reduction technique using the overlapping coordinate patches collection. The local feature will be focused on LLE compared to the linear dimensionality reduction method such as principal component analysis and nonlinear dimensionality technique Isomap etc, LLE tries to construct a graph used to represent the data x_i in the high-dimensional dataset. LLE can establish the nonlinear mapping from the high-dimensional dataset to low-dimensional one even in the non-convex manifold by using the local point in combination with the corresponding neighborhood point to represent the manifold. At the same time, the weight of the linear combination is remained unchanged.

Assume that the weight w is calculated by linear combination with the given x_i and the corresponding points in the neighborhood. Then the hyperplane related to outline points can be fitted by the LLE in the local manifold. Thus the linear combination w is invariant for transformation, rotation and scaling such that the local feature in the low-dimensional remained well. The cost function about the LLE is defined by:

$$\varphi(x) = \sum_{i \in \Lambda} \left\| x_i - \sum_{j=1,...n} \omega_{ij} x_{ij} \right\|^2, subject\ to\ \left\| x^{(k)} \right\|^2 = 1, k = 1, ..., n \tag{2.2}$$

where $x^{(k)}$ represents the kth column corresponding to variable x_i in the fundamental solution matrix except the trivial solution $x=0$.

Roweis and Saul[15] proposed that the minimizing the cost function in the low-dimensional space can be implemented via the eigenvectors corresponding to the smallest non-zero eigenvalue related to the matrix $(I - W)^T (I - W)$. The W is a sparse matrix with size $n \times n$, and the reconstructed weight is equal to the element at the ith row and jth column. I is a unit matrix with same size as W. Essentially, the LLE is a manifold algorithm about the linear combination [16, 17], the successful example in practical application such as hyperanalytic

solution [18], auditory localization [19] etc, and the negative examples such as the visualization of the synthesis of biomedical datasets [20].

2.2 Strategy Design

The neighbor point selection is a critical factor for the performance of the LLE, and the improper number of the neighbor point will bring the instability topology of the manifold algorithm. For instance, the performance of the LLE is instable in the computer vision[21]. One possible explanation is that the "short circuit phenomenon" appears in the LLE learning. In addition, the performance of LLE will be degraded when the data with high-dimension is "overlapping" due to the covariance constraints [22]. Typically, the quality of the established mapping has potential influence on the neighbor point selection. However, this method has high computational cost and brings the lower computational speed. So this paper proposes an adaptive neighborhood selection technique to avoid the "short circuit phenomenon" based on the principle that the shortest path in low-density can be established easily because the largest neighborhood points results of the geodesic distance cannot be replaced by the shortest path. Thus the sparse matrix about the weight value can be adjusted via the information produced by the neighborhood points having "short circuit". The computing speed for searching is very fast so that the adjustment of the sparse matrix is very fast as well. The computational complexity and the information storage is about O(nd), in other word, it is suitable for the practical application. The proposed strategy for improved LLE (ILLE) is given in detail as follows.

The manifold denoising based on the given graph is constructed by the submanifold. The distribution probability of the data with noise in a submanifold for utilized sample can be computed by

$$P_X(x) = (2\pi\sigma^2)^{-\frac{k}{2}} \int_{\mathcal{M}} e^{-\frac{\|x-i(\theta)\|^2}{2\sigma^2}} p(\theta) dV(\theta) \tag{2.3}$$

where θ is the parameter of distribution function, $i(\theta)$ and σ are the mean and variance of the variable x. $p(\theta)$ is the kernel function. $V(\theta)$ is the volume integral element. The Gaussian measure related to the outlined formula is equivalent to the heat kernel $p_t(x,y) = (4\pi t)^{-\frac{k}{2}} \exp\left(-\frac{\|x-y\|^2}{4t}\right)$, which is essentially a diffusion process on R^k. The distribution of the submanifold is unknown and calculated by the true data-generating probability measurement formula in [23]. Based on the data preprocessing, the performance of the unsupervised learning algorithm LLE will be improved by the manifold denosing algorithm.

Assume the given real-valued $N \times D$ dataset, and the corresponding neighbors have locally linear patch. Characterize the local geometry information at each patch, and initial the $N \times k$ sparse marix, where k is specified as a proper real-valued number used to reconstruct the weight based on the data and its neighbors.

Error function is set in order to adjust the number of the neighbor points k until the error reaches the minimum, which is defined by

$$error_{rec}(w) = \sum_i \left(x_i - \sum_j w_{ij} x_j\right)^2 \tag{2.4}$$

The connection strength w_{ij} at the jth data corresponding to the ith reconstruction is calculated. $w_{ij}=0$ if there is not any affiliation between x_j and the set of neighbor point x_i, the rows in the weight matrix satisfying $\sum_j w_{ij} = 1$.

The High-dimensional observation xi is mapped into a low-dimensional variable yi based on the proper k in steps to minimize the following cost function

$$error_{rec}(w) = \sum_i \left(y_i - \sum_j w_{ij} y_j \right)^2 \tag{2.5}$$

The optimization processing is finished and the proposed approach is fast convergence and performance robust.

3. Intrinsic Dimensionality Estimation

The intrinsic dimension for the reduced low-dimension manifold needed be estimated by the proper method because the improper dimension will destroy or lose the feature of the high-dimension manifold. In [24], Christopher figured out that the intrinsic dimension can be defined via a scale-dependent quantity method. For instance, the Nystrom method is a typical technique used for the numerical approximation in manifold algorithm. In this section, the correlation dimension is used for the intrinsic dimension estimation, which is similar to fractal dimensions [25] used in fractal geometry. If the finite set $\zeta_n = \{x_1, x_2, ..., x_n\}$ in metric space X, let

$$C_n(r) = \alpha \sum_{i=1,...,n} \sum_{j=i+1,...,n} I_{\left\{ \|x_i - x_j\|^2 < r \right\}} \tag{3.1}$$

where $\alpha = \dfrac{2}{n(n-1)}$, n is the length of the given dataset. I_A is the corresponding index set.

The integral is given by $C(r) = \lim_{r \to 0} C_n(r)$ for a countable subset $S = \{x_1, x_2, ...\} \subset X$. If $C(r)$ can be derived, then the correlation dimension of ζ_n is defined by $CD_{corr} = \lim_{r \to 0} \dfrac{\log C(r)}{\log r}$. Thus, the intrinsic dimension is estimated before the nonlinear mapping is established.

4. Numerical Simulation

In this section, the famous high-dimensional manifold [26, 27] such as Swiss, changing swiss, broken swiss, helix and twinpeaks are used for the performance verification of the ILLE (Improved LLE) respectively. In addition, three methods related to the dimension estimation are given in order to compare the dimension estimation under different criteria. The original size of the original dataset is set as: 5000*3, the detailed information is given in table 1.

Dataset	CorrD	NearND	GMST	TLLE(ET)	HessianLLE(ET)	ILLE(ET)
Swiss	2	1	2	13.2486	**12.7917**	9.2968
Cswiss	2	1	2	13.4927	12.7648	9.6872
Bswiss	2	1	2	12.1695	12.7083	9.1383
helix	2	1	2	12.2348	12.5893	9.5459
twinpeaks	2	1	2	12.9661	12.9476	9.0787

Table 1: Performance Comparison

where the CorrD represents the outlined correlation dimension, NearND denotes the methods based on the nearest neighbor dimension, GMST indicates the method based on the analysis of the geodesic minimum spanning tree.TLLE(ET) is the elapsed time in traditional LLE in seconds, ILLE(ET) is the elapsed time by the improved LLE in seconds, Cswiss and

Bswis are the Changing swiss and Broken swiss respectively. The intrinsic dimension estimation is calculated by the following formula,

$$\hat{CD}_{corr} = \left\lceil \frac{\log C(r_2) - \log C(r_1)}{\log r_2 - \log r_1} \right\rceil = \left\lceil \frac{-3.3802 + 5.9292}{-2.3683 + 3.6910} \right\rceil = \lceil 1.9271 \rceil = 2 \quad (4.1)$$

where $\lceil \ \rceil$ represents the upper bound of the given number in integer. The three-dimensional (3D) visualization about the swiss, broken swiss data set and the corresponding two-dimensional (2D) dimensionality reduction results by the ILLE are given in Figs. 1-4, respectively.
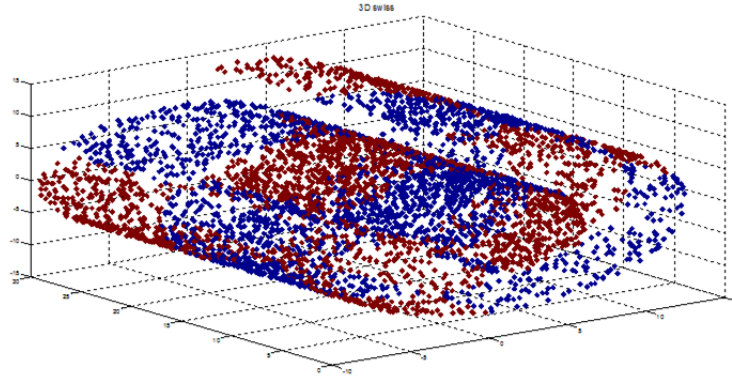

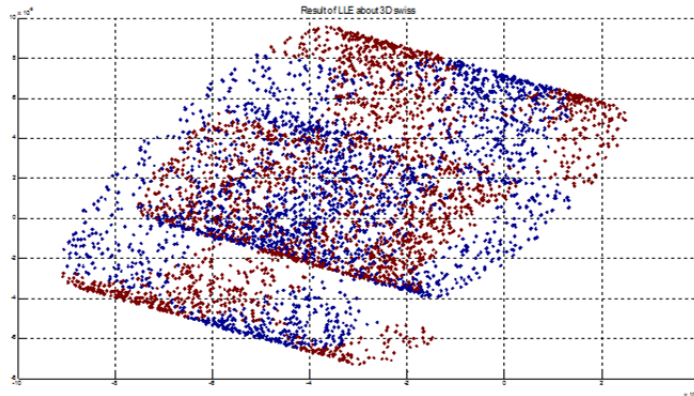**Figure 1:** 3D Visualization about the Swiss Dataset

Figure 2: 2D Dimensionality Reduction Results by the ILLE for Swiss Dataset
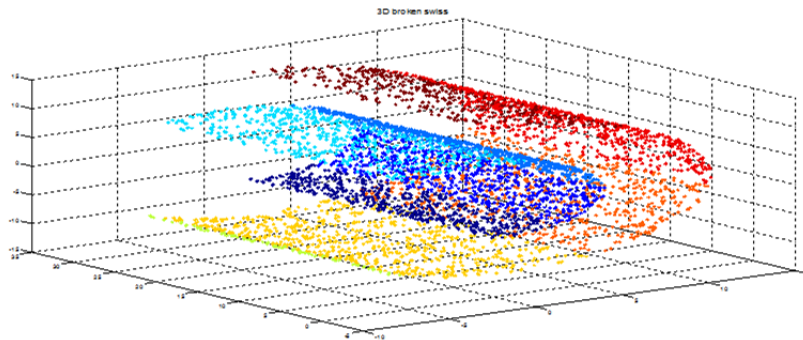

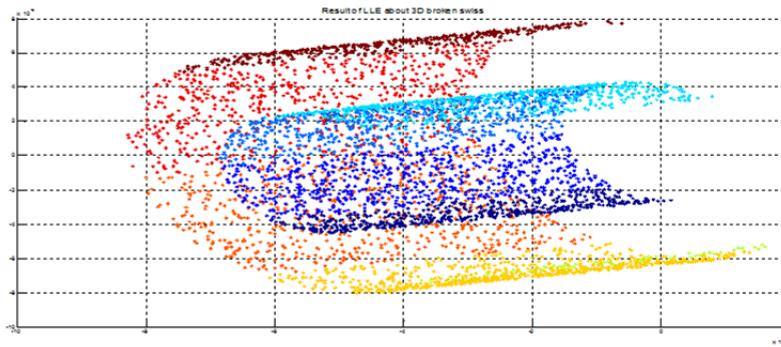**Figure 3:** 3D visualization about the broken Swiss dataset

Figure 4: 2D Dimensionality Reduction Results by the ILLE for Broken Swiss Dataset

Based on the simulation results in Table 1, the performance of the ILLE is better than TLLE, and the account of average cost time uses ILLE is reduced by about 27.10% compared to the TLLE in general, which demonstrates that the ILLE is an effective approach for improving the performance of TLLE based on the famous dataset.

5. Conclusion

This paper proposed an ILLE by using the sparse matrix and the weights related to each point. Firstly, the fundamental theory analysis is given to illustrate the basical introduction for the further analysis. Secondly, ILLE and the associated theoretical analysis is presented, the number of neighborhood nodes selection method is presented in order to improve the computational efficiency and performance of the TLLE; therefore, the performance of TLLE is improved by overcoming the shortcoming of TLLE. Thirdly, the correlation dimension technique is used to estimate the intrinsic dimension of the high-dimension manifold. Fourthly, the numerical simulation based on the famous high-dimensional manifold is used to verify the effectiveness and practicability of the proposed strategy. Finally, this paper is summerised in the last section. We believe that the proposed results have significant reference value on theoretical analysis and actual aspect, and it can be extended for other high-dimensional manifold algorithm reduction.

References

[1] Guillemard, M. and A. Iske, *Curvature analysis of frequency modulated manifolds in dimensionality reduction.* Calcolo, 2011. 48(1): p. 107--125.

[2] Zhang, Y. and Z. Zhou, *Multilabel dimensionality reduction via dependence maximization.* ACM Transactions on Knowledge Discovery from Data (TKDD), 2010. 4(3): p. 14.

[3] Thomaz, C.E. and G.A. Giraldi, *A new ranking method for principal components analysis and its application to face image analysis.* Image and Vision Computing, 2010. 28(6): p. 902--913.

[4] Wei, H., et al., *Multidimensional scaling analysis for passive moving target localization with TDOA and FDOA measurements.* Signal Processing, IEEE Transactions on, 2010. 58(3): p. 1677--1688.

[5] Mijovic, B., et al., *Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis.* Biomedical Engineering, IEEE Transactions on, 2010. 57(9): p. 2188--2196.

[6] Dehak, N., et al., *Front-end factor analysis for speaker verification.* Audio, Speech, and Language Processing, IEEE Transactions on, 2011. 19(4): p. 788--798.

[7] Alzate, C. and J.A. Suykens, *Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. 32(2): p. 335--347.

[8] Ogawa, T. and M. Haseyama, *Missing intensity interpolation using a kernel PCA-based POCS algorithm and its applications.* Image Processing, IEEE Transactions on, 2011. 20(2): p. 417--432.

[9] Sun, B., et al., *Feature fusion using locally linear embedding for classification.* Neural Networks, IEEE Transactions on, 2010. 21(1): p. 163--168.

[10] Wang, J., et al., *Locality-constrained linear coding for image classification.* 2010, IEEE. p. 3360--3367.

[11] Xie, L., et al., *Laplacian eigenmaps for automatic story segmentation of broadcast news.* Audio, Speech, and Language Processing, IEEE Transactions on, 2012. 20(1): p. 276--289.

[12] Si, S., D. Tao and K. Chan, *Evolutionary cross-domain discriminative hessian eigenmaps. Image Processing,* IEEE Transactions on, 2010. 19(4): p. 1075--1086.

[13] Lu, J. and Y. Tan, *Regularized locality preserving projections and its extensions for face recognition.* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2010. 40(3): p. 958--963.

[14] Chi, J. and M.M. Crawford, *Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity.* Geoscience and Remote Sensing Letters, IEEE, 2013. 10(4): p. 711--715.

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

[16] He, X., et al., *Neighborhood preserving embedding.* 2005, IEEE. p. 1208--1213.

[17] Kokiopoulou, E. and Y. Saad, *Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007. 29(12): p. 2143--2156.

[18] Chang, H., D. Yeung and Y. Xiong, *Super-resolution through neighbor embedding.* 2004, IEEE. p. I--I.

[19] Duraiswami, R. and V.C. Raykar, *The manifolds of spatial hearing.* 2005, IEEE. p. iii--285.

[20] Lim, I.S., et al., *Planar arrangement of high-dimensional biomedical data sets by isomap coordinates.* 2003, IEEE. p. 50--55.

[21] Jenkins, O.C. and M.J. Mataric, *Deriving action and behavior primitives from human motion data.* 2002, IEEE. p. 2551--2556.

[22] van der Maaten, L.J., E.O. *Postma and H.J. van den Herik, Dimensionality reduction: A comparative review.* Journal of Machine Learning Research, 2009. 10(1-41): p. 66--71.

[23] Hein, M. and M. Maier, *Manifold denoising.* 2006. p. 561--568.

[24] Burges, C., *Dimension reduction: A guided tour.* Machine Learning, 2009. 2(4): p. 275--365.

[25] Lapidus, M.L. and M. Van Frankenhuijsen, *Fractal geometry, complex dimensions and zeta functions: geometry and spectra of fractal strings.* 2012: Springer Science & Business Media.

[26].Seung, H.S. and D.D. Lee, *The manifold ways of perception.* Science, 2000. 290(5500): p. 2268--2269.

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