

Advanced in Control Engineering and Information Science

Locally Linear Embedding by Linear Programming

Zhijie Xu^a, Jianqin Zhang^b, Zhidan Xu^c, Zhigang Chen^d*^a

^a*School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China*

^b*School of Surveying and Mapping Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China*

^c*Basis Science Department, Harbin Commercial University, Harbin 150028, China*

^d*School of Mechanical-electronic and Automobile Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China*

Abstract

Dimensionality reduction has always been one of the most challenging tasks in the field of data mining. As a nonlinear dimensionality reduction method, locally linear embedding (LLE) has drawn more and more attention and applied widely in face image processing and text data processing. But this method is usually sensitive to noise, which limits its application in many fields. In this paper, we propose a locally linear embedding algorithm by linear programming (LLE by LP), and the experiments demonstrate the effectiveness of the approach in reducing the sensitivity to noise.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS 2011]

Keywords: locally linear embedding, linear programming, nonlinear dimensionality reduction.

1. Introduction

In the field of data mining and pattern recognition, dimensionality reduction has always been one of the most important and challenging tasks. Traditional methods to perform dimensionality reduction are mainly linear, such as principle component analysis (PCA), factor analysis and independent component analysis (ICA) [1, 2]. In early years, many nonlinear methods have been proposed to perform

* Corresponding author.

E-mail address: zhijieXu@163.com.

dimensionality reduction, such as kernel principal component analysis (KPCA), kernel linear discriminate analysis (KLDA), and principal curves (HS) [3, 4]. Recently, a new unsupervised learning technique for nonlinear mapping-manifold learning has captured the attention of many researchers in the field of machine learning and cognitive sciences. The major algorithms include isometric mapping (ISOMAP) and locally linear embedding (LLE) [5, 6]. The approaches can be used for discovering the intrinsic distribution and geometry structure of nonlinear high dimensional data effectively. LLE is simple to implement, and its optimizations do not involve local minima. Although LLE demonstrates good performance on a number of artificial and realistic data sets, its weakness can not be ignored. In this paper, we shall pay attention to the problem of sensitivity to noise, and propose a locally linear embedding algorithm by linear programming. Then experiments are performed on the well-known data sets scurve and swiss roll, and the results of LLE and LLE by LP are compared and discussed. The experiments demonstrate the effectiveness of the approach.

2. Locally linear embedding

Locally linear embedding (LLE) developed by Roweis and Saul [6] in 2000 is a promising method for the problem of nonlinear dimensionality reduction of high-dimensional data. Unlike classical linear dimensionality reduction methods, LLE provides information that can reveal the intrinsic manifold of data. It assumes that each data point and its neighbors lie on a locally-linear patch, and then applies this patch in a low space to generate data configuration. LLE recovers global non-linear structures from locally-linear fits. Here we review the LLE algorithm in its most basic form and more details can be found in [1].

LLE maps an input data set $X = \{x_1, x_2, \dots, x_n\}$, $x_i \in R^d$ globally to an output data set $Y = \{y_1, y_2, \dots, y_n\}$, $y_i \in R^m$, where $m \ll d$. Assuming the data lies on a nonlinear manifold which locally can be approximated linearly, the algorithm has three sequential steps:

Step 1. Determining neighbors: The K closest neighbors are selected for each point using a distance measure such as the Euclidean distance.

Step 2. Calculating reconstruction weights: In order to determine the value of the weights, the reconstruction errors are measured by the cost function:

$$\varepsilon(W) = \sum_{i=1}^n \left\| x_i - \sum_{j=1}^n W_{ij} x_j \right\|_2 \quad (2.1)$$

The weights W_{ij} are determined by minimizing the cost function defined in Equation (2.1) subject to two constraints: $\sum_j W_{ij} = 1$ and $W_{ij} = 0$ if x_j is not a neighbor of x_i . The weights are then determined by a least-squares minimization of the reconstruction error.

Step 3. Finding lower-dimensional embedding Y : Define a cost embedding function:

$$\Phi(Y) = \sum_{i=1}^n \left\| y_i - \sum_{j=1}^n W_{ij} y_j \right\|_2 \quad (2.2)$$

The lower-dimensional coordinate y_i are computed by minimizing the cost function defined in Equation (2.2) subject to two constraints: $\sum_i y_i = 0$ and $\sum_i y_i y_i^T / N = I$, where I is a $m \times m$ identity matrix. The above optimization problem can be solved by transforming it to an eigenvalue problem.

3. Locally linear embedding by linear programming

When LLE is used for dimensionality reduction of high-dimensional data, mapping quality is directly affected by the weight matrix. The weight matrix is quite sensitive to noise of the input data set X . Especially when the eigenvalues of $X^T X$ are small, even the desired results can not be obtained. In this section, we will present a robust LLE algorithm which accomplishes denoising and nonlinear dimensionality reduction simultaneously.

Given the input data set $X = \{x_1, x_2, \dots, x_n\}$, $x_i \in R^d$, we would like to denoise the input data and reduce its dimension simultaneously. Let us introduce an intermediate variable \bar{x}_i , which describes the denoised data. In order to calculate the manifold reconstruction weights, we should minimize the distance between \bar{x}_i and the linear manifold constructed by x_i 's neighbors, which is modeled as follows:

$$\min_{W_{ij}} d_1(\bar{x}_i, \sum_{j=1}^n W_{ij} x_j) \tag{3.1}$$

where d_1 is a distance function. We have to constrain \bar{x}_i to be close to x_i , otherwise the denoised data becomes completely independent from the original data. This requirement can be formulated as:

$$\min_{\bar{x}_i} d_2(\bar{x}_i, x_i) \tag{3.2}$$

where d_2 is also a distance function. Summarizing the above two aspects, we construct the optimization problem as follows:

$$\min_{\bar{x}_i, W_{ij}} \nu d_1(\bar{x}_i, \sum_{j=1}^n W_{ij} x_j) + d_2(\bar{x}_i, x_i) \tag{3.3}$$

Considering that most properties of x_i stay unchanged in \bar{x}_i , we choose l_1 -norm as d_2 . For d_1 , the l_∞ -norm is chosen for convenience of problem solving. Let $\alpha = \bar{x}_i - x_i$, then the optimization problem (3.3) can be formulated as:

$$\min_{\alpha, W_{ij}} \nu \left\| x_i + \alpha - \sum_{j=1}^n W_{ij} x_j \right\|_\infty + \frac{1}{d} \|\alpha\|_1 \tag{3.4}$$

where ν is a constant to determine the extent of noise. We can formulate (3.4) as:

$$\begin{aligned} \min_{\alpha, W_{ij}, \xi} \quad & \frac{1}{d} \sum_{i=1}^d |\alpha_i| + \nu \xi \\ \text{s.t.} \quad & \left| x_{ik} + \alpha_k - \sum_{j=1}^n W_{ij} x_{jk} \right| \leq \xi, k = 1, \dots, d \end{aligned} \tag{3.5}$$

where x_{ik} is the k th element of x_i , and α_k is the k th element of α .

Let $\alpha = \alpha^+ - \alpha^-$, $\alpha_k^+, \alpha_k^- \geq 0, k = 1, \dots, d$, then (3.5) can be formulated as the following linear programming problem:

$$\begin{aligned}
 & \min_{\alpha^+, W_{ij}, \xi} \quad \frac{1}{d} \sum_{i=1}^d (\alpha_i^+ + \alpha_i^-) + v\xi \\
 & \text{s.t.} \quad \left| x_{ik} + \alpha_k^+ - \alpha_k^- - \sum_{j=1}^n W_{ij} x_{jk} \right| \leq \xi, \\
 & \quad \alpha_k^+, \alpha_k^- \geq 0, k = 1, \dots, d
 \end{aligned} \tag{3.6}$$

In sum, the algorithm has three sequential steps:

Step 1. Selecting neighbors: The K closest neighbors are determined for each point using a distance measure such as the Euclidean distance.

Step 2. Calculating reconstruction weights: Calculate reconstruction weights W by solving the linear programming problem (3.6) subject to the constraint $\sum_j W_{ij} = 1$:

Step 3. Finding lower-dimensional embedding Y using the weights calculated in Step 2: Define a cost embedding function:

$$\Phi(Y) = \sum_{i=1}^n \left\| y_i - \sum_{j=1}^n W_{ij} y_j \right\|_2 \tag{3.7}$$

The lower-dimensional coordinate y_i are computed by minimizing the cost function defined in Equation (3.7) subject to two constraints: $\sum_i y_i = 0$ and $\sum_i y_i y_i^T / N = I$, where I is a $m \times m$ identity matrix. The above optimization problem can be solved by transforming it to an eigenvalue problem.

4. Experiments

Experiments are performed on the well-known data sets scurve and swiss roll. The number of data points of each data set is 2000, and the dimension is 3, and the intrinsic dimensionality is 2. Figure 1(a) shows the experimental data we use.



Fig. 1. (a) experimental data; (b) experimental result of scurve and swiss roll using LLE

When using LLE for dimensionality reduction, two parameters will have to be set: the intrinsic dimensionality and the number of neighbors K . In our experiments, the intrinsic dimensionality is 2. For the number of neighbors, we set it from 5 to 25. The results show that when the parameter is set 7 for scurve and 9 for swiss roll, the performance is best of all. The LLE mapping results are shown in figure 2. The resulting embedding shows that LLE can unraveled the underlying two dimensional structure successfully in spite of some deformation.

In order to test the effectiveness of locally linear embedding by linear programming, we impose additive Gaussian noise on the data. Specifically, $\sigma = 0.02$ for the first element of scurve data set, and $\sigma = 0.5$ for the first and second elements of swiss roll data set. Then we map the 3-D data with additive noise to 2-D using LLE and LLE by LP. The experimental results are shown in figure 2. In the

experiment, the number of neighbors is set 23, and $V = 1/3$ for scurve data set, and $V = 2/3$ for swiss roll data set.



Fig.2. (a) experimental results of LLE and LLE by LP for scurve data with noise; (b) experimental results of LLE and LLE by LP for swiss roll data with noise

From the experimental results, we notice that LLE is sensitive to noise of the input data set. When the input data set is corrupted by noise, the mapping quality is bad. Especially for swiss roll data set, distinct parts of the data set are mapped on top of each other. The mapping using LLE by LP looks better than that of original LLE. Although some distortion is introduced, the mapping clearly exhibits the two degrees of freedom. In sum, the sensitivity to noise is reduced in LLE by LP and the mapping quality is basically satisfactory.

5. Conclusion

In this paper, we investigated the use of locally linear embedding for nonlinear dimensionality reduction. In order to solve the problem of sensitivity to noise, we presented locally linear embedding algorithm by linear programming. The effectiveness of our method is demonstrated by experiments on scurve data set and swiss roll data set, which shows its encouraging performance.

References

- [1]P.A. Devijver and J. Kittler. *Pattern recognition, a statistical approach*. Prentice-Hall, London, 1982.
- [2]R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*. John Wiley & Sons, New York, NY, 2nd edition, 2001.
- [3]T.G. Dietterich. Ensemble learning. In : *The Handbook of Brain Theory and Neural Networks* , 2nd Edition. Cambridge , MA : MIT Press , 2002
- [4]T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 1989, 84:502-516.
- [5]J.B. Tenenbaum, V. de Silva, J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000 , 90 (5500) : 2319-2323
- [6]L.K. Saul and S.T. Roweis. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000 ,290:2323-2326.

Acknowledgements

The paper is sponsored by National Natural Science Foundation of China (51004005) and Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality (PHR201107123).