

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.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات