



Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Dimensionality reduction: An interpretation from manifold regularization perspective

Mingyu Fan^a, Nannan Gu^b, Hong Qiao^{c,*}, Bo Zhang^d

^a College of Mathematics & Information Science, Wenzhou University, Zhejiang Province 325035, China

^b School of Statistics, Capital University of Economics and Business, Beijing 100070, China

^c State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

^d LSEC and Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:

Received 8 April 2013

Received in revised form 22 January 2014

Accepted 4 March 2014

Available online xxxx

Keywords:

Dimensionality reduction

Manifold regularization

Feature mapping

Manifold learning

Out-of-sample extrapolation

ABSTRACT

In this paper, we propose to unify various dimensionality reduction algorithms by interpreting the Manifold Regularization (MR) framework in a new way. Although the MR framework was originally proposed for learning, we utilize it to give a unified treatment for many dimensionality reduction algorithms from linear to nonlinear, supervised to unsupervised, and single class to multi-class approaches. In addition, the framework can provide a general platform to design new dimensionality reduction algorithms. The framework is expressed in the form of a regularized fitting problem in a Reproducing Kernel Hilbert Space. It consists of one error part and two regularization terms: the complexity term and the smoothness term. The error part measures the difference between the estimated (low-dimensional) data distribution and the true (high-dimensional) data distribution or the difference between the estimated and targeted low-dimensional representations of data, the complexity term is a measurement of the complexity of the feature mapping for dimensionality reduction, and the smoothness term reflects the intrinsic structure of data. Based on the framework, we propose a Manifold Regularized Kernel Least Squares (MR-KLS) method which can efficiently learn an explicit feature mapping (in the semi-supervised sense). Experiments show that our approach is effective for out-of-sample extrapolation.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Dimensionality Reduction (DR) refers to the process of extracting essential low-dimensional structures from high-dimensional data. It is motivated by the growing amounts of high-dimensional data and the fact that many of the features in high-dimensional data are relevant with each other such that they can be represented by a few ones. Consider, for instance, the gray-scale images of an object taken under fixed lighting conditions with a moving camera. Each of such an image would intrinsically be represented by a brightness value and two camera orientation measures (e.g., up–down and left–right angles). DR offers a powerful tool to attack the ‘curse of dimensionality’ for various applications, such as face recognition

* Corresponding author. Tel.: +86 1062636367.

E-mail addresses: fanmingyu@amss.ac.cn, fanmingyu@wzu.edu.cn (M. Fan), nngu@cueb.edu.cn (N. Gu), hong.qiao@mail.ia.ac.cn (H. Qiao), b.zhang@amt.ac.cn (B. Zhang).

<http://dx.doi.org/10.1016/j.ins.2014.03.011>

0020-0255/© 2014 Elsevier Inc. All rights reserved.

[20,24], biometrics [28,33,40,54], multimedia information retrieval [18], document clustering [5], pose estimation and tracking [44,38,45], and is of central importance in pattern recognition and machine learning.

In the past decades, there have emerged various kinds of methods to extract the intrinsic low-dimensional representation of data. Among them, Principal Component Analysis (PCA) [23] and Linear Discriminant Analysis (LDA) [12] have been the two most classic and well-known methods due to their relative simplicity and effectiveness. PCA finds a subspace in which the data projection has a maximum variance, while LDA attempts to achieve maximum class discrimination by maximizing inter-class distances and minimizing intra-class distances simultaneously. Both methods are referred to as linear DR methods since they can discover the linear structure of data.

As for nonlinear DR methods, manifold learning has been the main focus and has given a satisfactory performance of discovering intrinsic nonlinear structures of data. Traditional manifold learning algorithms include Isometric feature mapping (Isomap) [39], Locally Linear Embedding (LLE) [36], Laplacian Eigenmap (LE) [1], Diffusion maps [27,34], Hessian LLE (HLLE) [10], Local Tangent Space Alignment (LTSA) [57] and Maximum Variance Unfolding (MVU) [46–48]. In all the above methods, data are assumed to distribute on an intrinsically low-dimensional manifold and the low-dimensional embedding of the data set is then calculated based on different intuitions.

DR algorithms are usually designed based on specific motivations, and different algorithms may be fit for different application situations. Common properties and intrinsic differences among these algorithms are not completely clear. Therefore, it will be more informative to provide a unified framework for better interpreting the common properties and intrinsic differences of DR algorithms.

In this paper, we make the following two main contributions to dimensionality reduction.

- We propose to utilize the Manifold Regularization (MR) framework which was originally presented for learning from labeled and unlabeled examples [2], to give a unified treatment for various DR algorithms. We can use this framework to understand and explain many of the state-of-art DR algorithms from a systematic view.
- The proposed framework can be used as a general platform to design new DR algorithms. Based on the framework, we propose a new DR approach in the least squares sense. The proposed approach has three features: (1) it can produce an explicit nonlinear feature mapping, (2) it works in the semi-supervised way, and (3) different kinds of regularization terms and prior knowledge can be incorporated into this approach, which makes it flexible to use in many tasks, such as data visualization and data classification.

The rest of the paper is organized as follows. Section 2 introduces the MR framework and its new interpretation for dimensionality reduction. In Section 3, the framework is utilized to unify many of the popular DR methods. In Section 4, based on the framework, we propose the Manifold Regularized Kernel Least Squares (MR-KLS) approach. Then, from the MR-KLS approach, and by utilizing the LLE reconstruction coefficients, the LLE-KRLS algorithm is derived for the out-of-sample extrapolation of learned manifolds. In Section 5, experiments are conducted on several benchmark data sets, and the effectiveness of our LLE-KRLS algorithm is shown. Finally, some conclusions are given in Section 6.

2. The MR framework for dimensionality reduction

In this paper, we propose to interpret the MR framework in a new way so as to give a unified treatment for DR algorithms. Thus, in this section, we first review the MR framework and then give a new interpretation of the MR framework for dimensionality reduction.

2.1. The MR framework

The Manifold Regularization (MR) [2] framework was proposed for learning based on the theory of Reproducing Kernel Hilbert Spaces (RKHSs). Given l labeled examples $\{(\mathbf{x}_i, y_i^{label})\}_{i=1}^l$ and $(N-l)$ unlabeled examples $\{\mathbf{x}_i\}_{i=l+1}^N$, where $\mathbf{x}_i$ ($\forall i \in \{1, \dots, N\}$) are the data points and y_i^{label} is the corresponding class label of $\mathbf{x}_i$ ($\forall i \in \{1, \dots, l\}$) the MR framework can be expressed as

$$f^* = \arg \min_{f \in \mathcal{H}_K} \left\{ \frac{1}{l} \sum_{i=1}^l V(\mathbf{x}_i, y_i^{label}, f) + \gamma_K \|f\|_K^2 + \gamma_I \|f\|_I^2 \right\}. \quad (1)$$

Here, f is the desired classification function, V is a loss function such as the squared loss $V(\mathbf{x}_i, y_i^{label}, f) = (y_i^{label} - f(\mathbf{x}_i))^2$ or the hinge loss $\max\{0, 1 - y_i^{label} f(\mathbf{x}_i)\}$, γ_K and γ_I are two given regularization parameters.

In the MR framework (1), the first term of the objective function is the error term that measures the loss between the given class label and the label gained by the desired classification function f , the second term is the complexity regularization term that measures the complexity of f in the RKHS $\mathcal{H}_K$ associated with certain positive semi-definite kernel $k(\cdot, \cdot)$, and the third term is the smoothness regularization term that measures the smoothness of f with respect to the geometric distribution of data. In [2], the third term $\|f\|_I^2$ is defined as

Download English Version:

<https://daneshyari.com/en/article/6858114>

Download Persian Version:

<https://daneshyari.com/article/6858114>

[Daneshyari.com](https://daneshyari.com)