Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

Data Clustering by Laplacian Regularized ℓ^1-Graph

Yingzhen Yang[1], Zhangyang Wang[1], Jianchao Yang[2],
Jiangping Wang[1], Shiyu Chang[1], Thomas S. Huang[1]
[1] Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
[2] Adobe Research, San Jose, CA 95110, USA
{yyang58, zwang119, jwang63, chang87, huang}@ifp.uiuc.edu, jiayang@adobe.com

Abstract

ℓ^1-Graph has been proven to be effective in data clustering, which partitions the data space by using the sparse representation of the data as the similarity measure. However, the sparse representation is performed for each datum separately without taking into account the geometric structure of the data. Motivated by ℓ^1-Graph and manifold leaning, we propose Laplacian Regularized ℓ^1-Graph (LRℓ^1-Graph) for data clustering. The sparse representations of LRℓ^1-Graph are regularized by the geometric information of the data so that they vary smoothly along the geodesics of the data manifold by the graph Laplacian according to the manifold assumption. Moreover, we propose an iterative regularization scheme, where the sparse representation obtained from the previous iteration is used to build the graph Laplacian for the current iteration of regularization. The experimental results on real data sets demonstrate the superiority of our algorithm compared to ℓ^1-Graph and other competing clustering methods.

Introduction

In many real applications, high-dimensional data always reside on or close to a intrinsically low dimensional manifold embedded in the high-dimensional ambient space. Clustering the data according to its underlying manifold structure is important and challenging in machine learning. Moreover, the effectiveness of graph-based clustering methods, such as spectral clustering, motivates researchers to build the graph which reflects the manifold structure of the data. ℓ^1-graph (Cheng et al. 2010), which builds the graph weight matrix by reconstructing each datum from the remaining data and the noise term through sparse representation, has been shown to be robust to noise and capable of finding datum-adaptive neighborhood for the graph construction. Also, the sparse manifold clustering method (Elhamifar and Vidal 2011) points out that the sparse graph is useful for recovering the intrinsic manifold structure in the data.

However, ℓ^1-graph performs sparse representation for each datum separately without considering the geometric information and manifold structure of the entire data. Previous research shows that the data representations which respect the manifold structure of the data produce superior results in various clustering and classification tasks (Belkin, Niyogi, and Sindhwani 2006; He et al. 2011). Inspired by ℓ^1-graph and manifold learning, we propose Laplacian Regularized ℓ^1-Graph (LRℓ^1-Graph) for data clustering in this paper. In accordance with the manifold assumption, the sparse representations of LRℓ^1-Graph are regularized using the proper graph Laplacian. As a result, the sparse representations are smoothed along the geodesics on the data manifold where nearby datums have similar sparse representations. Furthermore, motivated by the fact that the sparse representations of the ℓ^1-Graph lead to a pairwise similarity matrix for spectral clustering with satisfactory empirical performance, we propose an iterative regularization scheme which utilizes the regularized sparse representations from the previous iteration to build the graph Laplacian for the current iteration of regularization for the LRℓ^1-Graph. The iterative regularization scheme produces superior clustering performance shown in our experimental results.

This research is supported in part by ONR Grant N00014-12-1-0122; and in part by National Science Foundation Grant DBI 10-62351.

The Proposed Regularized ℓ^1-Graph

Given the data $\boldsymbol{X} = \{x_1\}_{l=1}^{n} \in \mathbb{R}^d$, ℓ^1-graph seeks for the robust sparse representation for each datum by solving the ℓ_1-norm optimization problem

$$\|x_i - \boldsymbol{B}\boldsymbol{\alpha}^i\|_2^2 + \lambda\|\boldsymbol{\alpha}^i\|_1, i = 1, 2, \ldots, n \tag{1}$$

where matrix $\boldsymbol{B} = [x_1, \ldots, x_n, \mathbf{I}] \in \mathbb{R}^{d\times(n+d)}$, λ is the weighting parameter controlling the sparsity of the sparse representation, $\mathbf{I}$ is the identity matrix, $\boldsymbol{\alpha}^i = \{\boldsymbol{\alpha}_j^i\}_{j=1}^{n+d}$ is the sparse representation for x_i under the dictionary $\boldsymbol{B}$. We denote by $\boldsymbol{\alpha}$ the $(n+d)\times n$ coefficient matrix with elements $\boldsymbol{\alpha}_{ij} = \boldsymbol{\alpha}_i^j, 1 \le i \le n+d, 1 \le j \le n$. To avoid trivial solution that $\boldsymbol{\alpha}_{ii} = 1$, it is required that $\boldsymbol{\alpha}_{ii} = 0$ for $1 \le i \le n$. Let $G = (\boldsymbol{X}, \mathbf{W})$ be the ℓ^1-graph where the data $\boldsymbol{X}$ are the vertices, $\mathbf{W}$ is the graph weight matrix and $\mathbf{W}_{ij}$ indicates the similarity between x_i and x_j. Interpreting $\boldsymbol{\alpha}_{ji}$ as the directed similarity between x_i and x_j, ℓ^1-graph (Cheng et al. 2010) builds the symmetric pairwise similarity matrix $\mathbf{W}$ using the sparse representations as below:

$$\mathbf{W} = (\boldsymbol{\alpha}_{[1:n]} + \boldsymbol{\alpha}_{[1:n]}^T)/2 \tag{2}$$

where $\boldsymbol{\alpha}_{[1:n]}$ is the first n rows of $\boldsymbol{\alpha}$. W is then feeded into the spectral clustering method to produce the clustering result.

The pairwise similarity matrix using the sparse representations $\boldsymbol{\alpha}$ plays an essential role for the performance of ℓ^1-graph based clustering. In order to obtain the sparse representations that account for the geometric information and manifold structure of the data, we employ the manifold assumption (Belkin, Niyogi, and Sindhwani 2006) which in our case requires that nearby data points x_i and x_j in the data manifold exhibit similar sparse representations with respect to the dictionary $\boldsymbol{B}$. In other words, $\boldsymbol{\alpha}$ varies smoothly along the geodesics in the intrinsic geometry. Given a pairwise similarity matrix $\hat{\mathbf{W}}$, the sparse representations $\boldsymbol{\alpha}$ that capture the geometric structure of the data according to the manifold assumption should minimize the following objective function $\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\mathbf{W}}_{ij}\|\boldsymbol{\alpha}^i-\boldsymbol{\alpha}^j\|_2^2 = \mathrm{Tr}\left(\boldsymbol{\alpha}\mathrm{L}\boldsymbol{\alpha}^{\mathrm{T}}\right)$ where $\mathbf{L}=\hat{\mathbf{D}}-\hat{\mathbf{W}}$ is the graph Laplacian matrix, and $\hat{\mathbf{D}}$ is a diagonal matrix given by $\hat{\mathbf{D}}_{ii}=\sum_{j=1}^{n}\hat{\mathbf{W}}_{ij}$. Incorporating the regularization term $\mathrm{Tr}\left(\boldsymbol{\alpha}\mathrm{L}\boldsymbol{\alpha}^{\mathrm{T}}\right)$ into (1), we obtain the objective function for LRℓ^1-graph as below:

$$\min_{\boldsymbol{\alpha}}\sum_{i=1}^{n}\|x_i-\boldsymbol{B}\boldsymbol{\alpha}^i\|_2^2+\lambda\|\boldsymbol{\alpha}^i\|_1+\gamma\mathrm{Tr}(\boldsymbol{\alpha}\mathrm{L}\boldsymbol{\alpha}^{\mathrm{T}}) \quad (3)$$
$$s.t.\quad \boldsymbol{\alpha}_{ii}=0, 1\le i\le n$$

As suggested in the manifold regularization framework (Belkin, Niyogi, and Sindhwani 2006), the pairwise similarity matrix $\hat{\mathbf{W}}$ is constructed by Gaussian kernel. We use coordinate descent to optimize the objective function (3). Since (3) is convex, the global minimum is guaranteed. In each step of coordinate descent, we minimize (3) with respect to $\boldsymbol{\alpha}^i$ with fixed $\boldsymbol{\alpha}^{-i}=\left[\boldsymbol{\alpha}^1,...\boldsymbol{\alpha}^{i-1},\boldsymbol{\alpha}^{i+1},...\boldsymbol{\alpha}^n\right]$. Namely, we optimize the following objective function in terms of $\boldsymbol{\alpha}^i$ by coordinate descent:

$$\min_{\boldsymbol{\alpha}^i}\|x_i-\boldsymbol{B}\boldsymbol{\alpha}^i\|^2+\lambda\|\boldsymbol{\alpha}^i\|_1+\gamma\mathbf{L}_{ii}(\boldsymbol{\alpha}^i)^T\boldsymbol{\alpha}^i+2(\boldsymbol{\alpha}^i)^T\sum_{j\neq i}L_{ij}\boldsymbol{\alpha}^j$$
$$s.t.\quad \boldsymbol{\alpha}_{ii}=0 \quad (4)$$

(4) can be optimized using the efficient feature-sign search algorithm (Lee et al. 2006). In addition, ℓ^1-Graph uses the sparse representations to build the pairwise similarity matrix (2) for spectral clustering and obtains satisfactory clustering results. It inspires us to propose an iterative regularization scheme, where the regularized sparse representation from the previous iteration is used to construct the graph Laplacian for the current iteration of regularization. Algorithm 1 describes the learning procedure for LRℓ^1-Graph.

Experimental Results

The coefficient matrix $\boldsymbol{\alpha}$ obtained from the learning procedure for LRℓ^1-Graph is used to construct the pairwise similarity matrix $\mathbf{W}$ by (2), and then $\mathbf{W}$ is feeded into the spectral clustering algorithm to produce the clustering result. We compare our algorithm to K-means (KM), Spectral Clustering (SC), ℓ^1-Graph and Sparse Manifold Clustering and Embedding (SMCE). The clustering results on several real data sets, i.e. UCI Wine, UCI Breast Tissue (BT) and ORL face database, are shown in Table 1 where the clustering performance is measured by Accuracy (AC) and the Normalized Mutual Information (NMI). ORL-k denotes the first k samples in the ORL face database. We use fixed empirical value $\lambda=0.1, \gamma=30, M=2$ throughout the experiments, and tune λ between $[0.1,1]$ for ℓ^1-Graph and SMCE. We observe that LRℓ^1-Graph outperforms other clustering methods by our proposed iterative regularization scheme.

Algorithm 1 Learning Procedure for LRℓ^1-Graph

Input:
The data set $\boldsymbol{X}=\{x_1\}_{l=1}^n$, the number of iterations M for iterative regularization, the initial graph Laplacian $\mathbf{L}=\hat{\mathbf{D}}-\hat{\mathbf{W}}$ computed with Gaussian kernel, the parameters λ and γ.

1: $m=1$, initialize the coefficient matrix $\boldsymbol{\alpha}=\mathbf{0}$
2: **while** $m\le M$ **do**
3: Use coordinate descent algorithm with current $\mathbf{L}$ to optimize (3) and obtain the coefficient matrix $\boldsymbol{\alpha}$.
4: Build the graph Laplacian $\mathbf{L}=\mathbf{D}-\mathbf{W}$ where $\mathbf{W}=(\boldsymbol{\alpha}_{[1:n]}+\boldsymbol{\alpha}_{[1:n]}^T)/2$, $\mathbf{D}$ is a diagonal matrix such that $\mathbf{D}_{ii}=\sum_j\mathbf{W}_{ij}$. $m=m+1$.
5: **end while**

Output: The coefficient matrix (i.e. the sparse representations) $\boldsymbol{\alpha}$.

Table 1: Clustering Results on Real Data Sets

Data Set	Measure	KM	SC	ℓ^1-Graph	SMCE	LRℓ^1-Graph
Wine	AC	0.7022	0.5618	0.6629	0.7135	**0.8371**
	NMI	0.4287	0.3522	0.4070	0.4317	**0.5903**
BT	AC	0.3396	0.4057	0.4434	0.4528	**0.7547**
	NMI	0.3265	0.3563	0.3104	0.4131	**0.6437**
ORL-100	AC	0.5817	0.3813	0.7100	0.7200	**0.9113**
	NMI	0.6188	0.5154	0.7653	0.7517	**0.9011**
ORL-200	AC	0.5310	0.3812	0.7267	0.7050	**0.7873**
	NMI	0.6654	0.5631	0.8239	0.8114	**0.8617**
ORL	AC	0.5304	0.4479	0.6958	0.7050	**0.7547**
	NMI	0.7299	0.6710	0.8446	0.8362	**0.8754**

References

Belkin, M.; Niyogi, P.; and Sindhwani, V. 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research* 7:2399–2434.

Cheng, B.; Yang, J.; Yan, S.; Fu, Y.; and Huang, T. S. 2010. Learning with l1-graph for image analysis. *IEEE Transactions on Image Processing* 19(4):858–866.

Elhamifar, E., and Vidal, R. 2011. Sparse manifold clustering and embedding. In *NIPS*, 55–63.

He, X.; Cai, D.; Shao, Y.; Bao, H.; and Han, J. 2011. Laplacian regularized gaussian mixture model for data clustering. *Knowledge and Data Engineering, IEEE Transactions on* 23(9):1406–1418.

Lee, H.; Battle, A.; Raina, R.; and Ng, A. Y. 2006. Efficient sparse coding algorithms. In *NIPS*, 801–808.