Sparse Subspace Clustering with Entropy-Norm

Liang Bai [1] **Jiye Liang** [1]

Abstract

Sparse subspace clustering (SSC) and spectral clustering (SC) are both state-of-the-art methods to identify complex clusters in high-dimensional input space. However, there are few researches to discuss the relation between them. Therefore, in this paper, we provide an explicit theoretical connection between them from the perspective of learning a data similarity matrix. We show that spectral clustering with Gaussian kernel can be viewed as sparse subspace clustering with entropy-norm (SSC+E). Compared to SSC, SSC+E can obtain a sparse, analytical, symmetrical and nonnegative similarity matrix. Besides, SSC+E makes use of Gaussian kernel to compute the sparse similarity matrix of objects, which can avoid the complex computation of the sparse optimization program of SSC. Finally, we provide the experimental analysis to compare the efficiency and effectiveness of sparse subspace clustering and spectral clustering on ten benchmark data sets. The theoretical and experimental analysis can well help users for the selection of high-dimensional data clustering algorithms.

1. Introduction

Clustering is an important problem in statistical multivariate analysis, data mining and machine learning (Han & Kamber, 2001). The goal of clustering is to group a set of objects into clusters so that the objects in the same cluster are highly similar but remarkably dissimilar with objects in other clusters (Jain, 2008). To tackle this problem, various types of clustering algorithms have been developed in the literature (e.g., (Aggarwal & Reddy, 2014) and references therein), including partitional, hierarchical, density-based, grid-based clustering, etc.

*Equal contribution [1]School of Computer and Information Technology, Shanxi University. Correspondence to: Liang Bai <bailiang@sxu.edu.cn>, Jiye Liang <ljy@sxu.edu.cn>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

Recently, increasing attention has been paid to clustering high-dimensional data which is ubiquitous in real-world data mining applications, such as image processing, text analysis, and bioinformatics et al. Sparsity is an accompanying phenomenon of high-dimensional data, which leads to "curse of dimensionality", i.e., all pairs of points tend to be almost equidistant from one another. It is a special challenge for clustering high-dimensional data. In order to solve this problem, lots of clustering algorithms have been developed in the literature (e.g., (Parsons et al., 2004; Elhamifar & Vidal, 2013) and references therein). Among them, spectral clustering and sparse subspace clustering are two state-of-the-art methods to effectively separate the high-dimensional data in accordance with the underlying subspace. Spectral clustering (Shi & Malik, 2000; Ng et al., 2001) is a representative of graph-based clustering, which first converts a data set into a graph or a data similarity matrix and then uses a graph cutting method to identify clusters. However, the clustering results of the spectral clustering are sensitive to the converted graph. In the classical spectral clustering algorithm, the graph is often constructed by kernel functions (Dhillon et al., 2007) or k-nearest neighbors (KNN) (Zhu et al., 2014). Besides, some scholars developed graph-learning methods to obtain a high-quality graph from the data set. For example, Nie et al. proposed a clustering algorithm with adaptive neighbors (Nie et al., 2014), which learns the data similarity matrix by assigning the adaptive and optimal neighbors for each data point based on the local connectivity.

Sparse subspace clustering can also be seen as a special spectral clustering. It makes use of self representation of the data to construct the sparse similarity graph and apply spectral clustering on such graph to obtain the final clustering result. In (Elhamifar & Vidal, 2009), Elhamifar and Vidal presented the SSC algorithm with $\mathcal{L}_1$-norm in detail. Furthermore, several variants of SSC have been proposed to find out the sparse representation of the data under different assumptions. Wang and Xu proposed noisy SSC to handle noisy data that lie close to disjoint or overlapping subspaces (Wang & Xu, 2016). Yang et al. proposed SSC with $\mathcal{L}_0$-norm (Yang et al., 2016). Liu et al. proposed a low rank representation of all data jointly by using the structured sparsity loss (Liu et al., 2013). Hu et al. investigated theoretically the grouping effect for self-representation based approaches and presented a smooth representation model (H. Hu & Zhou, 2014). Although the existing SSC methods already have good theoretical and practical contributions, they still need to improve some deficiencies. For example, since their optimization program needs many iterations, their the computational cost is very expensive, which is more than $O(n^3)$ even though the fast solver is used, where n is the number of objects on a data set. Besides, they can not guarantee the symmetry and nonnegativity of the obtained sparse similarity matrix which is required when implementing spectral clustering.To enhance the efficiency of SSC, several scalable sparse subspace clustering algorithms have been proposed in (Peng et al., 2013; You et al., 2016; Matsushima & Brbic, 2019; Zhang et al., 2019).

Although scholars have provided lots of studies on spectral clustering and sparse subspace clustering, the relation between them is rarely discussed. Therefore, in this paper, we provide an explicit theoretical connection between them. we propose a sparse subspace clustering model with entropy-norm. In this optimization model, we transform a sparse subspace clustering problem into an optimization problem of learning a sparse similarity matrix and uses the Entropy-norm as the regularization term. We derive its optimal solution which is equivalent to Gaussian kernel as the sparse representation. Thus, we can conclude that spectral clustering with Gaussian kernel can be viewed as sparse subspace clustering with entropy-norm (SSC+E). Compared to SSC, SSC+E can avoid the complex computation of the sparse optimization program to obtain a sparse, analytical, symmetrical and nonnegative solution. Finally, we analyze the efficiency and effectiveness of sparse subspace clustering and spectral clustering on ten benchmark data sets. The theoretical and experimental analysis provided by this paper can well guide users to the selection of high-dimensional data clustering algorithms.

The outline of the rest of this paper is as follows. Section 2 introduces spectral clustering and sparse subspace clustering. Section 3 presents the theoretical relation between them. Section 4 shows the experimental analysis of the comparisons with them. Section 5 concludes the paper with some remarks.

2. Spectral clustering and sparse subspace clustering

Let X be a $m \times n$ data matrix with n objects and m attributes, x_i be the ith column of X which is used to represent the ith object. The optimization problem of the spectral clustering is described as follows.

$$\min_H \Theta = Tr(H^T L H), s.t., H^T H = I, \tag{1}$$

where $L = D - W$ is a Laplacian matrix, W is a $n \times n$ data similarity matrix, D is a diagonal matrix whose entries are column (or row, since W is symmetric) sums of W, and H is a $n \times k$ membership matrix. W is often defined based on Gaussian kernel as follows.

$$w_{ij} = \exp\left(-\frac{||x_i - x_j||^2}{\gamma}\right), \tag{2}$$

where γ is a kernel parameter. Besides, we know that $\hat{W} = D^{-1/2} W D^{-1/2}$ is the normalized similarity matrix of W. In this case, the spectral clustering becomes the normalized spectral clustering. The spectral clustering problem is the standard trace minimization problem which is solved by the matrix H which contains the first k eigenvectors of L as rows.

The formulation of sparse subspace clustering is

$$\min_Z F = \mathcal{L}(X, XZ) + \lambda \Omega(Z), \tag{3}$$

where $\lambda > 0$ is the tradeoff factor, $\mathcal{L}(X, XZ)$ is a loss function which wishes $X = XZ$ and $\Omega(.)$ is a regularization term which is used to sparsify Z. If $X = XZ$ is required and $\Omega(Z) = ||Z||_1$, the optimization problem is the classical SSC problem, i.e.,

$$\min_Z ||Z||_1, s.t., X = XZ, diag(Z) = 0. \tag{4}$$

In the noisy SSC algorithm which allows some tolerance for inexact representation, the optimization problem is defined as

$$\min_Z ||X - XZ||_F^2 + \lambda ||Z||_1, s.t., diag(Z) = 0. \tag{5}$$

Besides, in order to make the data lie in a union of subspaces, the constraint $\mathbf{1}^T Z = \mathbf{1}^T$ is added to the sparse subspace problem (Elhamifar & Vidal, 2013).

The optimization problem of SSC and noisy SSC can be solved efficiently using convex programming tools. With the reconstruction coefficient matrix Z, the sparse similarity matrix of objects is computed by

$$W = \frac{|Z| + |Z^T|}{2}, \tag{6}$$

Finally, with the similarity matrix W as the input, the final clustering result is obtained by conducting standard spectral clustering.

According to the above introductions, we can see that the main difference between spectral clustering and sparse subspace clustering is the definition of W. In the spectral clustering, W is directly computed by the kernel function, whose computational complexity is $O(n^2)$. In the sparse subspace clustering, W is computed by learning the similarity matrix, whose computational complexity is more than $O(n^3)$.

3. Sparse subspace clustering with entropy-norm

In this section, we analyze the theoretical connection between spectral clustering and sparse subspace clustering from the perspective of learning the data similarity matrix W. In the analysis, we first transform a sparse subspace clustering problem into an optimization problem of learning a sparse similarity matrix. Furthermore, we uses the Entropy-norm as the regularization term and derives its optimal solution to show the relation between spectral clustering and sparse subspace clustering. In the following, we provide the theoretical analysis in detail.

According to the definition of W in the sparse subspace clustering, we can conclude that if Z is a symmetrical and nonnegative matrix, W is equal to Z. In this case, the minimization problem F can be converted as follows

$$\min_{Z} F, s.t., Z = Z^T, Z \geq 0, diag(Z) = 0. \tag{7}$$

According to Eq.(7), we can see that the sparse subspace clustering problem can be viewed as learning a similarity matrix. However, it is very complex for the optimization program of SSC to obtain a symmetric and nonnegative Z. We know that the SSC algorithm can not guarantee the optimization solution Z is symmetric and nonnegative. Besides, Z obtained by it is not an analytical solution.

Therefore, we select information entropy as the regularization term of the objective function F and propose a sparse subspace clustering with entropy-norm to solve this problem. The objective function F is re-defined as

$$\min_{Z} F = \mathcal{L}(X, XZ) + \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij} \ln z_{ij}. \tag{8}$$

When using the Lagrangian multiplier to minimize F, we can obtain

$$\begin{aligned} &\frac{\partial F}{\partial z_{ij}} = \frac{\partial \mathcal{L}}{\partial z_{ij}} + \lambda\left(\ln z_{ij} + 1\right) = 0 \\ \Rightarrow &z_{ij} = \beta \exp\left(-\frac{f_{ij}}{\lambda}\right) \end{aligned} \tag{9}$$

where $\beta = \exp(-1)$ and $f_{ij} = \frac{\partial \mathcal{L}}{\partial z_{ij}}$. According to Eq.(9), we can conclude that z_{ij} is nonnegative, and if $f_{ij} = f_{ji}$, $z_{ij} = z_{ji}$.

Furthermore, we add the constraint $\mathbf{1}^T Z = \mathbf{1}^T$ to the optimization problem. When using the Lagrangian multiplier to minimize F with the constraint, we can obtain

$$\begin{aligned} &\min_{Z} F = \mathcal{L}(X, XZ) + \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij} \ln z_{ij} \\ &\quad + \gamma \sum_{i=1}^{n}\left(\sum_{h=1}^{n} z_{ih} - 1\right) \\ \Rightarrow &\frac{\partial F}{\partial z_{ij}} = \frac{\partial \mathcal{L}}{\partial z_{ij}} + \lambda\left(\ln z_{ij} + 1\right) + \gamma = 0 \\ \Rightarrow &\exp\left(-\frac{\lambda+\gamma}{\lambda}\right) = \frac{1}{2}\sum_{h=1}^{n} \exp\left(-\frac{f_{ih}}{\lambda}\right) \\ \Rightarrow &z_{ij} = \frac{\exp\left(-\frac{f_{ij}}{\lambda}\right)}{\sum_{h=1}^{n} \exp\left(-\frac{f_{ih}}{\lambda}\right)}. \end{aligned} \tag{10}$$

However, if $f_{ij} = f_{ji}$, we can not guarantee z_{ij} is equal to z_{ji}. In order to improve this problem, we relax the constraint $\mathbf{1}^T Z = \mathbf{1}^T$ and replace it with a new constraint $\sum_{h=1}^{n} z_{ih} + \sum_{h=1}^{n} z_{hj} = 2$ for $1 \leq i, j \leq n$. It notes that the sum of any row and column of Z is 2. When using the Lagrangian multiplier to minimize F with the new constraint, we can obtain

$$\begin{aligned} &\min_{Z} F = \mathcal{L}(X, XZ) + \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij} \ln z_{ij} \\ &\quad + \gamma \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{h=1}^{n} z_{ih} + \sum_{h=1}^{n} z_{hj} - 2\right) \\ \Rightarrow &z_{ij} = \frac{2\exp\left(-\frac{f_{ij}}{\lambda}\right)}{\sum_{h=1}^{n}\left[\exp\left(-\frac{f_{ih}}{\lambda}\right) + \exp\left(-\frac{f_{hj}}{\lambda}\right)\right]}. \end{aligned} \tag{11}$$

According to the above equations, we can conclude that if $f_{ij} = f_{ji}$, $z_{ij} = z_{ji}$. Besides, in order to reduce the computational complexity of z_{ij}, we wish f_{ij} is irrelevant with z_{ij}. Thus, we assume $\mathcal{L}(X, XZ)$ is a linear function with z_{ij}, i.e.,

$$\mathcal{L}(X, XZ) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} z_{ij} + b_{ij}.$$

In this case, $f_{ij} = a_{ij}$ and $a_{ij} = a_{ji}$. Thus, it is a key issue to define a_{ij}.

Based on the constraint $X = XZ$, we have

$$x_i = \sum_{j=1, j\neq i}^{n} z_{ij} x_j, \tag{12}$$

for $1 \leq i \leq n$. According to the equation, we can see that z_{ij} should reflect the similarity between x_i and x_j. The more similar they are, the higher the value of z_{ij} should be. Therefore, we assume a_{ij} is a distance metric between x_i

and x_j, i.e., $a_{ij} = d(x_i, x_j)$. According to the symmetry of the distance metric, we have $a_{ij} = a_{ji}$. Next, we discuss how to use the constraint $X = XZ$ to define $d(x_i, x_j)$.

For x_i, $1 \leq j \leq n$, we have

$$x_i = \sum_{j=1, i\neq j} z_{ij} x_j \Rightarrow x_i^T x_i = \sum_{j=1, i\neq j} z_{ij} x_i^T x_j. \quad (13)$$

Based on the new constraint, we have

$$\left(\sum_{j=1}^{n} z_{ji} + \sum_{j=1}^{n} z_{ij}\right) x_i^T x_i = 2 \sum_{j=1, i\neq j} z_{ij} x_i^T x_j. \quad (14)$$

Therefore, the constraint $x_i - \sum_{j=1, i\neq j}^{n} z_{ij} x_j = \mathbf{0}$ is transformed into

$$\left(\sum_{j=1}^{n} z_{ji} + \sum_{j=1}^{n} z_{ij}\right) x_i^T x_i - 2 \sum_{j=1, i\neq j} z_{ij} x_i^T x_j = 0. \quad (15)$$

Furthermore, we have

$$\sum_{i=1}^{n} \left(\left(\sum_{j=1}^{n} z_{ji} + \sum_{j=1}^{n} z_{ij}\right) x_i^T x_i - 2 \sum_{j=1, i\neq j} z_{ij} x_i^T x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} ||x_i - x_j||^2 z_{ij}. \quad (16)$$

Therefore, the optimization problem can be rewritten as

$$\min_{Z} Q = \sum_{i=1}^{n} \sum_{j=1}^{n} ||x_i - x_j||^2 z_{ij} + \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij} \ln z_{ij}, \quad (17)$$

s.t., $Z = Z^T, Z \geq 0, diag(Z) = 0$ and $\sum_{h=1}^{n} z_{ih} + \sum_{h=1}^{n} z_{hj} = 2$ for $1 \leq i, j \leq n$. In this case, $d(x_i, x_j)$ is defined as Euclidean distance. According to Eqs.(9), if $d(x_i, x_j)$ is Euclidean distance, Z on Eq.(9) is equivalent to Gaussian kernel, i.e.,

$$z_{ij} = \exp\left(-\frac{||x_i - x_j||^2}{\lambda}\right). \quad (18)$$

According to Eq.(18), the spectral clustering with Gaussian kernel can be viewed as a sparse subspace clustering with entropy-norm. Furthermore, we use $d(x_i, x_j)$ instead of f_{ij} in Eq.(11). In this case, Z is

$$z_{ij} = \frac{2\exp\left(-\frac{||x_i - x_j||^2}{\lambda}\right)}{\sum_{h=1}^{n}\left[\exp\left(-\frac{||x_i - x_h||^2}{\lambda}\right) + \exp\left(-\frac{||x_j - x_h||^2}{\lambda}\right)\right]}. \quad (19)$$

According to Eq. (20), we have the following relation

$$Z \leq D^{-1/2} W D^{-1/2}. \quad (20)$$

In this case, Z is a lower bound of the normalized Gaussian kernel and used to approximate it.

Furthermore, based on the objective function F on Eq.(8), we discuss the selection of the regularization term. If we define the regularization term $\Omega(Z) = ||Z||_2$, the objective function Q becomes

$$\sum_{i=1}^{n} \sum_{j=1}^{n} ||x_i - x_j||^2 z_{ij} + \lambda ||Z||_2. \quad (21)$$

In this case, the optimization problem Q is equivalent to that of Clustering with Adaptive Neighbors (Nie et al., 2014) which is proposed by Nie et al.. The authors improved the spectral clustering and used the optimization model to learn the data similarity matrix. The optimal solution Z is computed by

$$z_{ij} = -\frac{||x_i - x_j||^2}{2\lambda} + \eta_i. \quad (22)$$

where $\eta_i = \frac{\sum_{h=1}^{n}\left(||x_i - x_h||^2\right)}{2n\lambda} + \frac{1}{n}$. However, since each η_i may be different, the non-negativeness of Z depends on η_i. If the regularization term $\Omega(Z) = ||Z||_1$, the optimal solution Z is computed by

$$z_{ij} = \begin{cases} 1, j = \arg\min_{t=1} ||x_i - x_t||^2, \\ 0, otherwise. \end{cases} \quad (23)$$

In this case, the solution is trivial, since only its nearest object can be used to represent it. According to Eqs.(22) and (23), we can see the advantages of the entropy-norm, compared to other norms. According to the above theoretical analysis, we can get the following conclusions:

- From the perspective of learning a data similarity matrix, the spectral clustering with Gaussian kernel can be viewed as a sparse subspace clustering with entropy-norm (SSC+E);
- Compared to SSC, SSC+E can directly compute the data similarity matrix by Gaussian kernel, which can reduce the computational cost.
- The data similarity matrix obtained by SSC+E can be guaranteed to be sparse, analytical, symmetrical and nonnegative.

4. Experiment analysis

In the experiments, we analyze and compare the effectiveness and efficiency of the sparse subspace clustering algorithm (SSC) (Elhamifar & Vidal, 2009), the spectral clustering with adaptive neighbors (CAN) (Nie et al., 2014), and the spectral clustering with Gaussian kernel (SSC+E)

(Ng et al., 2001). Differently from SSC and CAN which compute a data similarity matrix by the optimization methods, SSC+E uses Gaussian kernel to compute the similarity matrix. The codes of these algorithms have been publicly shared by their authors.

The experiments are conducted on an Intel i9-7940X CPU@3.10HZ and 128G RAM. We carry out these algorithms on 10 benchmark data sets (Bache & Lichman; Cai). It is worth noting that we did not select large-scale data sets to test these algorithms in our experiments. The main reason is that the computational cost of the SSC algorithm is very high. For example, it needs more than 60 hours on the data set Landsat Satellite which includes 6,435 points. Furthermore, we employ two widely-used external indices (Aggarwal & Reddy, 2014), i.e., the normalized mutual information (NMI) and the adjusted rand index (ARI), to measure the similarity between a clustering result and the true partition on a data set. If the clustering result is close to the true partition, then its NMI and ARI values are high.

Table 1. Description of data sets: Number of Data Objects (n), Number of Dimensions (m), Number of Clusters (k).

Data set	n	m	k
Iris	150	4	3
Wine	178	13	3
Heart Statlog	569	30	2
Yale	165	1024	15
ORL	400	1024	40
Banknote	1,372	4	2
COIL	1,440	1024	20
Isolet	1,560	617	26
Handwritten Digits	5,620	63	10
Landsat Satellite	6,435	36	6

Before the comparisons, we need to set some parameters for these algorithms as follows. We set the number of clusters k is equal to its true number of classes on each of the given data sets. For the parameter λ, we test each algorithm with different λ values which are selected in the set $\{\lambda_1 = \frac{\delta}{50}, \lambda_2 = \frac{\delta}{40}, \lambda_3 = \frac{\delta}{30}, \lambda_4 = \frac{\delta}{20}, \lambda_5 = \frac{\delta}{10}, \lambda_6 = \delta\}$, where δ is the covariance of a data set. Besides, the SSC and CAN algorithms need to set the number of the nearest neighbors K. We set K to 10 in our experiments.

We first compare the effectiveness of the three algorithms with different λ values on these benchmark data sets. The comparison results are shown in Fig.1. According to the figures, we see that the performance of the SSC+E algorithm is superior to SSC and CAN on these tested data sets, except Heart statlog. Due to the fact that the number of the tested data sets is very limited, we can not conclude that the effectiveness of the data similarity matrices by Gaussian kernel are better than SSC and CAN. However, our experimental results illustrate that Gaussian kernel is a good choice to rapidly obtain a good sparse representation for a data set.

Furthermore, we compare the efficiency of the three algorithms with different λ values on these benchmark data sets. The comparison results are shown in Table 2. According to the table, we see that the SSC algorithm need very expensive computational costs, compared to SSC+E and CAN. We also can observe that the clustering speed of the SSC+E algorithm is slightly faster than CAN on these tested data sets. The main reason is that the SSC+E algorithm does not need learn the sparse representation but directly compute it by Gaussian kernel. According to the above analysis, we can conclude that the SSC+E algorithm can better balance the effectiveness and efficiency of obtaining a high-quality clustering results, compared to the SSC and CAN algorithms.

5. Conclusions

In this paper, we have analyzed the theoretical connection between spectral clustering and sparse subspace clustering from the perspective of learning the data similarity matrix. We have shown that the spectral clustering with Gaussian kernel can be viewed as a sparse subspace clustering with entropy-norm, which is called SSC+E. We have analyzed the advantages and disadvantages of SSC+E and SSC. Compared to SSC, the SSC+E algorithm can rapidly obtain a sparse, analytical, symmetrical and nonnegative similarity matrix. Finally, we have compared the efficiency and effectiveness of sparse subspace clustering and spectral clustering on ten benchmark data sets. The experimental results show that the spectral clustering with Gaussian kernel is still a good choice to rapidly obtain a high-quality clustering results.

References

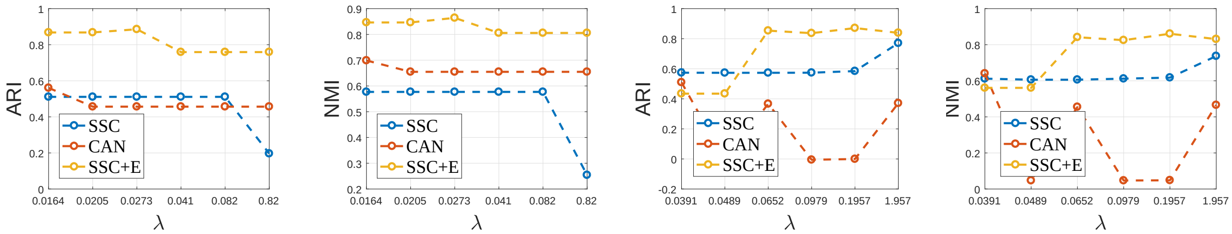
Aggarwal, C. C. and Reddy, C. K. (eds.). *Data Clustering: Algorithms and Applications*. CRC Press, 2014. ISBN 978-1-46-655821-2.

Bache, K. and Lichman, M. Uci machine learning repository. http://archive.ics.uci.edu/ml.

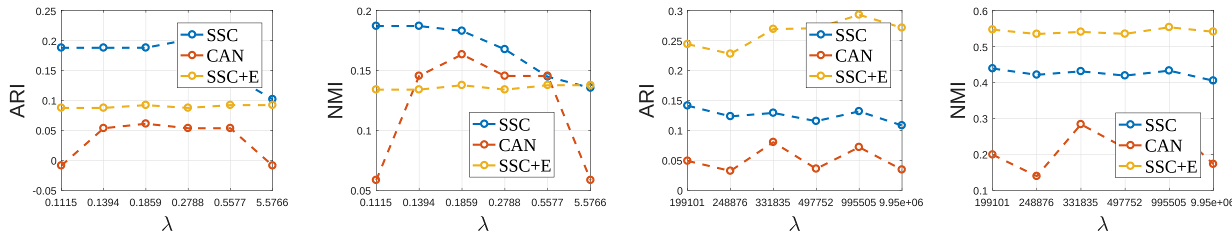
Cai, D. Codes and datasets for feature learning. http://www.cad.zju.edu.cn/home/dengcai/Data/data.html.

Dhillon, I. S., Guan, Y., and Kulis, B. Weighted graph cuts without eigenvectors: A multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1944–1957, 2007.

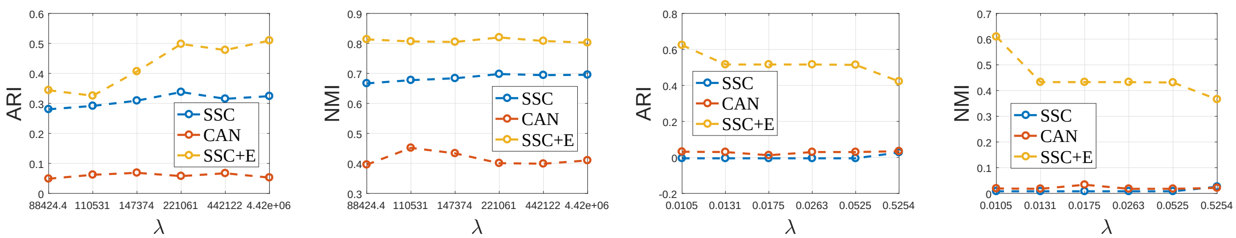
Elhamifar, E. and Vidal, R. Sparse subspace clustering.



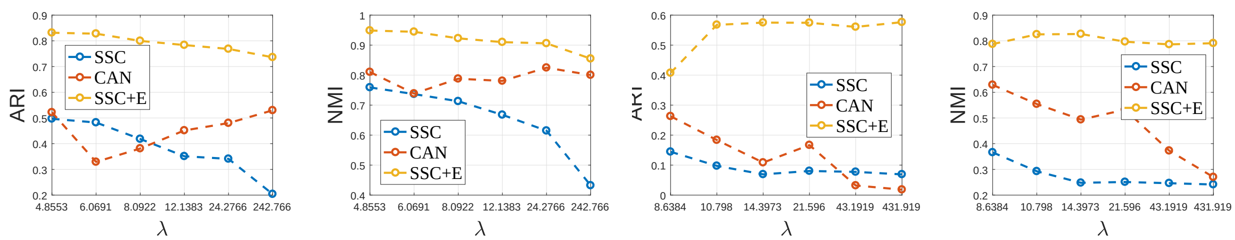
(a) ARI against λ on data set Iris (b) NMI against λ on data set Iris (c) ARI against λ on data set Wine (d) NMI against λ on data set ORL



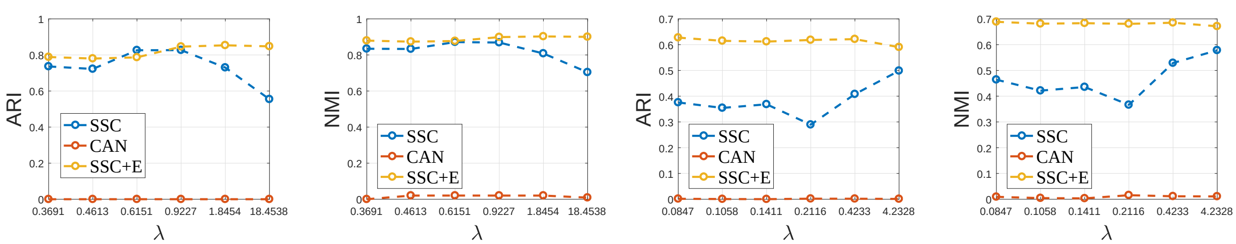
(e) ARI against λ on data set Heart statlog (f) NMI against λ on data set Heart statlog (g) ARI against λ on data set Yale (h) NMI against λ on data set Yale



(i) ARI against λ on data set ORL (j) NMI against λ on data set ORL (k) ARI against λ on data set Banknote (l) NMI against λ on data set Banknote



(m) ARI against λ on data set COIL (n) NMI against λ on data set COIL (o) ARI against λ on data set Isolet (p) NMI against λ on data set Isolet



(q) ARI against λ on data set Handwritten Digits (r) NMI against λ on data set Handwritten Digits (s) ARI against λ on data set Landsat Satellite (t) NMI against λ on data set Landsat Satellite

Figure 1. Clustering accuracies of different algorithms.

Table 2. Clustering speeds (seconds) of different algorithms

Data set	Algorithm	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
Iris	SSC	31.727	32.176	31.724	31.445	30.836	32.336
	CAN	0.078	0.069	0.046	0.078	0.047	0.047
	SSC+E	0.031	0.047	0.032	0.032	0.016	0.016
Wine	SSC	37.975	37.897	38.007	40.100	39.366	39.132
	CAN	0.094	0.078	0.125	0.094	0.078	0.094
	SSC+E	0.031	0.016	0.015	0.015	0.015	0.015
Heart Statlog	SSC	59.689	58.674	59.923	60.267	59.767	61.689
	CAN	0.156	0.188	0.188	0.156	0.156	0.125
	SSC+E	0.015	0.015	0.016	0.016	0.015	0.015
Yale	SSC	204.686	203.968	204.749	202.843	204.328	201.687
	CAN	0.125	0.219	0.125	0.125	0.156	0.140
	SSC+E	0.047	0.047	0.031	0.062	0.047	0.047
ORL	SSC	2245.200	2258.955	2258.564	2237.354	2266.546	2265.959
	CAN	0.625	0.671	0.671	0.687	0.641	0.703
	SSC+E	0.218	0.203	0.204	0.188	0.219	0.203
Banknote	SSC	387.909	390.128	392.174	386.707	391.206	393.673
	CAN	5.343	4.702	4.624	4.452	4.264	4.749
	SSC+E	0.187	0.187	0.218	0.187	0.188	0.187
COIL	SSC	70766.438	70601.720	69062.015	67736.005	63388.664	54684.533
	CAN	4.434	6.264	5.905	3.546	3.359	2.952
	SSC+E	0.534	0.547	0.563	0.563	0.578	0.594
Isolet	SSC	56079.227	55180.180	54446.940	64799.694	72375.284	76137.546
	CAN	7.545	4.187	7.467	7.108	7.436	4.405
	SSC+E	0.657	0.703	0.657	0.734	0.766	0.734
Handwritten Digits	SSC	13990.391	13495.460	13250.659	12951.093	12793.112	14684.592
	CAN	216.325	154.261	148.512	148.73	140.842	196.876
	SSC+E	3.702	3.577	3.702	3.625	3.843	3.999
Landsat Satellite	SSC	14387.362	14463.630	14510.508	15019.646	15736.703	14685.439
	CAN	265.11	317.706	204.983	318.269	315.863	316.786
	SSC+E	4.467	4.374	4.452	4.562	4.452	4.545

In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

Elhamifar, E. and Vidal, R. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11): 2765–2781, 2013.

H. Hu, Z. Lin, J. F. and Zhou, J. Smooth representation clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3834–3841, 2014.

Han, J. and Kamber, M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2001.

Jain, A. K. Data clustering: 50 years beyond k-means. In Daelemans, W., Goethals, B., and Morik, K. (eds.), *Machine Learning and Knowledge Discovery in Databases*, pp. 3–4, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., and Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

Matsushima, S. and Brbic, M. Selective sampling-based scalable sparse subspace clustering. In *Conference and Workshop on Neural Information Processing Systems*, 2019.

Ng, A. Y., Jordan, M. I., and Weiss, Y. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pp. 849–856. MIT Press, 2001.

Nie, F., Wang, X., and Huang, H. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data miningng*, pp. 977–986, 2014.

Parsons, L. R., Liu, H., Parsons, L., Haque, E., and Liu, H. Subspace clustering for high dimensional data: a review. In *ACM SIGKDD Explorations Newsletter*, pp. 90–105, 2004.

Peng, X., Zhang, L., and Yi, Z. Scalable sparse subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 430–437, 2013.

Shi, J. and Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

Wang, Y. and Xu, H. Noisy sparse subspace clustering. *Journal of Machine Learning Research*, 17:1–14, 2016.

Yang, Y., Feng, J., Jojic, N., Yang, J., and Huang, T. L0-sparse subspace clustering. In *14th European Conference on Computer Vision*, pp. 731–747, 2016.

You, C., Robinson, D. P., and Vidal, R. Scalable sparse subspace clustering by orthogonal matching pursuit. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3918–3927, 2016.

Zhang, T., Ji, P., Harandi, M., bing Huang, W., and Li, H. Neural collaborative subspace clustering. *International Conference on Machine Learning*, 2019.

Zhu, X., Loy, C. C., and Gong, S. Constructing robust affinity graphs for spectral clustering. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1450–1457, 2014.