[WORK IN PROGRESS] Research Report - Spectral Clustering

Sourabh Antani

Mathematics Department,
Southern Methodist University

June 3, 2021

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - k-means [1] [2]

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - k-means [1] [2]
  - BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)[3]

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - k-means [1] [2]
  - BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)[3]
  - Mean-shift [4]

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - *k*-means [1] [2]
  - BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)[3]
  - Mean-shift [4]
  - DBSCAN [5]

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - k-means [1] [2]
  - BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)[3]
  - Mean-shift [4]
  - DBSCAN [5]
- Linear Algebra based

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - *k*-means [1] [2]
  - BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)[3]
  - Mean-shift [4]
  - DBSCAN [5]
- Linear Algebra based
  - Non-Negative Matrix Factorization [6] [7] [8]

Clustering Algorithms

Below are some examples of clustering algorithms used in practice, aside from spectral clustering.

- Euclidean distance and spatial density based:
  - *k*-means [1] [2]
  - BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)[3]
  - Mean-shift [4]
  - DBSCAN [5]
- Linear Algebra based
  - Non-Negative Matrix Factorization [6] [7] [8]
  - PCA Based clustering [9]

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph
- $\mathbf{V} = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$: Set of vertices

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph
- $\mathbf{V} = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$: Set of vertices
- **E**: Set of weighted edges with weight w_{ij} for edge between v_i & v_j. $w_{ij} = 0$ iff vertices v_i and v_j are not connected. $w_{ij} = w_{ji}$.

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph
- $\mathbf{V} = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$: Set of vertices
- **E**: Set of weighted edges with weight w_{ij} for edge between v_i & v_j. $w_{ij} = 0$ iff vertices v_i and v_j are not connected. $w_{ij} = w_{ji}$.
- $\mathbf{W} = (\mathbf{w_{ij}})_{\mathbf{i},\mathbf{j}=\mathbf{1},\ldots,\mathbf{n}}$: Weighted *adjacency matrix*

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph
- $\mathbf{V} = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$: Set of vertices
- **E**: Set of weighted edges with weight w_{ij} for edge between v_i & v_j. $w_{ij} = 0$ iff vertices v_i and v_j are not connected. $w_{ij} = w_{ji}$.
- $\mathbf{W} = (\mathbf{w_{ij}})_{\mathbf{i,j=1,\ldots,n}}$: Weighted *adjacency matrix*
- $\mathbf{d_i} = \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{w_{ij}}$: Degree of a vertex v_i.

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph
- $\mathbf{V} = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$: Set of vertices
- **E**: Set of weighted edges with weight w_{ij} for edge between v_i & v_j. $w_{ij} = 0$ iff vertices v_i and v_j are not connected. $w_{ij} = w_{ji}$.
- $\mathbf{W} = (\mathbf{w_{ij}})_{\mathbf{i,j=1,\ldots,n}}$: Weighted *adjacency matrix*
- $\mathbf{d_i} = \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{w_{ij}}$: Degree of a vertex v_i.
- **D**: Degree matrix = $diag(d_1, \ldots, d_n)$.

Spectral Clustering

Graph Terminology (1/2)

- $\mathbf{G} = (\mathbf{V}, \mathbf{E})$: Undirected graph
- $\mathbf{V} = \{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$: Set of vertices
- **E**: Set of weighted edges with weight w_{ij} for edge between v_i & v_j. $w_{ij} = 0$ iff vertices v_i and v_j are not connected. $w_{ij} = w_{ji}$.
- $\mathbf{W} = (\mathbf{w_{ij}})_{\mathbf{i,j=1,\ldots,n}}$: Weighted *adjacency matrix*
- $\mathbf{d_i} = \sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{w_{ij}}$: Degree of a vertex v_i.
- **D**: Degree matrix = $diag(d_1, \ldots, d_n)$.
- $\mathbf{W(A, B)} = \sum_{\mathbf{i \in A, j \in B}} \mathbf{w_{ij}}$

Spectral Clustering

Graph Terminology (2/2)

- $\bar{\mathbf{A}}$: Complement of A.

Spectral Clustering

Graph Terminology (2/2)

- $\bar{\mathbf{A}}$: Complement of A.
- $\mathbf{1}_A$: Indicator vector = a vector with entries $f_i = 1$ if $v_i \in A$, $f_i = 0$ otherwise.

- $\bar{\mathbf{A}}$: Complement of A.
- $\mathbf{1}_A$: Indicator vector = a vector with entries $f_i = 1$ if $v_i \in A$, $f_i = 0$ otherwise.
- $|\mathbf{A}|$: Number of vertices in A, one measure of size of A

Spectral Clustering

Graph Terminology (2/2)

- $\bar{\mathbf{A}}$: Complement of A.
- $\mathbf{1}_A$: Indicator vector = a vector with entries $f_i = 1$ if $v_i \in A$, $f_i = 0$ otherwise.
- $|\mathbf{A}|$: Number of vertices in A, one measure of size of A
- $\mathbf{vol}(\mathbf{A}) := \sum_{\mathbf{i} \in \mathbf{A}} \mathbf{d_i}$: Sum of edge weights, another measure of size of A

- $\bar{\mathbf{A}}$: Complement of A.
- $\mathbf{1}_A$: Indicator vector = a vector with entries $f_i = 1$ if $v_i \in A$, $f_i = 0$ otherwise.
- $|\mathbf{A}|$: Number of vertices in A, one measure of size of A
- $\mathbf{vol}(\mathbf{A}) := \sum_{\mathbf{i} \in \mathbf{A}} \mathbf{d_i}$: Sum of edge weights, another measure of size of A
- $A \subset V$: Connected subset if any two vertices in A can be joined by a path such that all intermediate points also lie in A

Spectral Clustering

Graph Terminology (2/2)

- $\bar{\mathbf{A}}$: Complement of A.
- $\mathbf{1}_A$: Indicator vector = a vector with entries $f_i = 1$ if $v_i \in A$, $f_i = 0$ otherwise.
- $|\mathbf{A}|$: Number of vertices in A, one measure of size of A
- $\mathbf{vol(A)} := \sum_{\mathbf{i \in A}} \mathbf{d_i}$: Sum of edge weights, another measure of size of A
- $A \subset V$: Connected subset if any two vertices in A can be joined by a path such that all intermediate points also lie in A
- Connected Component: A is called **Connected Component** if it is a connected subset such that A and $\bar{A}$ are disjoint

Spectral Clustering

Graph Terminology (2/2)

- $\bar{\mathbf{A}}$: Complement of A.
- $\mathbf{1}_A$: Indicator vector = a vector with entries $f_i = 1$ if $v_i \in A$, $f_i = 0$ otherwise.
- $|\mathbf{A}|$: Number of vertices in A, one measure of size of A
- $\mathbf{vol}(\mathbf{A}) := \sum_{\mathbf{i} \in \mathbf{A}} \mathbf{d_i}$: Sum of edge weights, another measure of size of A
- $A \subset V$: Connected subset if any two vertices in A can be joined by a path such that all intermediate points also lie in A
- Connected Component: A is called **Connected Component** if it is a connected subset such that A and $\bar{A}$ are disjoint
- **Partitions** of a graph :Non-empty sets $A_1, \ldots, A_k$ form partition of graph if $A_i \cap A_j = \emptyset$ and $A_1 \cup \cdots \cup A_k = V$

Unnormalized	Normalized (Symmetric)	Normalized (Random Walk)
$L = D - W$	$L_{sym} = D^{-1/2} L D^{-1/2}$ $= I - D^{-1/2} W D^{-1/2}$	$L_{rw} = D^{-1} L$ $= I - D^{-1} W$
Symmetric	Symmetric	Not-Symmetric
u - eigenvector	$D^{1/2} u$ - eigenvector	u - eigenvector
Positive-semidefinte, Non-Negative Real-Valued Eigenvalues		
0 is an eigenvalue with multiplicity equal to no. of connected components		

Spectral Clustering

Graph Cuts

- The goal of clustering is to maximize the similarity within cluster and minimize the similarity between clusters.

Spectral Clustering

Graph Cuts

- The goal of clustering is to maximize the similarity within cluster and minimize the similarity between clusters.
- Minimizing the between cluster similarity equals to minimizing the objective function

$$\mathrm{cut}(A_1, \ldots, A_k) := \frac{1}{2}\sum_{i-1}^{k} W(A_i, \bar{A}_i)$$

- The goal of clustering is to maximize the similarity within cluster and minimize the similarity between clusters.
- Minimizing the between cluster similarity equals to minimizing the objective function

$$\mathrm{cut}(A_1, \ldots, A_k) := \frac{1}{2}\sum_{i-1}^{k} W(A_i, \bar{A}_i)$$

- To maximize the within-cluster similarity, we define two objective functions

$$RatioCut(A_1, \ldots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \bar{A}_i)}{|A_i|} = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|} \tag{1}$$

$$Ncut(A_1, \ldots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \bar{A}_i)}{vol(A_i)} = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)} \tag{2}$$

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.
begin

Construct a similarity graph and its weighted adjacency matrix W
Compute the unnormalized Laplacian L
Compute the first k eigenvectors $u_1, \ldots, u_k$ of L
Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns
For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i^{th} row of U
Cluster the points $(y_i), i = 1, \ldots, n$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$

end
Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{x_j | y_j \in C_i\}$

Spectral Clustering

Algorithms - Normalized Spectral Clustering - Shi & Malik (2000) [10]

This algorithm uses generalized eigenvectors of L, which are the eigenvectors of normalized random-walk Laplacian L_{rw}

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct

begin

- Construct a similarity graph and its weighted adjacency matrix W
- Compute the unnormalized Laplacian L.
- Compute the first k generalized eigenvectors $u_1, \ldots, u_k$ of generalized eigenproblem $Lu = \lambda Du$.
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i^{th} row of U.
- Cluster the points $(y_i), i = 1, \ldots, n$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

end

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{x_j | y_j \in C_i\}$

Spectral Clustering

Algorithms - Normalized Spectral Clustering - Ng, Jordan & Weiss (2002) [11]

This algorithm uses the eigenvectors of normalized symmetric Laplacian L_{sym}. Note that if $D^{1/2}u$ is an eigenvector of L_{sym} if u is an eigenvector of L hence an additional normalization step is needed.

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

begin

Construct a similarity graph and its weighted adjacency matrix W

Compute the normalized Laplacian L_{sym}

Compute the first k eigenvectors $u_1, \ldots, u_k$ of L_{sym}

Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns

Form the matrix $T \in \mathbb{R}^{n \times k}$ from U by normalizing the norm to 1, $t_{ij} = u_{ij}/(\sum_k u_{ik}^2)^{1/2}$

For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i^{th} row of U

Cluster the points $(y_i), i = 1, \ldots, n$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$

end

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{x_j | y_j \in C_i\}$

Spectral Clustering

Numerical Experiments

Mixture of 4 Gaussians in $\mathbb{R}^1$ with a Gaussian kernel $e^{-|x_i - x_j|^2/(2\sigma^2)}$ with $\sigma = 1$ as similarity function.

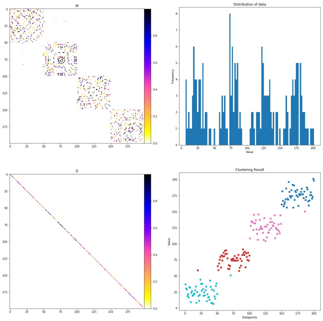


Figure: Clustering results for 200 points in $\mathbb{R}^1$

Spectral Clustering

Numerical Experiments

Points in $\mathbb{R}^2$ with $||x - y||_2$ as similarity function

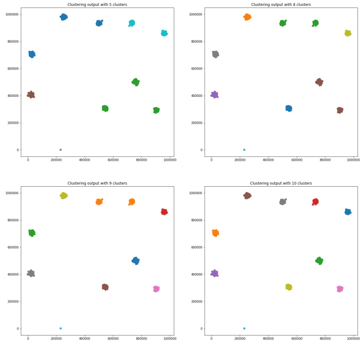
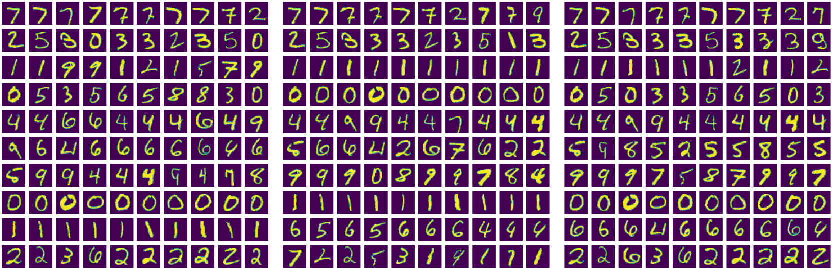


Figure: Clustering results for 1351 points in $\mathbb{R}^2$

Spectral Clustering

Numerical Experiments

Unsupervised Image Classification



(a) Count of pixels that are non-zero in both images as similarity measure

(b) Inverse of 2-norm of the difference vector between two images as similarity measure

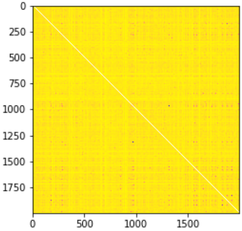
(c) Count of pixels that are zero or non-zero in both images as similarity measure

Figure: 2000 images from MNIST data-set were clustered into 10 clusters. In the above figure, each row of images represent a cluster and show the first 10 images in that cluster

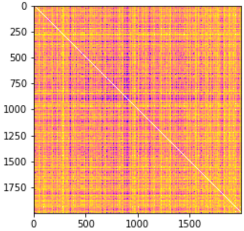
Spectral Clustering

Numerical Experiments

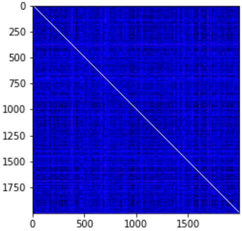
Unsupervised Image Classification



(a) Count of pixels that are non-zero in both images as similarity measure



(b) Inverse of 2-norm of the difference vector between two images as similarity measure

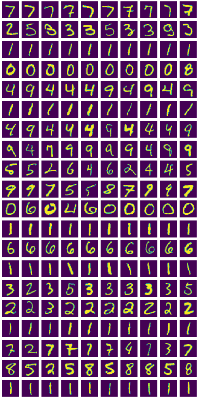


(c) Count of pixels that are zero or non-zero in both images as similarity measure

Figure: Weighted adjecency matrix for 2000 images from MNIST data-set with similarity functions described under the images

Unsupervised Image Classification

(a) Count of pixels that are non-zero in both images as similarity measure



(b) Inverse of 2-norm of the difference vector between two images as similarity measure

J. Macqueen.

Some methods for classification and analysis of multivariate observations.

In *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

S. Lloyd.

Least squares quantization in pcm.

IEEE Transactions on Information Theory, 28(2):129–137, 1982.

Tian Zhang, Raghu Ramakrishnan, and Miron Livny.

Birch: An efficient data clustering method for very large databases.

In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, SIGMOD '96, pages 103–114, New York, NY, USA, 1996. Association for Computing Machinery.

Yizong Cheng.

Mean shift, mode seeking, and clustering.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(8):790–799, 1995.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu.

A density-based algorithm for discovering clusters in large spatial databases with noise.

pages 226–231. AAAI Press, 1996.

William H. Lawton and Edward A. Sylvestre.

Self modeling curve resolution.

Technometrics, 13(3):617–633, 1971.

Pentti Paatero, Unto Tapper, Pasi Aalto, and Markku Kulmala.

Matrix factorization methods for analysing diffusion battery data.

Journal of Aerosol Science, 22:S273–S276, 1991.

Chris Ding, Xiaofeng He, and Horst D. Simon.

On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering, pages 606–610.

Proceedings of the 2005 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics, April 21, 2005.

Nian Zhang, Keenan Leatham, Jiang Xiong, and Jing Zhong.

Pca-k-means based clustering algorithm for high dimensional and overlapping spectra signals.

In *2018 Ninth International Conference on Intelligent Control and Information Processing (ICICIP)*, pages 349–354, 2018.

J. Shi and J. Malik.

Normalized cuts and image segmentation.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):888–905, 2000.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss.

On spectral clustering: Analysis and an algorithm.

Proc. NIPS, pages 849–856, 2001.

G. H. Golub and C. F. Van Loan.

Matrix Computations.

Johns Hopkins University Press, Baltimore, MD, 3rd edition, 1996.

B. N. Parlett.

The Symmetric Eigenvalue Problem.

Number 20 in Classics in Applied Mathematics. SIAM, Philadelphia, PA, 1998.

G. W. Stewart and J. G. Sun.

Matrix perturbation theory.

Academic Press, Boston, MA, 1990.

G. Cheung, E. Magli, Y. Tanaka, and M. K. Ng.

Graph spectral image processing.

Proceedings of the IEEE, 106(5):907–930, 2018.