Politecnico di Torino
1859
Dipartimento di Scienze Matematiche "G. L. Lagrange"
DISMA
ECCELLENZA 2018·2022

Homework Spectral Clustering

Vittorio Gallo, matricola 300829
Sergio A. Angelini, matricola 300873
Filippo Grobbo, matricola 305723

Course of *Computational Linear Algebra for Large Scale Problems*
Academic year 2022 - 2023

Problem Overview

Algorithm 1

function Spectral clustering

Input

x_i: set of points ($x_i \in R^n$, i=1,..,N)

Output

c_j: set of clusters (j=1,..,K)

Define a similarity function

Compute the similarity graph

Construct the Laplacian matrix of the graph

Compute the smallest eigenvalues according to some criterion

Construct the matrix U that contains as columns eigenvectors corresponding to the smallest eigenvalues

Consider the rows of U as a new set of points y_i (i=1,..,N) and compute Kmeans on them

Associate the clusters obtained for y_i to the corresponding x_i (1 to 1 correspondence)

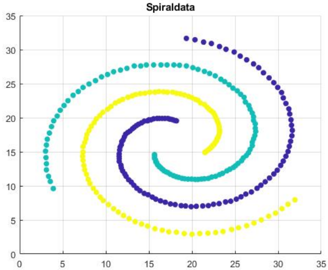
end function

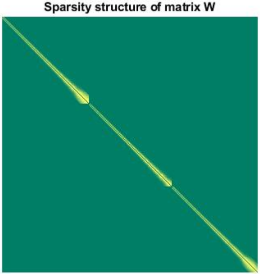
Proposed Approach

For this project we used the following elements:

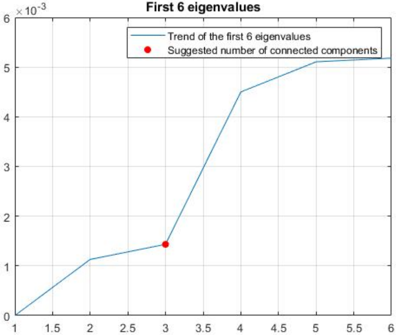
- Circle and Spiral datasets,
- Similarity function $s_{ij} = exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right)$ with $\sigma = 1$,
- K-nearest-neighborhood similarity graph,
- Elbow method for computation of the smallest eigenvalues,
- Function KNN_similarity_graph implemented by us.

Spiral dataset example

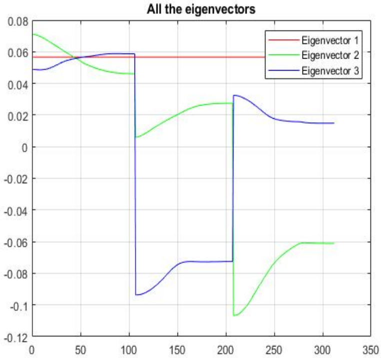


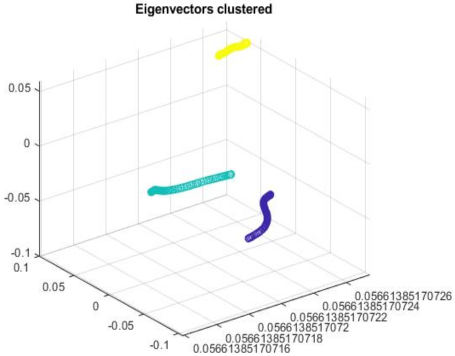


1. Plot of Spiral dataset
2. Spy of matrix W
3. Elbow method

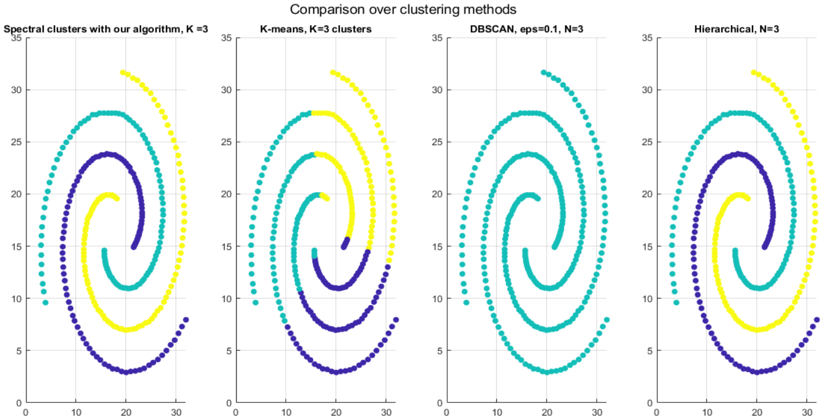


Eigenvector analysis

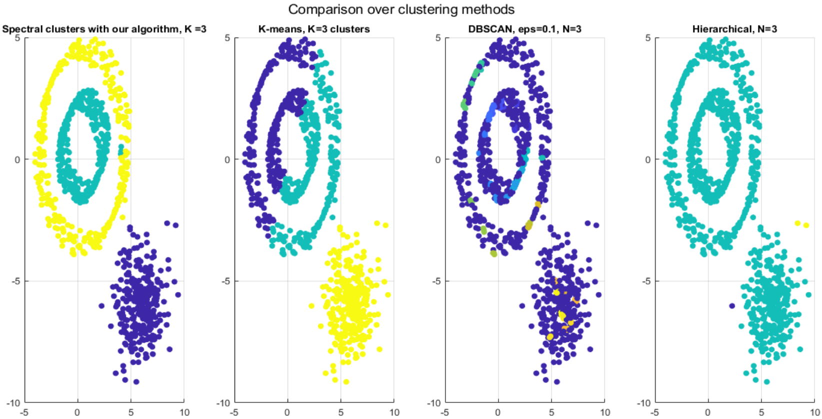




Results for Spiral 2-D dataset



Results for Circle 2-D dataset

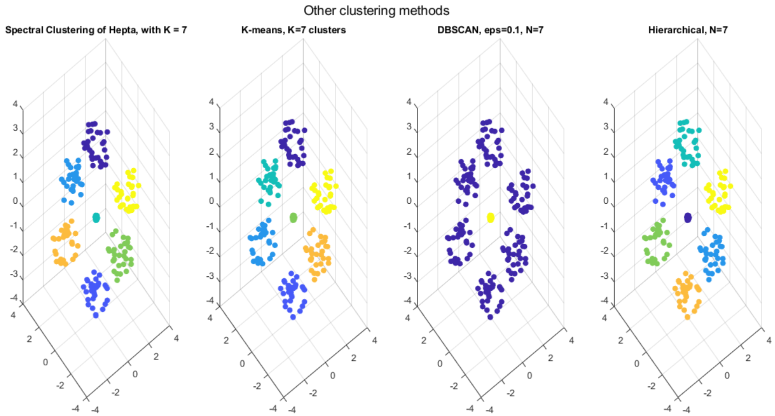


Results for 3-D datasets

Generally we can say that our algorithm about Spectral Clustering is accurate, while the suggestion function for the eigenvalue elbow method can be improved.

	Atom	**EngyTime**	**GolfBall**	**Hepta**	**Target**	**Tetra**	**TwoDiamonds**
SC functions with L	yes	yes	yes	yes	no	yes	yes
SC functions with L_s	yes	yes	yes	yes	no	yes	yes
KNN	10	2	10	10	10	10	10
comp_conn	4	3	3	9	6	6	4
rescaling	yes	no	no	no	yes	no	no
real number of clusters	2	2	1	7	6	4	2
suggested number of clusters	incorrect	correct	correct	correct	correct	correct	incorrect

Results for Hepta 3-D dataset



Normalized symmetric Laplacian matrix L_s: important results

- Zero is eigenvalue for both the matrices L_s and L and has the same multiplicity in both cases.
- The eigenvectors corresponding to the zero eigenvalues have the form $v_0 = D^{\frac{1}{2}} \mathbb{1}_{comp\_conn}$, where $\mathbb{1}_{comp\_conn}$ is the vector with the value 1 on the indices corresponding to one connected component and zero elsewhere.
- The relation between eigenvalues of the two matrices is:

$$\lambda_s = \lambda_L \left[\frac{1 - \mathcal{R}_{D^{-\frac{1}{2}} A D^{-\frac{1}{2}}}(v)}{\mathcal{R}_D(x) - \mathcal{R}_A(x)} \right] \tag{1}$$

where x is the corresponding eigenvector of L, v is the corresponding eigenvector of L_s and $\mathcal{R}_Q(y)$ is the Rayleigh coefficient of matrix Q w.r.t. the normalized vector y.

Grazie per l'attenzione

1859

Politecnico
di Torino