Spectral Clustering

Tapesh Yadav

Final Presentation
E0 270: Machine Learning
Instructor: Ambedkar Dukkipati

Outline

Graph notation and similarity graphs

- Clustering problem can be reformulated as follows:

Problem

Find a partition of the graph such that the edges between different groups have a very low weight (which means that points in different clusters are dissimilar from each other) and the edges within a group have high weight (which means that points within the same cluster are similar to each other).

Different similarity graphs

ϵ neighbourhood Graph

- Connect all points whose pairwise distances are smaller than ϵ.
- Unweighted

k-nearest neighbor graphs (knn)

- Connect vertex v_i with vertex v_j if v_j is amongst k nearest neighbours of v_i.
- Use **mutual k nearest neighbour** method or **k nearest neighbour.** to get undirected graph.

The fully connected graph:

- Here we simply connect all points with positive similarity with each other, and we weight all edges by s_{ij}.
- Gaussian similarity function, $s(x_i, x_j) = exp\left(\frac{||x_i - x_j||^2}{2\sigma^2}\right)$.

Graph Laplacians

- The main tools for spectral clustering are graph Laplacian matrices.
- **Unormalized** graph laplacian, $L = D - W$.
- **Normalized** graph laplacians:

$$L_{sym} = D^{-1/2} L D^{-1/2}$$

$$L_{rw} = D^{-1} L$$

Unnormalized spectral clustering

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

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

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

Normalized spectral clustering according to Shi and Malik (2000)

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

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

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

Normalized spectral clustering according to Ng, Jordan, and Weiss (2002)

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

- Construct a similarity graph by one of the ways mentioned previously. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L.
- Compute the first k eigenvectors $u_1, ..., u_k$ of L_{sym}.
- Let $U \in \mathbb{R}^{nxk}$ be matrix containing the vectors $u_1, ..., u_k$ as columns.
- Form the matrix $T \in \mathbb{R}^{nxk}$ from U by normalizing the rows to norm 1.
- For $i = 1, ..., n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of T.
- Cluster the points $(y_i)_{i=1,...,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, ..., C_k$.

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

Cut, RatioCut and Ncut

for a partition A,B of V, define

$$cut(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

We mention here three objective functions, minimizing which broadly result in clusters having minimized intercluster distance.

- $cut(A_1, A_2, \ldots, A_k) = \sum_{i=1}^{k} cut(A_i, \bar{A}_i)$
- $RatioCut(A_1, ..., A_k) = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{|A_i|}$ (Explains unnormalised spectral clustering).
- $Ncut(A_1, ..., A_k) = \sum_{i=1}^{k} \frac{cut(A_i, \bar{A}_i)}{vol(A_i)}$ (Explains normalized spectral clustering).

Other point of views

- Spectral clustering (Ncut) can be explained using **random walk on graphs** ($P = D^{-1}W \implies \mathrm{L}_{rw} = I - P$).
- **Perturbation theory** also gives a beautiful explanation.

Figure 1 from paper

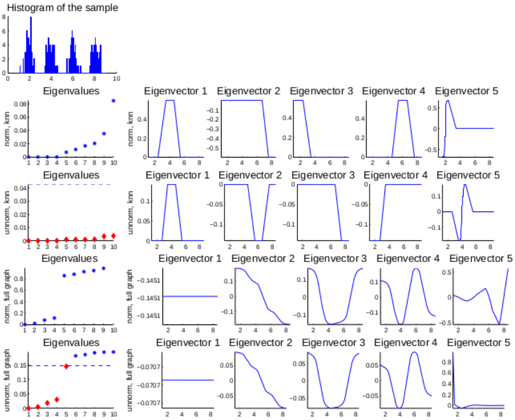
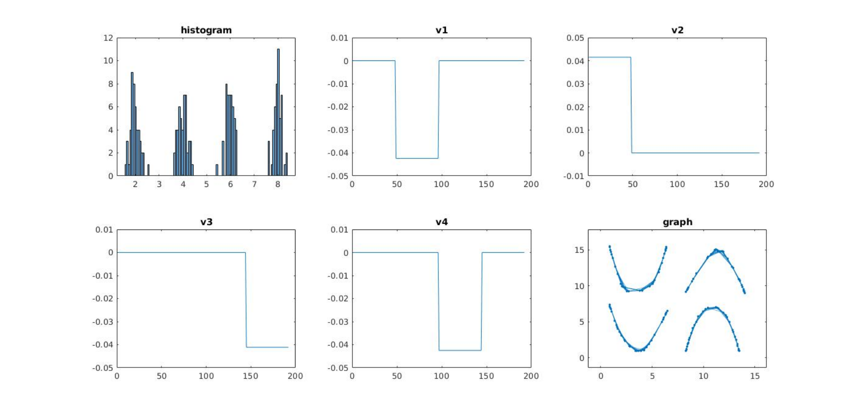


Figure 1: Toy example for spectral clustering where the data points have been drawn from a mixture of four Gaussians on $\mathbb{R}$. Left upper corner: histogram of the data. First and second row: eigenvalues and eigenvectors of L_{rw} and L based on the k-nearest neighbor graph. Third and fourth row: eigenvalues and eigenvectors of L_{rw} and L based on the fully connected graph. For all plots, we used the Gaussian kernel with $\sigma = 1$ as similarity function. See text for more details.

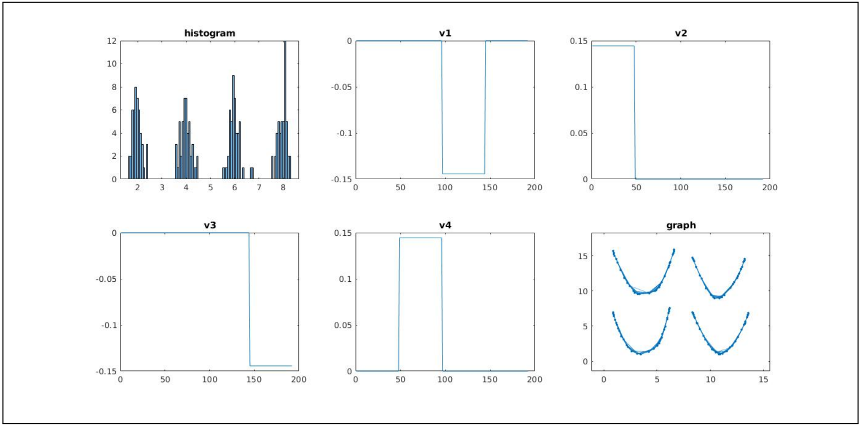
Reproducing Figure 1 with 196 data points

Figure: normalized, knn



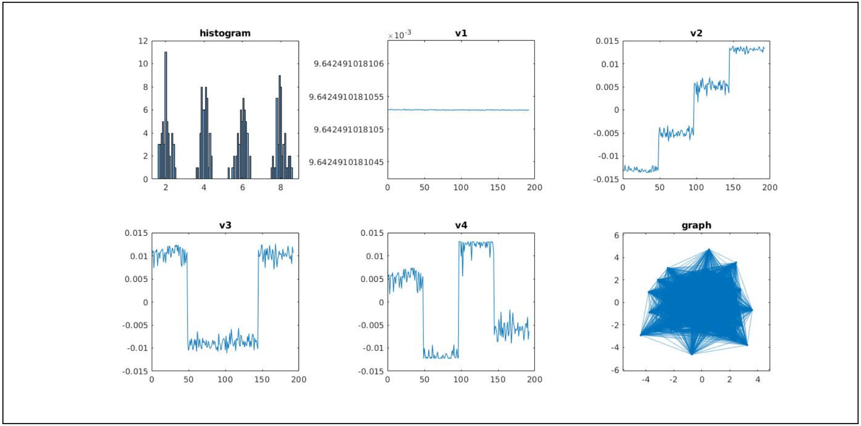
Reproducing Figure 1 with 196 data points

Figure: unnormalized, knn



Reproducing Figure 1 with 196 data points

Figure: normalized, full graph



Reproducing Figure 1 with 196 data points

Figure: unnormalized, full graph

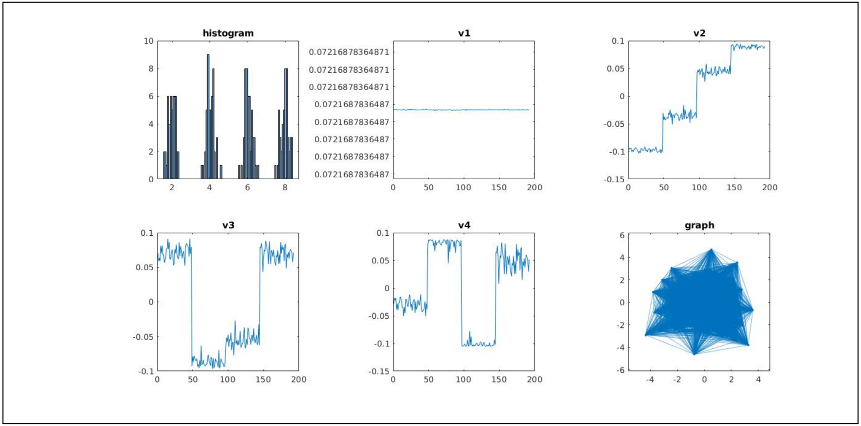


Figure 2 from paper

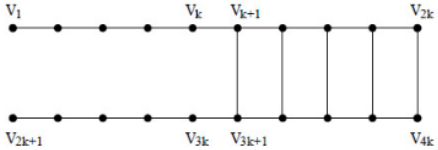
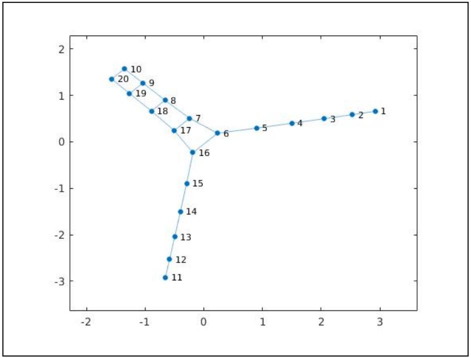


Figure 2: The cockroach graph from Guattery and Miller (1998).

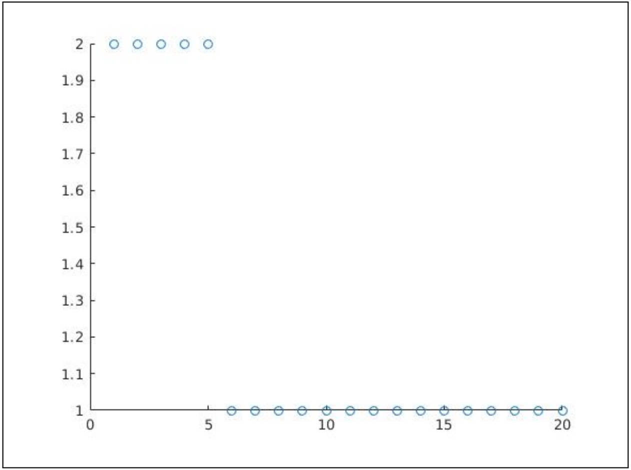
Reproducing Figure 2

Figure: Cockroach graph



Reproducing Figure 2

Figure: kmeans clustering



Reproducing Figure 2

Figure: sign based clustering

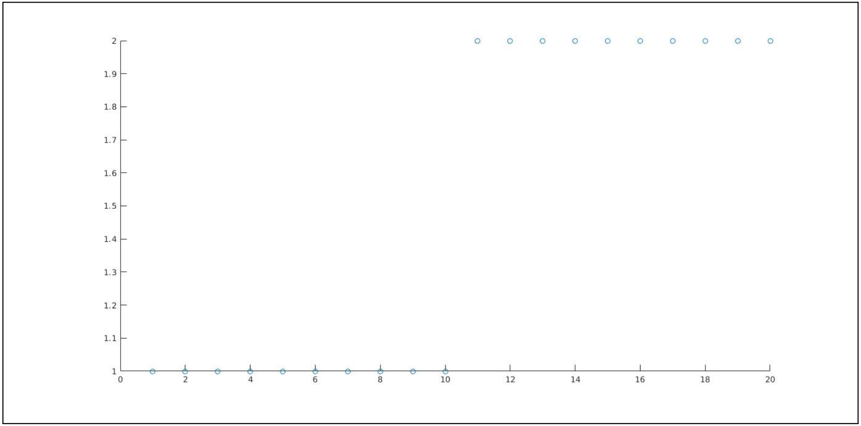


Figure 3 from Paper

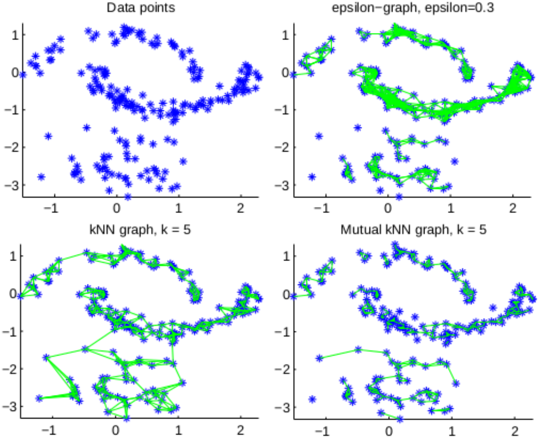


Figure 3: Different similarity graphs, see text for details.

Reproducing Figure 3

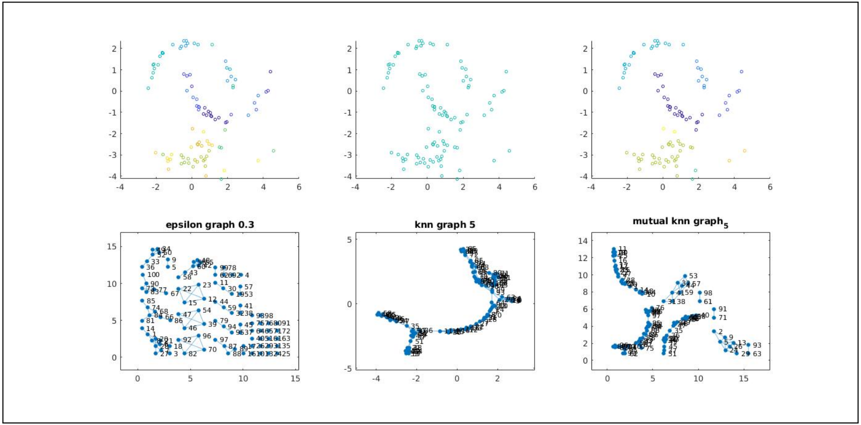


Figure 4 from Paper

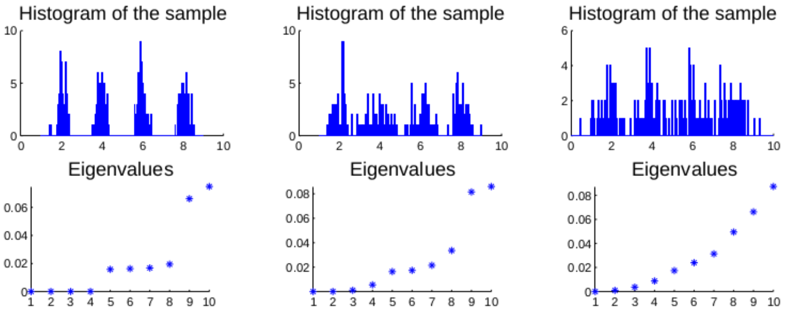


Figure 4: Three data sets, and the smallest 10 eigenvalues of L_{rw}. See text for more details.

Reproducing Figure 4

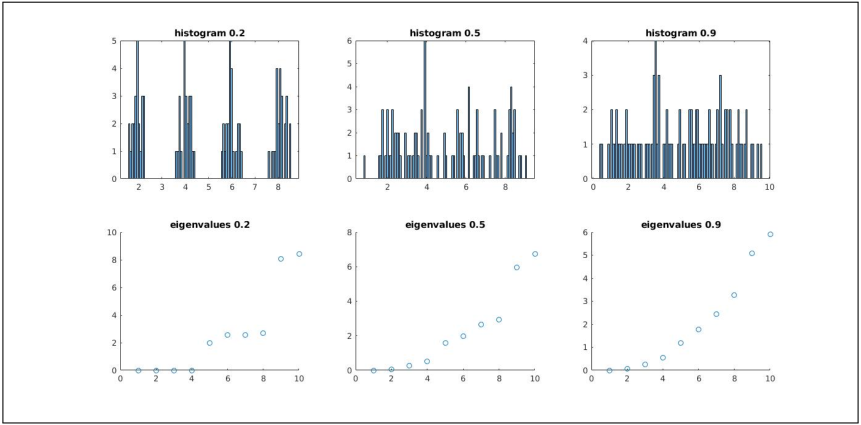


Figure 5 from Paper

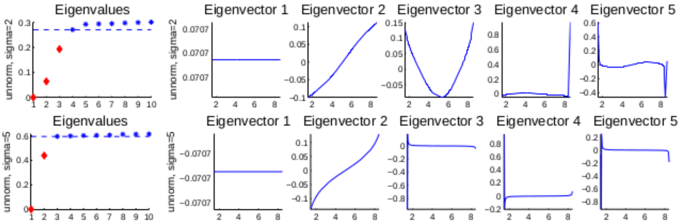
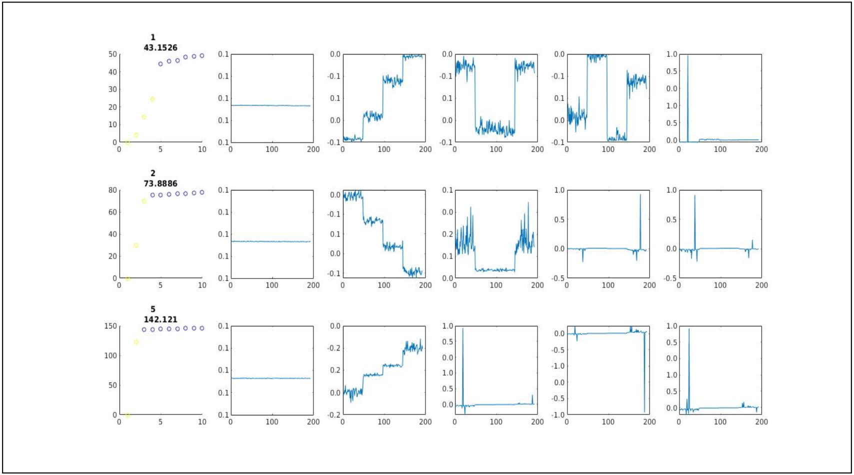


Figure 5: Consistency of unnormalized spectral clustering. Plotted are eigenvalues and eigenvectors of L, for parameter $\sigma = 2$ (first row) and $\sigma = 5$ (second row). The dashed line indicates $\min d_j$, the eigenvalues below $\min d_j$ are plotted as red diamonds, the eigenvalues above $\min d_j$ are plotted as blue stars. See text for more details.

Reproducing Figure 5



Using different metrics for kmeans

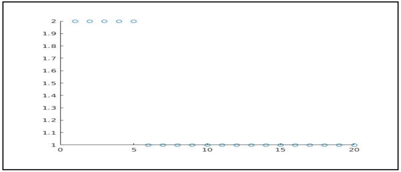


Figure: squared norm metric

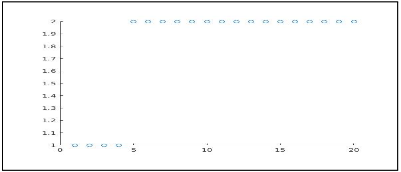


Figure: cityblock metric

Higher Eigenvectors of Laplacian

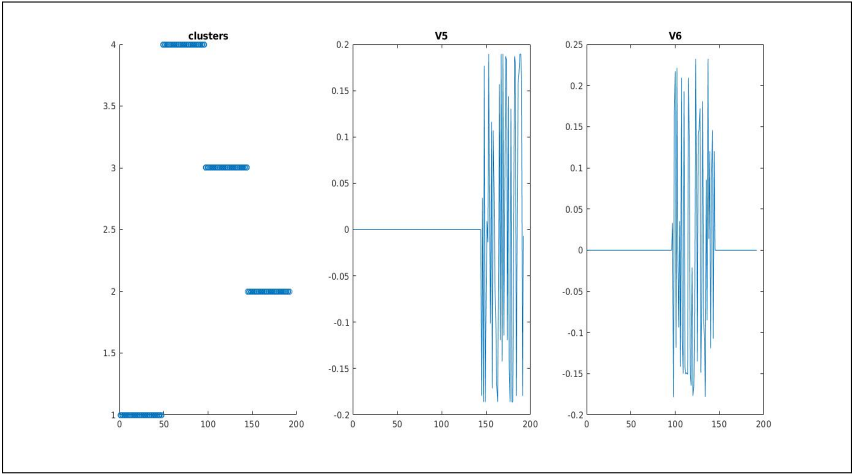
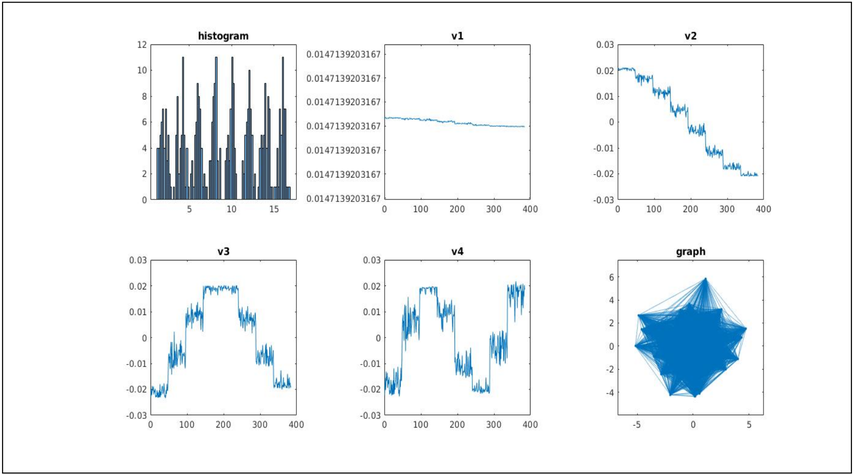


Figure: Higher Eigenvectors (5th and 6th) for unnormalized knn

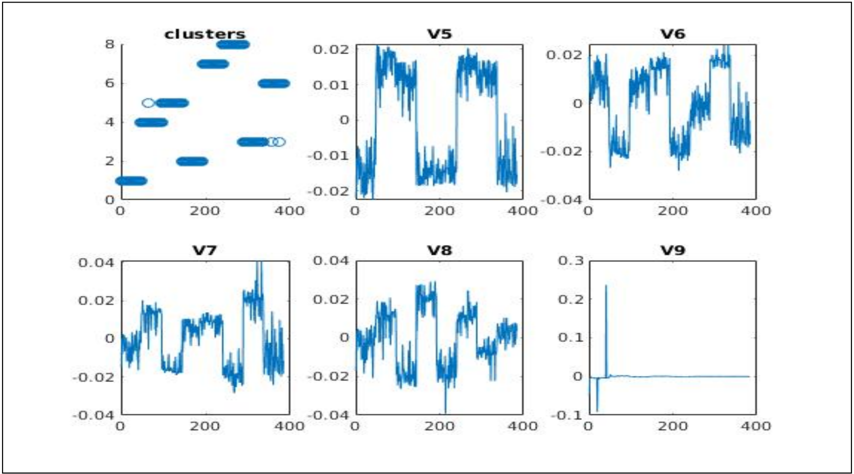
8 clusters: normalized, full graph

Figure: Data histogram and first 4 eigenvectors



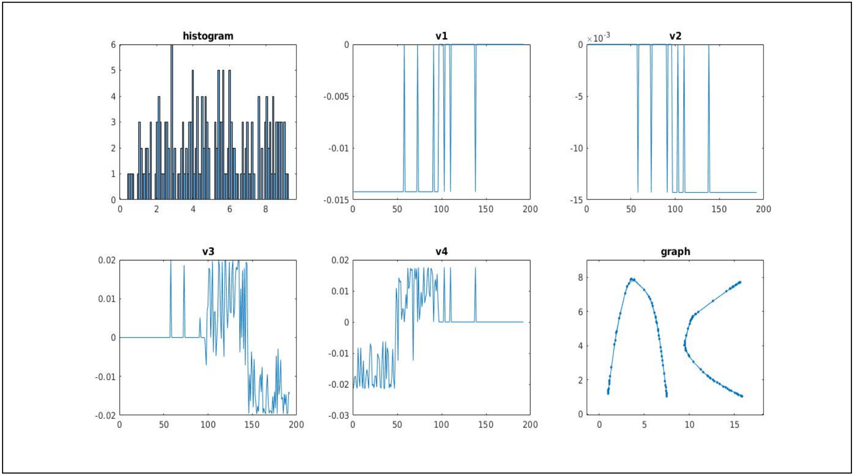
8 clusters: normalized, full graph

Figure: Clusters and last four eigenvectors



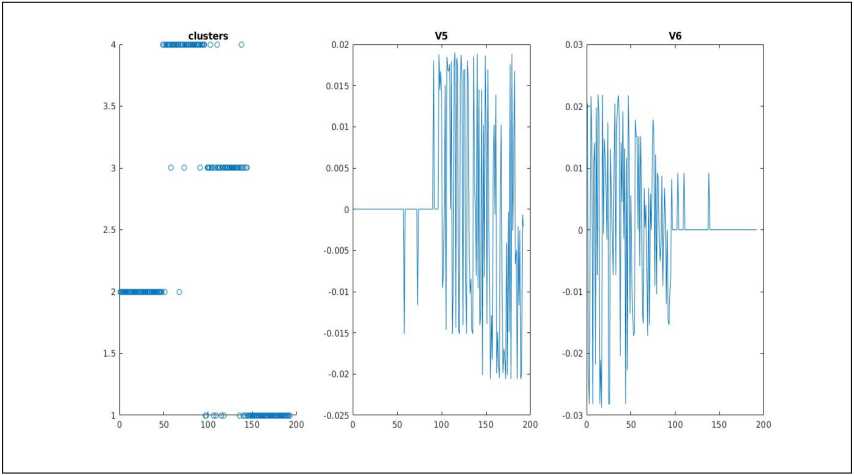
High variance input data

Figure: Data histogram and first 4 eigenvectors



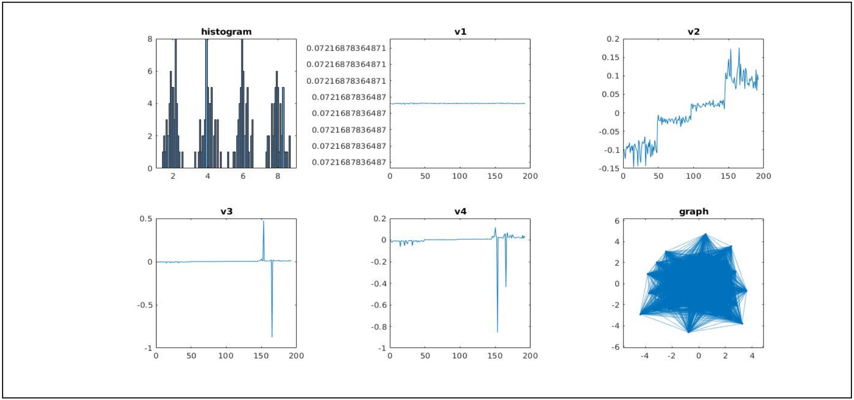
High variance input data

Figure: Clusters and last two eigenvectors



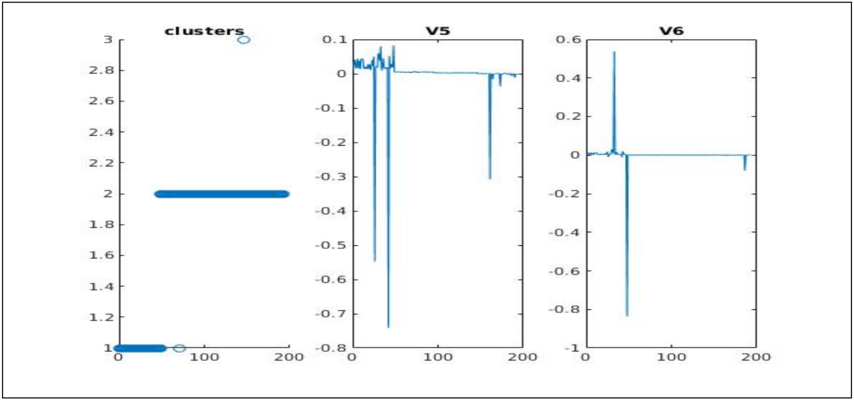
Dirac delta like behaviour of bigger eigenvectors for full graph, unnormalised clustering

Figure: Data histogram and first 4 eigenvectors



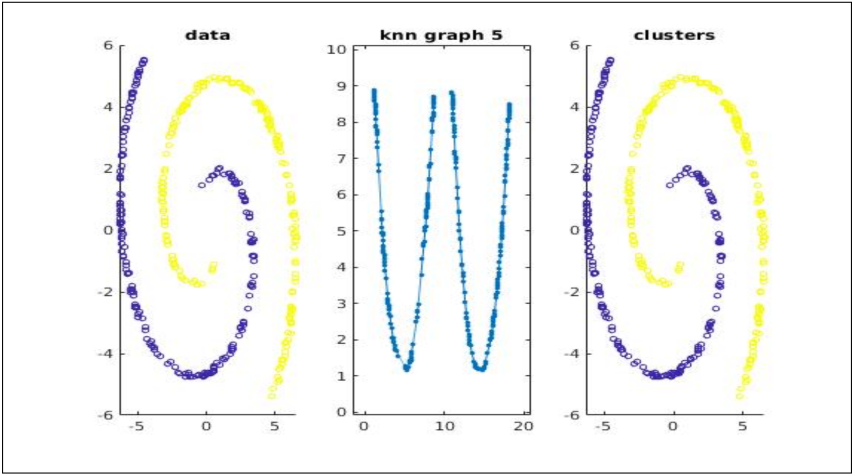
Dirac delta like behaviour of bigger eigenvectors for full graph, unnormalised clustering

Figure: Clusters and last two eigenvectors



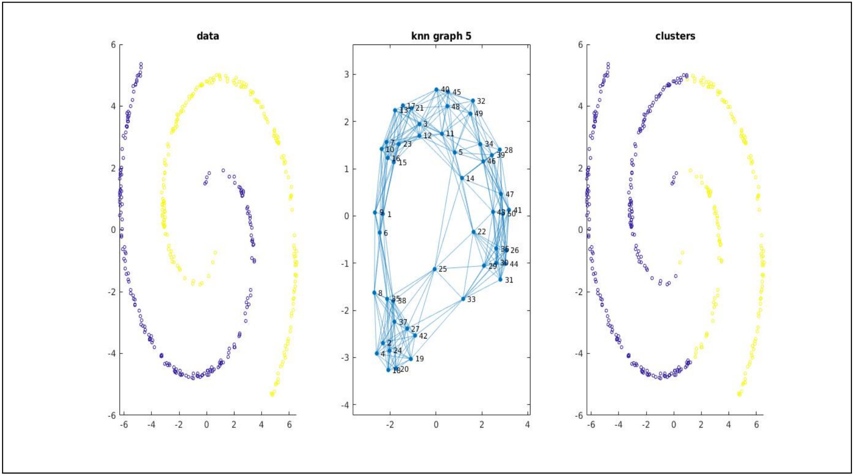
Clustering on spiral dataset

Figure: Clustering using unnormalized, mutual knn



Clustering on spiral dataset

Figure: Clustering using unnormalized, epsilon neighbour



References

Ulrike von Luxburg, *A Tutorial on Spectral Clustering*, http://arxiv.org/abs/0711.0189, 2007