Worksheet on "Spectral Clustering"

PRML – CS5691 (Jul–Nov 2023)

October 10, 2023

1. Write the laplacian matrix of this Graph .

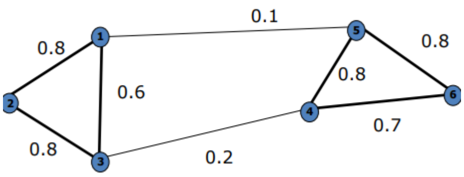

Figure 1:

Solution: First we need the adjacency matrix of this graph =

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0	0.8	0.6	0	0.1	0
x_2	0.8	0	0.8	0	0	0
x_3	0.6	0.8	0	0.2	0	0
x_4	0	0	0.2	0	0.8	0.7
x_5	0.1	0	0	0.8	0	0.8
x_6	0	0	0	0.7	0.8	0

Figure 2:

Then we find the degree matrix of this graph =

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.5	0	0	0	0	0
x_2	0	1.6	0	0	0	0
x_3	0	0	1.6	0	0	0
x_4	0	0	0	1.7	0	0
x_5	0	0	0	0	1.7	0
x_6	0	0	0	0	0	1.5

Figure 3:

So the laplacian matrix becomes =

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.5	-0.8	-0.6	0	-0.1	0
x_2	-0.8	1.6	-0.8	0	0	0
x_3	-0.6	-0.8	1.6	-0.2	0	0
x_4	0	0	-0.2	1.7	-0.8	-0.7
x_5	-0.1	0	0	-0.8	1.7	-0.8
x_6	0	0	0	-0.7	-0.8	1.5

Figure 4:

2. You have a dataset of four data points in two-dimensional space:
Data Point 1: (1, 2)
Data Point 2: (2, 3)
Data Point 3: (3, 4)
Data Point 4: (4, 5)
Perform spectral clustering to group these data points into two clusters. The similarity matrix has been computed using a Gaussian kernel with a bandwidth parameter of $\sigma = 1$, resulting in the following similarity matrix:

```
S = [[1.0, 0.9, 0.6, 0.3],
     [0.9, 1.0, 0.8, 0.5],
     [0.6, 0.8, 1.0, 0.9],
     [0.3, 0.5, 0.9, 1.0]]
```

Figure 5:

Perform the following steps of spectral clustering:

i) Compute the (unnormalized) Laplacian matrix L.
ii) Compute the eigenvalues and eigenvectors of L.
iii) Select the eigenvector corresponding to the smallest eigenvalue, normalize the eigenvector so that it is unit-norm, and check if applying a threshold on it will reveal the two clusters?
iv) Select the eigenvector corresponding to the second smallest eigenvalue of L, normalize this eigenvector, and check if choosing a threshold of 0.7 will reveal two similar clusters in the dataset?

In the last two parts, provide the clusters obtained after performing spectral clustering and explain your reasoning for the cluster assignments based on the eigenvector and threshold.

Solution: Compute the unnormalized Laplacian matrix L:

The unnormalized Laplacian matrix L can be computed as L = D - S, where D is the degree matrix and S is the similarity matrix.

The degree matrix D is a diagonal matrix where each element D[i][i] is the sum of the elements in the i-th row of the similarity matrix S.

For this dataset, D and L are as follows:

```
D = [[2.5, 0, 0, 0],
     [0, 2.2, 0, 0],
     [0, 0, 2.3, 0],
     [0, 0, 0, 2.7]]

L = [[1.0, -0.9, -0.6, -0.3],
     [-0.9, 1.0, -0.8, -0.5],
     [-0.6, -0.8, 1.0, -0.9],
     [-0.3, -0.5, -0.9, 1.0]]
```

Figure 6:

Compute the eigenvalues and eigenvectors of L:

Calculate the eigenvalues and eigenvectors of the Laplacian matrix L.

FILL ANsWER for smallest eigen value and corresp. eigen vector.

The second smallest non-zero eigenvalue is approximately 0.1722, and the corresponding eigenvector is
$[0.5475, 0.4665, -0.5475, -0.4315]$.
Select the eigenvector corresponding to the second smallest non-zero eigenvalue.

We've already identified the eigenvector as [0.5475, 0.4665, -0.5475, -0.4315].

Normalize this eigenvector:

Normalize the eigenvector by dividing it by its Euclidean norm:

Normalized eigenvector [0.664, 0.568, -0.664, -0.523]

Perform spectral clustering using the normalized eigenvector and a threshold of 0.7:

Assign each data point to Cluster 1 if the corresponding value in the normalized eigenvector is greater than or equal to 0.7; otherwise, assign it to Cluster 2.

Cluster assignments:

Data Point 1: Cluster 1
Data Point 2: Cluster 1
Data Point 3: Cluster 2
Data Point 4: Cluster 2

3. Let G be a simple undirected graph over n nodes, with d_i denoting the degree of the ith node. If the eigen values of the graph Laplacian L of G are ordered from the smallest to the largest (e.g., second smallest eigenvalue is λ_2), then show that $\lambda_2 \leq \frac{n}{n-1}\bar{d} \leq \lambda_n$. Here, $\bar{d} = \frac{1}{n}\sum_i d_i$ is the average degree of a node.

Solution: see `https://web.mit.edu/~jadbabai/www/ESE680/Laplacian_Thesis.pdf` - end of Pg 9 for soln.

4. Prove the following : Let G = (V, E) be a graph, and let i and j be vertices of degree one that are both connected to another vertex k. Then, the vector v given by

$$
\boldsymbol{v}(u) = \begin{cases} 1 & u = i \\ -1 & u = j \\ 0 & otherwise, \end{cases}
$$

Figure 7:

is an eigenvector of the Laplacian of G of eigenvalue 1.

Solution: One can immediately verify that LGv = v . The existence of this eigenvector implies that v(i) = v(j) for every eigenvector v of a different eigenvalue. `https://www.cs.yale.edu/homes/spielman/561/2009/lect02-09.pdf`

5. Prove the following : The graph Sn has eigenvalue 0 with multiplicity 1, eigenvalue 1 with multiplicity $n - 2$, and eigenvalue n with multiplicity 1.

Solution: To prove this statement, we can consider a specific graph called the star graph Sn and analyze its eigenvalues. The star graph Sn is a graph with n vertices, where one central node is connected to all other nodes (leaves) but the leaves are not connected to each other. It is known for its unique eigenvalue structure.
Here's a proof of the eigenvalues of the star graph Sn:

Eigenvalue 0 with Multiplicity 1:

The star graph Sn has one isolated vertex, which is not connected to any other vertices. Therefore, its adjacency matrix will have a zero in the corresponding diagonal element. This results in an eigenvalue of 0 with multiplicity 1.

Eigenvalue 1 with Multiplicity n - 2:

The remaining n - 2 vertices are all connected to the central node. These connections create a regular subgraph, which means the adjacency matrix of this subgraph has eigenvalues that are equal to 1. Since there are n - 2 such vertices, you have (n - 2) eigenvalues equal to 1.

Eigenvalue n with Multiplicity 1:

The sum of the degrees of all vertices in Sn is equal to n, and this graph is regular except for the isolated vertex. Therefore, the sum of the eigenvalues is equal to the sum of the degrees, which is n. Since we have one eigenvalue equal to 0 and (n - 2) eigenvalues equal to 1, the remaining eigenvalue must be n.

So, to summarize:

Eigenvalue 0 with multiplicity 1. Eigenvalue 1 with multiplicity n - 2. Eigenvalue n with multiplicity 1. This concludes the proof of the eigenvalues of the star graph Sn.

see lemma 2.4.3 `https://www.cs.yale.edu/homes/spielman/561/2009/lect02-09.pdf`