
Graph Spectral Clustering Quiz

Last Name	First Name	SID
-----------	------------	-----

Rules:

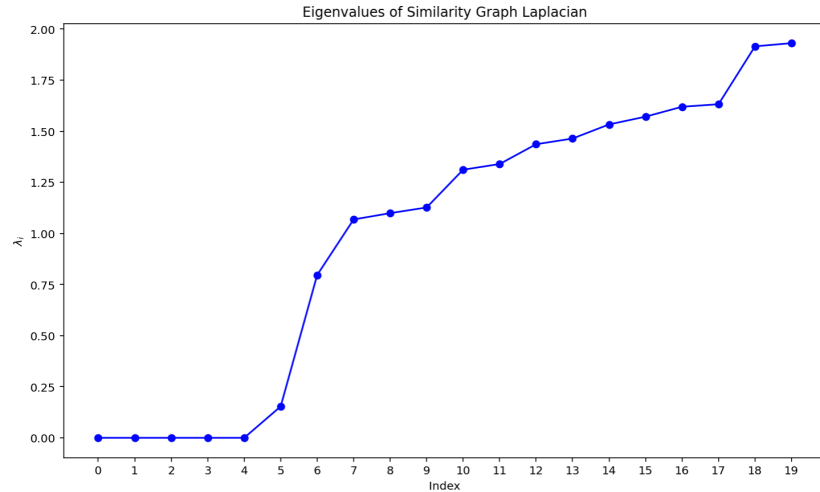
- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- You have 30 minutes to complete the quiz.
- Collaboration with others is strictly prohibited.
- You may not reference your notes, the textbook, and any material that can be found through the course website. You may not use Google to search up general knowledge.
- For any clarifications you have, please create a private Piazza post. We will have a Google Doc that shows our official clarifications.

1 True/False [1 point each]

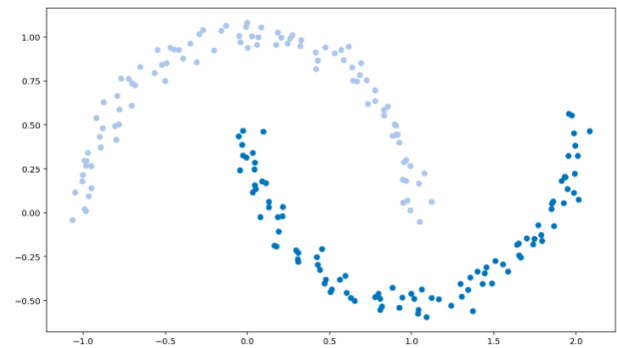
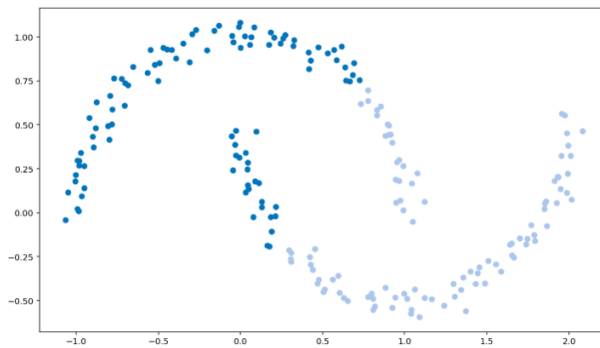
- 1) The largest eigenvectors of the similarity graph Laplacian matrix should be used to construct the eigenvector matrix that will eventually be used for clustering.
 - (a) True
 - (b) False
- 2) When performing k-means in the graph spectral clustering algorithm, k-means should be performed on the columns of the eigenvector matrix of the Laplacian matrix.
 - (a) True
 - (b) False
- 3) Graph spectral clustering builds on a centroid-based clustering algorithm to achieve more complex decision boundaries between clusters.
 - (a) True
 - (b) False
- 4) All entries in the similarity matrix used in graph spectral clustering are nonnegative.
 - (a) True
 - (b) False
- 5) The goal of spectral graph clustering is to separate points into clusters such that intra-cluster (within each cluster) similarity is minimized and inter-cluster (between different clusters) similarity is maximized.
 - (a) True
 - (b) False

2 Multiple Choice [2 points each]

- 1) Which matrix has diagonal entries that correspond to the sum of the weights of edges adjacent to each vertex?
 - (a) Degree matrix
 - (b) Adjacency matrix
 - (c) Laplacian matrix
 - (d) None of the above
- 2) Looking at the plot below for eigenvalue growth rate, which of the following values of k would be best to choose?

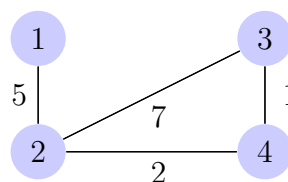


- (a) 3
 (b) 5
 (c) 7
 (d) 9
- 3) For the two plots shown below, one is the result of spectral clustering with a Gaussian similarity function with $\sigma = 0.1$, and another is the result of spectral clustering with a Gaussian similarity function with $\sigma = 1$. Label the plots with the corresponding σ value of the Gaussian similarity function used.



3 Short Answers [3 points each]

- 1) Find the Laplacian matrix for this graph:



- 2) What is the role of a similarity function in graph spectral clustering? List 2 other examples of valid similarity functions.
- 3) Explain why it is always possible to compute the first k eigenvectors of the Laplacian matrix in spectral clustering.
- 4) What effect does decreasing σ have on the RBF kernel?
- 5) Why do we prefer using k-means++ over k-means in general? Give two reasons.
- 6) Fill in the blanks to complete the pseudocode of the normalized graph spectral clustering algorithm.

Return clusters A_1, \dots, A_k

Given data $X \in \mathbb{R}^{n \times d}$, number of desired clusters k

- (a) Construct a _____ $S \in \mathbb{R}^{n \times n}$ according to some well defined metric, such as the Gaussian kernel or Euclidean distance.
- (b) Compute $L_n = \text{_____}$, the symmetric normalized graph Laplacian of S .
- (c) Compute the first k _____ of L , where the first k _____ correspond to _____.
- (d) Construct a matrix H for L with _____, making sure to normalize the _____ to have norm 1.
- (e) Run _____ on _____ to cluster the data.