HW 5 – 4803, Spring 2023

Each problem is worth 10 points

1 Part I – theoretical problems

1. From the Book "An Introduction to Statistical Learning"– 10.10 Exercise 2

2-3. From the Book "An Introduction to Statistical Learning"– 12.6 Exercises 1 and 3

4. Consider an n-complete graph, with vertex $\{v_1, v_2, \ldots, v_n\}$. Every vertex is connected with the other $n-1$ vertices, and the connected edge has weight 1.

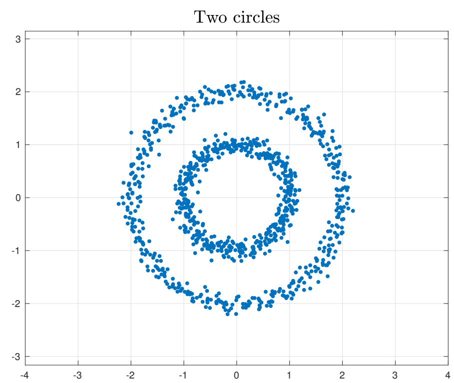
   1. What is the graph Laplacian?
   2. What are the eigenvalues and eigenvectors of the graph Laplacian?

2 Part II – programming

1-2. From the Book "An Introduction to Statistical Learning"– 10.10 Exercises 6 and 7

3. From the Book "An Introduction to Statistical Learning"– 12.6 Exercise 10

4. Generate the following data set near two circles. The circles are centered at the origin and have radius 1 and 2 respectively.



We will generate data in the following steps:

- First generate 200 samples on each circle. Let us parameterize the circles as $x_1(t) = r \cos t$; $x_2(t) = r \sin t$ where $r = 1, 2$ respectively. For each circle, 200 uniform samples of $t \in [0, 2\pi)$ give rise to 200 points on the circle.
- Add Gaussian noise to each sample above. The noise vector is $[n_1 \ n_2]$ where $n_1, n_2 \sim \text{Normal}(0, \sigma^2)$, where $\sigma = 0.05$.

(a) Apply K means with $K = 2$, and display the clustering results.

(b) Apply spectral clustering and cluster this data set into 2 clusters. Construct the ϵ-neighborhood graph, and display the clustering results with three choices of ϵ. Try a large ϵ, a proper ϵ, and a small ϵ. What is a good range of ϵ such that we can cluster the two circles?

(c) Apply spectral clustering and cluster this data set into 2 clusters. Construct the k- nearest neighbor graph, and display the clustering results with three choices of k. Try a large k, a proper k, and a small k. What is a good range of k such that we can cluster the two circles?

(d) Repeat the experiments in (b) and (c) where the data set has larger noise: $\sigma = 0.2$.