401-4944-20L Mathematics of Data Science: Problem Set 2

(Spring 2020)

Afonso S. Bandeira
bandeira@math.ethz.ch
https://people.math.ethz.ch/~abandeira/

April 12, 2020

This homework is optional and it won't be graded. If you want to discuss a solution (to make sure it is correct) or want to ask questions about a problem stop by office hours or write a TA or myself an email and we can schedule a time to talk. Date is of last update (e.g. correction of typos)

Try not to look up the answers, you'll learn much more if you try to think about the problems without looking up the solutions. If you need hints, feel free to email me.

Problem 2.1 *For a graph G let $h(G)$ denote its Cheeger constant and $\lambda_2(\mathcal{L}_G)$ the second smallest eigenvalue of its normalized graph Laplacian. Recall that Cheeger inequality guarantees that*

$$\frac{1}{2}\lambda_2(\mathcal{L}_G) \leq h_G \leq \sqrt{2\lambda_2(\mathcal{L}_G)}.$$

This exercise shows that this inequality is tight (at least up to constants).

1. *Construct a family of graphs for which $\lambda_2(\mathcal{L}_G) \to 0$ and for which there exists a constant $C > 0$ for which, for every G in the family,*

$$h_G \leq C\lambda_2(\mathcal{L}_G)$$

2. *Construct a family of graphs for which $\lambda_2(\mathcal{L}_G) \to 0$ and for which there exists a constant $c > 0$ for which, for every G in the family*

$$h_G \geq c\sqrt{\lambda_2(\mathcal{L}_G)}$$

Problem 2.2 *Given a graph G show that the dimension of the nullspace of L_G corresponds to the number of connected components of G.*

Problem 2.3 *Given a set of points $x_1, \ldots, x_n \in \mathbb{R}^p$ and a partition of them in k clusters $S_1, \ldots, S_k$ recall the k-means objective*

$$\min_{S_1,\ldots,S_k} \min_{\mu_1,\ldots,\mu_k} \sum_{l=1}^{k} \sum_{i \in S_i} \|x_i - \mu_l\|^2 .$$

Show that this is equivalent to

$$\min_{S_1,\ldots,S_k} \sum_{l=1}^{k} \frac{1}{|S_l|} \sum_{i,j \in S_l} \|x_i - x_j\|^2 .$$

Diffusion Maps and other embeddings

Problem 2.4 *The ring graph on n nodes is a graph where node $1 < k < n$ is connected to node $k-1$ and $k+1$ and node 1 is connected to node n. Derive the two-dimensional diffusion map embedding for the ring graph (if the eigenvectors are complex valued, try creating real valued ones using multiplicity of the eigenvalues). Is it a reasonable embedding of this graph in two dimensions?*