COMPSCI 514: Problem Set 4

Due: Monday December 5, 11:59pm in Gradescope.

Instructions:

- You are allowed to, and highly encouraged to, work on this problem set in a group of up to three members.
- Each group should **submit a single solution set**: one member should upload a pdf to Gradescope, marking the other members as part of their group in Gradescope.
- You may talk to members of other groups at a high level about the problems but **not work through the solutions in detail together**.
- You must show your work/derive any answers as part of the solutions to receive full credit.

1. Line of Best Fit (6 points)

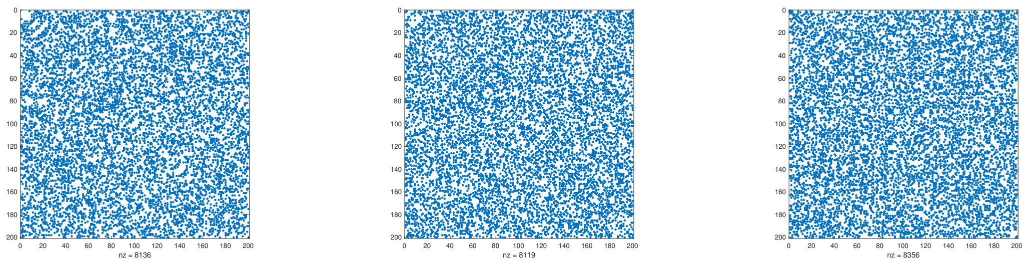
1. (2 points) Consider invertible $\mathbf{A} \in \mathbb{R}^{d\times d}$ with SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Prove that $\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$.
2. (2 points) Consider any $\mathbf{A} \in \mathbb{R}^{n\times d}$ with SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. One of the most classic data fitting methods, least squares regression is: given a vector $\vec{y} \in \mathbb{R}^n$, find:

$$\vec{\beta}_* \in \arg\min_{\vec{\beta}\in\mathbb{R}^d} \|\mathbf{A}\vec{\beta} - \vec{y}\|_2^2. \tag{1}$$

   The rows of $\mathbf{A}$ represent d-dimensional data points, the entries of $\vec{y}$ represent observations at these points, and $\mathbf{A}\vec{\beta}_*$ is the 'line of best fit', which attempts to fit these observations as closely as possible with a linear function of the rows. Prove that $\vec{\beta}_* = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T\vec{y}$ satisfies equation (1) above. Avoid using any calculus in your proof. **Hint:** The solution will involve a projection matrix.
3. (2 points) Describe in a few sentences how part (2) relates to part (1). What is $\|\mathbf{A}\vec{\beta}_* - \vec{y}\|_2^2$ when $\mathbf{A}$ is square and invertible?

2. Uncovering Graph Structure (9 points)

Consider the three graph adjacency matrices shown below. and available in the `graphs.mat` file. One of these graphs is a completely random graph on 200 nodes – where each edge is added independently with probability p_1. The other is a graph with two connected components. Each connected component is a completely random graphs on 100 nodes, where each edge is added independently with probability p_2. Finally, the third graph is a stochastic block model graph with intra and inter community connection probabilities p_3 and q_3. The connection probabilities are set so that all graphs have roughly the same average degree.



(a) Graph A (b) Graph B (c) Graph C

1. (3 points) Identify which graph is which. Include a printout of any code you use and a plot(s) to justify your answer. **Hint:** Consider reordering the nodes according to the values that they are assigned in the second eigenvector of the adjacency matrix.

2. (3 points) In terms of p_1, p_2, p_3, and q_3, what at the largest and second largest eigenvalues of the expected adjacency matrices of these three graphs?

3. (3 points) Use parts (1) and parts (2) to estimate the values for p_1, p_2, p_3, and q_3. Include a printout of any that code you use.

3. The Power of Message Passing (20 points)

Consider an undirected, unweighted, d-regular graph on n nodes. I.e., a graph where every node has degree d. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be its adjacency matrix. Think of each node in the graph as a user, and the edges as representing communication links between the users.

1. (3 points) Prove that the maximum magnitude eigenvalue of $\mathbf{A}$ is equal to d. **Hint:** First exhibit an eigenvector with eigenvalue d. Then show that this is the maximum magnitude eigenvalue by showing that no eigenvector can have eigenvalue $> d$.

2. (2 points) Prove that if the graph is *disconnected* then $\mathbf{A}$ actually has two orthogonal eigenvectors both with eigenvalue d. **Hint:** Here it just suffices to exhibit the eigenvectors and check their corresponding eigenvalues.

3. (2 points) Prove that if the graph is *bipartite* then $\mathbf{A}$ has an eigenvector with eigenvalue $-d$. **Hint:** Here it just suffices to exhibit this eigenvector and check that its corresponding eigenvalue equals $-d$.

Consider now that setting where user has some initial value z_i and they want to estimate the average value $\mu = \frac{1}{n}\sum_{i=1}^{n} z_i$. Consider the following simple distributed averaging process: each user sets their initial estimate of the average to $\mu_i^{(0)} = z_i$. Then, at each step, each user sends its current estimate of the average $\mu_i^{(t)}$ to all of its neighbors in the network. Each user then updates their estimate to be the average of their neighbors' estimates. I.e., they set $\mu_i^{(t+1)} = \frac{1}{d}\sum_{j \in \mathcal{N}(i)} \mu_j^{(t)}$. Here $\mathcal{N}(i)$ denotes the set of neighbors of the i^{th} node.

4. (2 points) Write this averaging process as a linear algebraic equation involving $\mathbf{A}$, the vector of estimates at time t, $\vec{\mu}^{(t)} \in \mathbb{R}^n$, and the vector of estimates at time t, $\vec{\mu}^{(t+1)} \in \mathbb{R}^n$.

5. (2 points) Show that we can write $\vec{\mu}^{(0)} = \mu \cdot \vec{1} + c_2\vec{v}_2 + \ldots + c_n\vec{v}_n$, where $\vec{1} \in \mathbb{R}^n$ is the all ones vector, $\vec{v}_2, \ldots, \vec{v}_n \in \mathbb{R}^n$ are orthonormal eigenvectors of $\mathbf{A}$, which are all orthonormal to $\vec{1}$, and $c_2, \ldots, c_n$ are some coefficients.

6. (2 points) Show that similarly, we can write $\vec{\mu}^{(t)} = \mu \cdot \vec{1} + \left(\frac{\lambda_2}{d}\right)^t c_2\vec{v}_2 + \ldots + \left(\frac{\lambda_n}{d}\right)^t c_n\vec{v}_n$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of $\mathbf{A}$.

7. (2 points) Argue that as long as all eigenvalues except the largest have magnitude $< d$, then for all i, $\lim_{t\to\infty} \mu_i^{(t)} = \mu$. I.e., we converge to a state where all nodes know the true mean μ.

8. (2 points) By parts (2) and (3), we know that if the graph is disconnected or bipartite, then it has multiple eigenvalues with magnitude d, and thus the above result does not hold. Describe intuitively, why the estimates are not guaranteed to converge to the true mean μ in both these cases.

9. (3 points) Describe and analyze a modification of the averaging protocol that leads to it converging even when the graph is bipartite. **Hint:** Consider adding a self-loop to each node.

4. Graph Resistance (8 points)

Given a graph G with Laplacian $\mathbf{L} \in \mathbb{R}^{n\times n}$, the *effective resistance* between two nodes i and j can be defined as:

$$r_{i,j} = \frac{1}{\min_{\vec{x}\in\mathbb{R}^n : x_i=1, x_j=0} \vec{x}^T\mathbf{L}\vec{x}} .$$

This is the same effective resistance you may have learned about in introductory physics, if we view the graph as a network of unit resistors.

1. (2 points) Let $z_1 = 1$ and $z_n = 0$ and consider the function

$$f(z_2, z_3, \ldots, z_{n-1}) = (z_1 - z_2)^2 + (z_2 - z_3)^2 + (z_3 - z_4)^2 + \ldots + (z_{n-1} - z_n)^2 .$$

   Find a setting of the values $z_2, z_3, \ldots, z_{n-1}$ such that $f(z_2, z_3, \ldots, z_{n-1}) = 1/(n-1)$. Prove that for all $z_2, z_3, \ldots, z_{n-1} \in \mathbb{R}$, $f(z_2, z_3, \ldots, z_{n-1}) \geq 1/(n-1)$.

2. (2 points) Suppose the only edges in G is a path of $n-1$ edges between node 1 and node n. What is the value of $r_{1,n}$?

3. (2 points) Now suppose that the edges in G consist of t paths between node 1 and node n. These paths doesn't share any nodes except for node 1 and node n. The paths have lengths $\ell_1, \ell_2, \ldots, \ell_t$. What is the value of $r_{1,n}$?

4. (2 points) Let $c_{i,j}$ be the minimum number of edges that need removed such that node i and node j are disconnected. Prove that $r_{i,j} \geq 1/c_{i,j}$.