ELL888: Advanced Machine Learning / AIL723: Graph Machine Learning

Assignment 1

Date of release: 13th Febrruary & **Last date:** 12th March 2024.

1. You are free to discuss the assignment problems with other students in the class. But all your answers/-codes should be produced independently without looking at/referring to anyone else's answers/codes.
2. The submission directory should contain a report in .pdf format (answers/ derivations and results) and code. You should zip your directory and name the resulting file as "yourentrynumber-firstname-lastname.zip".

1 Spectral Clustering [40 points]

Given a graph $G = \{\mathcal{V}, \mathcal{E}\}$, the K-way *normalized minCUT* problem (simply referred to as *minCUT*) is the task of partitioning $\mathcal{V}$ in K disjoint subsets by removing the minimum volume of edges. The problem is equivalent to maximizing

$$\frac{1}{K}\sum_{k=1}^{K}\frac{\text{links}\mathcal{V}_k}{\text{degree}\mathcal{V}_k} = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{i,j\in\mathcal{V}_k}\mathcal{E}_{i,j}}{\sum_{i\in\mathcal{V}_k, j\in\mathcal{V}/\ \mathcal{V}_k}\mathcal{E}_{i,j}}, \tag{1}$$

where the numerator counts the edge volume within each cluster, and the denominator counts the edges between the nodes in a cluster and the rest of the graph. Let $C \in \{0,1\}^{N\times K}$ be a cluster assignment matrix so that $Ci, j = 1$ if node i belongs to cluster j, and 0 otherwise.

1. **[5 points]** Formulate the *minCUT* problem as a optimization problem.
2. **[10 points]** Show that the optimization problem is NP-hard, recast it in a relaxed continuous formulation that can be solved in polynomial time and guarantees a near-optimal solution.
3. **[25 points]** Write Python code* for the following tasks for Cora ($K = 7$) and Citeseer ($K = 6$) datasets.
   (a) Plot the degree distribution on an appropriate scale.
   (b) Perform spectral analysis of unnormalized, normalized and random-walk Laplacians. State insights you observed.
   (c) Perform **spectral clustering** and evaluate the performance of your algorithm in terms of standard metric(s) and give the space and time complexity.
   (d) Give the possible approaches (implement) to scale spectral clustering for large N with $\mathcal{O}(N^2)$ or $\mathcal{O}(N)$. Also, compare your results with what you have obtained in (c).

$\star$ Do not use in-built libraries; write your own code (you can use in-build libraries for data processing).

2 Graph Learning [40 points]

Formulate (and implement*) graph learning problem via the Probabilistic Graphical Model (GGM) and smooth signal model. Present comparison and analysis using the MNIST dataset.

$\star$ Do not use in-built libraries; write your own code (you can use in-build libraries for data processing).

3 Gaussian Graphical Models [20 points]

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector (not necessarily Gaussian) with mean $\boldsymbol{\mu}$ and covariance matrix Σ. The partial correlation matrix R of $\mathbf{X}$ is a $d \times d$ matrix where each entry $R_{ij} = \rho(X_i, X_j|\mathbf{X}_{-ij})$ is the partial correlation between X_i and X_j given the $d-2$ remaining variables $\mathbf{X}_{-ij}$. Let $\Theta = \Sigma^{-1}$ be the inverse covariance matrix of $\mathbf{X}$.

We will prove the relation between R and Θ, and how Θ characterizes conditional independence in Gaussian graphical models.

1. **[5 points]** Show that

$$\begin{pmatrix} \Theta_{ii} & \Theta_{ij} \\ \Theta_{ji} & \Theta_{jj} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}[e_i] & \mathrm{Cov}[e_i, e_j] \\ \mathrm{Cov}[e_i, e_j] & \mathrm{Var}[e_j] \end{pmatrix}^{-1} \tag{2}$$

   for any $i, j \in [d]$, $i \neq j$. Here e_i is the residual resulting from the linear regression of $\mathbf{X}_{-ij}$ to X_i, and similarly e_j is the residual resulting from the linear regression of $\mathbf{X}_{-ij}$ to X_j.

2. **[10 points]** Show that

$$R_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}}\sqrt{\Theta_{jj}}} \tag{3}$$

3. **[5 points]** From the above result and the relation between independence and correlation, we know $\Theta_{ij} = 0 \iff R_{ij} = 0 \Longleftarrow X_i \perp X_j \mid \mathbf{X}_{-ij}$. Note the last implication only holds in one direction.

   Now suppose $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ is jointly Gaussian. Show that $R_{ij} = 0 \implies X_i \perp X_j \mid \mathbf{X}_{-ij}$.