Assignment 2: ML & GCN

Due: Nov. 20, 2023.
The assignment should be submitted in PDF format with the filename "Assignment2-Your Name-Your Student No." to TA (Defang Chen, email: defchern@zju.edu.cn).

1. PAC Bound

In class, we have derived the generalization bound for the case where the loss function values are confined to the interval $[0,1]$. Please generalize this conclusion to the case where the loss function is bounded with $[C_1, C_2]$.

That is, given a training set $S = \{(x_i, y_i)\}_{i=1}^{m}$ sampled i.i.d from the distribution D, let define the true risk as $L_D(h) = \underset{(x,y)\sim D}{\mathbf{E}}[l(h,x,y)]$, the empirical risk used for model training as $L_S(h) = \frac{1}{m}\sum_{i=1}^{m} l(h, x_i, y_i)$, $l(h,x,y)$ denote the loss function with $l(h,x,y) \in [C_1, C_2]$. For any finite hypothesis space of $\mathcal{H}$, and for any learned function $\tilde{h} = \mathrm{argmin}_{h\in\mathcal{H}} L_S(h)$, with probability $1-\delta$, what is the generalization bound for $L_D(\tilde{h})$? Please provide a detailed derivation.

2. Laplacian matrix

Given a graph with adjacent matrix A, Laplacian matrix is defined as $\mathrm{L} = \mathrm{D} - \mathrm{A}$, where D is diagonal matrix with i-th diagonal element is the degree of i-th node.

1. Please prove Laplacian matrix L is positive-semidefinite.
2. Let define normalized Laplacian matrix as $\hat{L} = D^{-1/2} L D^{-1/2}$, please derive the upper/lower bound of the eigenvalues of $\hat{L}$.