CPSC 483/583: Deep Learning on Graph-structured Data **Fall 2022**

Homework 2

Due 11:59pm ET Thursday October 6 2022

This problem set should be completed individually.

General Instructions

These questions require thought, but do not require long answers. Please be as concise as possible. You are allowed to take a maximum of 1 late period (see the course website or slides about the definition of a late period).

Submission instructions: You should submit your answers in a PDF file. LaTeX is highly preferred due to the need of formatting equations.

Submitting answers: Prepare answers to your homework in a single PDF file. Make sure that the answer to each sub-question is on a *separate page*. The number of the question should be at the top of each page.

Honor Code: When submitting the assignment, you agree to adhere to the Yale Honor Code. Please read carefully to understand what it entails!

Homework survey: After submitting your homework, please fill out the Homework 2 Feedback Form. 0.5% overall extra credit will be given if you take time to reflect on the nature of the problems and possibly provide feedbacks for the benefit of future offering of this course, by filling in the form for each written and coding assignment.

Questions

1 Invariance and Equivariance of GNNs

Permutation Matrix A permutation π of m element is defined by $\begin{pmatrix} 1 & 2 & \cdots & m \\ \pi(1) & \pi(2) & \cdots & \pi(m) \end{pmatrix}$, where the i-th element is permuted to $\pi(i)$. The permutation matrix $P_\pi = (p_{ij})$ is defined by $p_{ij} = 1$ if $j = \pi(i)$ and $p_{ij} = 0$ otherwise. For example, the permutation matrix P_π corresponding to the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$ is $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$.

An important property of permutation matrices is that they are orthogonal matrices (i.e., $P_\pi P_\pi^\top = I$). Therefore, we have $P_\pi^{-1} = p_\pi^\top$.

Invariance and Equivariance Consider $X \in \mathbb{R}^{n \times d}$, a function $f(X)$ is **permutation invariant** if, for all permutation matrices $P \in \mathbb{R}^{n \times n}$: $f(PX) = f(X)$ holds. Accordingly, we say that $f(X)$ is **permutation equivariant** if, for all permutation matrices $P \in \mathbb{R}^{n \times n}$: $f(PX) = Pf(X)$ holds.

1. If there is a permutation P applied to the nodes, the original features $X \in \mathbb{R}^{n \times d}$ is changed to $PX \in \mathbb{R}^{n \times d}$, where n and d are the number of nodes and feature dimensions. Node permutations also accordingly act on the edges. Write out how the adjacency matrix $A \in \mathbb{R}^{n \times n}$ changes accordingly after node permutation.
2. Consider the message passing function of a graph convolutional layer: $H^{(l+1)} = \sigma(AH^{(l)}W_l^\top)$, $W_l^\top$ is the weight matrix. Prove that a graph convolutional layer is permutation equivariant.
3. Prove that if each layer of the network is permutation equivariant, then the network as a whole is permutation equivariant. (**Hint:** consider each layer as a function, then, stacking layers is equivalent to function composition. Prove the resulting composite function is permutation equivariant)

2 Laplacian Matrix over Graph

The graph Laplacian matrix is the 'discrete' version of the Laplacian operator over graphs. You might remember from multi-variable calculus, that in Euclidean space, the Laplace operator is the divergence of the gradient of a function, i.e. $\Delta f = \mathrm{div}(\nabla(f))$, where f is a function that maps n-dimensional real vector to a real number: $f : \mathbb{R}^d \mapsto \mathbb{R}$. Here, we want to extend this definition of Laplace operator to functions defined on graphs (instead of Euclidean space).

Step1: functions over the graph: A function on graph f is defined as a mapping from every node in the graph to a real vector space: $f : V \mapsto \mathbb{R}^d$. It can be 1-dimensional or a vector. In our demonstration, we use columns to represent the 1-dimensional function values over the nodes. That means, we define f as $f : V \mapsto \mathbb{R}$.

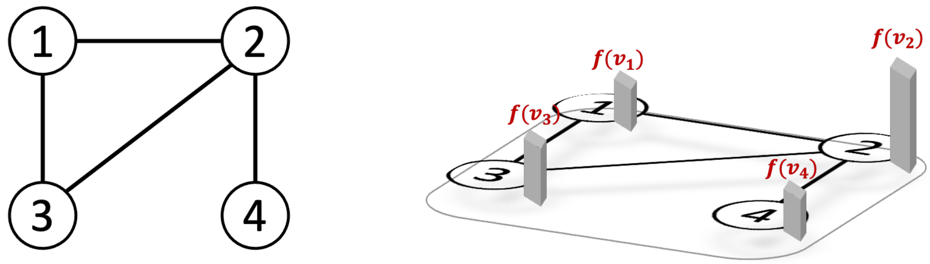

Figure 1: Graph and its associated function

Step2: define the gradient: The gradient operator gives the derivative of the function along each direction. In a graph, gradients are defined over each edge. Thus, gradient of the function along the edge $e = (v_i, v_j)$ is given by $\nabla f|_e = f(v_i) - f(v_j)$.

Step3: define the divergence of graph gradient: In a graph, the gradients over edges associated with a node construct a vector field over this node. Divergence at this node gives net outward flux of a vector field, that is, $\operatorname{div}(\nabla f)(v_i) = \sum_{(v_i,v_j)\in\mathcal{E}}(f(v_i) - f(v_j))$.

1. Graph Laplacian operator L is defined by the divergence of graph gradient. When applying L to a function f defined on all vertices of a graph, we have

$$Lf(v_i) = \operatorname{div}(\nabla f)(v_i)$$

   Actually, applying the Laplacian operator L to the function f is equivalent to a matrix product with the vector $\mathbf{f}$ (a vector with $f(v_i)$ as the i-th entry). Use the degree matrix D and adjacency matrix A to represent the Laplacian matrix.

2. Write out the Laplacian matrix of the graph in Figure 1.

3. Normalized Laplacian matrix $\tilde{L} = D^{-1/2}LD^{-1/2}$, and normalized adjacency matrix $\tilde{A} = D^{-1/2}AD^{-1/2}$. Figure out the relation between $\tilde{L}$ and $\tilde{A}$. Rewrite the message passing function of a graph convolutional layer $H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W_l^\top)$ with normalized Laplacian matrix $\tilde{L}$.

4. Dirichlet energy of a graph is defined as $E(\mathbf{f}) = \operatorname{trace}(\mathbf{f}^\top L\mathbf{f})$. Assuming that all node features are 1-dimensional, prove that $E(\mathbf{f}) = \mathbf{f}^\top L\mathbf{f} = \frac{1}{2}\sum_{(v_i,v_j)\in\mathcal{E}}(f(v_i) - f(v_j))^2$.

5. Over-smoothing phenomenon means the node embeddings $f(v_i)$ tend to become similar and indistinguishable, which is common after many layers of message passing in a GNN. Dirichlet energy is widely used to measure the degree of over-smoothing. Explain why the over-smoothing phenomenon corresponds to a small value of Dirichlet energy (high-level intuition is sufficient).