Solal Danan
solal.danan@polytechnique.edu

Lab session # 6
ALTEGRAD 2024

12/03/24

1 Question 1

In a message passing layer, the embedding of a node v_i is updated by aggregating information from its own embedding and the embeddings of its neighbors $\mathcal{N}(v_i)$. Formally, the embedding $z_i^{(t+1)}$ at layer $t+1$ is computed as:

$$z_i^{(t+1)} = \text{Aggregate}\Big(z_i^{(t)}, \{z_j^{(t)} \mid j \in \mathcal{N}(v_i)\}\Big),$$

where:

- $z_i^{(t)}$ is the embedding of node v_i at layer t,
- Aggregate$(\cdot)$ is a function (attention-weighted sum in the lab).

In our model, the embedding of a node v_i is updated in a message-passing layer using attention scores. Specifically, the embedding $z_i^{(t+1)}$ at layer $t+1$ is computed as:

$$z_i^{(t+1)} = \sum_{j \in \mathcal{N}(v_i)} \alpha_{ij}^{(t+1)} W^{(t+1)} z_j^{(t)},$$

where:

- $W^{(t+1)}$ is a trainable weight matrix,
- $\alpha_{ij}^{(t+1)}$ is the attention score assigned to neighbor v_j of v_i at layer $t+1$.

The embedding of v_1 at the second layer is:

$$z_1^{(2)} = \alpha_{12}^{(2)} W^{(2)} z_2^{(1)} + \alpha_{13}^{(2)} W^{(2)} z_3^{(1)}.$$

The embedding of v_4 at the second layer is:

$$z_4^{(2)} = \alpha_{42}^{(2)} W^{(2)} z_2^{(1)} + \alpha_{43}^{(2)} W^{(2)} z_3^{(1)} + \alpha_{45}^{(2)} W^{(2)} z_5^{(1)} + \alpha_{46}^{(2)} W^{(2)} z_6^{(1)}.$$

Since $z_1^{(1)} = z_4^{(1)}$, the initial embeddings of v_1 and v_4 are identical. Additionally:

- For v_1, the neighbors are v_2 and v_3, contributing embeddings $z_2^{(1)}$ and $z_3^{(1)}$.
- For v_4, the neighbors are v_2, v_3, v_5, v_6. Since $z_2^{(1)} = z_6^{(1)}$ and $z_3^{(1)} = z_5^{(1)}$, the embeddings $z_5^{(1)}$ and $z_6^{(1)}$ replicate the information already provided by $z_2^{(1)}$ and $z_3^{(1)}$.

However, the attention scores $\alpha_{ij}^{(t+1)}$ are computed based on the pairwise embeddings of nodes, and the additional neighbors v_5 and v_6 for v_4 could affect the weighting of $z_2^{(1)}$ and $z_3^{(1)}$. As a result, the aggregated embeddings for v_4 may not be identical to those of v_1.
In conclusion, while v_1 and v_4 share equivalent node embeddings at layer 1 and have overlapping neighbor contributions, the presence of additional neighbors for v_4 introduces differences in attention weighting. Thus:

$$z_1^{(2)} \neq z_4^{(2)}.$$

2 Question 2

The embedding update equation for the first layer of the GNN is given by $Z^{(1)} = f\big((A \odot T^{(1)}) X W^{(1)}\big)$, where X is the node feature matrix, $W^{(1)}$ is a trainable weight matrix, $T^{(1)}$ contains the attention coefficients, and f is a non-linear activation function.
GNNs rely on a combination of node features and graph structure to learn meaningful representations. If X is uniform, the initial embeddings for all nodes will be identical, and the message-passing process will fail

to produce sufficiently diverse node embeddings to differentiate between classes. In this scenario, the GNN relies entirely on the graph structure, represented by the adjacency matrix A, to compute node embeddings. However, the adjacency matrix alone does not encode class labels or the relationship between nodes and their respective classes. Without distinct features, the model cannot leverage node-level information.
The GNN model we implemented employs a graph attention mechanism, which computes attention coefficients α_{ij} for edges between nodes v_i and v_j. These coefficients are determined based on the feature vectors of the connected nodes. If all node features are identical, the attention mechanism cannot differentiate between neighbors, resulting in uniform attention scores:

$$\alpha_{ij} = \alpha_{ik}, \quad \forall j, k \in \mathcal{N}(v_i).$$

This uniformity eliminates the ability of the attention mechanism to selectively aggregate meaningful information from neighbors, significantly reducing the model's expressiveness.
Thus, the GNN's performance will be severely degraded in this scenario because the model cannot distinguish nodes based on their features.

3 Question 3

The rows of the given matrix Z correspond to nodes of three graphs. Specifically:

- Rows $1, 2, 3$ belong to graph G_1,
- Rows $4, 5, 6, 7$ belong to graph G_2,
- Rows $8, 9$ belong to graph G_3.

We compute the representations $z_{G_1}, z_{G_2}, z_{G_3}$ for each graph using the following readout functions:

(i) Sum

The sum of node features for each graph is:

$$z_{G_1} = \sum_{i=1}^{3} Z[i,:] = \begin{bmatrix} 2.2 + 0.2 + 0.5, & -0.6 + 1.8 + 1.1, & 1.4 + 1.5 - 1.0 \end{bmatrix} = \begin{bmatrix} 2.9, & 2.3, & 1.9 \end{bmatrix}$$

$$z_{G_2} = \sum_{i=4}^{7} Z[i,:] = \begin{bmatrix} 3.4, & 1.9, & 4.3 \end{bmatrix}, \quad z_{G_3} = \sum_{i=8}^{9} Z[i,:] = \begin{bmatrix} 1.8, & 1.2, & 1.6 \end{bmatrix}$$

(ii) Mean

The mean of node features for each graph is:

$$z_{G_1} = \frac{1}{3} \sum_{i=1}^{3} Z[i,:] = \begin{bmatrix} \frac{2.9}{3}, & \frac{2.3}{3}, & \frac{1.9}{3} \end{bmatrix} = \begin{bmatrix} 0.97, & 0.77, & 0.63 \end{bmatrix}$$

$$z_{G_2} = \frac{1}{4} \sum_{i=4}^{7} Z[i,:] = \begin{bmatrix} 0.85, & 0.48, & 1.08 \end{bmatrix}, \quad z_{G_3} = \frac{1}{2} \sum_{i=8}^{9} Z[i,:] = \begin{bmatrix} 0.9, & 0.6, & 0.8 \end{bmatrix}$$

(iii) Max

The maximum of node features for each graph is:

$$z_{G_1} = \max_{i=1}^{3} Z[i,:] = \begin{bmatrix} \max(2.2, 0.2, 0.5), & \max(-0.6, 1.8, 1.1), & \max(1.4, 1.5, -1.0) \end{bmatrix} = \begin{bmatrix} 2.2, & 1.8, & 1.5 \end{bmatrix}$$

$$z_{G_2} = \max_{i=4}^{7} Z[i,:] = \begin{bmatrix} 2.2, & 1.8, & 1.5 \end{bmatrix}, \quad z_{G_3} = \max_{i=8}^{9} Z[i,:] = \begin{bmatrix} 2.2, & 1.8, & 1.5 \end{bmatrix}$$

The **Sum** and **Mean** readout functions are the best choices for distinguishing these graphs because they provide unique representations for each graph (G_1, G_2, G_3). However, the **Max** function produces the same representation for all graphs, failing to distinguish between them.
Among these, the **Sum** function is particularly effective because it preserves more variance across the graph representations, making it more suitable for tasks where distinguishing between graphs is critical.

4 Question 4

We have two graphs, G_1 and G_2, where G_1 is a cycle with 4 nodes (C_4), and G_2 is a cycle with 8 nodes (C_8). The augmented adjacency matrices for these graphs are $\tilde{A}_{G_1}$ and $\tilde{A}_{G_2}$, where $\tilde{A} = A + I$, given as follows:

$$\tilde{A}_{G_1} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}, \quad \tilde{A}_{G_2} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

The initial node features X are defined as $X = \mathbf{1}_{n_{\text{nodes}} \times 1}$, where n_{nodes} is the number of nodes in the graph.

First Message Passing Layer

The output of the first layer is given by $Z^{(1)} = \text{ReLU}(\tilde{A} X W^{(1)})$, where $W^{(1)}$ is a learnable matrix of size $d_{\text{input}} \times d_{W^{(1)}}$.
1. For G_1: Each node in the cycle C_4 is equivalently connected. Since the initial features are identical, the contribution of each node to $Z^{(1)}_{G_1}$ will be the same. As a result, all rows of $Z^{(1)}_{G_1}$ are identical.
2. For G_2: The structure is an extension of G_1 to C_8. Each node in C_8 plays the same role as those in C_4, leading to $Z^{(1)}_{G_2}$ containing blocks similar to $Z^{(1)}_{G_1}$, repeated due to the doubled number of nodes.

Second Message Passing Layer

The output of the second layer is given by: $Z^{(2)} = \text{ReLU}(\tilde{A} Z^{(1)} W^{(2)})$, where $W^{(2)}$ is a learnable matrix of size $d_{W^{(1)}} \times d_{W^{(2)}}$.
- For G_2, the propagation through $\tilde{A}_{G_2}$ replicates the behavior of G_1, but the contribution accumulates across the duplicated structure. This results in $Z^{(2)}_{G_2}$ being essentially a doubled version of $Z^{(2)}_{G_1}$.

Readout Function

The final representation for a graph G is given by $z_G = \sum_{i=1}^{n_{\text{nodes}}} Z^{(2)}_i$, where $Z^{(2)}_i$ is the i-th row of $Z^{(2)}$.
- For G_1: Summing the rows of $Z^{(2)}_{G_1}$ produces a vector z_{G_1} of dimension $1 \times d_{W^{(2)}}$.
- For G_2: Due to the extended structure of G_2, every row of $Z^{(2)}_{G_2}$ contributes identically, and the overall sum is doubled compared to G_1. Thus, $z_{G_2} = 2 \times z_{G_1}$.