Raphael Razafindralambo
raphael.razafin@gmail.com

Lab session # 7
ALTEGRAD 2023

12/05/23

1 Question 1

Assume a new training example x_t and the current hidden state h_t. An LSTM cell contains the following layers:

- Forget gate layer $f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$
- Input gate layer $i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$
- Candidate values computation layer: $\tilde{c}_t = \sigma(W_c x_t + U_c h_{t-1} + b_c)$
- output gate layer: $o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$

where σ is the sigmoid function. The equations for the cell state, candidate cell state and the final output are

- $c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$
- $h_t = tanh(c_t) \odot o_t$
- $y_t = \text{softmax}(W h_t)$

where $\odot$ denotes the Hadamard product. It's clear that the LSTM operator is not permutation invariant softmax is not permutation invariant. Thus I would not recommend to use LSTM on sets.

2 Question 2

The differences are the following:

- DeepSets takes a set of elements as an input, whereas GNNs consider graphs which are sets of elements where some are linked to each other.
- GNNs can incorporate relational information between elements through message passing (neighborhood). DeepSets, on the other hand, treats each element independently before aggregation.
- DeepSets is a permutation invariant structure, there is no notion of order.

3 Question 3

1) We consider a stochastic block model with n nodes and $r = 2$ blocks.
In a homophilic cluster structure, there are many edges within the same community and fewer between different communities. With two communities, the matrix $P_{homo} \in \mathcal{M}_{2\times 2}$ can be defined as

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \tag{1}$$

where $\forall j \neq k \; p_{ii} \gg p_{jk}$. For exemple,

$$P_{homo} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \tag{2}$$

In the contrary, in a heterophilic structure, most edges are between different communities, and fewer are within the same community, we can define for example

$$P_{hetero} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \tag{3}$$

which satisfies $\forall j \neq k \; p_{ii} \ll p_{jk}$.
2) We consider $n = 20, r = 4$, and

$$\mathbf{P}_{ij} = \begin{cases} 0.8 & \text{if } i = j \\ 0.05 & \text{otherwise} \end{cases}.$$

Let Z be a random variable modeling the number of edges between nodes of different blocks of a stochastic block model.

- Each block has 5 nodes. The total number N of possible pairs between two nodes from different blocks is given by the number a combinations of two nodes chosen from $n = 20$, minus the combinations within the same blocks. So $N = \binom{n}{2} - r \times \binom{5}{2} = \binom{20}{2} - 4 \times \binom{5}{2} = 190 - 4 \times 10 = 150$.
- The probability of an edge between two nodes from different blocks is $p = 0.05$.

Thus, the number of edges between two nodes of different blocks of a stochastic block model satisfies $Z \sim \mathcal{B}(N, p)$. Indeed, Z is the outcome of a sequence of N independent Bernoulli steps of parameter p in which we consider a pair of nodes of different blocks and decide if we link them or not. Consequently the expected degree of a node is $\mathbb{E}[N] = Np = 150 \times 0.05 = 7.5$.

4 Question 4

We can consider the Frobenius norm given for all matrix $A \in \mathcal{M}_{n\times n}$ by:

$$||A||_F = \sqrt{Tr(AA^*)} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}.$$

Hence we can define the following loss:

$$\mathcal{L} = ||A - \hat{A}||_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij} - \hat{a}_{ij}|^2}.$$

This approach is suitable for scenarios where these weights can take on a continuous range of values.

References