SIMON Hugo
hugo.simon@telecom-paris.fr

Lab session # 7
ALTEGRAD 2021

08/02/2022

1 Question 1

They are several ways to achieve a almost perfect summation with our given architecture.

- One simple way to do this is to take **Embedding** $= id$, $W_1 = \varepsilon$, $W_2 = \frac{1}{\varepsilon}$ and $b_1 = b_2 = 0$.
  Then for ε sufficiently small
  $$f(X) = \frac{1}{\varepsilon} \sum_{i=1}^{m} \tanh(\varepsilon x_i) \sim \frac{1}{\varepsilon} \sum_{i=1}^{m} \varepsilon x_i = \sum_{i=1}^{m} x_i$$
- Another way is to take **Embedding**$[x_i] = (\delta_{j \leq x_i})_{1 \leq j \leq 10}$ such that each number from 1 to 10 is embedded by a number of ones equal to its value. Then take $W_1 = \eta I_{10}$, $W_2 = \mathbb{1}^\top$ and $b_1 = b_2 = 0$.
  Then for η sufficiently large
  $$f(X) = \sum_{j=1}^{10} \sum_{i=1}^{m} \tanh(\eta \delta_{j \leq x_i}) \sim \sum_{i,j=1}^{m,10} \delta_{j \leq x_i} = \sum_{i=1}^{m} x_i$$

2 Question 2

The difference between DeepSets and Graph Neural Network (GNN) for Graph-Level Tasks is tight. If a set is interpreted in a GNN as a graph without edges, hence its adjacency matrix is $A = 0$, thus its message passing matrix is $\tilde{A} = I$. Therefore, message passing layers are not passing any messages across nodes but just learn parameters for each one i.e. consists in an embedding as in DeepSets.

3 Question 3

Nodes indices have no intrinsic information (they could have any distinct values we want) so it is not relevant to train a network as if they had. Instead, we learn the representation of the nodes, embedded in a vectorial space more meaningful for classical network architectures.

4 Question 4

Using DeepSets for predicting the next item of a session is a bad idea as we would lose crucial structural information which is ordering and repetitions of elements.

References