Denis Duval
denis.duval13@gmail.com

Lab session # 7
ALTEGRAD 2022

12/21/22

1 Question 1

Let X be the embedded features (the elements of the set after they have been passed to the embedding layer). Let W_1, W_2 and b_1, b_2 the weights and biases of the two fully connected layers. The output y of the model writes:

$$y = \tanh(XW_1 + b_1) \cdot 1 \cdot W_2 + b_2$$

where 1 is a vector full of ones. Let's take zero biases. Because here the number of different digits is 11 which is below the embedding dimension ($= 128$) we could think that the embedding layer will do a 1-hot encoding of the digits. Then W_1 could be a matrix with an identity block of size $h_1 \cdot h_1$ top left. Then the output before the fully connected layer is a vector which has an element at i equals to $\tanh(1)$ times the number of times the digit i is present in the set. That being said, W_2 could be just a vector full of $\frac{1}{\tanh(1)}$. With these parameters, the model exactly performs the task we wanted. Actually, it seems that there are multiple optimal parameters.

2 Question 2

For the ϕ function we can simply take the identity matrix as the weight matrix(the hidden dim is then equal to the input dim), take no biases and take a ReLU as a nonlinearity. For the ρ function let's simply take the vector $[1, 1]$ and no biases. The sets will be first transformed into $[1.2 \quad 0], [0 \quad 0.5]$ and $[0.2 \quad 0], [0.2 \quad 0.1]$ through ϕ. Then with the sum and the ρ function there are transformed into 1.7 and 0.5.

3 Question 3

We can use the DeepSets model with slight modifications. First, as the elements of our sets are graphs we need to convert them to vectors. One idea is to use graph embeddings for example with graph2vec. Another idea would be to use a GNN instead of the first MLP. Once the graphs have been vectorized, the rest of the architecture could remain the same. Eventually we should add a classification head on this model if the task is classification. It could be a Bayes Classifier if we have enough information on what type of classification we want to achieve or even other algorithms such as K-means.

4 Question 4

The message passing layers are permutation equivariant : in fact it is easy to see that a permutation of the lines of a matrix M will give the same permutation of the lines of the matrix MW. Then the readout function is the sum function so it's permutation invariant. So our whole model is permutation invariant which means that it has no sensitivity to the ordering of the amino acids though this information is important for the protein structure. One generic way to solve this problem is to add some positional encoding at the beginning of the model. We could also think to change the readout function for example the sum could be changed to a weighted sum with the weights being trainable or not.