ECE 684: Natural Language Processing
Seq2Seq Assignment

Mehak Arora

November 5, 2024

1 Question 1

Let $x_t = [x_a, x_b, x_c, x_d, x_e, x_s]$ denote the one-hot encoded input where each dimension corresponds to indicating the occurrence of the letters "a","b","c","d","e" or the stop token (in x_s) at the input at the current time step t.

All our information from the input has to be compressed and sent to the output through the hidden state vector. Therefore, let the hidden state be $h = [n_a, n_b, n_c, n_d, n_e, n_{tt}]$ where n_a, n_b, n_c, n_d, n_e represents the number of total counts of the letters "a","b","c","d","e" respectively, and n_{tt} is a dimension that is included to count the number of output tokens produced (since the dimension of the encoder's hidden state has to be equal to the dimensions of the decoder's hidden state).

Starting with the encoder's weights:

$$h_{t+1} = W_e \begin{bmatrix} x \\ h \end{bmatrix}$$

$$\begin{bmatrix} n_a^{t+1} \\ n_b^{t+1} \\ n_c^{t+1} \\ n_d^{t+1} \\ n_e^{t+1} \\ n_{tt}^{t+1} \end{bmatrix} = W_e \begin{bmatrix} x_a \\ x_b \\ x_c \\ x_d \\ x_e \\ x_s \\ n_a \\ n_b \\ n_c \\ n_d \\ n_e \\ n_{tt} \end{bmatrix} \tag{1}$$

Setting: W_e as

$$W_e = \begin{bmatrix} 100000100000 \\ 010000010000 \\ 001000001000 \\ 000100000100 \\ 000010000010 \\ 111111000000 \end{bmatrix}$$

$$\begin{bmatrix} n_a^{t+1} \\ n_b^{t+1} \\ n_c^{t+1} \\ n_d^{t+1} \\ n_e^{t+1} \\ n_{tt}^{t+1} \end{bmatrix} = \begin{bmatrix} x_a + n_a \\ x_b + n_b \\ x_c + n_c \\ x_d + n_d \\ x_e + n_e \\ 1 \end{bmatrix}$$

To set the decoder weights:

$$o_t = ReLU(W_o h_t^{'})$$
$$h_{t+1}^{'} = W_h h_t^{'} \tag{2}$$

We can see that the values of elements h^{t+1} are non-negative. If we set the values of W_o also as non-negative numbers, the relu non-linearity is irrelevant can be ignored (since it just reduces to an identity function $x = x$.

$$o_t = W_o h_t^{'} \tag{3}$$

Let's set w_o as

$$\begin{bmatrix} 100000 \\ -1-1-1-1-11 \end{bmatrix} \tag{4}$$

and set w_h as

$$\begin{bmatrix} 010000 \\ 001000 \\ 000100 \\ 000010 \\ 000000 \\ 000001 \end{bmatrix} \tag{5}$$

This will output the first element of $h_t^{'}$ as the first element of the output, and will output a 0 as the second element long as any of the first five elements of $h_t^{'}$ are greater than 0 (because of the relu). The w_h matrix will circular shift the hidden state to the left with every decoder step, keeping the last element as 1. Once all the 5 elements are popped at the output, the second element of the output will be 1.