HW8. Seq2Seq

Size of Input, Hidden State, and Output Vectors

Component	Description	Size
Input Vector	One-hot encoding of tokens `"a"`, `"b"`, etc.	**6**
Hidden State	Counts of tokens `"a"` to `"e"`	**5**
Output Vector	Count and end-of-sequence indicator	**2**

Final Weights

1. Encoding Weight Matrix W_e (size 5×11)

$W_e = [A \mid B]$, where:

- A (size 5×6):

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

- B (size 5×5):

$$B = I_5$$

2. Output Weight Matrix W_o and Bias b_o (sizes 2×5 and 2×1)

$$W_o = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}, \quad b_o = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

3. Hidden State Transition Matrix W_h (size 5×5)

$$W_h = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Encoding Phase: Counts the occurrences of each token "a" to "e" in the input sequence.

Decoding Phase: Outputs the counts in order and signals the end when all counts have been outputted.

Model Flow:

- **Input**: Sequence of tokens ending with ".".
- **Hidden State**: Accumulates counts during encoding.
- **Output**: Sequence of counts followed by an end-of-sequence indicator.

Explanation

Tokens: "a", "b", "c", "d", "e", "." (end-of-sequence token)

Input Encoding:

- **Input vector** x_t: One-hot vector of size **6** (representing each token).
- **Hidden state** s_t: Vector of size **5** (counts of "a" to "e").
- **Initial hidden state**: $s_0 = [0, 0, 0, 0, 0]^T$.

Encoding Function:

$$s_{t+1} = W_e \begin{bmatrix} x_t \\ s_t \end{bmatrix}$$

- **Purpose**: Update the hidden state s_t by incrementing the count corresponding to the input token and preserving existing counts.

How W_e Works:

- **For tokens "a" to "e"**:
  - The corresponding row in A adds 1 to the count in s_t.
- **For the EOS token "."**:
  - The last column in A is zeros, so counts remain unchanged.
- $B = I_5$ ensures the previous counts in s_t are carried over.

Decoding Functions:

$$\text{output}_v = \text{ReLU}(W_o s'_v + b_o)$$

$$s'_{v+1} = W_h s'_v$$

- **Purpose**: Output the counts in order and signal the end of the sequence.

How W_o and W_h Work:

- W_o extracts the count of the current token and determines if the sequence has ended.
  - **First row**: Outputs the current count $s'^{(1)}_v$.
  - **Second row**: Computes $-\sum_{i=1}^{5} s'^{(i)}_v + 1$, which is positive only when all counts are zero.
- **Bias b_o** adjusts the end-of-sequence indicator to be 1 when the sequence ends.
- W_h shifts the hidden state to the left, preparing s'_{v+1} for the next token's count.

Decoding Process Steps

1. **Initialization**:
   - Set $s'_0 = s_T$ (the final hidden state from the encoding phase).
2. **At each decoding step v**:
   - **Compute Output**:
     - $\text{output}_v^{(1)} = \text{ReLU}(s'^{(1)}_v)$ (current token count).
     - $\text{output}_v^{(2)} = \text{ReLU}(-\sum_{i=1}^{5} s'^{(i)}_v + 1)$ (end-of-sequence indicator).
   - **Update Hidden State**:
   - **Update Hidden State**:
     - $s'_{v+1} = W_h s'_v$ (shift counts for the next token).