Assignment: Handwritten LSTM for Next Word Generation

February 25, 2025

Instructions

This assignment requires you to manually compute each step in using an LSTM for next-word prediction. You must show all matrix multiplications with explicit calculations and dimensions.

- Handwritten Notes: All work must be handwritten and submitted as scanned PDF.
- Follow Class Notes: You may refer to the lecture notes for guidance.
- Explicit Computation: No assumptions—every step must be written explicitly.
- Dimensions Required: Each matrix operation should have clear dimensions mentioned.

Problem Statement

You are given an LSTM-based next-word generator with the following settings:

- Vocabulary size: $V = 5$
- Hidden state size: $h = 2$
- Token embedding size: $e = 2$
- Context size: $c = 3$

Your goal is to take 3 token IDs as context and pass them through:

1. A token embedding layer.
2. An LSTM layer.
3. A logits layer to generate the next token.

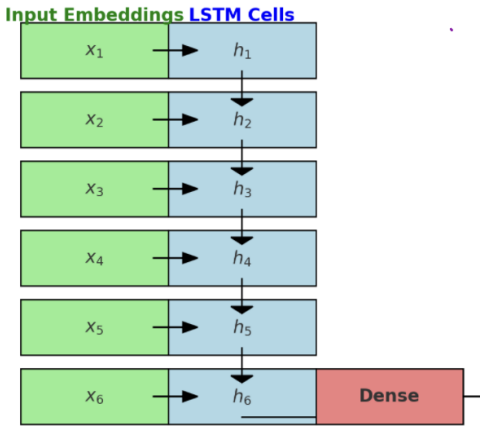

Figure 1: Your LSTM schematic might look something like this. However, since context size = 3, you will just have x1, x2, x3. The dimensions of h1, h2 and h3 will be 2. Dense means the final logits layer. You can assume vocab size = 5.

Steps to Follow

1. Token Embedding: Convert token IDs to embeddings. Use embedding dimension = 2. In class, we used an embedding dimension = 1.

2. LSTM Computations:
- Define LSTM parameters: W_f, W_i, W_o, W_c (weights) and b_f, b_i, b_o, b_c (biases).
- Compute gate activations:

$$
\begin{aligned}
f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\
i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\
o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\
\tilde{C}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) \\
C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\
h_t &= o_t \odot \tanh(C_t)
\end{aligned}
$$

- Compute all the above steps for each token in the context.

3. Logits Layer: Compute the output logits using the final hidden state h_T:

$$y = W_{\text{out}} h_T + b_{\text{out}}$$

4. Final Token Prediction: - Apply softmax to y to obtain probabilities. - Choose the token with the highest probability.

Submission Guidelines

- Submit a video recording of yourself explaining your notes to an undergraduate student. You can think of this as a class and you are teaching this class. You can make the video as long as possible.

- Submit a scanned PDF of your handwritten solutions.
- You can submit the video and PDF as a Drive link.
- Clearly label each step and ensure all matrices and dimensions are included.
- Your calculations must be legible and well-organized.

Team Size

You can have a maximum team size of 2 students per team. If you are not forming a team, you can submit the assignment individually.

Deadline

7 March 2024

- **Clarity and depth of video explanation** (50%)
- **Clarity and depth of hand written notes** (50%)

For any clarifications, please reach out before the deadline. Good luck!