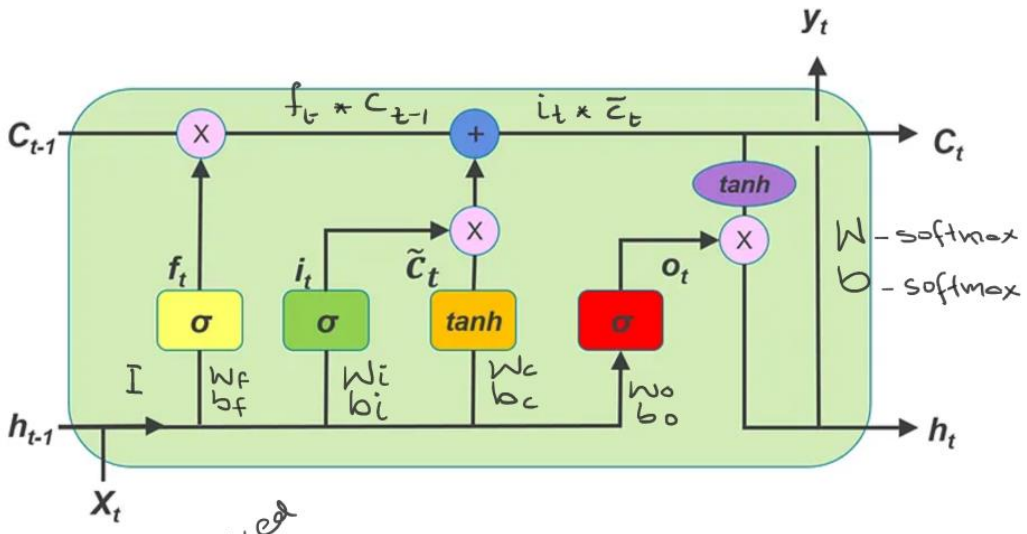
Written Assignment (LSTM)



- Reduced Equations for LSTM

$$I = [h_{t-1}, x_t]$$

$$f_t = sigmoid\ (W_f * I + B_f)$$

$$i_t = sigmoid\ (W_i * I + B_i)$$

$$o_t = sigmoid\ (W_o * I + B_o)$$

$$C_t^{\sim} = \tanh(W_c * I + b_c)$$

$$C_t = f_t * C_{t-1} + i_t * C_t^{\sim}$$

$$h_t = o_t * C_t$$

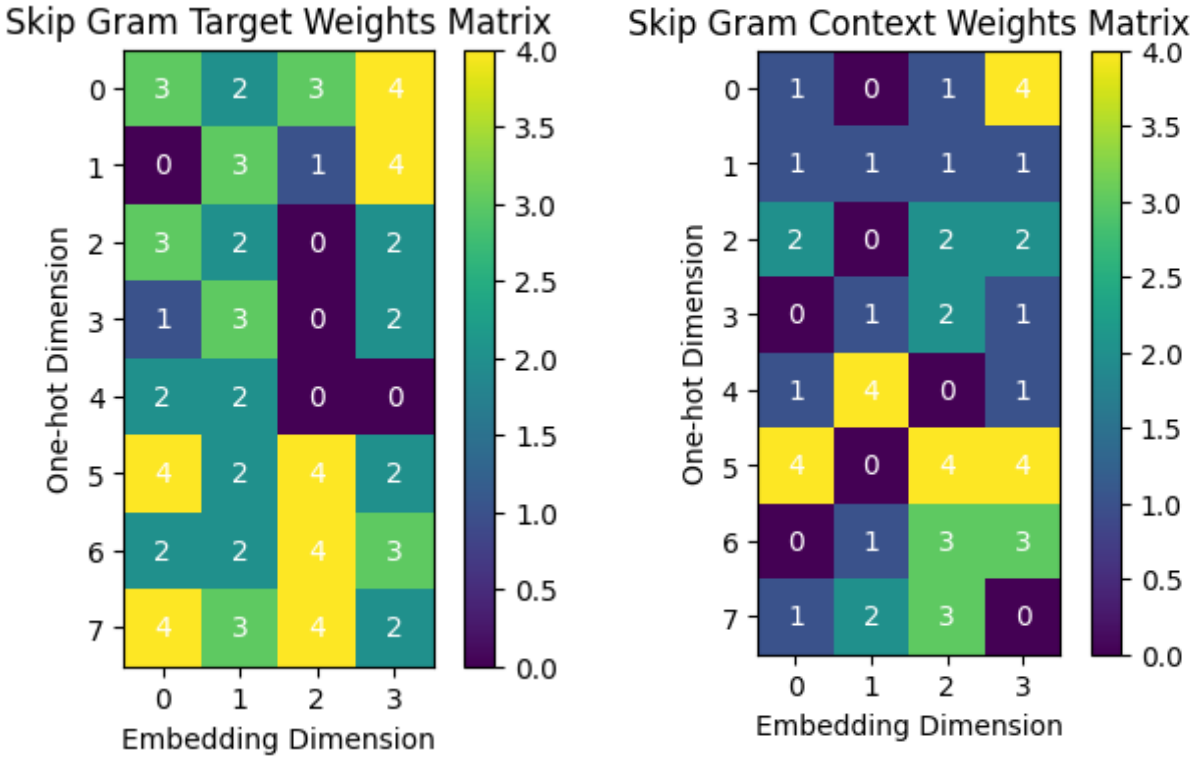
Vocabulary = {the, students, opened, their, laptops, books, exams, minds}

Note: One-hot Encoded vector for word 'the' will be given below you are required to compute for other words as well based on sequence of given vocabulary.

1	0	0	0	0	0	0	0

Task 1: Use the above calculated One-Hot Encoded Vector to calculate embeddings for the provided vocabulary using the Skip-gram model's Weight Matrix given below:

1. Embedding for the target words (Consider each word in vocabulary as target word).
2. Embedding for the surrounding context words (Consider each word in vocabulary as context word).
3. Average embedding for both target and context words.



Task 2: Utilize the three types of embeddings derived from the preceding task to complete the following sentence by predicting the subsequent word in the sequence:

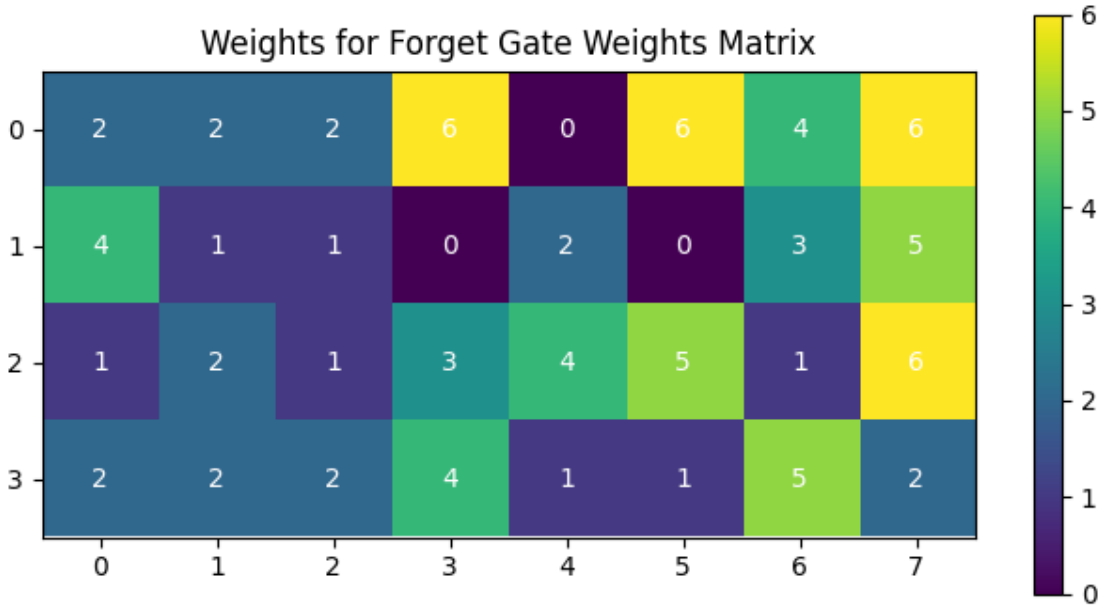
the students opened their _____?

Initially, employ target word embeddings to execute a forward pass for LSTM in order to complete the sentence. Subsequently, proceed with all steps for context embedding of each word and then for average embeddings.

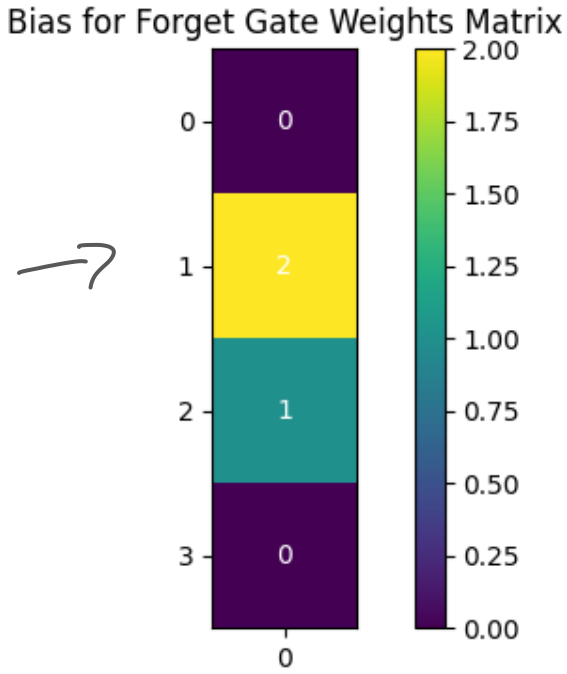
Utilize the provided weight and bias matrices to calculate the previous cell state (c_t) and hidden state (h_t) for each time step. Apply softmax to the output layer in the final time step to predict probabilities for the most probable next word. Express the word with the highest probability as a one-hot encoded vector and interpret it from the given vocabulary to fill in the blank. Incorporate all types of embeddings in this process and verify whether using average embedding yields identical results to using target and context embeddings. Assume that the previous hidden state and cell state are both initialized to zero.

Weight and Bias Matrices:

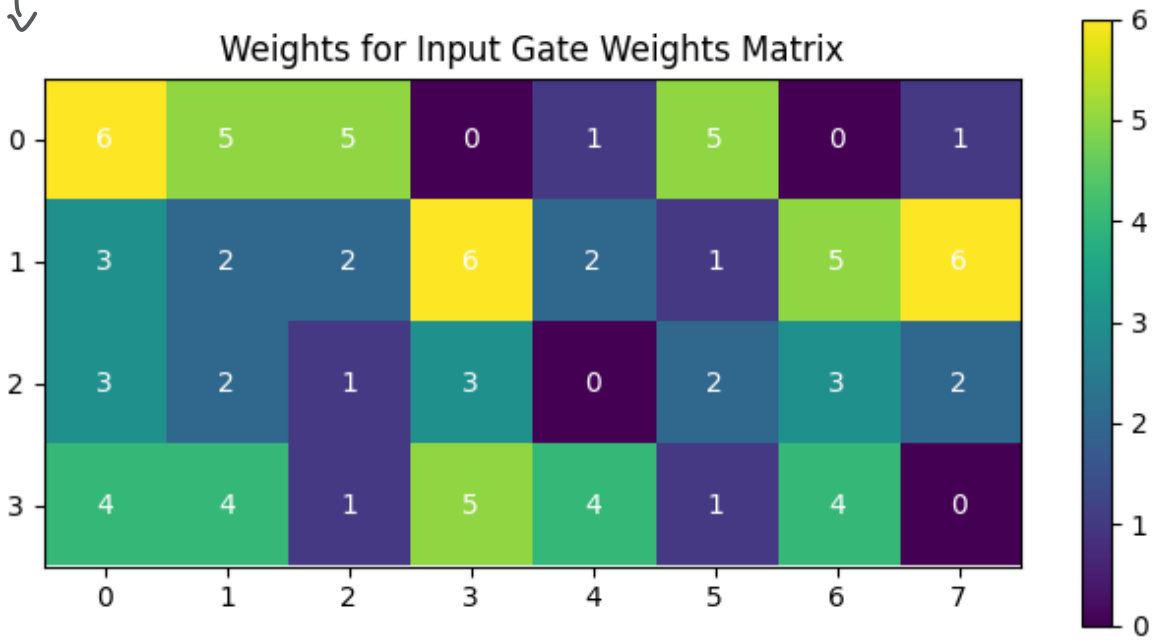
$W_f \rightarrow$



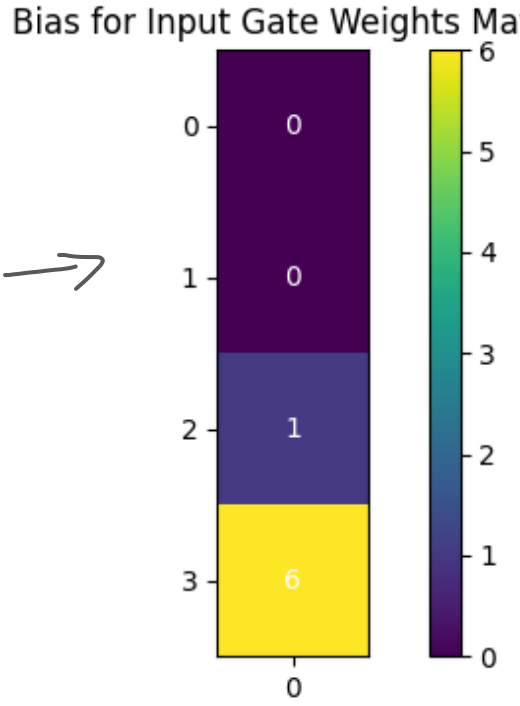
$b_f \rightarrow$



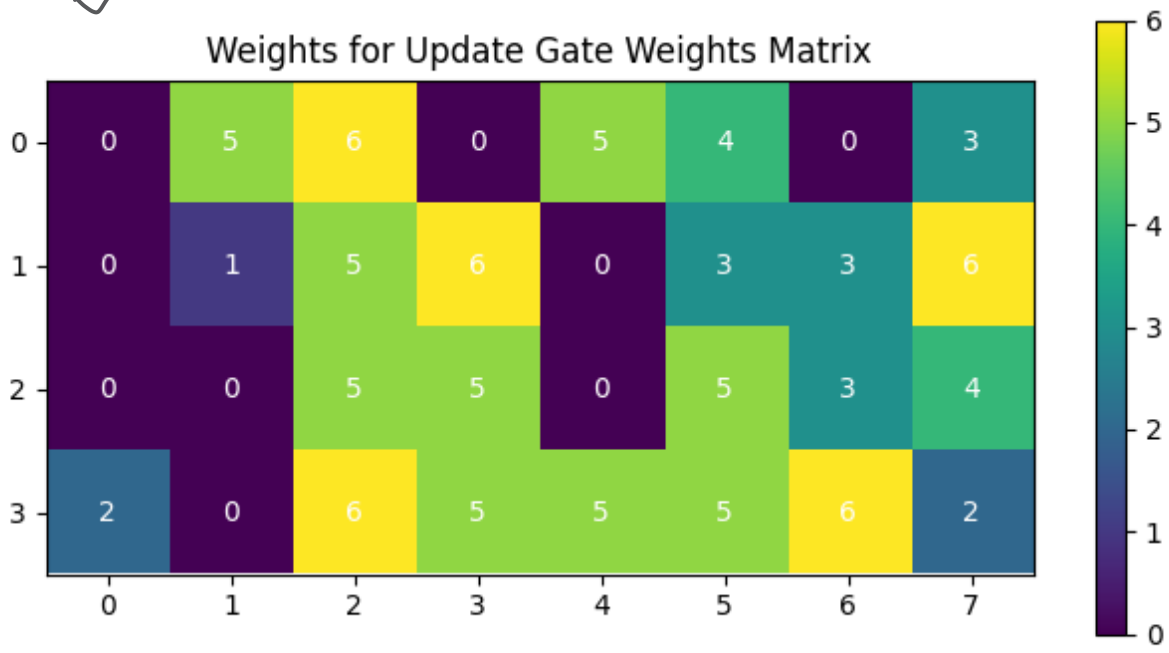
Wi

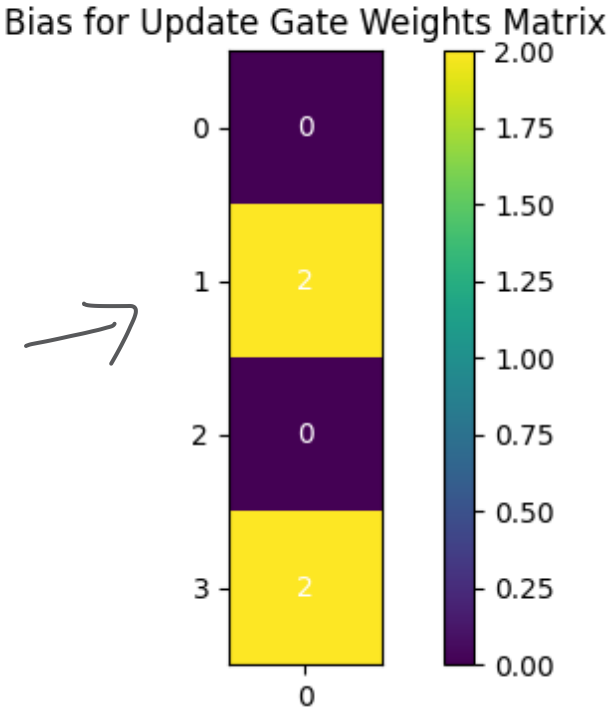

Weights for Input Gate Weights Matrix
0 6 5 5 0 1 5 0 1
1 3 2 2 6 2 1 5 6
2 3 2 1 3 0 2 3 2
3 4 4 1 5 4 1 4 0
0 1 2 3 4 5 6 7
6 5 4 3 2 1 0

bi →


Bias for Input Gate Weights Matrix
0 0
1 0
2 1
3 6
0
6 5 4 3 2 1 0

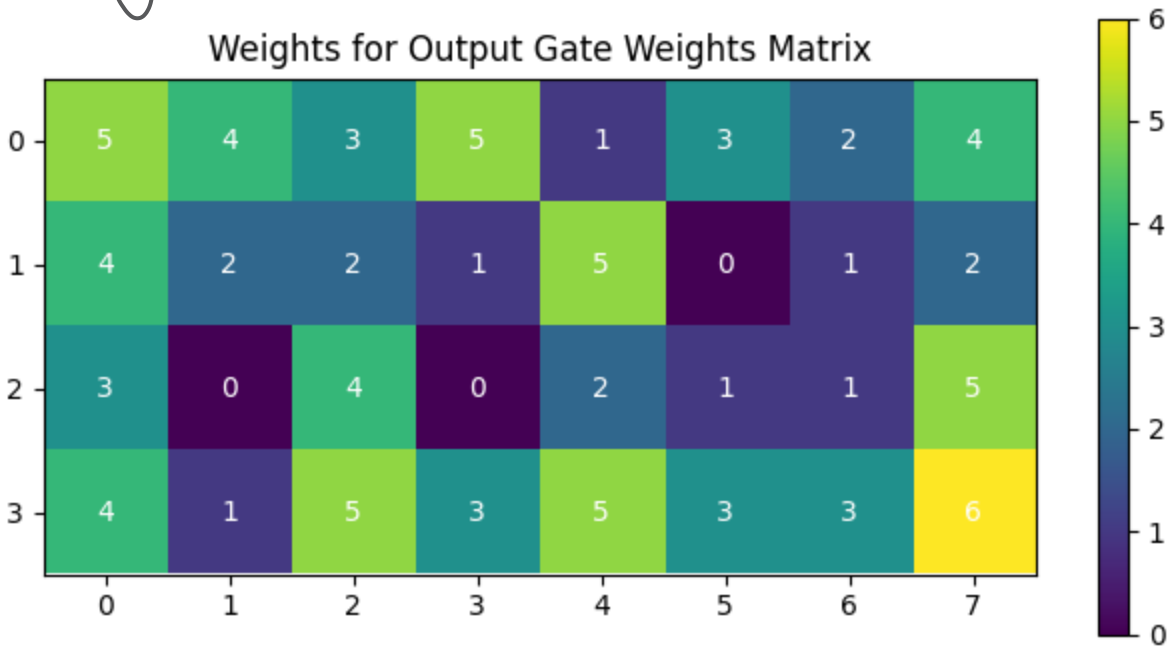
$W_c \rightarrow$

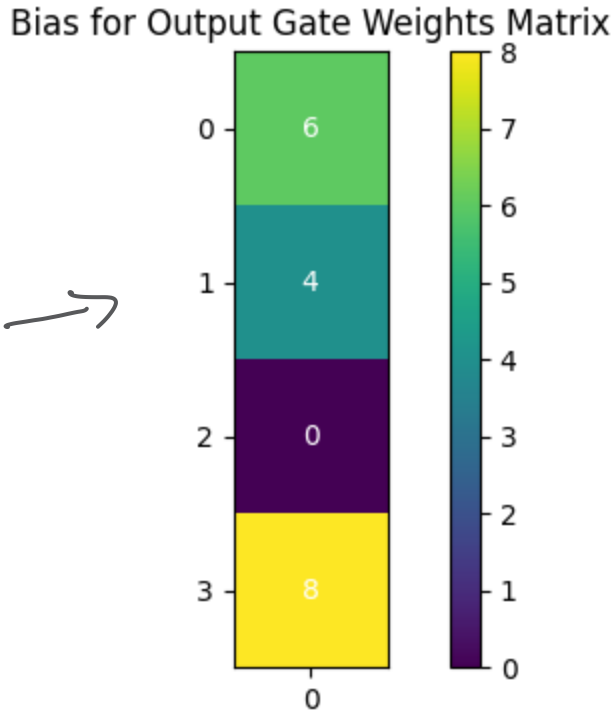




$b_c \rightarrow$

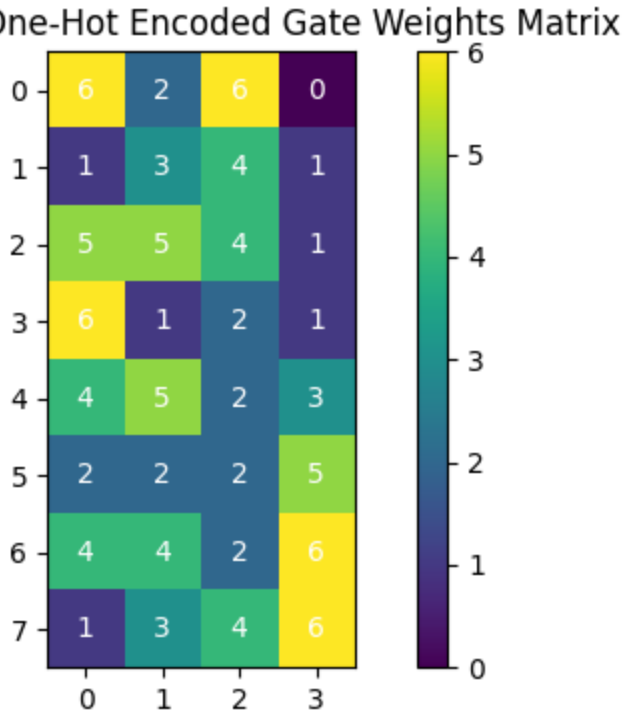
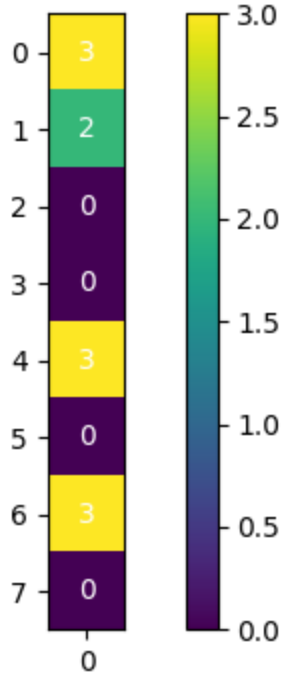
$W_o \rightarrow$





$b_o \rightarrow$

W_softmax →



b_softmax →