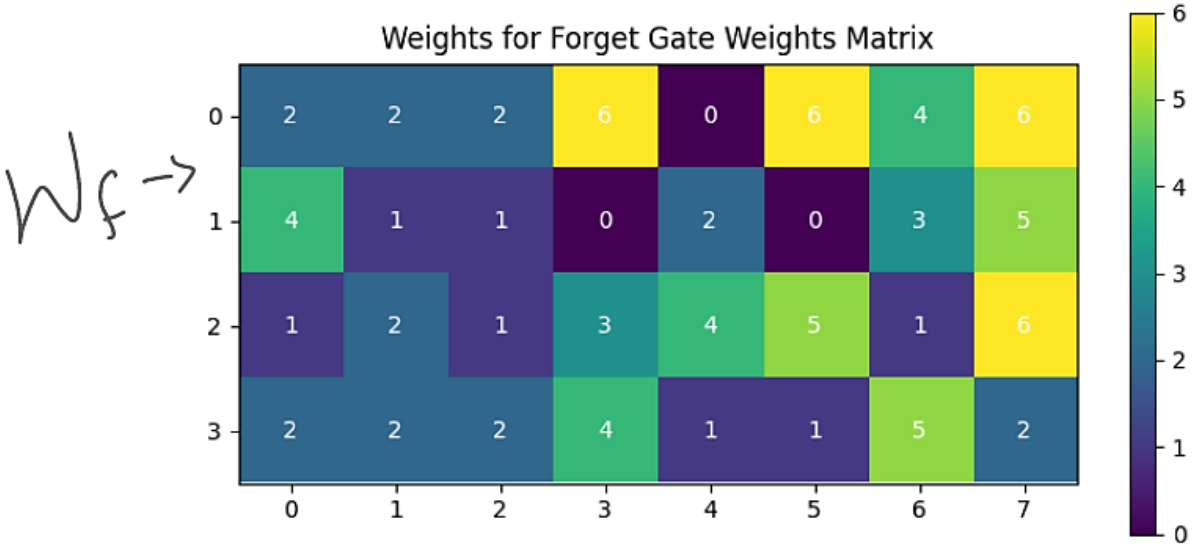
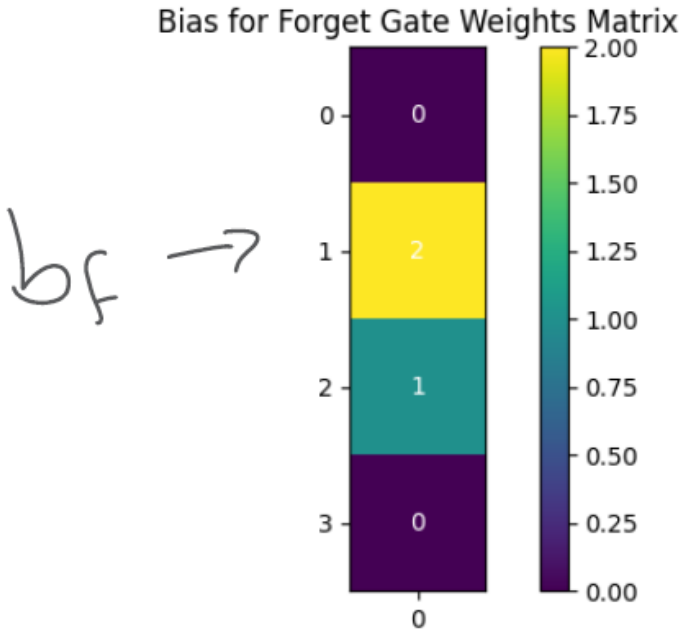
Utilize the given embeddings to complete the following sentence by predicting the subsequent word in the sequence:
the students opened their _____?

Employ given word embeddings to execute a forward pass for LSTM in order to complete the sentence.
Utilize the provided weight and bias matrices to calculate the previous cell state (ct) and hidden state (ht) for each time step.
Apply softmax to the output layer in the final time step to predict probabilities for the most probable next word. $W^{softmax} * h_t + B^{softmax}$
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.

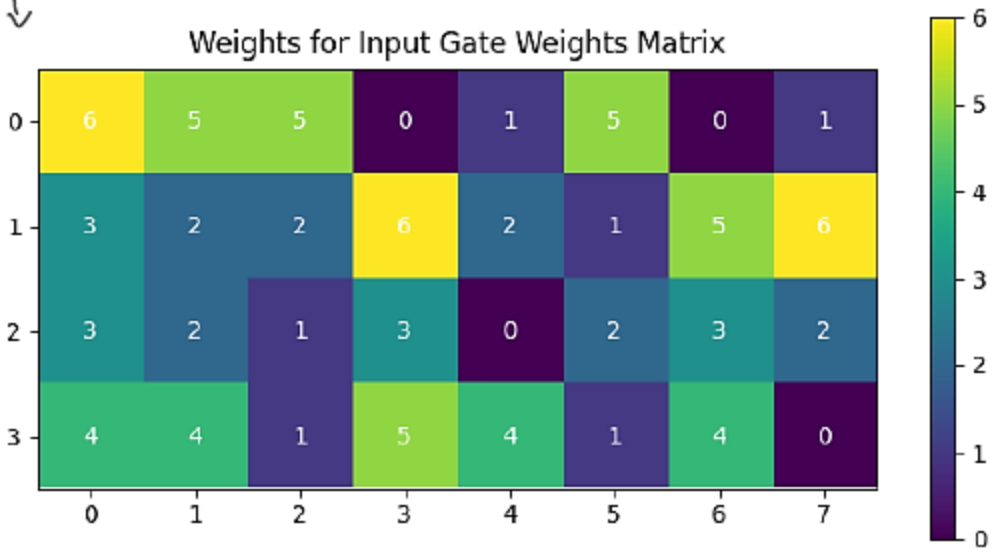
The	[3,2,3,4]
students	[0 3 1 4]
opened	[3 2 0 2]
their	[1 3 0 2]
laptops	[2 2 0 0]
books	[4 2 4 2]
exams	[2 2 4 3]
minds	[4 3 4 2]

Weight and Bias Matrices:

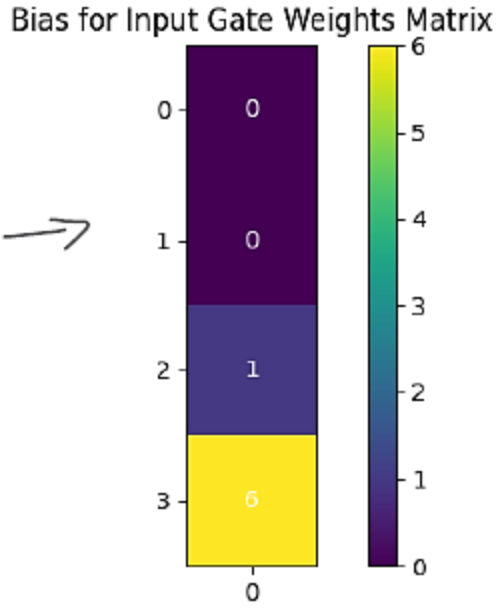



Bias for Forget Gate Weights Matrix
0
1
2
3
0
0
2
1
0
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00

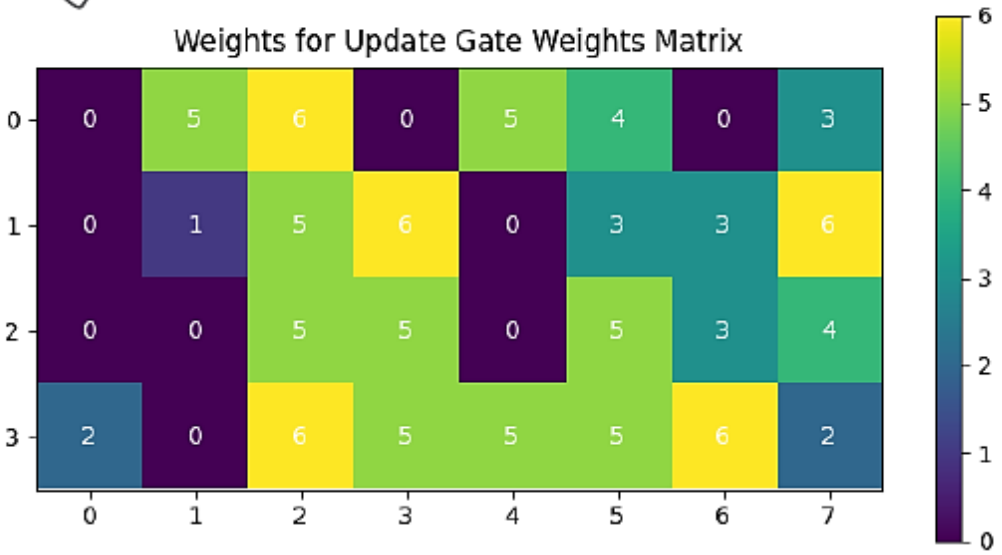
Wi

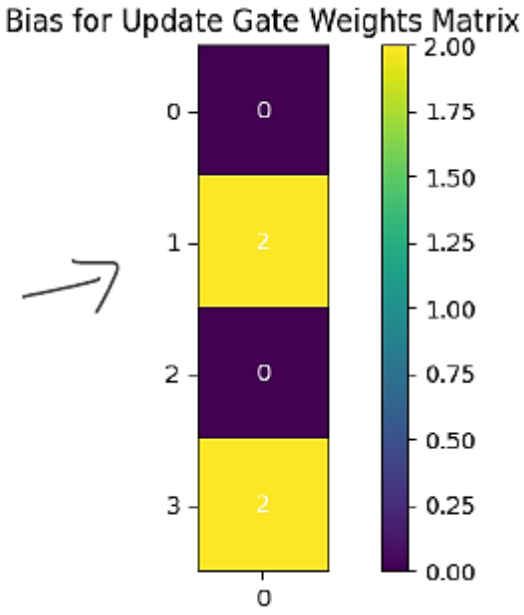

Weights for Input Gate Weights Matrix
0 6 5 5 0 1 5 0 1
1 3 2 2 6 2 1 5 6
2 3 2 1 3 0 2 3 2
3 4 4 1 5 4 1 4 0
0 1 2 3 4 5 6 7
6 5 4 3 2 1 0

bi →


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

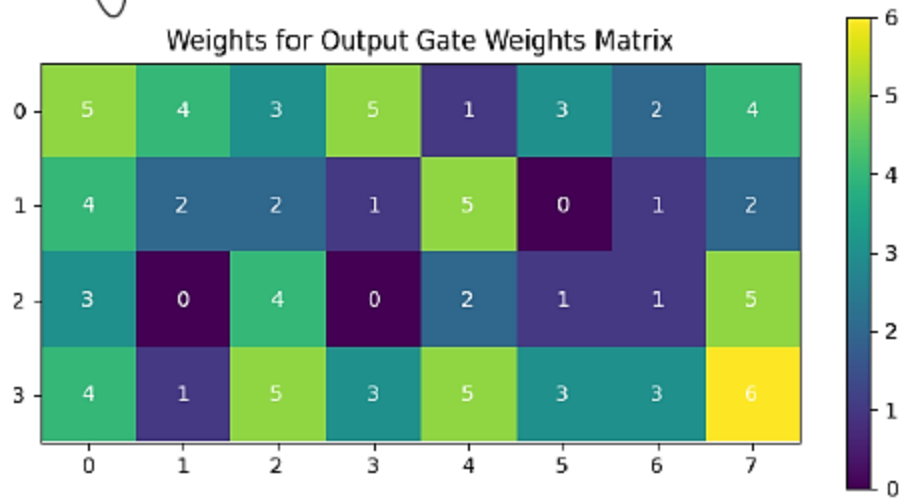
W_c →

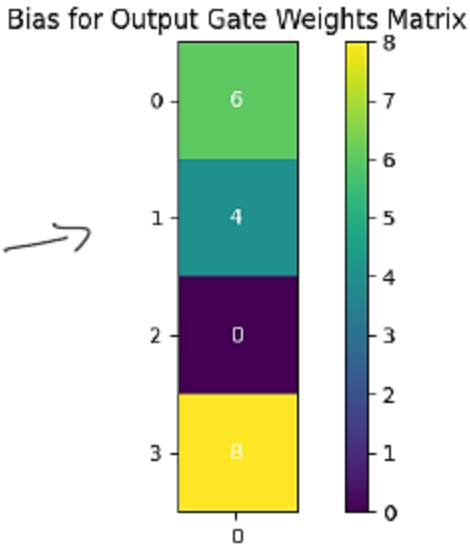

Weights for Update Gate Weights Matrix
0 5 6 0 5 4 0 3
0 1 5 6 0 3 3 6
0 0 5 5 0 5 3 4
2 0 6 5 5 5 6 2
0 1 2 3 4 5 6 7
6 5 4 3 2 1 0


Bias for Update Gate Weights Matrix
0
2
0
2
0 1 2 3
0
2.00 1.75 1.50 1.25 1.00 0.75 0.50 0.25 0.00

b_c →

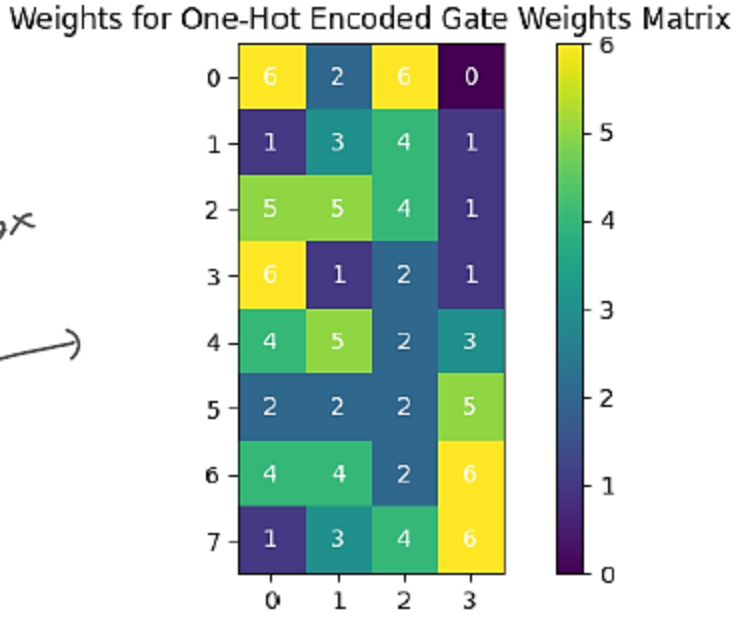
$W_o \rightarrow$



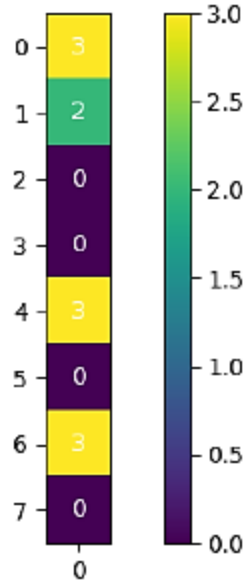


$b_o \rightarrow$

W_softmax →



b_softmax →



$C_t = [0, 0, 0, 0]$

$h_t = [0, 0, 0, 0]$

Target Embedding "The"

$I = [h_t, X_t]$

$= [0, 0, 0, 0, 3, 2, 3, 4] = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 3 \\ 4 \end{bmatrix}$

$$f_t = \sigma(W_f \times I * B_f)$$

$$f_t = \begin{bmatrix} 2, 2, 2, 6, 0, 6, 4, 6 \\ 4, 1, 1, 0, 2, 0, 3, 5 \\ 1, 2, 1, 3, 4, 5, 1, 6 \\ 2, 2, 2, 4, 1, 1, 5, 2 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

$4 \times 8 \qquad 8 \times 1 \qquad 4 \times 1$

$$\begin{bmatrix} 48 \\ 35 \\ 49 \\ 28 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 48 \\ 37 \\ 50 \\ 28 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0.99 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_t = \sigma(w_i \times I * B_i)$$

$$= \begin{bmatrix} 6, 5, 5, 0, 1, 5, 0, 1 \\ 3, 2, 2, 6, 2, 1, 5, 6 \\ 3, 2, 1, 3, 0, 2, 3, 2 \\ 4, 4, 1, 5, 4, 1, 4, 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 3 \\ 3 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix}$$

$$= \begin{bmatrix} 17 \\ 47 \\ 21 \\ 26 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix} = \sigma \begin{bmatrix} 17 \\ 47 \\ 22 \\ 32 \end{bmatrix} = \begin{bmatrix} 0.99 \\ 1 \\ 0.99 \\ 1 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$O_t = \sigma(W_o * I + B_a)$$

$$O_t = \begin{bmatrix} 31 \\ 26 \\ 31 \\ 54 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 8 \end{bmatrix} = \sigma \begin{bmatrix} 37 \\ 30 \\ 31 \\ 62 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_t = \tanh(W_c * I + B_c)$$

$$= \begin{bmatrix} 35 \\ 39 \\ 35 \\ 31 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix} = \tanh \begin{bmatrix} 35 \\ 41 \\ 35 \\ 53 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Update C_t:

$$C_t = f_t * C_{t-1} + i_t \times \tilde{C}_t$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Update h_t

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \tanh \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \end{bmatrix} = \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \end{bmatrix}$$

Now for "students"

$$C_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad h_t = \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \end{bmatrix} \quad I = \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ 0 \\ 3 \\ 1 \\ 4 \end{bmatrix}$$

$$f_t = \begin{bmatrix} 55.12 \\ 27.56 \\ 45.32 \\ 23.6 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} = 6 \begin{bmatrix} 55.12 \\ 29.56 \\ 46.32 \\ 23.6 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_t = \begin{bmatrix} 31.16 \\ 41.88 \\ 23.84 \\ 17.64 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix} = 6 \begin{bmatrix} 31.16 \\ 41.88 \\ 24.84 \\ 23.64 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$O_t = \begin{bmatrix} 59.92 \\ 15.84 \\ 29.32 \\ 45.88 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 8 \end{bmatrix} = 6 \begin{bmatrix} 45.92 \\ 19.84 \\ 29.32 \\ 53.88 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_t = \begin{bmatrix} 32.36 \\ 45.12 \\ 41.6 \\ 38.87 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix} = \tanh \begin{bmatrix} 32.36 \\ 47.12 \\ 41.6 \\ 40.87 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Update C_t

$$C_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

Update h_t

$$h_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * \tanh \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix}$$

Now for "Opened"

$$C_t = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} \quad h_t = \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix} \quad I = \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 3 \\ 2 \\ 0 \\ 2 \end{bmatrix}$$

$$f_t = \begin{bmatrix} 35.52 \\ 21.76 \\ 40.72 \\ 18.6 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} = \sigma \begin{bmatrix} 35.52 \\ 23.76 \\ 41.72 \\ 18.6 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_t = \begin{bmatrix} 30.36 \\ 32.48 \\ 16.64 \\ 27.44 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix} = \sigma \begin{bmatrix} 30.36 \\ 32.48 \\ 17.64 \\ 33.44 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$O_t = \begin{bmatrix} 33.32 \\ 27.64 \\ 24.72 \\ 45.48 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 8 \end{bmatrix} = \sigma \begin{bmatrix} 39.32 \\ 31.64 \\ 24.72 \\ 53.48 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_t = \begin{bmatrix} 39.56 \\ 29.52 \\ 27.6 \\ 41.48 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix} = \tanh \begin{bmatrix} 39.56 \\ 31.52 \\ 27.6 \\ 43.48 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Update C_t

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \times \begin{bmatrix}2\\2\\2\\2\end{bmatrix} + \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}3\\3\\3\\3\end{bmatrix}$$

Update h_t

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix} \times \tanh \begin{bmatrix}3\\3\\3\\3\end{bmatrix} = \begin{bmatrix}0.995\\0.995\\0.995\\0.995\end{bmatrix}$$

Now for "their"

$$C_t = \begin{bmatrix}3\\3\\3\\3\end{bmatrix} \quad h_t = \begin{bmatrix}0.995\\0.995\\0.995\\0.995\end{bmatrix} \quad I = \begin{bmatrix}0.995\\0.995\\0.995\\0.995\\-\\3\\8\\2\end{bmatrix}$$

$$f_t = \begin{bmatrix}41.94\\17.97\\37.97\\17.95\end{bmatrix} + \begin{bmatrix}0\\2\\1\\0\end{bmatrix} = 6 \begin{bmatrix}41.94\\19.97\\38.97\\17.95\end{bmatrix} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$i_t = \begin{bmatrix}33.92\\29.93\\18.95\\20.93\end{bmatrix} + \begin{bmatrix}0\\0\\1\\6\end{bmatrix} = 6 \begin{bmatrix}33.92\\29.93\\19.95\\26.93\end{bmatrix} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$O_t = \begin{bmatrix}34.915\\17.955\\21.965\\38.935\end{bmatrix} + \begin{bmatrix}6\\4\\0\\8\end{bmatrix} = 6 \begin{bmatrix}40.915\\21.955\\21.965\\46.935\end{bmatrix} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$\tilde{C}_t = \begin{bmatrix} 33.94 \\ 32.94 \\ 32.95 \\ 36.96 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix} = \tanh \begin{bmatrix} 33.94 \\ 34.94 \\ 32.95 \\ 38.96 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Update C_t

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix}$$

Update C_t

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix}$$

Update h_t

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \tanh \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 0.9993 \\ 0.9993 \\ 0.9993 \\ 0.9993 \end{bmatrix}$$

Now for Prediction:

W - softmax x h_t + b - softmax

$$\begin{bmatrix} 6,2,6,0 \\ 1,3,4,1 \\ 5,5,4,1 \\ 6,1,2,1 \\ 4,5,2,3 \\ 2,2,2,5 \\ 4,4,2,6 \\ 1,3,4,6 \end{bmatrix} \times \begin{bmatrix} 0.9993 & 0.9993 & 0.9993 & 0.9993 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \\ 3 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

(4×1) original

(8×4)

8×1

$$= \text{softmax}\left(\begin{bmatrix} 16.9902 \\ 10.9937 \\ 14.9895 \\ 9.993 \\ 16.9902 \\ 10.9923 \\ 18.988 \\ 13.990 \end{bmatrix}\right) = \begin{bmatrix} 0.104 \\ 0.0002 \\ 0.014 \\ 0.00009 \\ 0.104 \\ 0.00025 \\ 0.77 \\ 0.005 \end{bmatrix}$$

we can see from the vocabulary that 0.77 correlates with the word exams so,

the students opened their "exams"