Foundation for Advancement of Science & Technology

MAli Awan
21L-5648
NLP-Hand Assign.

Q#1:

vocab { the, students, opened, their, laptops, books, exams, minds }.

One hot encoded vector:

the : 10000000 students : 01000000 opened : 00100000
their : 00010000 laptops : 00001000 books : 00000100
minds : 00000010 exams : 00000001

(a) Embedding for target words

the : [3, 2, 3, 4] opened : [3, 2, 0, 2] laptop [2, 2, 0, 0]
student : [0, 3, 1, 4] their : [1, 3, 0, 2] books [4, 2, 4, 2]
exams : [2, 2, 4, 3] mind [4, 3, 4, 2]

(b) Embeddings for content word

the [1014] students [1111] opened [2022]
their [0121] laptop [1401] book [4044]
exam [0133] mind [1230]

(c) Embedding for both content and target word

the : [2 1 2 4] students [0.5 2 1 2.5] opened [2.5 1 1 2]
their : [0.5 2 1 1.5] laptop [1.5 3 0 0.5] books [4 1 4 3]
exams : [1 1.5 3.5 3] minds [2.5 2.5 3.5 1]

Foundation for Advancement of Science & Technology

Q No 2

LSTM:

① Target Embeddings:

$$f_t = \sigma(w_f \cdot [h_{t-1} + x_t] + b_f)$$

$$= \sigma\left(\begin{bmatrix} 2 & 2 & 2 & 6 & 0 & 6 & 4 & 6 \\ 4 & 1 & 1 & 0 & 2 & 0 & 3 & 5 \\ 1 & 2 & 1 & 3 & 4 & 5 & 1 & 6 \\ 2 & 2 & 2 & 4 & 1 & 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 43 \\ 35 \\ 49 \\ 28 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}\right) = \sigma\begin{bmatrix} 48 \\ 34 \\ 50 \\ 28 \end{bmatrix} \Rightarrow f_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_t = \sigma(w_i [h_{t-1} + x_t] + b_i)$$

$$= \sigma\left(\begin{bmatrix} 6 & 5 & 5 & 0 & 1 & 5 & 0 & 1 \\ 3 & 2 & 2 & 6 & 2 & 1 & 5 & 6 \\ 3 & 2 & 1 & 3 & 0 & 2 & 3 & 2 \\ 4 & 4 & 1 & 5 & 4 & 1 & 4 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix}\right)$$

$$= \sigma\begin{bmatrix} 17 \\ 47 \\ 22 \\ 32 \end{bmatrix} \Rightarrow i_t = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Foundation for Advancement of Science & Technology

$$\Rightarrow \tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right).$$

$$\sigma\left(\begin{bmatrix} 0 & 5 & 6 & 0 & 5 & 4 & 0 & 3 \\ 0 & 1 & 5 & 6 & 0 & 3 & 3 & 6 \\ 0 & 0 & 5 & 5 & 0 & 5 & 3 & 4 \\ 2 & 0 & 6 & 5 & 5 & 5 & 6 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix}\right)$$

$$\tanh\left(\begin{bmatrix} 35 \\ 39 \\ 35 \\ 51 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix}\right) = \tanh\left(\begin{bmatrix} 35 \\ 41 \\ 35 \\ 53 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

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

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

$$\Rightarrow O_t = \sigma\left(W_o \left[h_{t-1}, x_t\right] + b_o\right)$$

$$= \sigma \begin{bmatrix} 5 & 9 & 3 & 5 & 1 & 3 & 2 & 4 \\ 4 & 2 & 2 & 1 & 5 & 0 & 1 & 2 \\ 3 & 0 & 4 & 0 & 2 & 1 & 1 & 5 \\ 4 & 1 & 5 & 3 & 5 & 3 & 3 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 8 \end{bmatrix}$$

$$\sigma\left(\begin{bmatrix} 37 \\ 30 \\ 31 \\ 62 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_t = O_t \times \tanh(c_t)$$

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

$$f_{t+1} = \sigma(W_f \cdot [h_t, x_{t+1}] + b_f)$$

$$\Rightarrow \sigma \begin{bmatrix} 2 & 2 & 2 & 6 & 0 & 6 & 4 & 6 \\ 4 & 1 & 1 & 0 & 2 & 0 & 3 & 5 \\ 1 & 2 & 1 & 3 & 4 & 5 & 1 & 6 \\ 2 & 2 & 2 & 4 & 1 & 1 & 5 & 2 \end{bmatrix} * \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ 0 \\ 3 \\ 1 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow \sigma\left(\begin{bmatrix} 55.4 \\ 29.56 \\ 46.32 \\ 23.6 \end{bmatrix}\right) \Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_{t+1} = \sigma(W_i \cdot [h_{t-1}, x_{t+1}] + b_i)$$

$$\Rightarrow \sigma\left(\begin{bmatrix} 6 & 5 & 5 & 0 & 1 & 5 & 0 & 1 \\ 3 & 2 & 2 & 6 & 2 & 1 & 5 & 6 \\ 3 & 2 & 1 & 3 & 0 & 2 & 3 & 2 \\ 4 & 4 & 1 & 5 & 4 & 1 & 4 & 0 \end{bmatrix} \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ 0 \\ 3 \\ 1 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 31.16 \\ 41.88 \\ 24.84 \\ 21.96 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Foundation for Advancement of Science & Technology

$$\tilde{C}_{t+1} = \tanh\left(w_c \cdot [h_t, x_{t+1}] + b_c\right)$$

$$\Rightarrow \tanh\left(\begin{bmatrix} 0 & 5 & 6 & 0 & 5 & 4 & 0 & 3 \\ 0 & 1 & 5 & 6 & 0 & 3 & 3 & 6 \\ 0 & 0 & 5 & 5 & 0 & 5 & 3 & 4 \\ 2 & 6 & 6 & 5 & 5 & 5 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ 0 \\ 3 \\ 1 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix}\right)$$

$$\Rightarrow \tanh\left(\begin{bmatrix} 32.36 \\ 47.12 \\ 41.6 \\ 40.88 \end{bmatrix}\right) \Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+1} = f_{t+1} \times C_t + i_{t+1} \times \tilde{C}_{t+1}$$

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

$$O_{t+1} = \sigma\left(w_o [h_t, x_{t+1}] + b_o\right)$$

$$= \sigma\left(\begin{bmatrix} 5 & 4 & 3 & 5 & 1 & 3 & 2 & 4 \\ 4 & 2 & 2 & 1 & 5 & 0 & 1 & 2 \\ 3 & 0 & 4 & 0 & 2 & 1 & 1 & 5 \\ 4 & 1 & 5 & 3 & 5 & 3 & 3 & 6 \end{bmatrix} \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ 0 \\ 3 \\ 1 \\ 4 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 8 \end{bmatrix}\right)$$

$$= \sigma\begin{pmatrix} 45.92 \\ 19.84 \\ 29.32 \\ 53.88 \end{pmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Foundation for Advancement of Science & Technology

$$h_{t+1} = O_{t+1} \times \tanh(c_{t+1})$$

$$\Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix}$$

$$f_{t+2} = \sigma\left[w_f \cdot [h_{t+1}, x_{t+2}] + b_f \right]$$

$$= \sigma\left(\begin{bmatrix} 2 & 2 & 2 & 6 & 0 & 6 & 4 & 6 \\ 4 & 1 & 1 & 0 & 2 & 0 & 3 & 5 \\ 1 & 2 & 1 & 3 & 4 & 5 & 1 & 6 \\ 2 & 2 & 2 & 4 & 1 & 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 3 \\ 2 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right)$$

$$= \sigma\left(\begin{bmatrix} 35.52 \\ 23.76 \\ 21.72 \\ 18.6 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_{t+2} = \sigma(w_i \times [h_{t+1}, x_{t+2}] + b_i)$$

$$= \sigma\left[w_i \times \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 3 \\ 2 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right] \Rightarrow \sigma\left(\begin{bmatrix} 30.36 \\ 32.48 \\ 17.66 \\ 33.44 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{c}_{t+2} = \tanh(w_c \cdot [h_{t+1}, x_{t+2}] + b_c)$$

$$= \tanh\left((w_c) \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 3 \\ 2 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 2 \end{bmatrix} \right) \Rightarrow \tanh\left(\begin{bmatrix} 39.56 \\ 31.52 \\ 27.6 \\ 23.48 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+2} = f_{t+2} * C_{t+1} + i_{t+2} * \tilde{C}_t$$

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

$$o_{t+2} = \sigma(W_o [h_{t+1}, x_{t+2}] + b_o)$$

$$= \sigma\left([W_o] \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 3 \\ 6 \\ 2 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 6 \\ 8 \end{bmatrix}\right) = \sigma\left(\begin{bmatrix} 39.32 \\ 31.64 \\ 24.72 \\ 53.18 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{t+2} = o_{t+2} \times \tanh(C_{t+2})$$

$$h_{t+2} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix} = \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix}$$

$$f_{t+3} = \sigma(W_f \cdot [h_{t+2}, x_{t+3}] b_f)$$

$$= \sigma\left([W_f] \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \\ 1 \\ 3 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 6 \\ 2 \\ 6 \end{bmatrix}\right) = \sigma\left(\begin{bmatrix} 41.88 \\ 19.97 \\ 38.93 \\ 17.9 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_{t+3} = \sigma(W_i \times [h_{t+2}, x_{t+3}] + b_i)$$

$$= \sigma\left([W_i] \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \\ 1 \\ 3 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix}\right) \Rightarrow \sigma\left(\begin{bmatrix} 32.87 \\ 29.87 \\ 19.91 \\ 26.8 \end{bmatrix}\right) \Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Foundation for Advancement of Science & Technology

$$\tilde{C}_{t+3} = \tanh\left(W_c \cdot [h_{t+2}, x_{t+3}] + b_c\right)$$

$$= \tanh\left(\begin{bmatrix} 0 & \cdots & 6 & 3 \\ 0 & \cdots & 3 & 6 \\ 0 & \cdots & 3 & 4 \\ 2 & \cdots & 6 & 2 \end{bmatrix}\begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \\ 1 \\ 3 \\ 6 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 0 \\ 4 \end{bmatrix}\right) \Rightarrow \tanh\left(\begin{bmatrix} 33.89 \\ 37.88 \\ 32.9 \\ 38.87 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+3} = f_{t+3} \times C_{t+2} + i_{t+3} \times \tilde{C}_{t+3}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 2 \end{bmatrix}$$

$$O_{t+3} = \sigma\left(W_o [h_{t+2}, x_{t+3}] + b_o\right)$$

$$= \sigma\left(W_o \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \\ 1 \\ 3 \\ 6 \\ 2 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 6 \end{bmatrix}\right) \Rightarrow \sigma\left(\begin{bmatrix} 40.83 \\ 21.91 \\ 21.93 \\ 46.87 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{t+3} = O_{t+3} \times \tanh\left(C_{t+3}\right)$$

$$\Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix} \cong \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\hat{y} = \text{Softmax}\left(W \cdot h_{t+3} + B\right)$$

$$\text{Softmax}\left(\begin{bmatrix} 6 & 2 & 6 & 0 \\ 1 & 3 & 4 & 1 \\ 5 & 5 & 4 & 1 \\ 6 & 1 & 5 & 1 \\ 4 & 5 & 2 & 1 \\ 2 & 2 & 2 & 3 \\ 4 & 4 & 2 & 5 \\ 1 & 3 & 4 & 6 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 6 \\ 0 \\ 3 \\ 0 \\ 3 \\ 2 \end{bmatrix}\right)$$

Foundation for Advancement of Science & Technology

$$= \text{Softmax}\left(\begin{bmatrix} 17 \\ 11 \\ 15 \\ 10 \\ 14 \\ 11 \\ 19 \\ 14 \end{bmatrix}\right) \Rightarrow \begin{bmatrix} 0.10 \\ 0.0002 \\ 0.014 \\ 0.00007 \\ 0.10 \\ 0.0002 \\ 0.77 \\ 0.005 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

We get the following sentence.

the student opened their exams.

Foundation for Advancement of Science & Technology

② Content Embedding:-

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f).$$

$$= \sigma\left([w_f]\begin{bmatrix}0\\0\\0\\1\\0\\4\end{bmatrix} + \begin{bmatrix}0\\2\\1\\0\\1\end{bmatrix}\right) = \sigma\left(\begin{bmatrix}28\\27\\30\\14\end{bmatrix}\right) \Rightarrow \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$i_t = \sigma(w_i * [h_{t-1}, x_t] + b_i)$$

$$= \sigma\left([w_i]\begin{bmatrix}0\\0\\0\\1\\0\\4\end{bmatrix} + \begin{bmatrix}0\\0\\1\\0\\6\end{bmatrix}\right) = \sigma\left(\begin{bmatrix}5\\31\\12\\17\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$\tilde{c_t} = \tanh\left(w_c \cdot [h_{t-1}, x_t] + b_c\right)$$

$$\tanh\left(\begin{bmatrix}w_c\end{bmatrix}\begin{bmatrix}0\\0\\0\\0\\1\\0\\4\end{bmatrix} + \begin{bmatrix}0\\2\\0\\2\end{bmatrix}\right) = \tanh\left(\begin{bmatrix}17\\29\\19\\21\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c_t}$$

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

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o)$$

$$= \sigma\left([w_o]\begin{bmatrix}0\\0\\0\\0\\1\\0\\4\end{bmatrix} + \begin{bmatrix}6\\4\\0\\8\end{bmatrix}\right) = \sigma\left(\begin{bmatrix}24\\18\\23\\40\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$h_t = O_t \times \tanh(h)(c_t)$$

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

$$f_{t+1} = \sigma\left(w_f \cdot [h_t, x_{t+1}] + b_f\right)$$

$$= \sigma\left(\begin{bmatrix} w_f \end{bmatrix} \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ \vdots \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 25.12 \\ 16.56 \\ 23.32 \\ 16.0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_{t+1} = \sigma\left(w_i \left[h_t, x_{t+1}\right] + b_f\right)$$

$$= \sigma\left(\begin{bmatrix} w_i \end{bmatrix} \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ \vdots \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 19.16 \\ 23.88 \\ 17.84 \\ 25.47 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_{t+1} = \tanh\left(w_c \left[h_t, x_{t+1}\right] + b_c\right)$$

$$= \tanh\left(\begin{bmatrix} 20.16 \\ 23.12 \\ 19.6 \\ 29.8 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+1} = f_{t+1} \times C_t + i_{t+1} \times \tilde{C}_{t+1}$$

$$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} + \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

$$o_{t+1} = \sigma\left(w_o\left(h_t, x_{t+1}\right) + b_o\right).$$

$$= \sigma\left(\left[w_o\right] \begin{bmatrix} 0.76 \\ 0.76 \\ 0.76 \\ 0.76 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 6 \end{bmatrix}\right) = \sigma \begin{pmatrix} 28.92 \\ 18.84 \\ 17.37 \\ 34.88 \end{pmatrix}$$

$$= \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$h_{t+1} = o_{t+1} \times \tanh\left(C_{t+1}\right).$$

$$\Rightarrow \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \times \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix} = \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix}$$

$$f_{t+2} = \sigma\left(w_f \cdot \left[h_{t+1}, x_{t+2}\right] + b_f\right)$$

$$= \sigma\left(\left[w_f\right] \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 2 \\ 0 \\ 2 \\ 1 \end{bmatrix} + \begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 31.52 \\ 29.76 \\ 29.72 \\ 25.6 \end{bmatrix}\right) = \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$

$$i_{t+2} = \sigma\left(W_i \cdot [h_{t+1}, x_{t+2}] + b_i\right)$$

$$= \sigma\left(\begin{bmatrix} 19.36 \\ 38.43 \\ 19.67 \\ 35.47 \end{bmatrix}\right) \approx \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_{t+2} = \tanh\left(W_c \cdot [h_{t+1}, x_{t+2}] + b_c\right)$$

$$= \tanh\left(\begin{bmatrix} 26.52 \\ 31.52 \\ 23.6 \\ 40.48 \end{bmatrix}\right) \approx \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+2} = f_{t+2} \times C_{t+1} + i_{t+2} \times \tilde{C}_{t+2}$$

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

$$o_{t+2} = \sigma\left(W_o [h_{t+1}, x_{t+2}] + b_o\right)$$

$$= \sigma\left([W_o] \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 2 \\ 6 \\ 4 \end{bmatrix} + \begin{bmatrix} 6 \\ 4 \\ 0 \\ 6 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 36.32 \\ 28.67 \\ 22.76 \\ 48.48 \end{bmatrix}\right) \approx \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{t+2} = o_{t+2} \times \tanh\left(C_{t+2}\right)$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix} = \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix}$$

$$f_{t+3} = \sigma\left(w_f \cdot \left[h_{t+2}, x_{t+3}\right] + b_f\right)$$

$$= \sigma\left(\left(w_f\right)\begin{bmatrix}0.99\\0.99\\0.99\\0.99\\0\\1\\2\end{bmatrix} + \begin{bmatrix}0\\2\\1\\6\end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix}31.88\\18.94\\20.93\\22.4\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$i_{t+3} = \sigma\left(w_i \cdot \left[h_{t+2}, x_{t+3}\right] + b_i\right)$$

$$= \sigma\left(\left[w_i\right]\begin{bmatrix}0.99\\0.99\\0.99\\0.99\\0\\1\\2\end{bmatrix} + \begin{bmatrix}0\\0\\1\\6\end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix}21.84\\29.87\\19.91\\28.86\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$\tilde{C}_{t+3} = \tanh\left(w_c\left[h_{t+2}, x_{t+3}\right) + b_c\right]$$

$$= \tanh\left[\left(w_c\right)\begin{bmatrix}0.99\\0.99\\0.99\\0.99\\0\\1\\2\end{bmatrix} + \begin{bmatrix}0\\2\\0\\2\end{bmatrix}\right)$$

$$= \tanh\left[\begin{bmatrix}17.88\\28.83\\24.9\\33.87\end{bmatrix}\right] = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$C_{t+3} = f_{t+3} \times C_{t+2} + i_{t+3} \times \tilde{C}_{t+3}.$$

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

$$o_{t+3} = \sigma\left(w_o\left(h_{t+2}, x_{t+3}\right) + b_o\right)$$

$$= \sigma\left(\begin{bmatrix} 93.83 \\ 16.91 \\ 14.93 \\ 35.87 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_t = o_{t+3} + \tanh\left(c_{t+3}\right) \Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0.94 \\ 0.71 \\ 0.92 \\ 0.94 \end{bmatrix} = \begin{bmatrix} 0.94 \\ 0.94 \\ 0.94 \\ 0.49 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$y = \text{Softmax}\left(w \cdot h_{t+3} + B\right)$$

$$\text{Softmax}\left(\begin{bmatrix} 6 & 2 & 6 & 0 \\ 1 & 3 & 7 & 1 \\ 5 & 5 & 4 & 1 \\ 6 & 1 & 2 & 1 \\ 7 & 5 & 2 & 3 \\ 2 & 2 & 2 & 5 \\ 7 & 4 & 2 & 6 \\ 1 & 3 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 8 \\ 2 \\ 0 \\ 6 \\ 3 \\ 0 \\ 3 \\ 0 \end{bmatrix}\right)$$

$$\text{Softmax}\left(\begin{bmatrix} 17 \\ 11 \\ 16 \\ 16 \\ 17 \\ 11 \\ 19 \\ 14 \end{bmatrix}\right) \Rightarrow \begin{bmatrix} 0.10 \\ 0.0002 \\ 0.014 \\ 0.00009 \\ 0.10 \\ 0.0002 \\ 0.44 \\ 0.005 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

We got the following sentence

the student opened their exam.

Average Embedding:-

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$= \sigma\left(\begin{bmatrix} W_f \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 38 \\ 32 \\ 10 \\ 21 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$= \sigma\left(\begin{bmatrix} W_i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix} 41 \\ 39 \\ 17 \\ 23 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$= \tanh\left(\begin{bmatrix} 26 \\ 25 \\ 24 \\ 27 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

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

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

$$o_t = \sigma\left(w_o\left(h_{t-1}, x_t\right) + b_o\right)$$

$$= \sigma\left(\left[w_o\right]\begin{bmatrix}0\\0\\0\\2\\1\\4\end{bmatrix} + \begin{bmatrix}6\\7\\0\\8\end{bmatrix}\right)$$

$$= \sigma\left(\begin{bmatrix}31\\27\\27\\51\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$h_t = o_t \times \tanh\left(c_t\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \times \begin{bmatrix}0.76\\0.76\\0.76\\0.76\end{bmatrix} = \begin{bmatrix}0.76\\0.76\\0.76\\0.76\end{bmatrix}$$

$$f_{t+2} = \sigma\left(w_f \cdot \left[h_t, x_{t+1}\right] + b_f\right).$$

$$= \sigma\left(\left[w_f\right]\begin{bmatrix}0.76\\0.76\\0.76\\0.76\\0.5\\2\\1\\2.5\end{bmatrix} + \begin{bmatrix}0\\2\\1\\0\end{bmatrix}\right) = \sigma\left(\begin{bmatrix}40.12\\23.06\\37.32\\20.1\end{bmatrix}\right)$$

$$= \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$i_{t+1} = \sigma\left(w_i\left[h_t, x_{t+1}\right] + b_i\right)$$

$$= \sigma\left(\begin{pmatrix}25.16\\32.88\\19.87\\24.64\end{pmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$\tilde{c}_{t+1} = \tanh\left(w_c \cdot \left[h_t, x_{t+1}\right] + b_c\right)$$

$$= \tanh\left(\begin{bmatrix}26.36\\35.12\\30.6\\35.38\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$C_{t+1} = f_t * C_t + i_{t+1} * \tilde{C}_{t+1}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}$$

$$O_{t+1} = \left(w_o \left[h_t, x_{t+1} \right] b_o \right)$$

$$= \sigma \left(\begin{bmatrix} 37.42 \\ 19.37 \\ 21.32 \\ 27.38 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{t+1} = O_{t+1} * \tan h (C_{t+1}) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix} = \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \end{bmatrix}$$

$$f_{t+2} = \sigma \left(w_f \cdot \left[h_{t+1}, x_{t+2} \right] + b_f \right)$$

$$= \sigma \left(\left[w_f \right] \begin{bmatrix} 0.96 \\ 0.96 \\ 0.96 \\ 0.96 \\ 2.5 \\ 1 \\ 1 \\ 2 \end{bmatrix} * \begin{bmatrix} 0 \\ 2 \\ 1 \\ 6 \end{bmatrix} \right) = \sigma \left(\begin{bmatrix} 33.52 \\ 25.76 \\ 35.72 \\ 22.1 \end{bmatrix} \right)$$

$$* \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_{t+2} = \sigma \left(w_i \left[h_{t+1}, x_{t+2} \right] + b_i \right)$$

$$= \sigma \left(\begin{bmatrix} 24.86 \\ 35.48 \\ 18.64 \\ 37.44 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_{t+2} = \tanh \left(w_o \left(h_{t+1}, x_{t+2} \right] + b_o \right)$$

$$= \tanh\left(\begin{bmatrix} 23.06 \\ 31.52 \\ 25.6 \\ 41.98 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+2} = f_{t+2} \times C_{t+1} + i_{t+2} \times \hat{C}_{t+2}$$

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

$$O_{t+2} = \sigma\left(w_o\left(h_{t+1}, x_{t+2}\right) + b_o\right)$$

$$= \sigma\left(\left[\text{[illegible]}\right]\begin{bmatrix} 37.82 \\ 30.17 \\ 23.72 \\ 56.98 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{t+2} = O_{t+2} \times \tanh\left(C_{t+2}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix} = \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix}$$

$$f_{t+3} = \sigma\left(w_f \cdot \left[h_{t+2}, x_{t+3}\right] + b_f\right)$$

$$= \sigma\left(\left[w_f\right]\begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \\ 0.5 \\ 2 \\ 1 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 6 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$i_{t+3} = \sigma\left(w_i\left(h_{t+2}, x_{t+3}\right) + b_i\right)$$

$$= \sigma\begin{pmatrix} 27.87 \\ 29.87 \\ 19.91 \\ 27.96 \end{pmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\tilde{C}_{t+3} = \tanh\left(W_c\left[h_{t+2}, x_{t+3}\right] + b_c\right)$$

$$= \tanh\left(\begin{pmatrix} 25.89 \\ 31.88 \\ 28.9 \\ 36.3 \end{pmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$C_{t+3} = f_{t+3} \times C_{t+2} + i_{t+3} \times \tilde{C}_{t+3}$$

$$= \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix}$$

$$O_{t+3} = \sigma\left(W_o\left(h_{t+2}, x_{t+3}\right) + b_o\right)$$

$$= \sigma\begin{pmatrix} 37.33 \\ 19.41 \\ 18.43 \\ 41.37 \end{pmatrix} \Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$h_{t+3} = O_{t+3} \times \tanh\left(C_{t+3}\right)$$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0.99 \\ 0.99 \\ 0.99 \\ 0.99 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\hat{y} = \text{Softmax}\left(W h_{t+3} + B\right)$$

$$= \begin{bmatrix} 6 & 2 & 6 & 0 \\ 1 & 3 & 4 & 1 \\ 5 & 5 & 4 & 1 \\ 6 & 1 & 2 & 1 \\ 4 & 1 & 2 & 3 \\ 2 & 2 & 2 & 5 \\ 4 & 4 & 2 & 6 \\ 1 & 3 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1111 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \\ 3 \\ 6 \\ 3 \\ 1 \end{bmatrix}$$

$$\Rightarrow \text{Softmax}\begin{pmatrix} 17 \\ 11 \\ 15 \\ 10 \\ 17 \\ 11 \\ 19 \\ 14 \end{pmatrix} = \begin{bmatrix} 0.10 \\ 0.0002 \\ 0.014 \\ 0.00004 \\ 0.10 \\ 0.0002 \\ 0.77 \\ 0.0008 \end{bmatrix}$$

$$\approx \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

By using average embedding we got the following sentences

the student opened their exams.

This it yields the same results as we get from target & context embedding.