

"Solution"

Roll No: _____

Quiz 5 (NLP)

Name: _____

Q1: If our concern is to save Training Time with infinite memory in hand then which classifier you will pick between Transformer & LSTM. Briefly explain your answer? (No more than 40 words)

Transformers as Training its Encoder & Decoder will take only Time step = 1 but at cost of a lot of storage while LSTM takes Time step = 'n' & here 'n' is the sentence length.

Q2: Write Equation and compute Masked Attention on the given (Pre-Trained) embedding of words?

Note: Consider Q, K, and V = X

①

مخكب [0.5, -0.7]
شبه [-0.7, 0.5]
 $X = \begin{bmatrix} 0.5 & -0.7 \\ -0.7 & 0.5 \end{bmatrix}$
Masked Attention
② $\Rightarrow \text{Softmax}\left(\frac{Q \cdot K^T}{\sqrt{2}}\right) \cdot V$

③ $Q \cdot K^T \Rightarrow \begin{bmatrix} 0.74 & -0.7 \\ -0.7 & 0.74 \end{bmatrix}$
④ $\frac{Q \cdot K^T}{\sqrt{2}} = \begin{bmatrix} 0.52 & -0.49 \\ -0.49 & 0.52 \end{bmatrix}$
⑤ Masking $\Rightarrow \begin{bmatrix} 0.52 & -\infty \\ -0.49 & 0.52 \end{bmatrix}$

⑥ Applying Softmax
 $\begin{bmatrix} 1 & 0 \\ 0.27 & 0.73 \end{bmatrix}$
Masked Attention ($Q \cdot K^T$)
Softmax ($\frac{Q \cdot K^T}{\sqrt{2}}$)
 $\begin{bmatrix} 1 & 0 \\ 0.27 & 0.73 \end{bmatrix} \cdot \begin{bmatrix} 0.5 & -0.7 \\ -0.7 & 0.5 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.5 & -0.7 \\ -0.37 & 0.17 \end{bmatrix}$

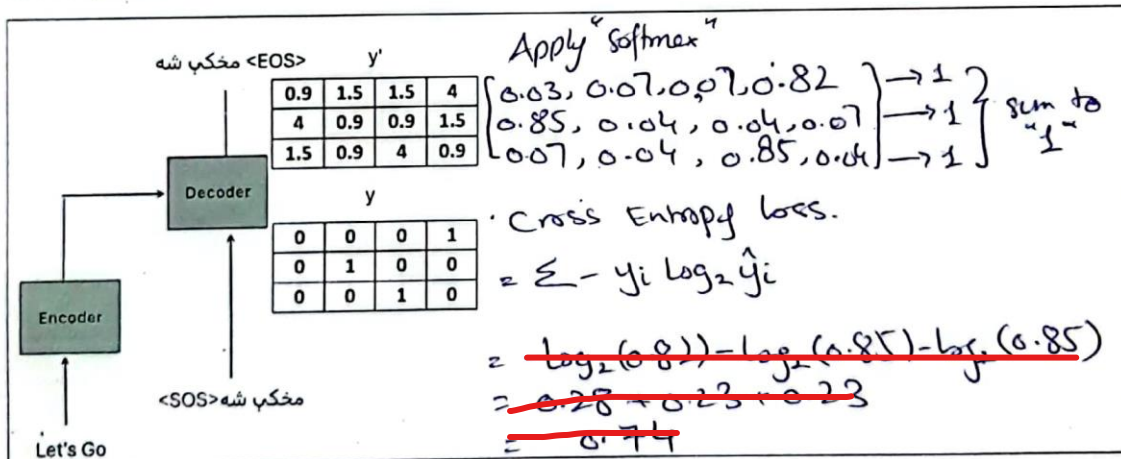
Q3: Compute BLEU Score for the following Machine Translation Sentence?

Example: Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: The cat the cat on the mat.

Q4: Compute Loss function on the following given scenario?



$(-\log_2(0.82) - \log_2(0.04) - \log_2(0.85))$