축하합니다! 통과하셨습니다!

받은 학점 80% **최신 제출물 학점** 80% **통과 점수:** 80% 이상

다음 항목으로 이동

1. A Transformer Network processes sentences from left to right, one word at a time. 1 / 1점

- (●) False
- ( ) True

더 보기

맞습니다
A Transformer Network can ingest entire sentences all at the same time.

2. Transformer Network methodology is taken from: (Check all that apply) 1 / 1점

- [ ] Convolutional Neural Network style of architecture.
- [ ] None of these.
- [x] Convolutional Neural Network style of processing.

✓ Correct

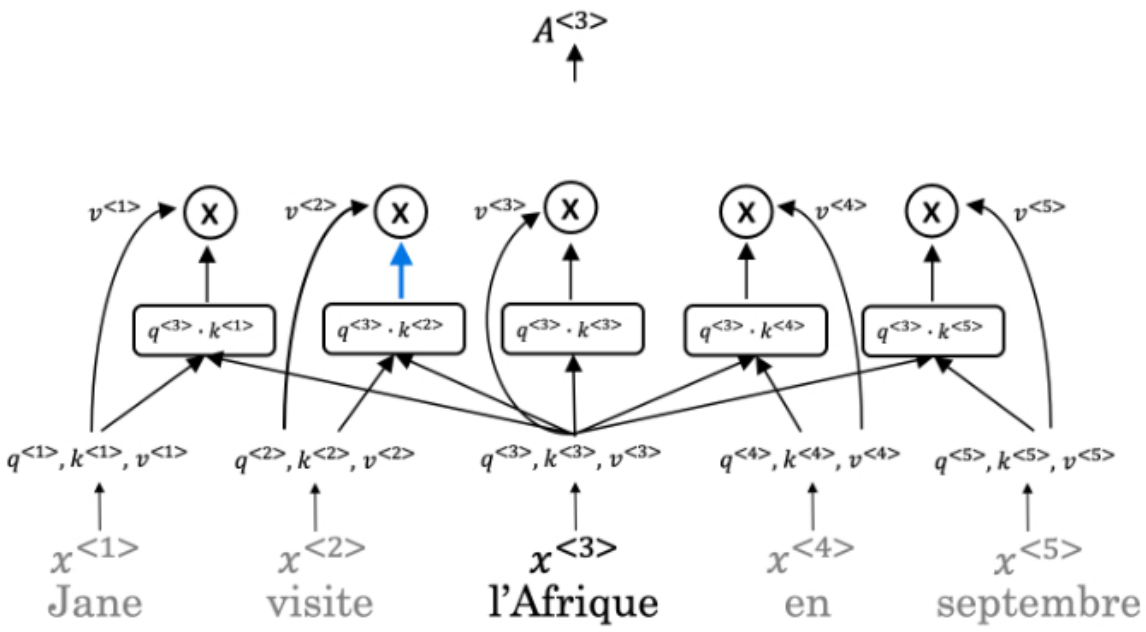
- [x] Attention mechanism.

✓ Correct

더 보기

맞습니다
Great, you got all the right answers.

3. The concept of *Self-Attention* is that: 1 / 1점



- (●) Given a word, its neighbouring words are used to compute its context by summing up the word values to map the Attention related to that given word.
- ( ) Given a word, its neighbouring words are used to compute its context by selecting the lowest of those word values to map the Attention related to that given word.
- ( ) Given a word, its neighbouring words are used to compute its context by taking the average of those word values to map the Attention related to that given word.
- ( ) Given a word, its neighbouring words are used to compute its context by selecting the highest of those word values to map the Attention related to that given word.

더 보기

맞습니다

4. Which of the following correctly represents *Attention*? 1 / 1점

- $${A(Q,K,V)} = (\frac{\exp(q*k^{j})}{\exp(q*k^{j}))*V^{i}$$
- $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q*k^{j})}{{\sum}_j\exp(q*k^{j}))}*{\sum}_i{v}^{i}$$
- $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q*k^{j})}{{\sum}_j\exp(q*k^{j}))}*V^{i}$$
- $${A(Q,K,V)} = {\sum}_i(\frac{\exp(q*v^{j})}{{\sum}_j\exp(q*v^{j}))}*K^{i}$$

더 보기

맞습니다
This is the correct Attention formula.

5. Which of the following statements represents Key (K) as used in the self-attention calculation? 1 / 1점

- K = qualities of words given a Q
- K = specific representations of words given a Q
- K = interesting questions about the words in a sentence
- K = the order of the words in a sentence

더 보기

맞습니다
The qualities of words given a Q are represented by Key (K).

6. $\boldsymbol{Attention(W_i^Q Q, W_i^K K, W_i^V V)}$ 1 / 1점

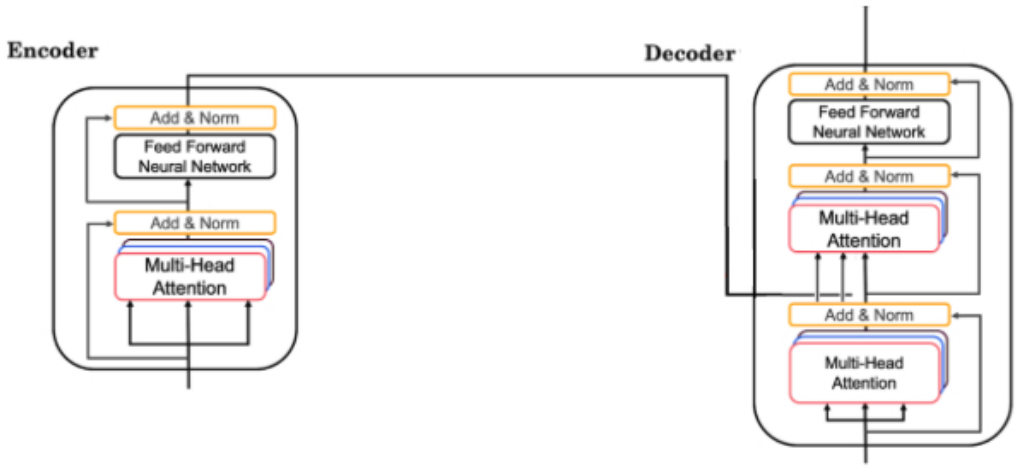
What does i represent in this multi-head attention computation?

- ○ The computed attention weight matrix associated with the order of the words in a sentence
- ○ The computed attention weight matrix associated with the *ith* "word" in a sentence.
- ◉ The computed attention weight matrix associated with the *ith* "head" (sequence)
- ○ The computed attention weight matrix associated with specific representations of words given a Q

↙↗ 더 보기

✓ 맞습니다
i here represents the computed attention weight matrix associated with the "head" (sequence).

7. Following is the architecture within a Transformer Network ***(without displaying positional encoding and output layers(s)).*** 0 / 1점



What is ***NOT*** necessary for the *Decoder's* second block of *Multi-Head Attention*?

- ○ Q
- ○ V
- ○ All of the above are necessary for the Decoder's second block.

K

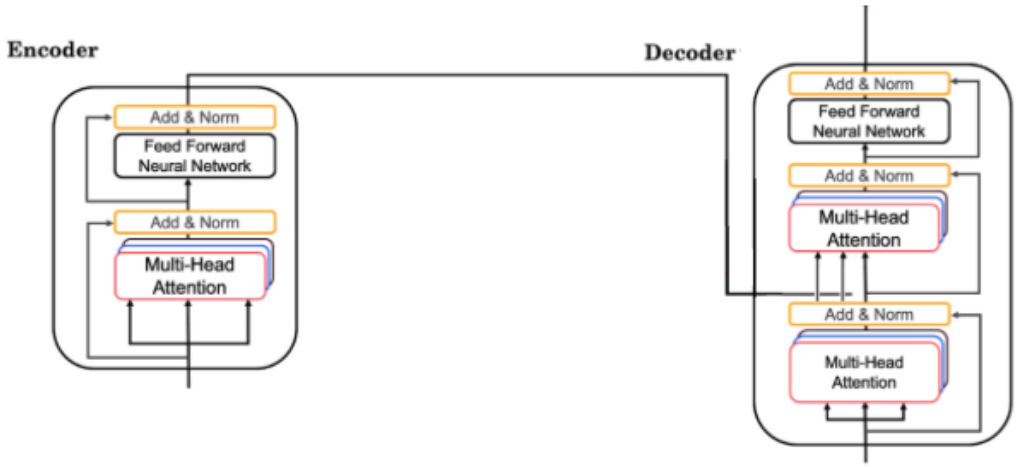
더 보기

틀립니다

The uses K from the for its second block of Multi-Head Attention. To revise the concept watch the lecture .

8. Following is the architecture within a Transformer Network ***(without displaying positional encoding and output layers(s)).***

0 / 1점



What does the output of the *encoder* block contain?

- Softmax layer followed by a linear layer.
- Contextual semantic embedding and positional encoding information
- Linear layer followed by a softmax layer.
- Prediction of the next word.

더 보기

틀립니다

Feedback: To revise the concept watch the lecture .

9. Which of the following statements is true? 1 / 1점

- (●) The transformer network differs from the attention model in that only the transformer network contains positional encoding.
- ( ) The transformer network is similar to the attention model in that both contain positional encoding.
- ( ) The transformer network differs from the attention model in that only the attention model contains positional encoding.
- ( ) The transformer network is similar to the attention model in that neither contain positional encoding.

더 보기

맞습니다
Positional encoding allows the transformer network to offer an additional benefit over the attention model.

10. Which of these is a good criterion for a good positional encoding algorithm? 1 / 1점

- [x] It should output a unique encoding for each time-step (word's position in a sentence).

  ✓ Correct

- [x] Distance between any two time-steps should be consistent for all sentence lengths.

  ✓ Correct

- [x] The algorithm should be able to generalize to longer sentences.

  ✓ Correct

- [ ] None of these.

더 보기

맞습니다

Great, you got all the right answers.