✓ **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

- ☑ Attention mechanism.

✓ Correct

- ☐ Convolutional Neural Network style of architecture.

- ☑ Convolutional Neural Network style of processing.
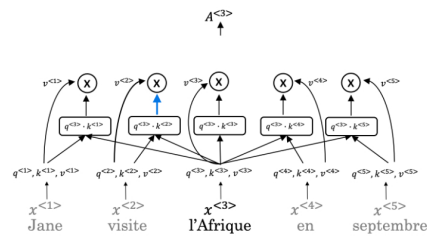
✓ Correct

- ☐ None of these.

[↗ 더 보기]

✓ **맞습니다**
Great, you got all the right answers.

---

3. What are the key inputs to computing the attention value for each word?

1/1점

- ⦿ The key inputs to computing the attention value for each word are called the query, key, and value.
- ◯ The key inputs to computing the attention value for each word are called the quotation, knowledge, and value.
- ◯ The key inputs to computing the attention value for each word are called the quotation, key, and vector.
- ◯ The key inputs to computing the attention value for each word are called the query, knowledge, and vector.

[↗ 더 보기]

✓ **맞습니다**
The key inputs to computing the attention value for each word are called the query, key, and value.

---

4. What letter does the "?" represent in the following representation of *Attention*?

1/1점

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

- ⦿ k
- ◯ q
- ◯ v
- ◯ t

[↗ 더 보기]

✓ **맞습니다**
k is represented by the ? in the representation.

---

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

1/1점

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

[↗ 더 보기]

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

---

6. $Attention(W_i^Q Q, W_i^K K, W_i^V V)$

1/1점

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

( ) The computed attention weight matrix associated with the ith "head" (sequence)

( ) The computed attention weight matrix associated with specific representations of words given a Q

( ) The computed attention weight matrix associated with the ith "word" in a sentence.

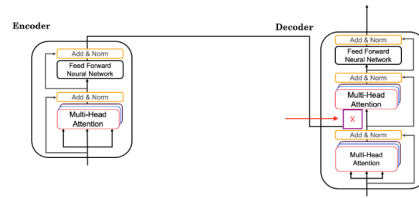( ) The computed attention weight matrix associated with the order of the words in a sentence

[↗ 더 보기]

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

---

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

What information does the *Decoder* take from the *Encoder* for its second block of *Multi-Head Attention*? (Marked X, pointed by the independent arrow)

(Check all that apply)

[ ] Q

[✓] K

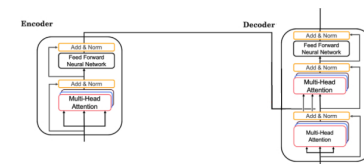✓ Correct

[✓] V

✓ Correct

[↗ 더 보기]

✓ 맞습니다
Great, you got all the right answers.

---

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

The output of the decoder block contains a softmax layer followed by a linear layer to predict the next word one word at a time.

(•) False

( ) True

[↗ 더 보기]

✓ 맞습니다
The output of the decoder block contains a linear layer followed by a softmax layer to predict the next word one word at a time.

---

9. Which of the following statements is true about positional encoding? Select all that apply.   0 / 1점

[✓] Positional encoding provides extra information to our model.

✓ Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

[ ] Positional encoding uses a combination of sine and cosine equations.

[✓] Positional encoding is important because position and word order are essential in sentence construction of any language.

✓ Correct
This is a correct answer, but other options are also correct. To review the concept watch the lecture *Transformer Network*.

[ ] Positional encoding is used in the transformer network and the attention model.

[↗ 더 보기]

⊗ 틀립니다
You didn't select all the correct answers

---

10. Which of these is *not* a good criterion for a good positional encoding algorithm?   0 / 1점

(•) It must be deterministic.

( ) Distance between any two time-steps should be consistent for all sentence lengths.

( ) The algorithm should be able to generalize to longer sentences.

○ It should output a common encoding for each time-step (word's position in a sentence).

더 보기

⊗ 틀립니다
This is a good criterion for a good positional encoding algorithm.