- **Transformers**
- **Lecture Notes (Optional)**
- **Quiz**
  - **테스트:** Transformers
    10개의 질문
- **Programming Assignment**
- **Transformer Applications - Ungraded Labs**
- **Conclusion**
- **References & Acknowledgments**

테스트테스트 • 30 min30 minutes

Transformers

과제 제출
기한년 9월 13일 오후 3:59 KST년 9월 13일 오후 3:59 KST
다시 시도해주십시오

성적 받기
통과 점수:80% 이상
성적
100%
피드백 보기
최고 점수가 유지됩니다.

탐색 확인

이 페이지에서 나가시겠습니까?
이 페이지에 머물기 이 페이지에서 나가기

Transformers
성적 평가 퀴즈 • 30 min

만료 년 9월 13일 오후 3:59 KST

축하합니다! 통과하셨습니다!
통과 점수: 80% 이상
학습 계속하기
성적
100%

Transformers

최신 제출물 성적
100%
1.
질문 1

A Transformer Network, like its predecessors RNNs, GRUs and LSTMs, can process information one word at a time. (Sequential architecture).

1 / 1점

○

True

◉ ○

False

맞습니다

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

2.
질문 2

Transformer Network methodology is taken from: (Check all that apply)

1 / 1점

□

None of these.

☑ □

Convolutional Neural Network style of processing.

맞습니다

□

Convolutional Neural Network style of architecture.

☑ □

Attention mechanism.

맞습니다
3.
질문 3

The concept of *Self-Attention* is that:

2

1 / 1점

○

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 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 highest 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 lowest of those word values to map the Attention related to that given word.

맞습니다
4.
질문 4

Which of the following correctly represents *Attention ?*

1 / 1점

○

$Attention(Q, K, V) = softmax(\frac{QV^T}{\sqrt{d_k}})K$Attention(Q, K, V) = softmax(dk

√

QV T)K

◉ ○

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

√

QKT)V

○

$Attention(Q, K, V) = min(\frac{QK^T}{\sqrt{d_k}})V$Attention(Q, K, V) = min(dk

√

QKT)V

○

$Attention(Q, K, V) = min(\frac{QV^T}{\sqrt{d_k}})K$Attention(Q, K, V) = min(dk

√

QV T)K

✓

맞습니다

5.

질문 5

Are the following statements true regarding Query (Q), Key (K) and Value (V) ?

Q = interesting questions about the words in a sentence

K = specific representations of words given a Q

V = qualities of words given a Q

1 / 1점

◉○

False

○

True

✓

맞습니다

Correct! Q = interesting questions about the words in a sentence, K = qualities of words given a Q, V = specific representations of words given a Q

6.

질문 6

ii here represents the computed attention weight matrix associated with the ithith “word” in a sentence.

1 / 1점

○

True

◉○

False

✓

맞습니다

Correct! ii here represents the computed attention weight matrix associated with the ithith “head” (sequence).

7.

질문 7

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

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

(Check all that apply)

1 / 1점

☑□

K

✓

맞습니다

☑□

V

✓

맞습니다

□□

Q

8.

질문 8

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

What is the output layer(s) of the *Decoder* ? (Marked YY, pointed by the independent arrow)

1 / 1점

○

Linear layer

○

Softmax layer

○

Softmax layer followed by a linear layer.

◉○

Linear layer followed by a softmax layer.

✓
맞습니다
9.
질문 9

Why is positional encoding important in the translation process? (Check all that apply)

1 / 1점
☑ □

Position and word order are essential in sentence construction of any language.

✓
맞습니다
☐ □

It helps to locate every word within a sentence.

☐ □

It is used in CNN and works well there.

☑ □

Providing extra information to our model.

✓
맞습니다
10.
질문 10

Which of these is a good criteria for a good positionial encoding algorithm?

1 / 1점
☑ □

It should output a unique encoding for each time-step (word’s position in a sentence).

✓
맞습니다
☑ □

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

✓
맞습니다
☑ □

The algorithm should be able to generalize to longer sentences.

✓
맞습니다
☐ □

None of the these.