

## 문제 + 정답

1. 다음 중 one-hot encoding의 단점이 아닌 것은?

- ① 차원의 저주 문제 발생
- ② 단어 간 유사도를 반영하기 어렵다
- ③ 의미 정보를 풍부하게 표현할 수 있다
- ④ 메모리 비효율적이다

정답: ③

2. LSTM의 forget gate가 수행하는 역할은?

- ① 새로운 정보를 생성한다
- ② 이전 정보를 삭제하거나 유지한다
- ③ 출력값을 정규화한다
- ④ 입력값을 변환한다

정답: ②

3. Seq2Seq 모델에서 Encoder의 역할은?

- ① 출력을 생성한다
- ② 입력 시퀀스를 고정된 벡터로 인코딩한다
- ③ 번역 문장을 디코딩한다
- ④ Attention을 계산한다

정답: ②

4. Self-Attention이란?

- ① 문장 간의 관계를 학습
- ② 문장 내부 단어 간의 관계를 학습
- ③ 입력과 출력 간의 관계를 학습
- ④ 단어 임베딩을 학습

정답: ②

5. Transformer가 RNN보다 발전한 이유로 옳지 않은 것은?

- ① 병렬 학습이 가능하다
- ② 장기 의존성 학습에 유리하다
- ③ Recurrence 구조가 존재한다
- ④ Self-Attention을 사용한다

정답: ③

6. Multi-Head Attention의 주된 목적은?

- ① 여러 Attention을 병렬로 학습해 다양한 문맥을 포착
- ② 단일 Attention을 반복 적용
- ③ Gradient를 줄인다

④ Decoder를 단순화한다

정답: ①

7. Transformer의 주요 응용 분야가 아닌 것은?

① 번역

② 요약

③ 이미지 분류

④ 대화 생성

정답: ③

8. LSTM의 3가지 게이트 이름과 각각의 역할을 쓰시오.

정답 :

Forget gate: 이전 cell state에서 무엇을 버릴지 결정.

Input gate: 현재 입력으로부터 얼마나 새 정보를 추가할지 결정.

Output gate: cell state에서 얼마를 hidden state로 내보낼지 결정.

9. Multi-Head Attention의 필요성을 설명하시오.

정답 :

하나의 attention head는 제한된 표현(한 관점)만 포착하므로, 여러 head를 병렬로 사용하면 문법적 관계·의미적 관계·위치 정보 등 다양한 관점을 동시에 모델링할 수 있어 표현력이 증가한다.

10. Transformer가 RNN보다 효율적인 이유를 병렬 처리 관점에서 설명하시오.

정답 :

RNN은 시점별로 순차 계산을 해야 하므로 병렬화가 제한된다.

Transformer의 self-attention은 모든 토큰의 상호작용을 동시 계산할 수

있어 GPU/TPU 등에서 대규모 병렬 처리가 가능하다.

11. 제시된 CNN의 한계점 중, 텍스트나 음성 같은 순차적 데이터 처리에 부적합한 가장 큰 이유는?

- 가) 데이터의 순서(Order) 정보를 무시함
- 나) 파라미터 효율성이 낮음
- 다) 지역적 특징 학습에 약함
- 라) 연산 속도가 너무 느림

답) 가

12. '자기 어텐션 (Self-Attention)'에 대한 설명으로 올바른 것은?

- 가) 두 개의 서로 다른 데이터 간 유사도를 반영
- 나) 하나의 입력 데이터 내부에서 요소(패치) 간 유사도를 반영
- 다) CNN의 합성곱 연산과 동일함
- 라) RNN의 순환 구조를 필수적으로 사용함

답) 나

13. ViT에서 분할된 패치 토큰의 '순서(위치 정보)'를 학습에 반영하기 위해 추가하는 것은?

답) 위치 인코딩 (Position Encoding)

14. ReLU의 문제점 중, 입력이 계속 음수여서 뉴런이 영구적으로 0만 출력하게 되는 현상을 무엇이라 하는가?

- 가) 기울기 폭발
- 나) 죽은 뉴런 문제 (Dead ReLU)
- 다) 정보 희석
- 라) 느린 수렴

답) 다

15. 드롭아웃이 과적합을 방지하는 직관적인 원리이자 장점은?

- 가) 모든 가중치를 0으로 만들
- 나) 네트워크가 특정 뉴런/특징에 과도하게 의존하지 않게 함
- 다) 학습률을 선형적으로 줄여줌
- 라) 모든 뉴런의 출력을 0으로 만들

답) 다

16. 다음 중 one-hot encoding의 단점이 아닌 것은?

- ① 차원의 저주 문제 발생
- ② 단어 간 유사도를 반영하기 어렵다
- ③ 의미 정보를 풍부하게 표현할 수 있다
- ④ 메모리 비효율적이다

정답: ③

17. LSTM의 forget gate가 수행하는 역할은?

- ① 새로운 정보를 생성한다
- ② 이전 정보를 삭제하거나 유지한다
- ③ 출력값을 정규화한다
- ④ 입력값을 변환한다

정답: ②

18. Seq2Seq 모델에서 Encoder의 역할은?

- ① 출력을 생성한다
- ② 입력 시퀀스를 고정된 벡터로 인코딩한다

③ 번역 문장을 디코딩한다

④ Attention을 계산한다

정답: ②

19. Self-Attention이란?

① 문장 간의 관계를 학습

② 문장 내부 단어 간의 관계를 학습

③ 입력과 출력 간의 관계를 학습

④ 단어 임베딩을 학습

정답: ②

20. Transformer가 RNN보다 발전한 이유로 옳지 않은 것은?

① 병렬 학습이 가능하다

② 장기 의존성 학습에 유리하다

③ Recurrence 구조가 존재한다

④ Self-Attention을 사용한다

정답: ③

21. Multi-Head Attention의 주된 목적은?

① 여러 Attention을 병렬로 학습해 다양한 문맥을 포착

② 단일 Attention을 반복 적용

③ Gradient를 줄인다

④ Decoder를 단순화한다

정답: ①

22. Transformer의 주요 응용 분야가 아닌 것은?

① 번역

② 요약

③ 이미지 분류

④ 대화 생성

정답: ③

23. Transformer가 RNN보다 효율적인 이유를 병렬 처리 관점에서 설명하시오.

정답 :

Transformer가 RNN보다 병렬 처리 관점에서 효율적인 이유.