

1. 머신러닝을 어떻게 정의할 수 있나요?

효율을 높이기 위해 컴퓨터의 학습 능력을 통하여 데이터의 패턴을 찾고 알고리즘을 개발하는 것

2. 머신러닝이 도움을 줄 수 있는 문제 유형 네 가지를 말해보세요.

- 1) 기존의 방식으로는 많은 수동 조정과 규칙이 필요한 문제
- 2) 전통적 방식으로는 해결 불가능한 복잡한 문제
- 3) 데이터가 유동적인 문제
- 4) 복잡한 문제와 대량의 데이터가 주어진 상황

3. 레이블된 훈련 세트란 무엇인가요?

훈련 데이터(x) 각각에 답(y)이 지정되어 있는 것

4. 가장 널리 사용되는 지도 학습 작업 두 가지는 무엇인가요?

분류, 회귀

5. 보편적인 비지도 학습 작업 네 가지는 무엇인가요?

군집, 시각화와 차원 축소, 이상치 탐지와 특이치 탐지, 연관 규칙 학습

6. 사전 정보가 없는 여러 지형에서 로봇을 걸어가게 하려면 어떤 종류의 머신러닝 알고리즘을 사용할 수 있나요?

강화 학습

7. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요?

비지도 학습 - 군집

8. 스팸 감지의 문제는 지도 학습과 비지도 학습 중 어떤 문제로 볼 수 있나요?

지도 학습 - 분류 (훈련 데이터의 레이블 이용)

9. 온라인 학습 시스템이 무엇인가요?

적은 양의 데이터를 사용해 점진적으로 훈련해 나가는 학습 방식

10. 외부 메모리 학습이 무엇인가요?

컴퓨터 주메모리에 들어갈 수 없는 대용량의 데이터를 처리하도록 설계된 알고리즘 (온라인 학습 알고리즘 사용)

11. 예측을 하기 위해 유사도 측정에 의존하는 학습 알고리즘은 무엇인가요?

사례 기반 학습

12. 모델 파라미터와 학습 알고리즘의 하이퍼파라미터 사이에는 어떤 차이가 있나요?

모델 파라미터는 데이터를 통해 자동으로 학습되는 값이며, 학습 알고리즘의 하이퍼파라미터는 학습 과정 제어를 목적으로 사용자가 사전에 설정하는 값이다.

13. 모델 기반 알고리즘이 찾는 것은 무엇인가요? 성공을 위해 이 알고리즘이 사용하는 가장 일반적인 전략은 무엇인가요? 예측은 어떻게 만드나요?

효용함수의 최대화 또는 비용함수의 최소화를 실행하는 모델 파라미터

새로 입력된 데이터를 학습한 모델을 통하여 출력값을 계산

14. 머신러닝의 주요 도전 과제는 무엇인가요?

충분하지 않은 양의 훈련 데이터, 대표성이 없는 훈련 데이터, 낮은 품질의 데이터, 관련이 없는 특성, 과대적합, 과소적합 상황

15. 모델이 훈련 데이터에서의 성능은 좋지만 새로운 샘플에서의 일반화 성능이 나쁘다면 어떤 문제가 있는 건가요? 가능한 해결책 세 가지는 무엇인가요?

훈련 데이터의 과대 적합 상황

검증, 하이퍼파라미터 튜닝과 모델 선택, 교차 검증

16. 테스트 세트가 무엇이고 왜 사용해야 하나요?

훈련된 모델의 성능을 평가하는 데이터셋

17. 검증 세트의 목적은 무엇인가요?

다양한 하이퍼파라미터 값을 통한 후보 모델 평가용으로 사용(홀드아웃 검증)

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

성능이 왜곡되는 문제 발생, 최종 성능을 평가해야하는 테스트 세트로 하이퍼파라미터를 결정하게 될 경우 모델 자체가 테스트 세트에 최적화 되어 과대 적합의 문제 상황이 발생한다.