

1 주차 과제 _ 안도원

1. 머신러닝을 어떻게 정의할 수 있나요?
→ 명시적인 프로그래밍 없이 컴퓨터가 학습하는 능력을 갖추게 하는 연구 분야
2. 머신러닝이 도움을 줄 수 있는 문제 유형 네 가지를 말해보세요.
→ 생산 라인에서 제품 이미지를 분석해 자동으로 분류하기
→ 뇌를 스캔하여 종양 진단하기
→ 자동으로 뉴스 기사를 분류하기
→ 챗봇 또는 개인 비서 만들기
3. 레이블된 훈련 세트란 무엇인가요?
→ 지도 학습에 사용되는, 답을 포함한 샘플 데이터의 집합
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 감지, 연관 규칙 학습
6. 사전 정보가 없는 여러 지형에서 로봇을 걸어가게 하려면 어떤 종류의 머신러닝 알고리즘을 사용할 수 있나요?
→ 강화 학습 알고리즘
7. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요?
→ k-means, DBSCAN, 계층 군집 분석(DCA) 등 군집 알고리즘
8. 스팸 감지의 문제는 지도 학습과 비지도 학습 중 어떤 문제로 볼 수 있나요?
→ 지도 학습
학습에 이용되는 데이터의 스팸 여부는 레이블로 볼 수 있다.
9. 온라인 학습 시스템이 무엇인가요?
→ 적은 양의 데이터(미니배치)를 사용해 점진적으로 훈련하는 것
10. 외부 메모리 학습이 무엇인가요?

→ RAM 뿐만이 아니라 외부 메모리의 데이터에 접근할 수 있는 상태에서 학습하는 것, 큰 데이터셋을 학습하는 데 도움됨.

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

→ 사례 기반 학습

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

→ 하이퍼파라미터는 훈련 과정에 변하는 모델 파라미터와 다르게 알고리즘 학습 모델을 정의하는데 사용된다.

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

→ 새로운 데이터를 예측할 수 있는 함수를 찾는다. 손실 함수 최소화

14. 머신러닝의 주요 도전 과제는 무엇인가요?

→ 충분하지 않은 양의 훈련 데이터

→ 대표성 없는 훈련 데이터

→ 낮은 품질의 데이터

→ 관련이 없는 특성

→ 훈련 데이터 과대/과소적합

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

→ 훈련 데이터 과대적합의 문제이다.

해결책:

→ 여러 규제를 적용

→ 모델 단순화

→ 교차 검증

16. 테스트 세트가 무엇이고 왜 사용해야 하나요?

➔ 훈련된 모델의 성능을 평가하기 위해 데이터셋의 일정 비율을 분리하여 사용하는 세트.

17. 검증 세트의 목적은 무엇인가요?

➔ 모델의 성능을 측정하기 위함임.

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

➔ 과대적합 문제