

1 주차 과제

1. 머신러닝을 어떻게 정의할 수 있나요?
A. 데이터를 스스로 학습해 성능을 향상시키는 기술
2. 머신러닝이 도움을 줄 수 있는 문제 유형 네 가지를 말해보세요.
A. 기존 솔루션으로 많은 수동 조정과 규칙이 필요한 문제, 전통적인 방식으로 해결 방법이 없는 복잡한 문제, 유동적인 환경, 복잡한 문제와 대량의 데이터에서의 통찰
3. 레이블된 훈련 세트란 무엇인가요?
A. 답을 포함하고 있는 훈련용 데이터
4. 가장 널리 사용되는 지도 학습 작업 두 가지는 무엇인가요?
A. 분류, 회귀
5. 보편적인 비지도 학습 작업 네 가지는 무엇인가요?
A. 군집, 시각화와 차원 축소, 이상치 탐지와 특이치 탐지, 연관 규칙 학습
6. 사전 정보가 없는 여러 지형에서 로봇을 걸아가게 하려면 어떤 종류의 머신러닝 알고리즘을 사용할 수 있나요?
A. 강화 학습
7. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요?
A. 군집 알고리즘
8. 스팸 감지의 문제는 지도 학습과 비지도 학습 중 어떤 문제로 볼 수 있나요?
A. 지도 학습
9. 온라인 학습 시스템이 무엇인가요?
A. 적은 양의 데이터를 사용해 점진적으로 훈련하는 시스템
10. 외부 메모리 학습이 무엇인가요?
A. 빅데이터 분석시 데이터의 양이 지나치게 커서 컴퓨터의 메모리로 감당되지 않는 경우 일부씩 학습하여 전체 데이터가 적용될 때까지 반복하는 학습 방법
11. 예측을 하기 위해 유사도 측정에 의존하는 학습 알고리즘은 무엇인가요?
A. 사례 기반 학습
12. 모델 파라미터와 학습 알고리즘의 하이퍼파라미터 사이에는 어떤 차이가 있나요?
A. 하이퍼파라미터는 훈련 과정에서 변하지 않음.
13. 모델 기반 알고리즘이 찾는 것은 무엇인가요? 성공을 위해 이 알고리즘이 사용하는 가장 일반적인 전략은 무엇인가요? 예측은 어떻게 만드나요?
A. 데이터에 적합한 모델을 찾는다. 모델을 미리 지정한 후 훈련 세트를 사용해 모델을 훈련시킴. 훈련된 모델에 새로운 데이터를 적용하여 이에 대한 예측을 실행한다.
14. 머신러닝의 주요 도전 과제는 무엇인가요?
A. 충분하지 않은 양의 훈련 데이터, 대표성 없는 훈련데이터, 낮은 품질의 데이터, 관련 없는 특성, 과대/과소적합한 데이터에 대한 패턴 학습
15. 모델이 훈련 데이터에서의 성능은 좋지만 새로운 샘플에서의 일반화 성능이 나쁘다면 어떤 문제가 있는 건가요? 가능한 해결책 세 가지는 무엇인가요?
A. 문제 - 과대적합. 해결책 - 데이터 양 늘리기, 모델 복잡도 줄이기, 학습 데이터의 noise 줄이기

16. 테스트 세트가 무엇이고 왜 사용해야 하나요?

A. 학습 후 테스트용 세트. 생성한 모델의 오차 비율과 성능을 확인하기 위해 사용

17. 검증 세트의 목적은 무엇인가요?

A. 학습 세트의 결과 평가

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

A. 과대 적합이 발생하고, 유의미한 테스트 결과를 얻을 수 없음.