

위티 머신러닝 세션 1주차 과제

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.

신용카드 부정 거래 감지하기, 자동으로 뉴스 기사 분류하기, 과거 구매 이력을 기반으로 고객이 관심을 가질 수 있는 상품 추천하기, 챗봇 또는 개인 비서 만들기

3. 레이블된 훈련 세트란 무엇인가요?

지도학습에서 주로 사용하고, 훈련 데이터에 레이블이라는 답을 포함하고 있는 데이터 세트

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 (에이전트가 환경을 관찰하여 행동을 실행하고 그 결과로 보상 혹은 벌점을 받는다)

7. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요?

군집

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

지도학습

9. 온라인 학습 시스템이 무엇인가요?

적은 양의 데이터(미니배치)를 사용해 점진적으로 훈련하는 것

10. 외부 메모리 학습이 무엇인가요?

대용량 데이터를 다룰 때 미니배치라는 작은 단위로 나누어 점진적으로 학습하는 방식, 보통 온라인 학습 알고리즘과 결합되어 사용됨

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 학습 데이터로부터 모델이 자동으로 학습하는 값이고, 하이퍼파라미터는 사용자가 모델 학습 전에 직접 설정하는 값으로, 학습 방식이나 모델의 구조를 조절한다. 또한, 학습 과정 중에 변하는 파라미터가 아님.

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

모델 기반 알고리즘이 찾는 것은 학습 과정에서 모델을 가장 잘 설명하는 최적 파라미터 값. 모델을 미리 지정한 후 훈련 세트를 사용하여 모델을 훈련시킨다. 훈련된 모델을 사용해 새로운 데이터에 대한 예측을 실행한다.

14. 머신러닝의 주요 도전 과제는 무엇인가요?

충분하지 않은 양의 훈련 데이터, 대표성 없는 훈련 데이터, 낮은 품질의 데이터, 관련이 없는 특성, 훈련 데이터 과대적합&과소적합

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

과대적합 문제, 해결책: 데이터 양 늘리기, 모델 복잡도 줄이기, 여러 규제 적용

16. 테스트 세트가 무엇이고 왜 사용해야 하나요?

훈련 세트를 제외한 나머지 데이터셋, 훈련에 사용되지 않은 데이터셋을 이용해 훈련된 모델의 성능을 평가해야 하기 때문에.

17. 검증 세트의 목적은 무엇인가요?

검증 세트는 훈련 세트의 일부로 만들어진 데이터셋이며 모델의 하이퍼파라미터를 조정하거나 모델을 선택하는 데 사용된다.

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

모델이 테스트 데이터에 과대적합되어 새로운 데이터에 대한 일반화 성능이 떨어질 수 있다.