

1 주차 과제

1. 머신러닝을 어떻게 정의할 수 있나요?

: 데이터를 통해 예측 모델을 학습하고, 그 모델이 더 나은 성능을 내도록 계수나 파라미터를 조정하는 과정

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

: 기존 솔루션으로 많은 수동 조정과 규칙이 필요한 문제, 전통적인 방식으로 해결 방법이 없는 복잡한 문제, 유동적인 환경인 문제, 복잡한 문제와 대량의 데이터에서 통찰을 얻고 싶을 때

3. 레이블된 훈련 세트란 무엇인가요?

: 답이 정해져있는 훈련 데이터 세트

4. 가장 널리 사용되는 지도 학습 작업 두 가지는 무엇인가요?

: 회귀, 분류

5. 보편적인 비지도 학습 작업 네 가지는 무엇인가요?

: 군집, 시각화와 차원축소, 이상치탐지와 특이치탐지, 연관규칙학습

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

: 강화학습, 온라인학습

7. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요?

: 비지도학습(군집)

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

: 지도학습

9. 온라인 학습 시스템이 무엇인가요?

: 적은 양의 데이터를 사용해 점진적으로 훈련하는 시스템

10. 외부 메모리 학습이 무엇인가요?

: 모델이 기존의 메모리만 사용하는 대신, 외부 저장소를 활용하여 더 많은 정보를 저장하고 학습하는 방식

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 모델이 학습하면서 자동으로 최적화되는 값, 하이퍼파라미터는 모델 학습 전에 사람이 설정해야 하는 값(훈련 과정에서 변하지 않음)

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

: 모델 기반 알고리즘이 찾는 것은 입력 데이터와 출력 간의 관계이며 성공을 위해 이 알고리즘이 사용하는 가장 일반적인 전략은 손실함수 최소화로, 주로 경사하강법이나 정규화 등의 기법을 통해 이루어짐. 예측을 만드는 방법은 학습된 모델을 사용하여 새로운 입력 데이터에 대한 예측을 생성하는 것임.

14. 머신러닝의 주요 도전 과제는 무엇인가요?

: 충분하지 않은 훈련 데이터, 대표성이 없는 훈련 데이터, 낮은 품질의 데이터, 관련이 없는 특성, 훈련 데이터 overfitting/underfitting

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

: 훈련 데이터 overfitting 문제. 해결책 : 정규화를 통해 과대적합 감소시키기, 교차검증, 더 많은 데이터 사용

16. 테스트 세트가 무엇이고 왜 사용해야 하나요?

: 모델 테스트용 데이터셋. 훈련된 모델의 성능을 평가하기 위해 사용함

17. 검증 세트의 목적은 무엇인가요?

: 모델을 평가하고, 하이퍼파라미터를 조정하기 위해 사용

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

: 테스트 세트는 모델의 최종 성능을 평가하기 위해 훈련과 검증 과정에서 한 번도 사용되지 않아야 하는데 하이퍼파라미터를 튜닝할 때 테스트 세트를 사용한다면 테스트 세트가 훈련 과정의 일부가 되어 모델의 일반화 능력을 제대로 평가할 수 없음. 모델의 overfitting, 객관적인 성능 평가의 왜곡 등이 발생할 수 있음.