

## 1주차 과제

1. 머신러닝을 어떻게 정의할 수 있나요?

머신러닝이란 명시적인 프로그래밍 없이 컴퓨터가 학습하는 능력을 갖추게 하는 연구 분야

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

스팸 필터링, 이미지 분류, 이상 탐지, 추천 시스템

3. 레이블된 훈련 세트란 무엇인가요?

각 데이터에 대한 정답(레이블)이 포함된 데이터 세트

4. 가장 널리 사용되는 지도 학습 작업 두 가지는 무엇인가요?

분류(Classification), 회귀(Regression)

5. 보편적인 비지도 학습 작업 네 가지는 무엇인가요?

군집(Clustering), 시각화(Visualization)와 차원 축소(Dimensionality Reduction), 이상치 탐지(Anomaly detection)와 특이치 탐지(novelty detection), 연관 규칙 학습(Association Rule Learning)

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

강화학습

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 훈련

10. 외부 메모리 학습이 무엇인가요?

훈련 데이터의 양이 메모리에 한 번에 담을 수 없을 정도로 커서, 데이터를 미니배치로 나누어 훈련시키는 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 훈련 과정에서 학습되는 값이고 하이퍼파라미터는 학습 알고리즘을 제어하는 설정값으로 훈련 전에 미리 설정한다.

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

효용 함수, 비용 함수. 학습된 모델을 통해 예측을 만든다.

14. 머신러닝의 주요 도전 과제는 무엇인가요?

충분하지 않은 데이터, 낮은 품질의 데이터, 대표성 없는 훈련 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 데이터에 대한 과대적합. 모델 복잡성 줄이기, 훈련 데이터 추가, 데이터 정규화

16. 테스트 세트가 무엇이고 왜 사용해야 하나요?

모델의 최종 성능을 평가하기 위해 사용된다. 훈련 데이터로는 모델을 평가할 수 없기 때문에 독립된 데이터가 필요하다.

17. 검증 세트의 목적은 무엇인가요?

하이퍼파라미터를 튜닝하기 위해 사용된다.

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

모델이 테스트 세트에 과대적합되어 일반화 성능을 평가하기 어렵다.