

## 1주차 과제

1. 머신러닝을 어떻게 정의할 수 있나요? – 데이터의 학습을 통해 패턴을 찾고, 그것을 기반으로 새로운 데이터에 대한 답이나 예측을 도출해내는 것
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 환경, 복잡한 문제와 대량의 데이터에서 통찰 얻기
3. 레이블된 훈련 세트란 무엇인가요? – 답이 정해진 데이터들
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 걸아가게 하려면 어떤 종류의 머신러닝 알고리즘을 사용할 수 있나요? – 강화 학습 (시행착오를 통해 보상을 최대화하는 행동을 배움)
7. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요? – 군집 (비지도 학습)
8. 스팸 감지의 문제는 지도 학습과 비지도 학습 중 어떤 문제로 볼 수 있나요? – 지도 학습 (분류, 이미 정답인 레이블이 있는 데이터를 학습하기 때문에)
9. 온라인 학습 시스템이 무엇인가요? – 적은 양의 데이터를 사용해 점진적으로 계속 훈련
10. 외부 메모리 학습이 무엇인가요? – 데이터가 너무 클 때 RAM이 아닌 저장장치에서 필요한 부분만 불러와서 학습하는 방식
11. 예측을 하기 위해 유사도 측정에 의존하는 학습 알고리즘은 무엇인가요? – 사례 기반 학습 알고리즘
12. 모델 파라미터와 학습 알고리즘의 하이퍼파라미터 사이에는 어떤 차이가 있나요? – 모델 파라미터는 학습 과정에서 조정됨, 하이퍼파라미터는 사용자가 직접 설정하는 값이고 학습 도중 변하지 않음
13. 모델 기반 알고리즘이 찾는 것은 무엇인가요? 성공을 위해 이 알고리즘이 사용하는 가장 일반적인 전략은 무엇인가요? 예측은 어떻게 만드나요? 모델을 미리 지정한 후 훈련 세트를 사용하여 모델을 훈련시킴. 적절한 모델 선택 후 손실 함수 최소화, 훈련된 모델을 사용해 새로운 데이터에 대한 예측 실행
14. 머신러닝의 주요 도전 과제는 무엇인가요? – 훈련 데이터 과적합, 훈련 데이터 부족, 대표성 없는 데이터, 낮은 품질의 데이터, 관련이 없는 특성
15. 모델이 훈련 데이터에서의 성능은 좋지만 새로운 샘플에서의 일반화 성능이 나쁘다면 어떤 문제가 있는 건가요? 가능한 해결책 세 가지는 무엇인가요? – 훈련 데이터 과적합, 단순한 모델 사용, 데이터를 더 많이 모으기, 정규화 사용
16. 테스트 세트가 무엇이고 왜 사용해야 하나요? – 모델이 잘 작동하는지 확인하기

위해 학습에 사용하지 않은 데이터

17. 검증 세트의 목적은 무엇인가요? – 모델이 잘 작동하는 지 확인하고 모델을 조정하는 것, 최종 성능 평가에 사용되는 테스트 세트와 다른 것이다

18. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요? – 모델이 테스트 세트에 과적합될 위험이 있음.