

1주차 과제

- 머신러닝을 어떻게 정의할 수 있나요? *머신러닝 프로그래밍 없이 데이터로부터 패턴을 학습하고 소고 성능을 향상시키는 기술.*
- 머신러닝이 도움을 줄 수 있는 문제 유형 네 가지를 말해보세요. *분류, 회귀, 군집, 이상 탐지*
- 레이블된 훈련 세트란 무엇인가요? *각 데이터 샘플에 정답이 포함된 훈련 데이터*
- 가장 널리 사용되는 지도 학습 작업 두 가지는 무엇인가요? *분류, 회귀*
- 보편적인 비지도 학습 작업 네 가지는 무엇인가요? *군집, 연관 분석, 차원 축소, 이상 탐지 & 특이점 탐지, 연관 규칙 학습*
- 사전 정보가 없는 여러 지형에서 로봇을 걸아가게 하려면 어떤 종류의 머신러닝 알고리즘을 사용할 수 있나요? *강화 학습*
- 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요? *군집*
- 스팸 감지의 문제는 지도 학습과 비지도 학습 중 어떤 문제로 볼 수 있나요? *지도 학습*
- 온라인 학습 시스템이 무엇인가요? *작은 양의 데이터를 이용하여 점진적으로 학습하는 것*
- 외부 메모리 학습이 무엇인가요? *대용량의 데이터를 배치 단위로 학습하는 방식.*
- 예측을 하기 위해 유사도 측정에 의존하는 학습 알고리즘은 무엇인가요? *서지 기반 학습*
- 모델 파라미터와 학습 알고리즘의 하이퍼파라미터 사이에는 어떤 차이가 있나요? *↓
자율로 학습되는 것 ↓
사람이 설정해야 하는 것*
- 모델 기반 알고리즘이 찾는 것은 무엇인가요? 성공을 위해 이 알고리즘이 사용하는 가장 일반적인 전략은 무엇인가요? 예측은 어떻게 만드나요? *데이터를 가장 잘 설명하는 수학적 모델 → 입력 데이터를 모델에 넣고, 학습한 패턴으로 미래값 예측
비용 함수 최소화 최소화하지 않은 양의 훈련 데이터, 대용량 없는 훈련 데이터.*
- 머신러닝의 주요 도전 과제는 무엇인가요? *불완전한 데이터, 라벨이 없는 특징, 훈련 데이터 오버피팅 & 언더피팅*
- 모델이 훈련 데이터에서의 성능은 좋지만 새로운 샘플에서의 일반화 성능이 나쁘다면 어떤 문제가 있는 건가요? 가능한 해결책 세 가지는 무엇인가요? *오버피팅 ① 더 많은 훈련 데이터 사용 ② 모델의 복잡도 ↓ ③ 정규화 적용*
- 테스트 세트가 무엇이고 왜 사용해야 하나요? *학습하지 않은 데이터로 구성된 데이터셋. 모델의 사전 성능을 평가하는 데 사용된다.*
- 검증 세트의 목적은 무엇인가요? *최적의 모델 찾기*
- 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요? *테스트 세트에 대한 성능이 높아진다.*