1. 머신러닝을 어떻게 정의할 수 있나요 ?

- 머신러닝은 데이터로부터 학습할 수 있는 시스템을 만드는 것.

- 학습이란 어떤 작업에서 주어진 성능 지표가 더 나아지는 것을 의미.

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

* 음식메뉴 추천,
* 롤 아이템 추천
* 지진 예측
* 선거 당선율 예측

1. 레이블된 훈련세트란 무엇인가요?

* 훈련 데이터에 원하는 답이 포함되어있는 훈련 세트.

1. 가장 널리 사용되는 지도 학습 작업 두 가지는 무엇인가요?

* 분류, 특성을 이용해 타깃 수치를 예측하는 것 = 회귀!

1. 보편적인 비지도 학습 작업 네 가지는 무엇인가요?

* 군집, 시각화, 차원 축소, 연관 규칙 학습.

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

* 강화학습.

1. 고객을 여러 그룹으로 분할하려면 어떤 알고리즘을 사용해야 하나요?

* 만약 그룹을 어떻게 정의할지 모른다면 비슷한 고객끼리 군집으로 나누기 위해 군집 알고리즘을 사용할 수 있다.
* 그러나 어떤 그룹이 있어야 할지 안다면 분류 알고리즘에 각 그룹에 대한 샘플을 주입합니다

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

* 지도 학습

1. 온라인 학습 시스템이 무엇인가요?

* 데이터를 순적으로 한 개씩 또는 미니배치라 부르는 작은 묶음 단위로 주입하여 시스템을 훈련시킨다. 매우 많은 양의 데이터를 훈련 시킬 수 있다.

1. 외부 메모리 학습이 무엇인가요?

* 컴퓨터 한 대의 메인 메모리에 들어갈 수 없는 아주 큰 데이터셋을 학습.

1. 예측을 하기 위해 유사도 측정에 의존하는 학습 알고리즘은 무엇인가요?

* 사례 기반 학습. 훈련 데이터를 기억하는 학습. 새로운 샘플이 주어지면 유사도 측정을 사용해 학습된 샘플 중에서 가장 비슷한 것을 찾아 예측으로 사용합니다.

1. 모델 파라미터와 학습 알고리즘의 하이퍼파라미터 사이에는 어떤 차이가 있나요?

* 모델 파라미터는 모델에 사용되는 파라미터이다
* 하이퍼 파라미터는 학습 알고리즘의 파라미터이다.

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

* 샘플들의 모델을 만들어 예측에 사용하는 것. 모델이 얼마나 좋은지 측정하는 호용함수, 얼마나 나쁜지 측정하는 비용 함수. 비용함수를 최소화 하는 모델 파라미터를 찾는다.
* 최적의 파라미터 찾는거.
* 파라미터를 구해 모델을 구현하고 그 모델에 값을 대입하여 예측한다.

1. 머신러닝의 주요 도전 과제는 무엇인가요?

* 나쁜 알고리즘, 나쁜 데이터

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

* 과대적합 되어있다.
* 파라미터 수가 적은 모델을 선택하거나, 훈련 데이터에 있는 특성 수를 줄이거나, 모델에 제약을 가하여 단순화 시킨다.
* 훈련 데이터를 많이 모은다.
* 훈련 데이터의 잡음을 줄입니다(예를들면 오류 데이터 수정과 이상치 제거)

1. 테스트 세트가 무엇이고 왜 사용해야 하나요?

* 테스트 세트는 실전에 배치되기 전에 모델이 새로운 샘플에 대해 만들 일반화 오차를 추정하기 위해 사용한다.

1. 검증 세트의 목적은 무엇인가요?

* 검증 세트는 모델을 비교하는 데 사용됩니다. 이를 사용해 가장 좋은 모델을 고르고 하이퍼파라미터를 튜닝합니다.

1. 훈련-개발 세트가 무엇인가요? 언제 필요하고 어떻게 사용해야 하나요?

* 훈련 데이터의 일부를 떼어내어 또다른 세트를 만드는 것입니다.
* 모델의 성능의 부족함에 있어서 원인을 찾을 수 없을 때
* 모델을 훈련한 다음 훈련-개발세트에서 평가한다.

1. 테스트 세트를 사용해 하이퍼파라미터를 튜닝하면 어떤 문제가 생기나요?

* 일반화 오차를 테스트 세트에서 여러 번 측정했으므로 모델과 하이퍼파라미터가 테스트 세트에 최적화된 모델을 만들었기 때문에 실제 서비스에서 예기치 못한 오차가 크게 발생할 수 도 잇다