

1. 데이터를 통해 컴퓨터가 학습하여 문제를 풀 수 있도록 하는 방법
2. 예외처리가 많은 문제, 전통적인 방식으로 해결이 어려운 문제, 새로운 데이터에 대한 적응이 필요할 때, 복잡한 문제와 대량의 데이터에서 통찰이 필요할 때
3. 특징(레이블)이 정해진 데이터들로, 지도 학습에서 사용된다. 컴퓨터한테 알려줄 답이 정해져 있는 데이터.
4. 분류, 회귀 이렇게 두가지이다. 분류는 말 그대로 데이터를 분류하는 작업이고, 회귀는 결과값의 수치를 예측하는 작업이다.
5. 일단 비지도 학습은 레이블이 없는, 답이 정해져 있지 않은 데이터를 통해 학습이라 한다. 군집화와 시각화, 차원 축소, 연관 규칙 학습이 있다.
6. 강화 학습이 좋을 것 같다. 두 발로 걷는 로봇이라면, 한 발을 내딛었을 때, 넘어지거나 임계치 이상으로 균형이 크게 흔들린다면 벌점, 임계치 이하로 균형이 흔들린다면 보상을 주면 될 것 같다.
7. 고객 한 명 마다 레이블이 되어있다면 지도 학습에서 분류작업을 쓸 것 같다. 레이블이 없다면 군집화를 쓸 것 같다.
8. 비지도 학습에 가까울 것 같다. 스팸 메일도 계속해서 발전하기에 새로운 패턴을 발견할 필요가 있어 보인다. 즉 레이블이 안된 데이터로 학습해야 할 것 같다.
9. 적은 양의 데이터를 사용해 점진적으로 훈련시키는 학습이다. 단점으로 나쁜 데이터가 주입되는 경우 성능이 떨어진다는 것과 지속적인 모니터링을 요구하는 것이 있다.

10. 데이터의 양이 지나치게 커서 미니배치로 나눈 후 일부 데이터를 학습시키는 방법이다.

11. 사례 기반 학습이다.

12. 모델 파라미터는 학습을 통해 자동으로 조정되는 값이다. 하이퍼 파라미터는 학습 전 사용자가 설정해야 하는 값이다.

13. 훈련된 모델을 사용해 새로운 데이터에 대한 예측을 한다. 데이터 분석 -> 모델 선택 -> 훈련 데이터로 훈련시키기 -> 새로운 데이터에 모델을 적용해 예측 -> 기도하기

14. 간단한 문제라도 대량의 데이터를 필요로 한다. 대표성이 없거나 품질이 낮은 데이터밖에 없다면 성능에 문제가 생긴다. 풀려는 문제와 관련이 있는 레이블이 필요할 수 있다. 과대/과소 적합이 나타날 수 있다.

15. 과대적합으로 보인다. 데이터의 양을 늘리거나 이상치를 제거한다. 불필요한 변수를 줄인다.

16. 모델을 테스트할 때 쓰인다. 전체 데이터셋에서 20%정도이다. 성능을 평가하기 위해 사용한다.

17. 교차검증을 통해 모델을 평가할 수 있다.

18. 과대적합이 발생할 수 있다. 다른 데이터에서는 성능이 저하되겠다.