

20.06.15 인공지능 머신러닝 딥러닝

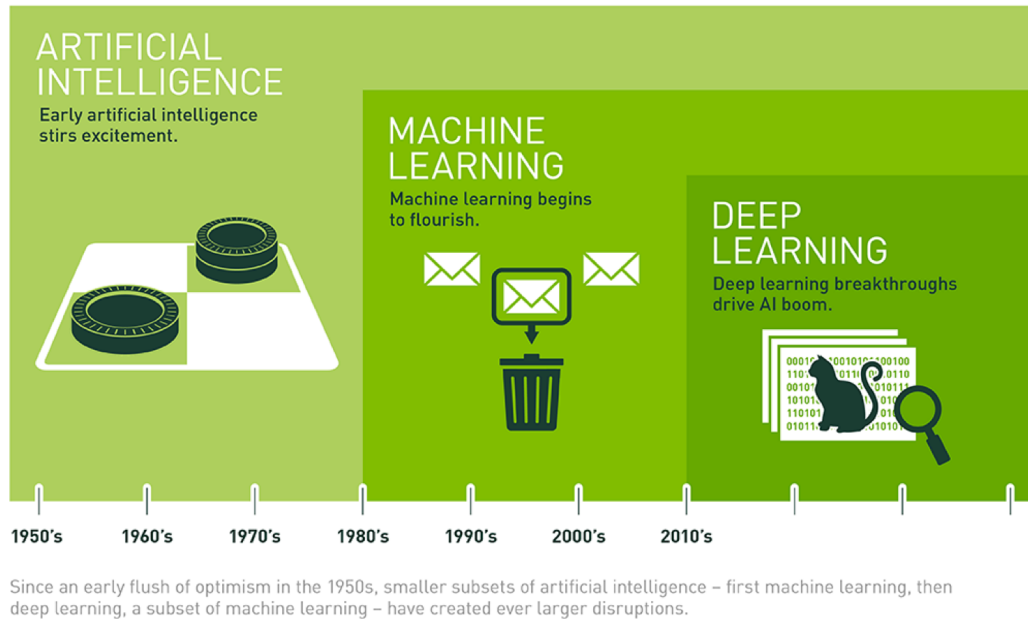
노트북: [TIL-MY]

만든 날짜: 2020-06-15 오전 10:56

수정한 날짜: 2020-08-07 오전 8:13

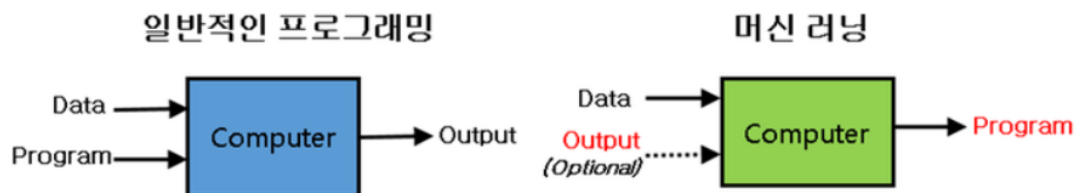
작성자: 황인범

URL: <http://solarisailab.com/archives/1785>



인공지능은 자체 규칙 시스템을 구축

이는 AI를 통해 컴퓨터가 사람에게 전적으로 의존했던 작업을 스스로 해결할 수 있음



머신러닝

대량의 데이터와 알고리즘을 통해 컴퓨터 그 자체를 학습시켜 작업수행방법을 익히는 것

분류에 사용할 데이터를 수동으로 제공해야 한다

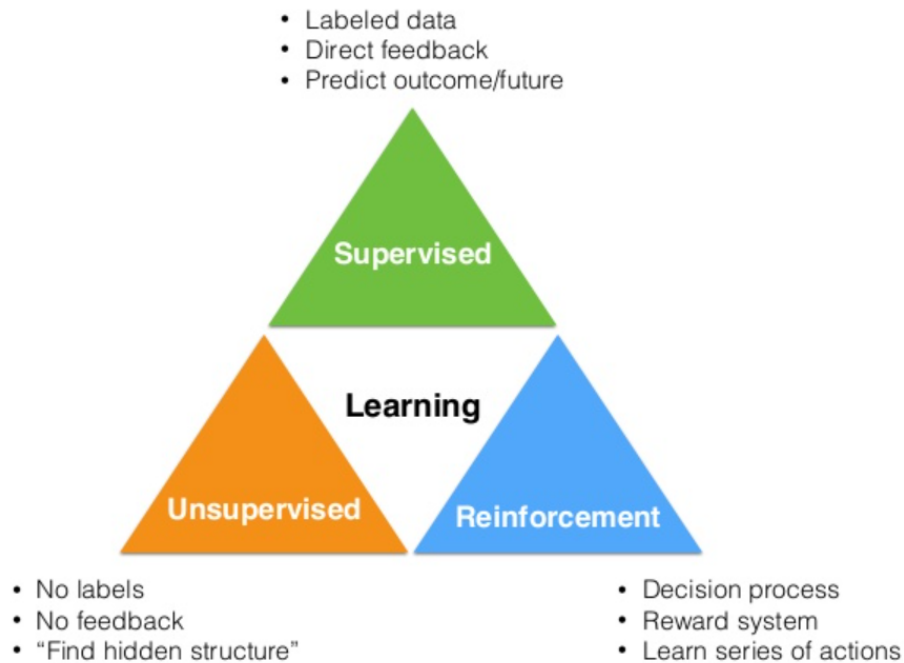
수 많은 데이터를 통해 주어진 입력에 대해 올바른 출력을 하게 하는 것

인공지능의 주요 아이디어의 일부를 가져와서 인간의 노에서 일어나는 의사결정과정을 모방하여 만든 신경망으로 현실 세계의 문제를 푼다.

기계가 학습하기 위해 주어진 데이터에서 특징을 추출하는과정에 여전히 사람이 개입

사람이 생각한 특징을 훈련

머신러닝(Machine Learning) 알고리즘 분류 - 지도학습(Supervised learning) / 비지도 학습(Unsupervised Learning) / 강화학습(Reinforcement Learning)



1.지도학습(supervised Learning) - 데이터에 대한 정답(레이블)을 알려주며 학습시키는 것 이것에는 크게 두가지로 분류 분류(classification) 와 회귀(regression)

1-1)분류(classification) - 이진분류 (예 또는 아니오) = discrete value(이산값) / 다중분류 어떤 데이터에 대해 여러 값 중 하나로 분류할 수 있는것 ex) 이 동물은 모야? 고양이 또는 사자

1-2)회귀(regression) - 어떤 데이터들의 특징을 토대로 값을 예측하는 것 = continuous value(연속값) (결과값은 실수가 가능) ex) 어디 동네에 평수면 집값이 얼마야? 어디면 얼마 어디면 얼마

단점:정답 집합을 확보하는 것

2.비지도 학습(unsupervised Learning) - 데이터에 대한 정답(레이블)을 따로 알려주지않고 비슷한 데이터들을 군집화 하는 것 .

ex) 고양이 병아리 기린 호랑이가 있을경우 비슷한 단위로 군집화한다. 이걸 클러스터링 (Clustering) 알고리즘이 있다 다리가 4개인 고양이 기린 호랑이로 / 목이 긴 기린은 다른 분류로 그런식으로 분류

이걸 feature(특성)에 따라 군집화 한다.

3.강화 학습(Reinforcement Learning) - 주어진 환경에서 현재 상태(state)에 대해 어떤 행동을 취하고 이로부터 어떤 보상(reward)을 얻으면서 학습을 진행한다. 동적인 상태에서 데이터를 수집하는 과정까지 포함되어 있는 알고리즘 단 행동 목록등은 사전에 정의가 되어야 한다.

최근에는 계산하는데신 신경망을 통해 근사값을 구함으로써 복잡한 문제를 해결할 수 있게 되었다.

지도 학습과 비지도 학습은 정적인 환경에서 이뤄지지만 강화학습은 동적인 상황에서 이뤄진다.

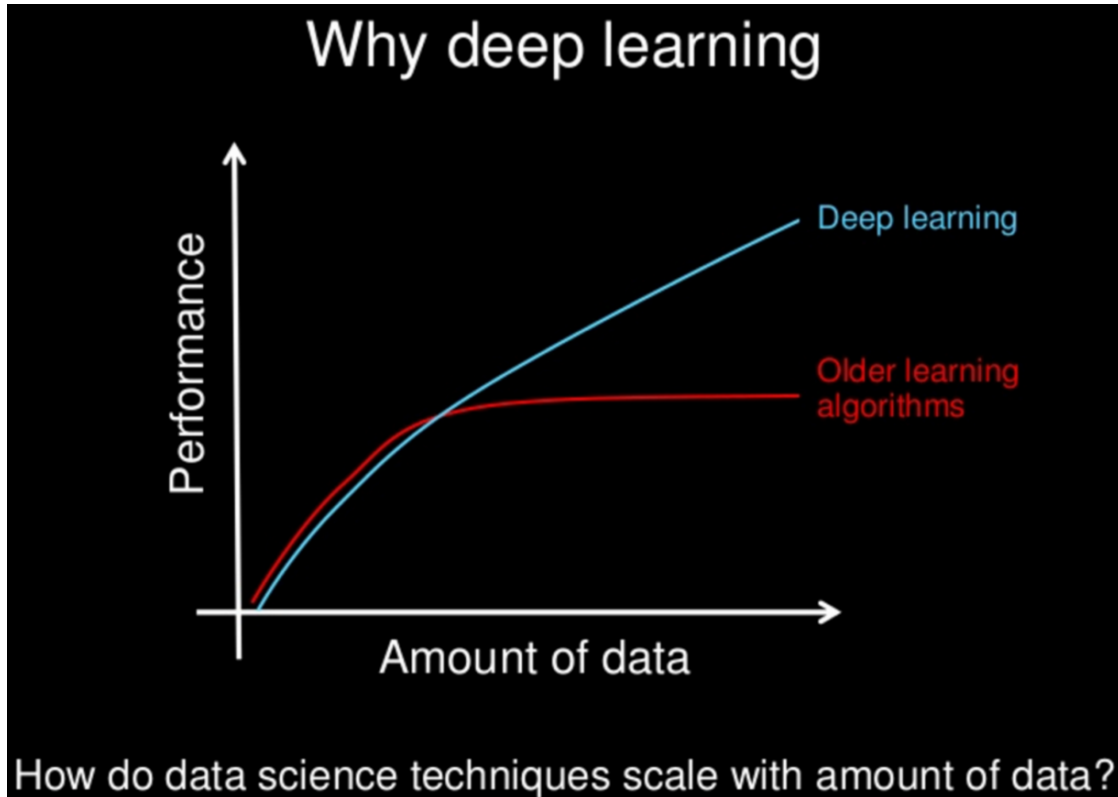
딥러닝

분류에 사용할 데이터를 스스로 학습할수 있다

머신 러닝의 도구와 기술의 일부에 더욱 초점을 맞춘다.
딥 러닝은 주어진 데이터를 그대로 입력데이터로 활용
데이터 자체에서 중요한 특징을 기계 스스로 학습

차이점

1. 데이터 의존도 - 데이터가 많을 수록 딥러닝의 성능은 올라간다. 머신러닝의 경우에는 일정량의 데이터에 수작업이 들어가기 때문에 적을 때는 이게 더 성능이 좋다.



2. 하드웨어 의존도 - 아까말했듯이 딥러닝은 컴퓨터를 학습시키는데 있어서 GPU가 사용 된다. 딥러닝 알고리즘은 본질적으로 많은 양의 행렬 곱셈을 수행한다.

3. Feature engineering - 데이터 복잡성을 줄이고 학습 알고리즘에서 패턴을 보다 잘보이게 하는 과정

머신러닝의 경우 대부분의 적용된 변수는 전문가가 식별한 다음 정보 영역 및 데이터 유형별로 손으로 코딩해야 한다.

딥러닝은 high-level features를 학습한다. 고차원적인 변수! 따라서 딥러닝은 모든 과제에서 새로운 변수 추출이라는 작업을 줄여준다.

4. 문제 해결 접근법 - #1. 머신러닝 - 문제를 여러 개의 파트로 쪼갬 후, 각각에 대한 답을 구하고 그 결과를 합치는 방법을 추천

#2. 딥러닝 - end-to-end 방식

5. 실행시간 - #1. 알고리즘 훈련일 경우 : 머신러닝(수초 ~ 수시간) > 딥러닝(오래걸림)

#2. test 일 경우 : 머신러닝 < 딥러닝

근데 이게 무조건 그런건 아니고 우세를 보이는 편이다.

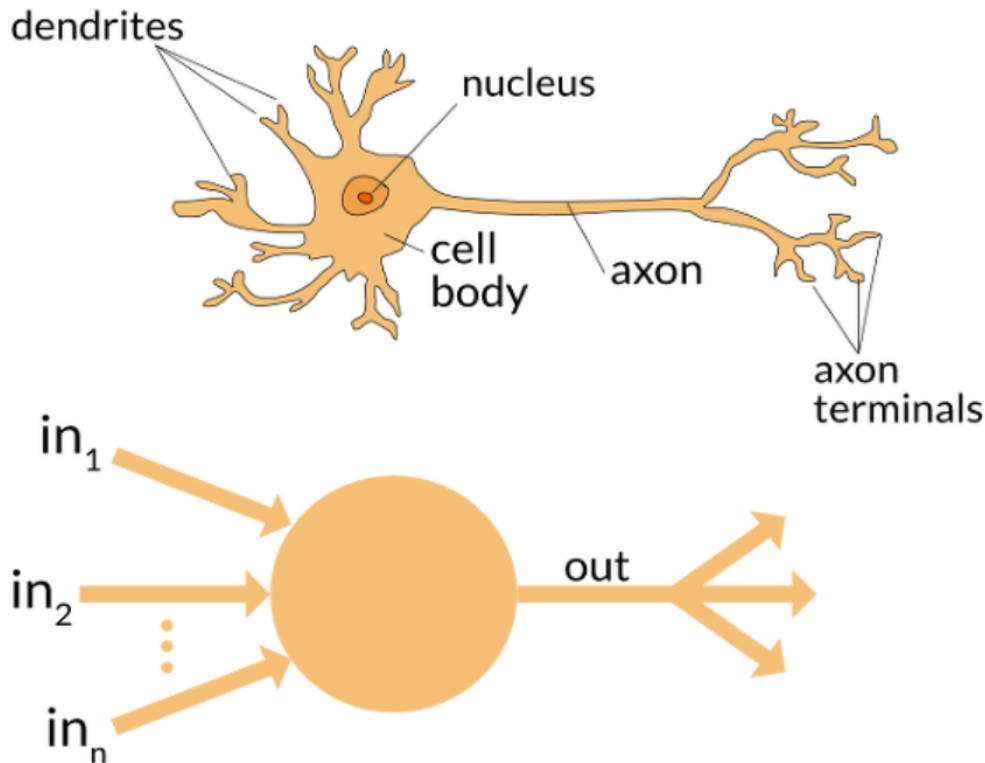
6. 해석력 - #딥러닝 - 딥러닝은 수학적으로 딥러닝의 어느 노드가 활성화되었는지를 알 수 있지만, 우리는 거기에 어떤 뉴런이 만들어 졌는지, 그리고

이들 뉴런레이어가 전체적으로 무엇을 하고 있는지 알지 못합니다. 그래서 결과를 해석하지 못한다.

#머신러닝 - 의사 결정 트리와 같은 기계 학습 알고리즘은 왜 선택한 알고리즘으로 선택했는지에 대한 명확한 규칙을 제공하므로 배후의 추론을 해석하기가 특히 쉽다. 따라서 결정 트리 및 선형 / 로지스틱 회귀와 같은 알고리즘은 주로 업계에서 해석 가능성을 위해 사용된다.

신경망(Neural network)

신경망이란 생물학에서 영감을 얻은 것으로 관측한 데이터로부터 컴퓨터가 학습을 하게 하는 프로그램 패러다임



각 뉴런은 다른 뉴런으로부터 입력을 받는다 -> 뉴런에 있는 각 입력들의 효과는 시냅틱 가중치에 의해 조절된다(플러스 or 마이너스) -> 상황에 맞게 조정되며 전체 망은 유용한 계산을 수행하도록 학습한다.

외부에서 입력되는 신호를 뇌에서 받아 각 가중치에 의해 계산을 한 후 최종적으로 특정한 행동을 하게 된다.

각 뉴런은 전체 입력값들을 모두다 결합하는 총합함수가 있다. 각각의 연결과 각각의 단위 자체에 임계함수 혹은 제한함수가 있어서 다른 뉴런으로 전파하기 전에 보낼 신호는 반드시 한계점을 초과해야 한다.

