

Machine Learning and Deep learning

Soyoung Park

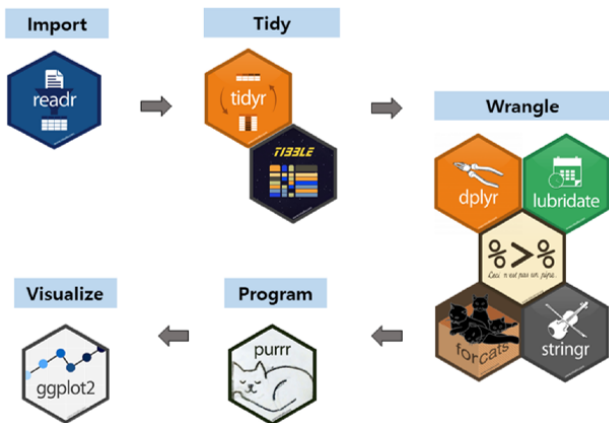
Pusan National University
Department of Statistics

soyoung@pusan.ac.kr

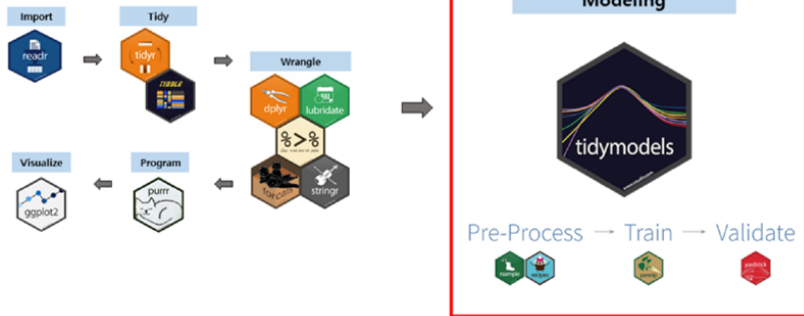
Data Science



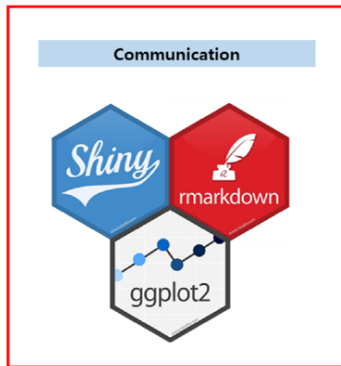
Learning data science steps I



Learning data science steps II

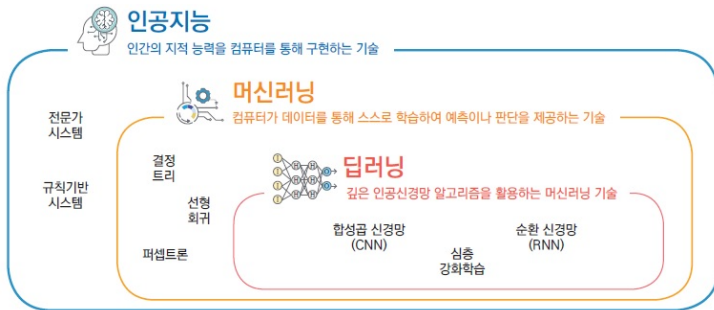


Learning data science steps III



통계적 머신러닝과 딥러닝

인공지능, 머신러닝, 딥러닝



¹<https://samstory.coolschool.co.kr/zone/story/modi/streams/76601>

인공지능, 머신러닝, 딥러닝

인공지능

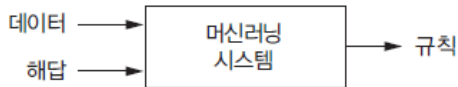
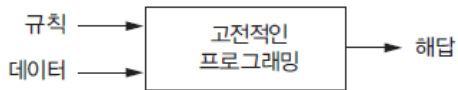
- 1950년대 출현 → 컴퓨터를 '생각'하게 만들수 있을지에 대한 노력으로 시작
-
-

인공지능, 머신러닝, 딥러닝

머신러닝을 시작하게한 질문들:

- 수행방법을 아는 모든 일을 넘어서,
- 프로그래머가 직접 데이터 처리 규칙을 작성하는 대신,

인공지능, 머신러닝, 딥러닝



보기 1.2 머신러닝: 새로운 프로그래밍 패러다임

인공지능, 머신러닝, 딥러닝

머신러닝

- 머신러닝 시스템은 명시적으로 프로그래밍되는것이 아니라
- 과업과 관련된 많은 examples를 접한 후, 이 속에서
- 예시:
- 고속 처리 하드웨어와 대규모 dataset의 활용이 가능해진 1990년대 이르러서야 인기를 끌기 시작

인공지능, 머신러닝, 딥러닝

Machine learning vs. Statistical learning

- 머신러닝은 베イズ분석과 같은 고전적인 통계분석에 효율적이지 못함
- 복잡한
- 머신러닝, 특히 딥러닝은
- 머신러닝 분야에서는 아이디어가 이론보다

인공지능, 머신러닝, 딥러닝

머신러닝 필요규칙

- 입력 데이터 점:
- 예상 출력 사례:
- 알고리즘의 성공적인 수행 여부를 측정하는 방법:

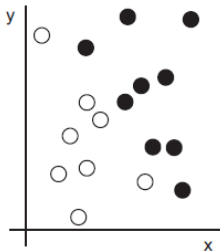
머신러닝

입력 데이터 점

- 머신러닝 모델은 입력 데이터를 의미있는 출력으로 변환
- 이는 입출력의 알려진 사례에 노출됨으로써
- 머신러닝과 딥러닝의 핵심과제는
- 즉, 예상하는 결과와 더 가깝게 접근할 수 있도록
- 표현(representation)이란 데이터를 대표(represent)하거나, 데이터를 부호화(encode)하기 위해 데이터를 색다르게 살펴보는 방법
- ex) 컬러 이미지는 RGB, HSV형식으로 부호화 가능

인공지능, 머신러닝, 딥러닝

점의 좌표 (x, y) 를 취해 그 점이 검은색인지, 흰색인지를 출력할 수 있는 알고리즘을 개발하려고 한다고 가정해보자.



보기 1.3 일부 표본 데이터

인공지능, 머신러닝, 딥러닝

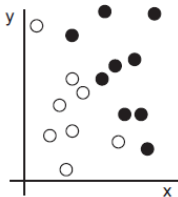
머신러닝 필요규칙

- 입력은
- 기대하는 결과는
- 알고리즘이 올바른 작업을 수행하는지 여부를 측정하는 방법은
예를들어, 올바르게 분류된 점의 백분율 일 수 있다.

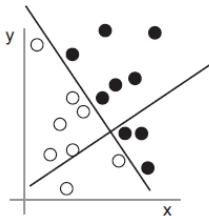
인공지능, 머신러닝, 딥러닝

흰색점과 검은색 점을 깔끔하게 구분하는 데이터의 새로운 표현 → 좌표 변환

1: 원자료

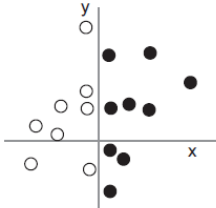


2: 좌표 변환



보기 1.4 좌표 변환

3: 더 나은 표현



점의 색에 관한 규칙:

인공지능, 머신러닝, 딥러닝

머신러닝

- 좌표변환의 방식을 체계적으로 찾아내고, 올바르게 분류된 점들의 백분율을 피드백으로 → 머신러닝
- 머신러닝에서 학습이란
- 모든 머신러닝 알고리즘은 데이터를 주어진 작업을 수행하기에 더 나은 표현으로 바꾸는 변환 방식을

인공지능, 머신러닝, 딥러닝

딥러닝(deep learning)

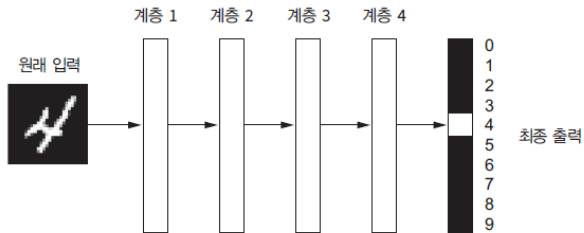
- 딥러닝은 머신러닝의 특정 하위 분야
- 표현을 더 의미있게 만들어가는
- 딥(deep)이라는 문구는 심층적인 이해수준을 언급하는것이 아님
- 딥(deep)은
- layered representations learning, hierarchical representations learning

인공지능, 머신러닝, 딥러닝

딥러닝과 통계학

- **통계적 머신러닝**은 통계학에서 전통적으로 다루는 모형과 기법을 바탕으로 한 데이터분석 기법
- **딥러닝**은 통계학의 응용영역을 인간의 인지능력 분야로 확장한 AI 기법
- 공통점:
- 딥러닝은
- 딥러닝은

인공지능, 머신러닝, 딥러닝



보기 1.5 숫자 분류를 위한 심층 신경망

인공지능, 머신러닝, 딥러닝

딥러닝(deep learning)

- 현대 딥러닝에서는 수십개에서 수백개에 이르는 연속적인 표현들이 포함되며, training data에 노출되면 모든것이 자동으로 학습됨
- 딥러닝에서 계층적 표현은 neural networks(신경망) 모델을 통해 학습됨
- 신경망이라는 생물학적 단어를 차용하였지만, 딥러닝이 뇌모델은 아님

인공지능, 머신러닝, 딥러닝

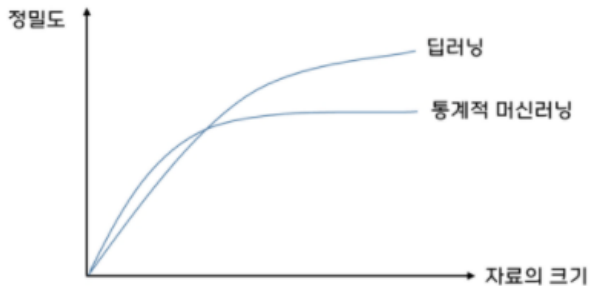
딥러닝의 기능적인 측면에서의 이해

-
-
-

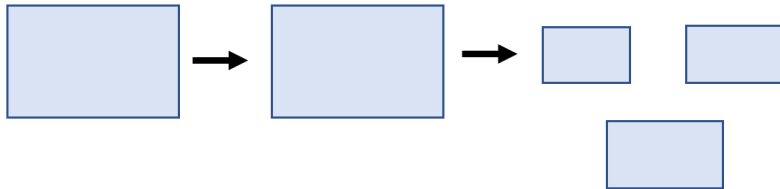
통계적 머신러닝과 딥러닝

구분	통계적 머신러닝	딥러닝
데이터의 크기		
분석자료 형태		
강점을 갖는 자료		
특성변수		
특성변수의 정규화 및 표준화		
모형		
최적화		
해석여부		
하드웨어		
실행요구시간		

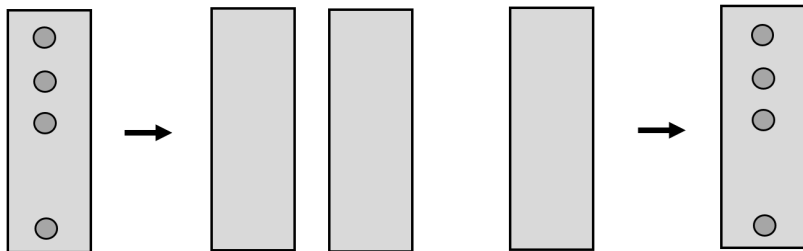
통계적 머신러닝과 딥러닝



머신러닝의 구조

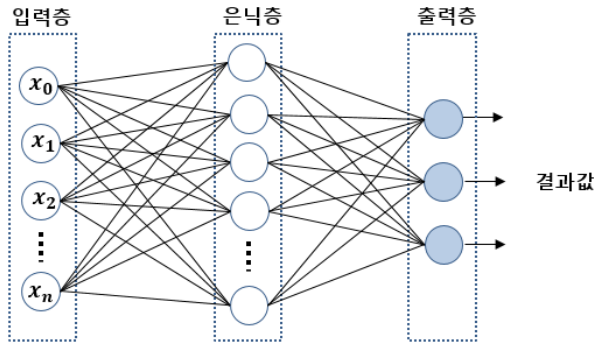


딥러닝의 구조



딥러닝 작동 방식

- 딥러닝은 특성변수(x)의 선형결합 및 비선형변환을 통해 목적변수(y)를 예측하는 모형
- 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)



¹<https://m.blog.naver.com/samsjang/221030487369>

인공지능, 머신러닝, 딥러닝

딥러닝의 업적

- 이미지를 사람이 분류하는 정도에 가깝게 분류
- 음성을 사람이 알아듣는 수준에 가깝게 인식
- 필기체를 사람이 인식하는 수준에 가깝게 인식
- 더 나은 기계 번역
- 텍스트를 더 자연스러운 음성으로 변환
- 구글 나우, 아마존 알렉사와 같은 디지털 비서 .. 등

딥러닝으로 무엇까지 할 수 있는지는 지금도 연구 대상! → idea 싸움

Tools for the data science

