

딥러닝 소개

머신러닝/딥러닝

- 머신러닝/딥러닝은 데이터 분석 및 인공지능 관련 기술을 획기적으로 발전시켰음.
- 사회, 경제, 산업의 거의 모든 분야에서 머신러닝/딥러닝의 활용이 점차 강조되어 왔음.
- 대형 언어 모델(LLM, Large language model)인 GPT(Generative pre-trained transformer)의 출현으로 딥러닝에 대한 관심이 보다 깊어졌음.
- 머신러닝/딥러닝 기술이 이미 많이 대중화되었음.

강의 주제

- 딥러닝의 기본 개념과 함께 다양한 딥러닝 기법을 최대한 직관적으로 전달하는 일에 집중
- 텐서플로우TensorFlow 2와 케라스Keras 활용

주교재

- 프랑소와 솔레의 [Deep Learning with Python\(2판\)](#) 참고

주요 내용

- 딥러닝 개념 소개
- 딥러닝 활용
 - 컴퓨터비전: 이미지 분석 및 분할
 - 시계열 예측
 - 자연어 처리: 텍스트 분류
 - 생성 신경망 모델: 문장, 이미지 등등 생성

전제 사항

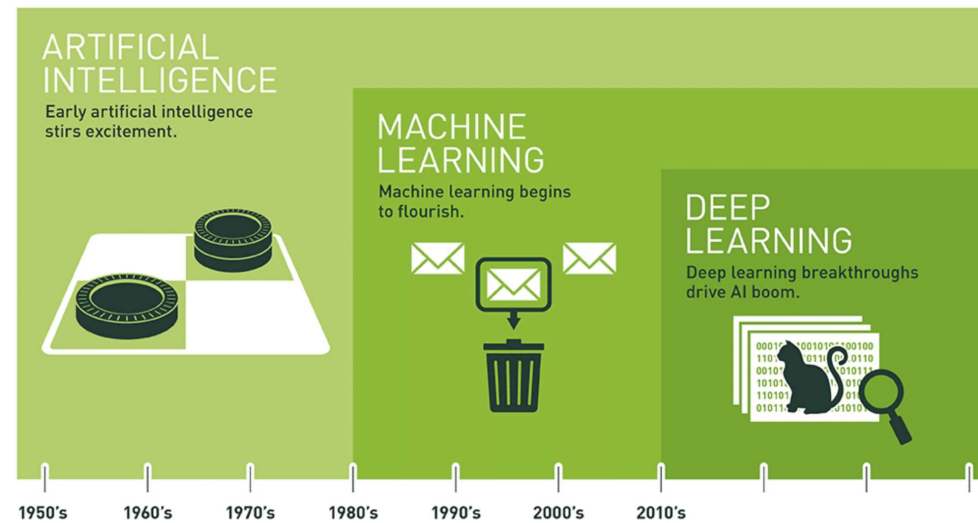
- 파이썬 왕기초
- 파이썬 데이터 분석
- 핸드즈온 머신러닝(3판)

인공지능, 머신러닝, 딥러닝

관계 1: 연구 분야 관점



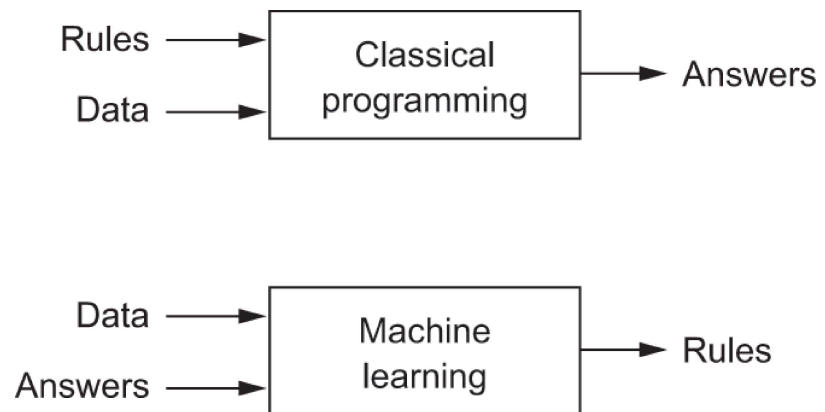
관계 2: 역사



인공지능

- 인공지능: 인간의 지적 활동을 모방하여 컴퓨터로 자동화하려는 시도. 머신러닝과 딥러닝을 포괄함.
- (1950년대) 컴퓨터가 생각할 수 있는가? 라는 질문에서 출발
- (1980년대까지) **학습**(러닝)이 아닌 모든 가능성을 논리적으로 전개하는 기법 활용
 - 서양장기(체스) 등에서 우수한 성능 발휘
 - 반면에 이미지 분류, 음성 인식, 자연어 번역 등 보다 복잡한 문제는 제대로 다루지 못함.
- (1990년대부터) 입력 데이터로부터 규칙을 스스로 찾아내도록 유도하는 머신러닝 기법이 유행하기 시작함.
 - 인공지능(AI) 분야의 주요 핵심 기법으로 자리잡음

전통적 프로그래밍 vs. 머신러닝



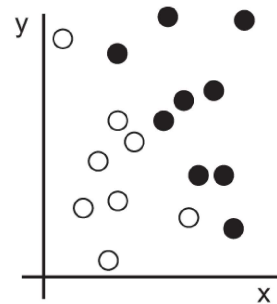
머신러닝 모델 학습의 필수 요소

- 입력 데이터셋(훈련셋)
- 타겟 데이터셋
- 모델 평가지표

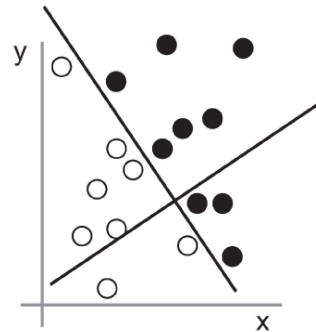
학습 규칙과 데이터 표현법

- 데이터 표현법 학습: 주어진 과제 해결에 가장 적절한 **데이터 표현법**을 모델 학습을 통해 알아냄.
- 예제: 좌표 변환

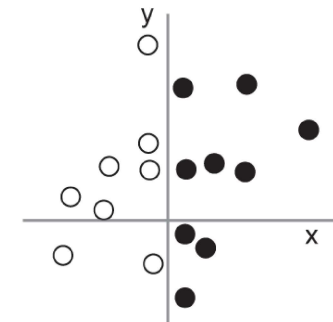
1: Raw data



2: Coordinate change

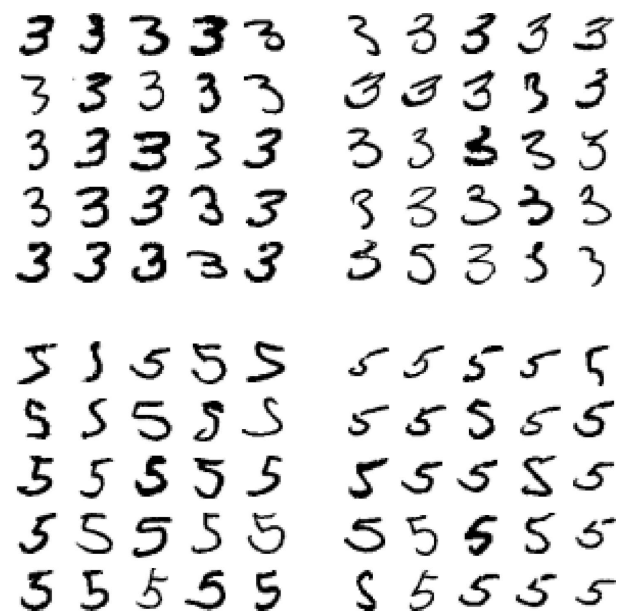


3: Better representation



데이터 변환 자동화

- 반면에 MNIST 손글씨의 경우는 좋은 표현법을 수동으로 찾는 일은 거의 불가능



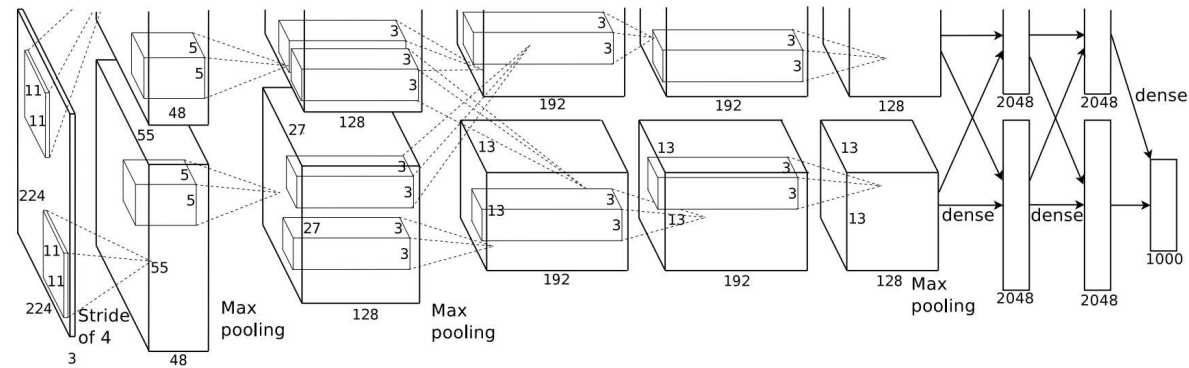
- 머신러닝 모델은 아래에 언급된 변환 등을 조합하여 문제 해결에 보다 도움이 되는 표현 변환 알고리즘을 찾아내려 시도함.
 - 회전
 - 이동
 - 사영projection
 - 잘라내기

가설 공간

- 주어진 문제의 해결에 가장 적절한 변환 알고리즘을 머신러닝 모델 스스로 알아내기는 기본적으로 불가능
- 대신에 데이터 표현법 변환 알고리즘을 어떻게 구현할 수 있는지 길안내 필요. 즉, 머신러닝 모델의 구성은 사람이 직접 지정
- 그러면 가능한 모든 알고리즘의 공간 내에서 최적의 데이터 변환 알고리즘을 데이터 학습을 통해 알아냄.
- 가설 공간: 머신러닝 모델이 변환 알고리즘 학습에 활용할 수 있는 알고리즘의 공간

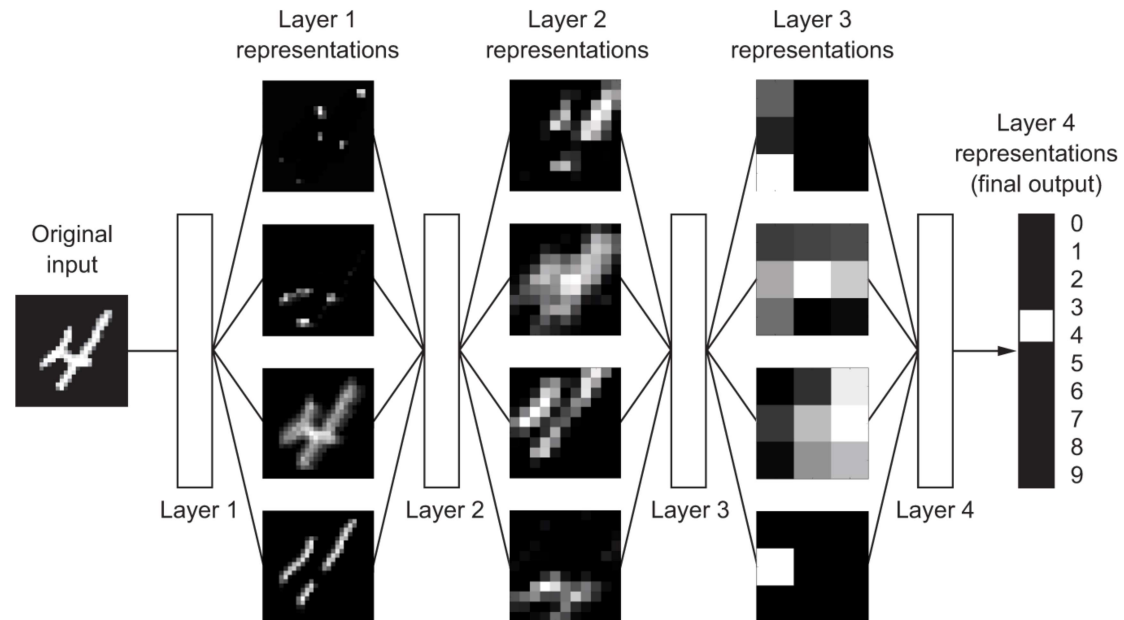
딥러닝 모델

- 딥러닝 모델은 세 개 이상의 층으로 구성된 **심층 신경망** deep neural network으로 구현
- ImageNet 분류 심층 신경망 모델 (2012년)



예제: 손글씨 숫자 인식

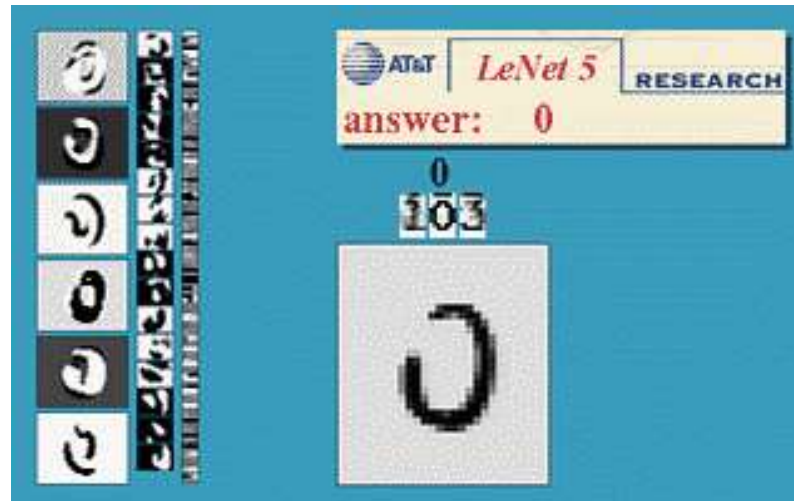
- 네 개의 층 사용
- 층마다 입력 데이터 변환
- 최종적으로 (사람은 이해할 수 없지만) 모델은 어떻게든 적절한 예측값을 계산할 수 있는 표현법으로 데이터 변환



머신러닝 역사

초창기 신경망

- 신경망의 기본 아이디어: 1950년대부터
- 최초의 성공적인 신경망 활용: 1989년 미국의 벨 연구소의 얀 르쿤^{Yann LeCun}이 **LeNet** 합성곱 신경망



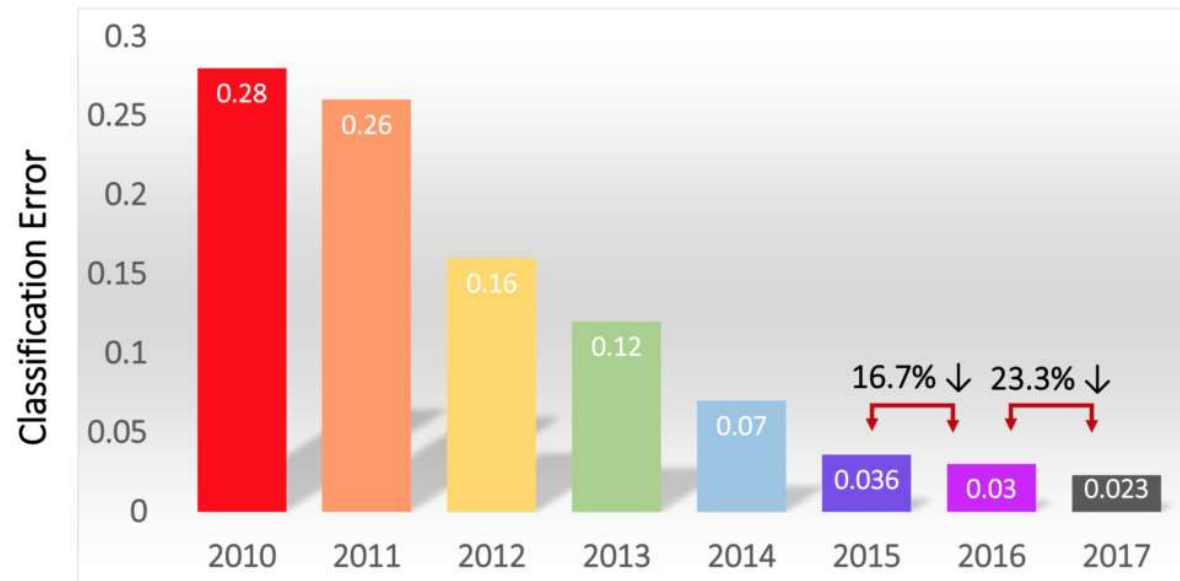
결정트리, 랜덤 포레스트, 그레이디언트 부스팅

- 2000년대: 결정트리_{decision tree}
- 2010년: 랜덤 포레스트_{random forest}
- 2014년: 그레이디언트 부스팅_{gradient boosting} 기법
- 현재까지도 딥러닝과 더불어 가장 많이 활용됨

딥러닝의 본격적 발전

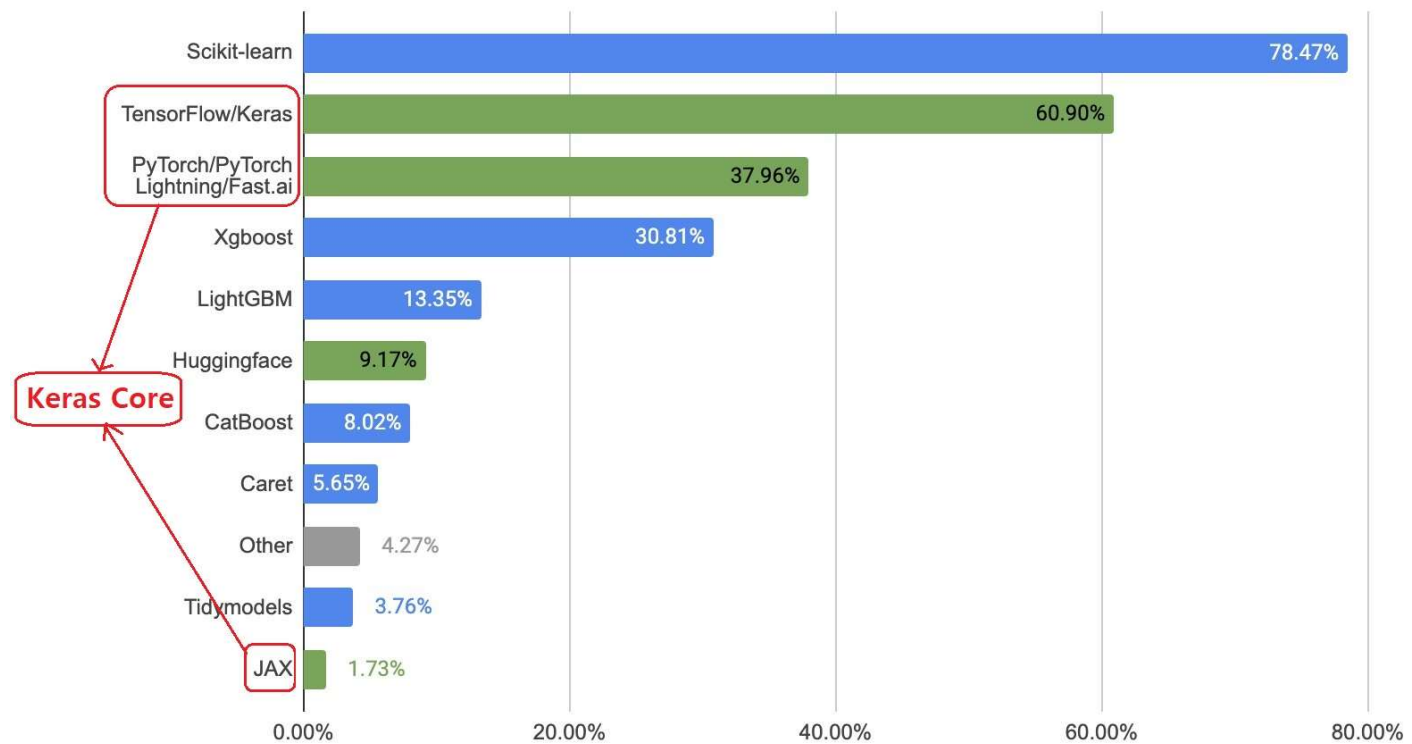
- 2011년: GPU를 활용한 딥러닝 모델 훈련이 시작
- 2012년: 이미지 분류 경진대회인 [이미지넷의 ILSVRC](#)의 2012년 대회에서 소개된 합성곱 신경망(CNN) 모델의 성능 매우 뛰어남

Classification Results (CLS)



최근 머신러닝 분야 동향

2022 Machine Learning & Data Science Survey by Kaggle: library usage (N=14,531)



딥러닝 발전 동력

- 하드웨어
- 데이터
- 알고리즘