4차산업혁명

- 인공지능 | 머신러닝 딥러닝

SSIDO 개발 운영진 | 심준식

INDEX

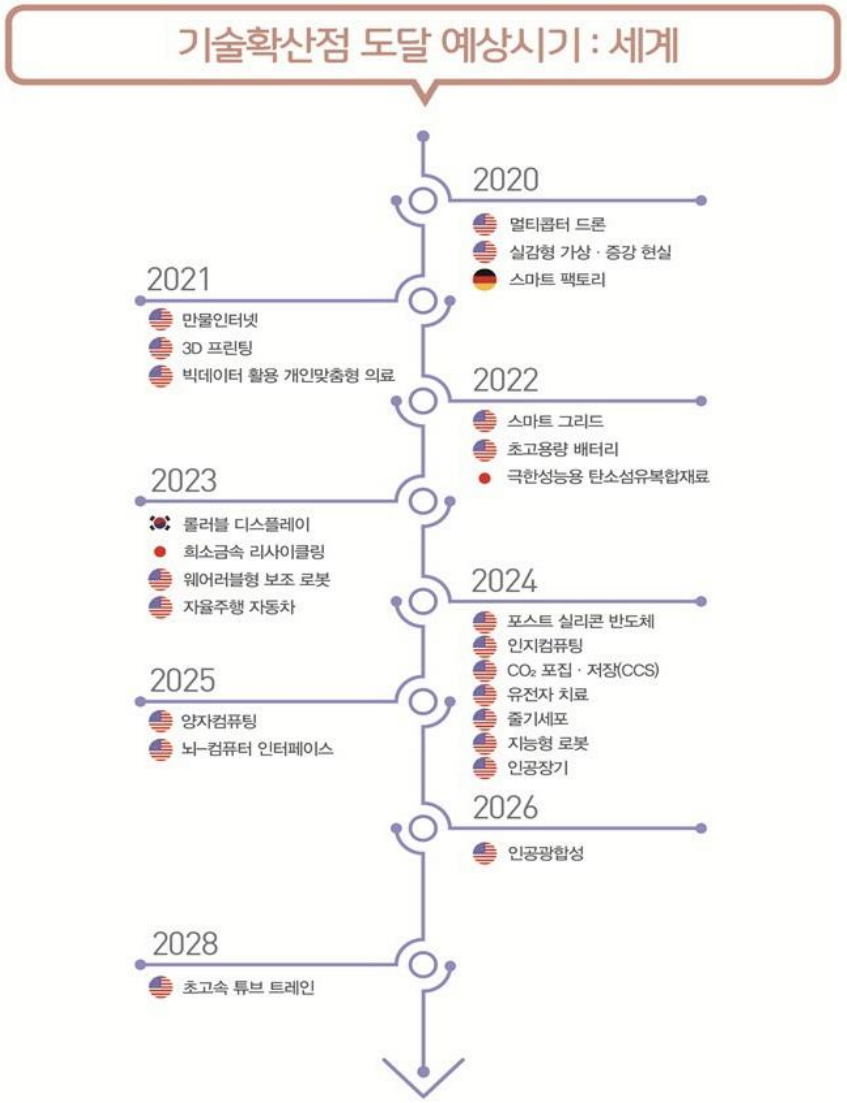
- 4차산업혁명
- 빅데이터와 인공지능
- 머신러닝
- 딥러닝

4차산업혁명

Fourth Industrial Revolution

01

02


기술확산점 도달 예상시기 : 세계
2020
멀티콥터 드론
실감형 가상 · 증강 현실
스마트 팩토리
2021
만물인터넷
3D 프린팅
빅데이터 활용 개인맞춤형 의료
2022
스마트 그리드
초고용량 배터리
극한성능용 탄소섬유복합재료
2023
롤러블 디스플레이
희소금속 리사이클링
웨어러블형 보조 로봇
자율주행 자동차
2024
포스트 실리콘 반도체
인지컴퓨팅
CO_2 포집 · 저장(CCS)
유전자 치료
줄기세포
지능형 로봇
인공장기
2025
양자컴퓨팅
뇌-컴퓨터 인터페이스
2026
인공광합성
2028
초고속 튜브 트레인

기술	기술확산점 정의	기술확산점 도달 예상시기 세계	국내
멀티콥터 드론	멀티콥터 드론의 운용 중 발생하는 사고율을 100만 비행시간당 2회 이하로 낮출 수 있는 안전운용 기술이 완성되는 시점	2020년	2024년
실감형 가상 · 증강 현실	게임 등 상호작용형 엔터테인먼트 시장에서 실감형 가상 · 증강 현실용 콘텐츠의 점유율이 11%가 되는 시점	2020년	2024년
스마트 팩토리	고도화된 스마트 팩토리의 비중이 전체 공장의 16%에 도달하는 시점	2020년	2025년
만물인터넷	만물인터넷 서비스의 가정 보급률이 11%가 되는 시점	2021년	2023년
3D 프린팅	3D 프린터의 일반 가정 보급률이 3%가 되는 시점	2021년	2024년
빅데이터 활용 개인맞춤형 의료	10만 명 이상의 개인별 의료정보가 국가적으로 통합되어 진료현장에 활용되는 시점	2021년	2025년
스마트 그리드	100만 명 이상의 광역 도시에 스마트 그리드 시스템을 구현한 시점	2022년	2024년
초고용량 배터리	1회 충전으로 800㎞를 주행할 수 있는 전기자동차의 상용모델이 국내에 출시되는 시점	2022년	2024년
극한성능용 탄소섬유복합재료	탄소섬유복합재료를 모든 외장재에 도입한 승용차의 첫 양산형 모델이 나오는 시점	2022년	2026년
롤러블 디스플레이	롤러블 컬러 디스플레이가 상용 모바일 제품에 최초로 적용되는 시점	2023년	2023년
희소금속 리사이클링	순환금속 리사이클링으로부터 공급되는 희소금속의 양이 희소금속 산업 수요의 8%가 되는 시점	2023년	2026년
웨어러블형 보조 로봇	하반신 마비 장애인의 보행을 보조하는 웨어러블 로봇의 렌탈 가격이 월 100만 원 이하가 되는 시점	2023년	2027년

기술	기술확산점 정의	기술확산점 도달 예상시기 세계	국내
자율주행 자동차	자율주행 자동차가 자동차 신차 판매의 12%를 점유하는 시점	2023년	2028년
포스트 실리콘 반도체	주요 반도체 제조 기업에서 포스트 실리콘 반도체를 초기 양산 시제품으로 생산하는 시점	2024년	2026년
인지컴퓨팅	인지컴퓨팅을 기반으로 사람을 이해하고 언어로 소통하는 맞춤형 개인비서 AI 서비스를 활용하는 인구의 비중이 스마트폰 사용 인구 중 16%에 도달하는 시점	2024년	2027년
CO_2 포집 · 저장 (CCS)	화력발전의 1%에 CCS가 적용되는 시점	2024년	2028년
유전자 치료	복합질환의 치료를 위한 2가지 이상의 유전자 치료제가 미국 FDA, 유럽 EMA, 일본 PMDA 등 허가 기관으로부터 의약품 범주의 시판 허가를 얻는 시점	2024년	2028년
줄기세포	특정난치병 10종 이상에 대하여 줄기세포를 활용한 치료 방법이 개발되어 임상치료에 적용되는 시점	2024년	2028년
지능형 로봇	네트워크 기반 지능형 로봇의 일반가정 보급률이 8%를 돌파하는 시점	2024년	2028년
인공장기	인체에 삽입되어 완전하게 독립적으로 기능하는 인공신장이 개발되어 인공신장 이식 건이 16%가 되는 시점	2024년	2029년
양자컴퓨팅	기상예측에 양자컴퓨팅을 처음으로 도입하는 시점	2025년	2031년
뇌-컴퓨터 인터페이스	사지마비 장애인의 16%가 뇌-컴퓨터 인터페이스 기술을 사용하는 시점	2025년	2032년
인공광합성	인공광합성 기술을 이용한 제품 생산이 기존 시장을 대체하는 비율이 3%가 되는 시점	2026년	2030년
초고속 튜브 트레인	시속 1,000㎞ 이상으로 운행하는 상용화된 초고속 튜브 트레인이 최초로 운행되는 시점	2028년	2033년

* 세계 기준으로 기술확산점 도달 예상시기가 빠른 순서대로 나열함

01

02

양자컴퓨팅

지능형 로봇

초고속
튜브 트레인

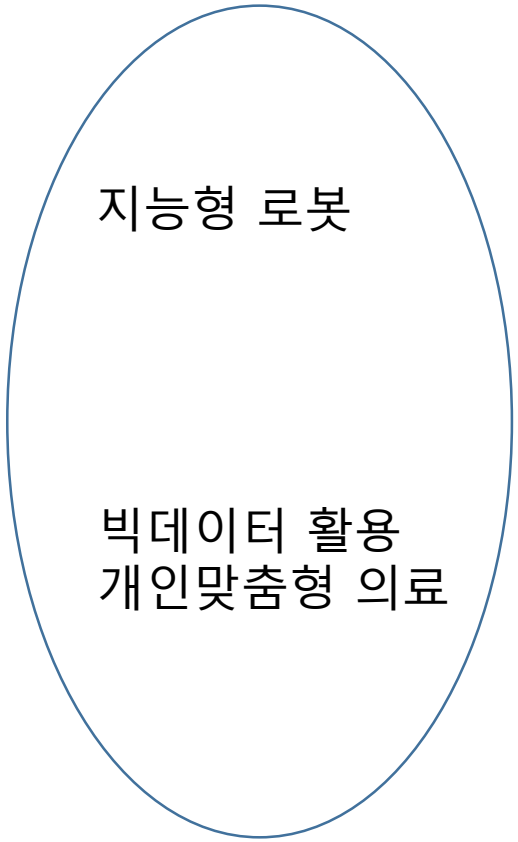
빅데이터 활용
개인맞춤형 의료

01

02

양자컴퓨팅

초고속
튜브 트레인



빅데이터와 인공지능

Big Data & A.I.

01

02

03

빅데이터 관심↑

01

02

03

Why ?

- 이 전에는 방대한 Raw Data를 담을 공간 부족
  - 빠르게 처리할 기술 부족

- 하지만 기술(하드웨어)의 발전으로 가능해짐
- 대부분의 생활이 디지털화 되어서 Raw Data 수집도 쉬워짐

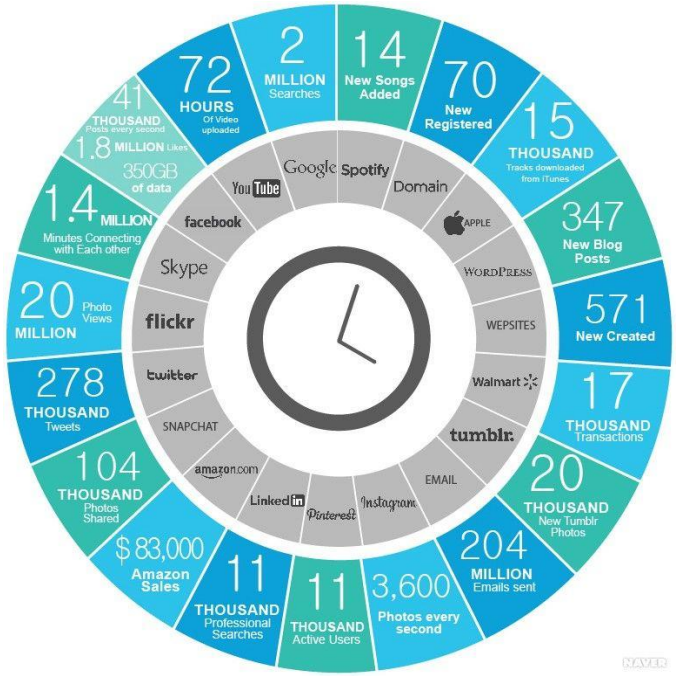
01

02

03

빅데이터 ?

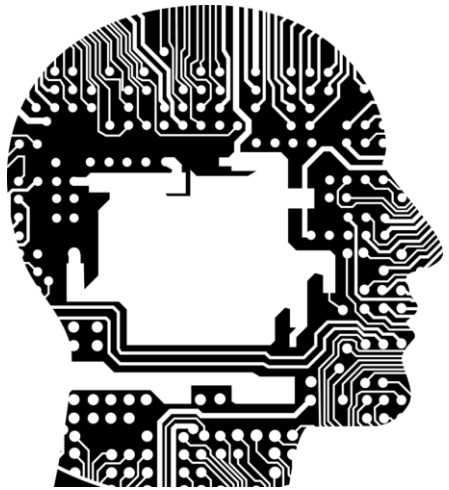
- 이름 그대로 굉장히 방대한 정형/비정형 데이터



01
02
03

인공지능 ?

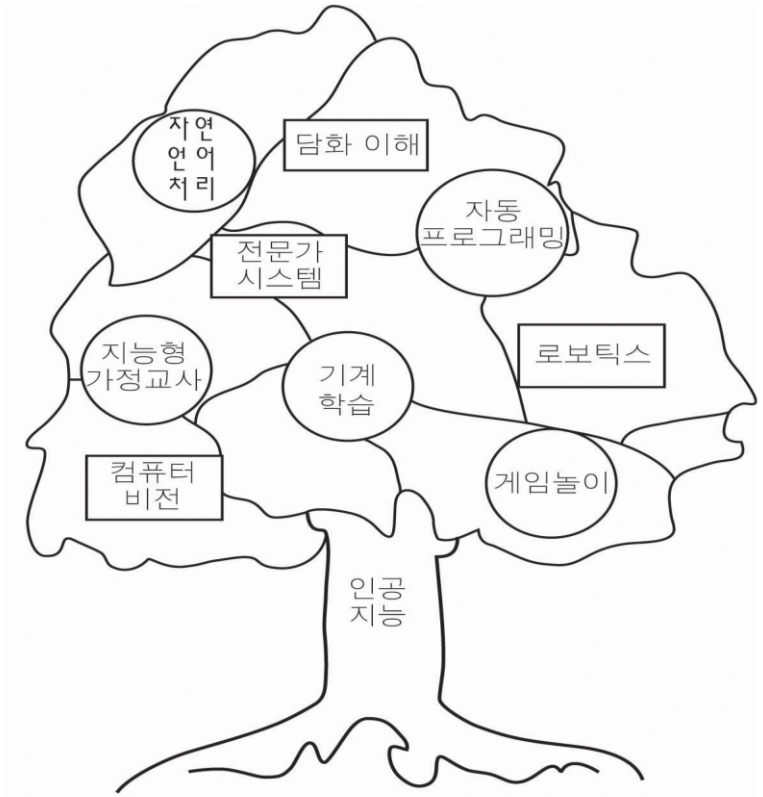
- 인공 + 지능 = 인공적으로 만들어진 지능



지능 ?

- 데이터(정보)들을 학습하고
  새로운 데이터가 들어왔을 때,
  자신만의 방식으로 '예측' 또는 '추론' 하는 것

- 즉, 컴퓨터가 인간의 지능적인 행동을
  모방할 수 있도록 하는 것


자연 언어 처리
담화 이해
자동 프로그래밍
전문가 시스템
지능형 가정교사
기계 학습
로보틱스
컴퓨터 비전
게임놀이
인공 지능
언어학, 컴퓨터 과학, 심리학, 철학, 전자공학, 경영과학

머신러닝

Machine Learning

01

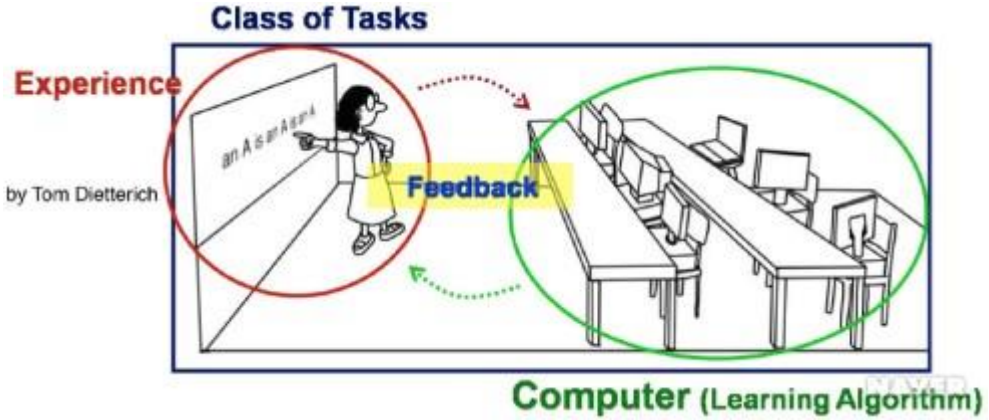
02

03

머신러닝 ?

- 머신러닝도 데이터 분석 기법

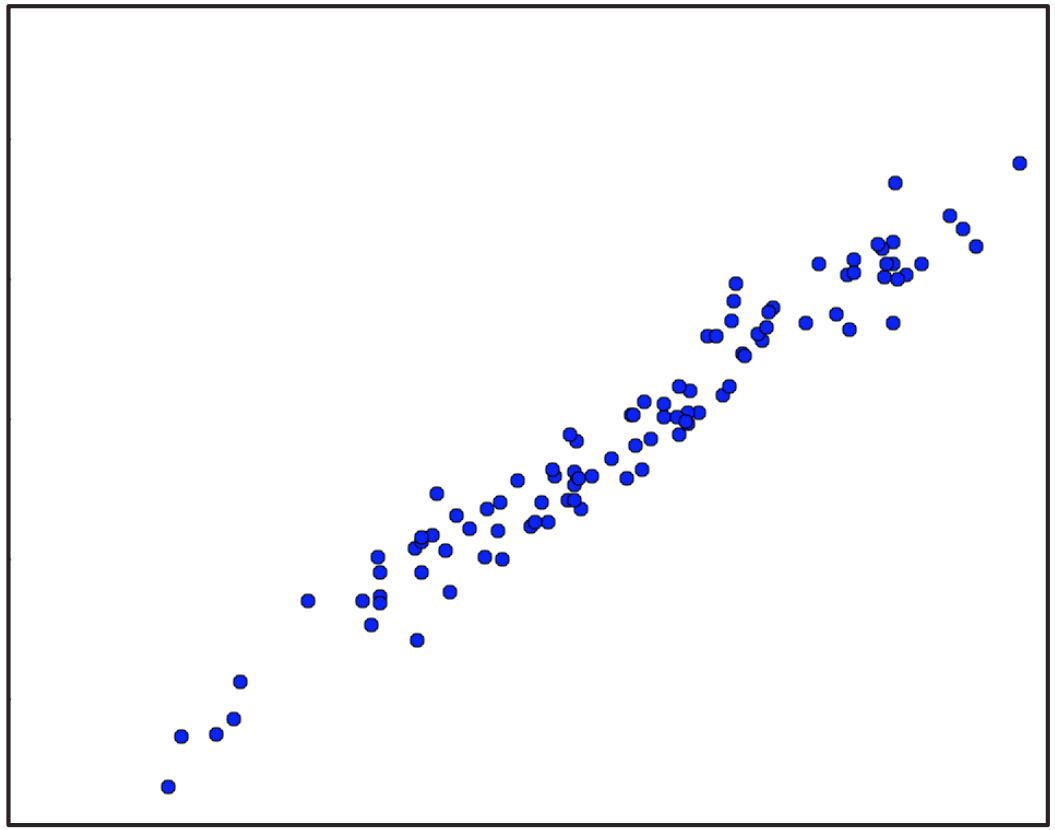
- 사람이 학습하듯이 컴퓨터에도 데이터들을 줘서 학습하게 함으로써 새로운 지식을 얻어내게 하는 것

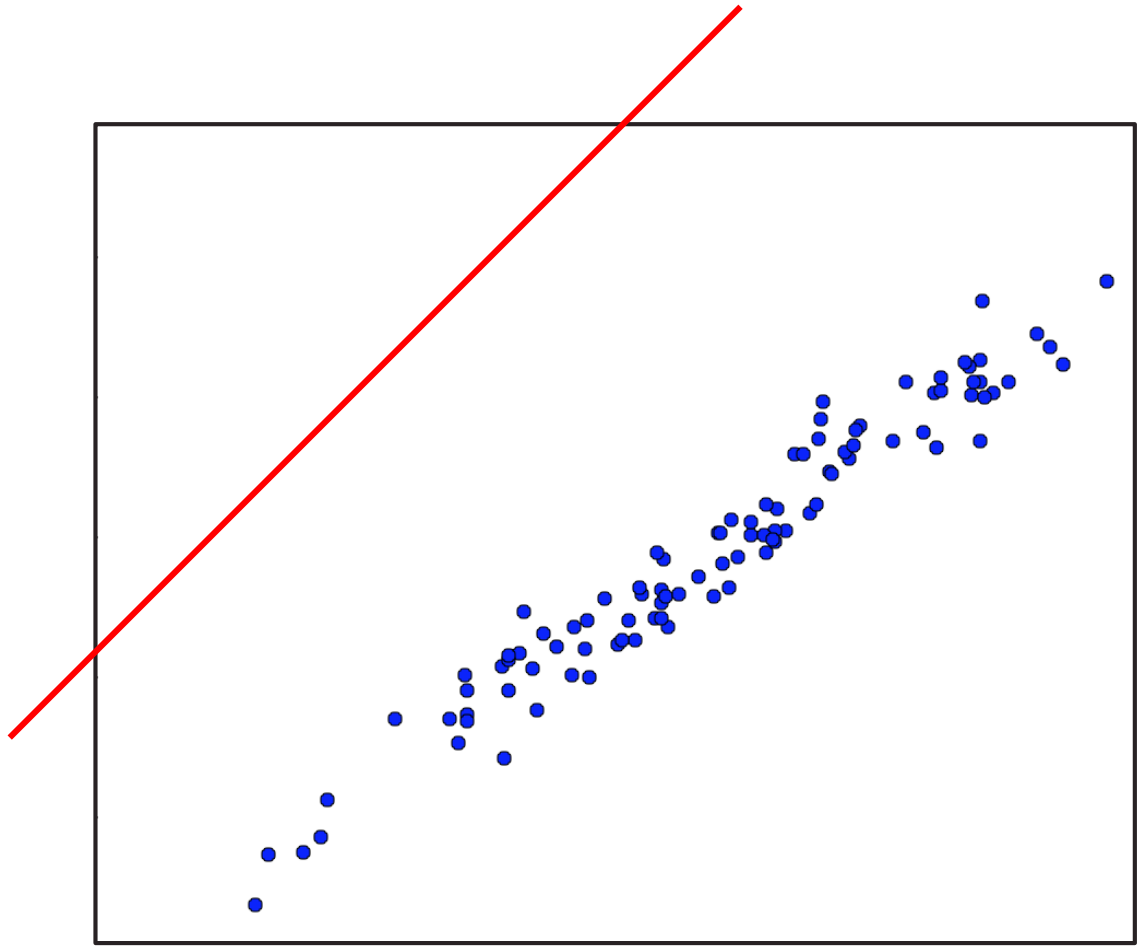


01

02

03

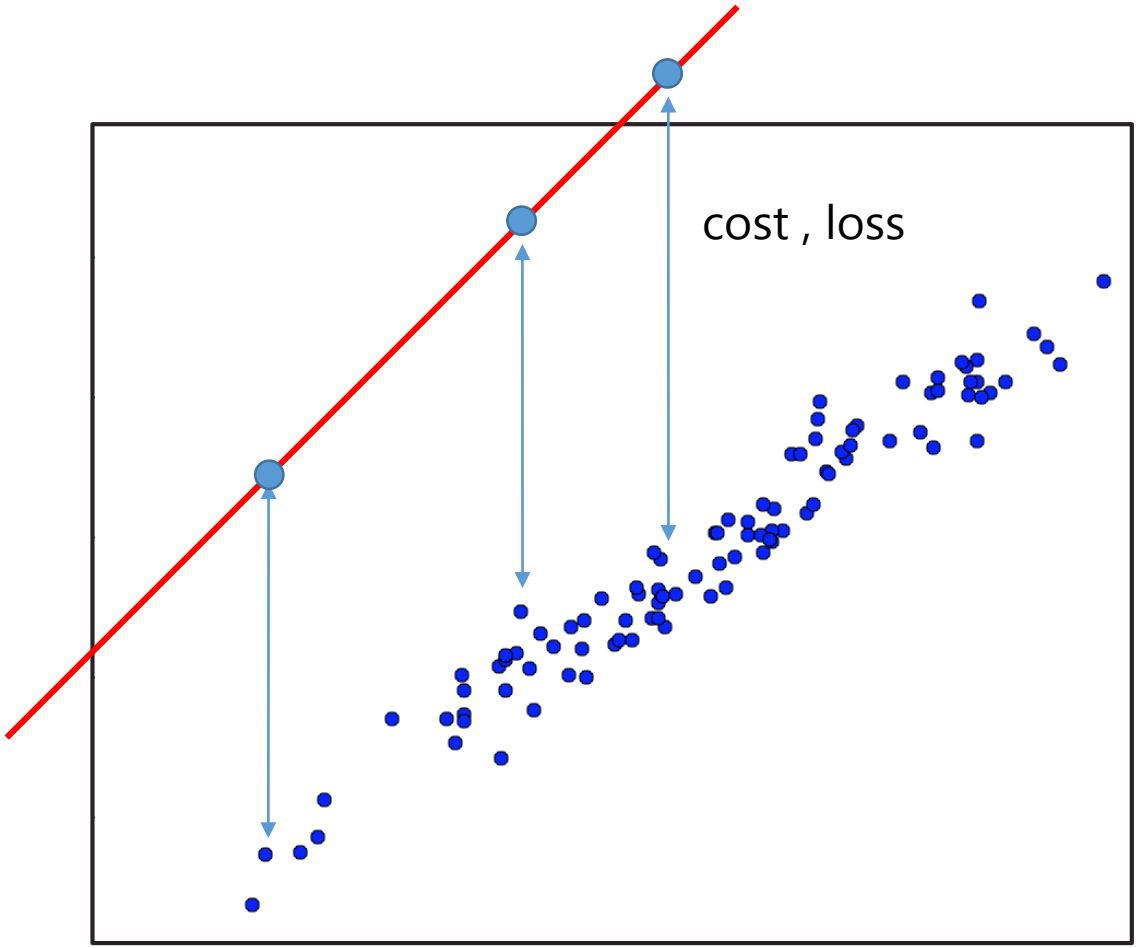




01

02

03



01

02

03

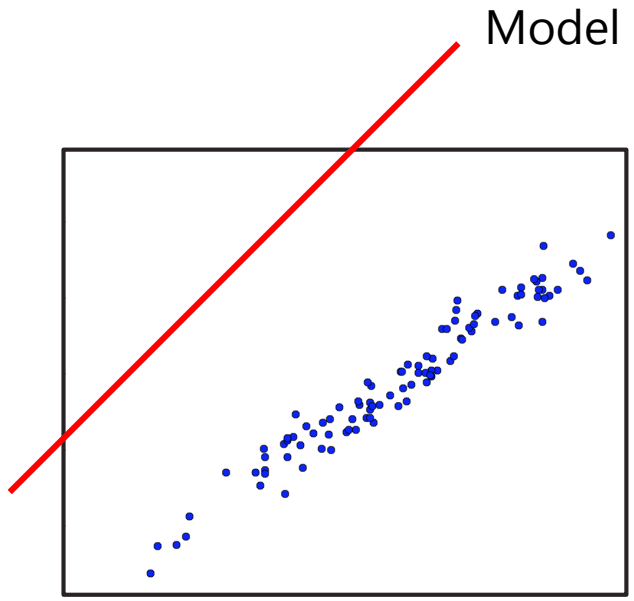
머신러닝

데이터들과 계속 비교하면서 cost 혹은 loss값을 줄여나가는 과정

01

02

03



$Y = a*X + b$

- 학습을 통해서 a값과 b값을 계산
- 실제 모델은 당연히 더욱 복잡

학습 방법

- 지도 학습(Supervised Learning)
- 비지도 학습(Unsupervised Learning)

지도 학습

- 학습 데이터가 (data,label)
- 즉, 정답을 제공
- Ex) 강아지사진과 고양이사진 구분
  (dog1.jpg , dog) , (dog2.jpg , dog) , ...
  (cat1,jpg , cat) , (cat2.jpg , cat) , ...

비지도 학습

- 학습 데이터가 (data)
- 즉, 정답을 제공 X
- Ex) 강아지사진과 고양이사진 구분
  (dog1.jpg) , (dog2.jpg) , ...
  (cat1,jpg) , (cat2.jpg) , ...

Batch Size

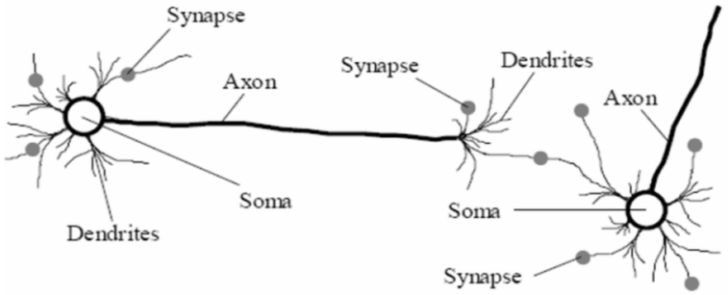
- 데이터를 몇 개의 묶음으로 비교할지 정하는 parameter
- 데이터 총 갯수 = 2000
  Batch size = 100
  20개씩 100묶음

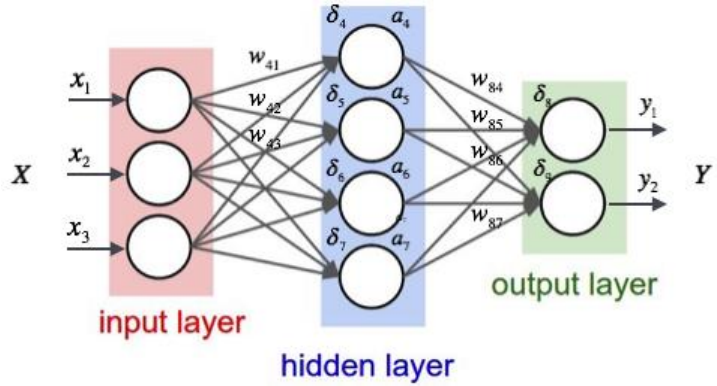
epoch

- 전체 훈련 데이터셋을 몇 번 반복해서 학습할지 정하는 parameter
- 에러(cost)를 충분히 줄이기 위해 보통 수백, 수천을 값으로 가짐

인공신경망?

- 생물학적 신경망(Neural Network)을 비슷하게 구현한 것





딥러닝

Deep Learning

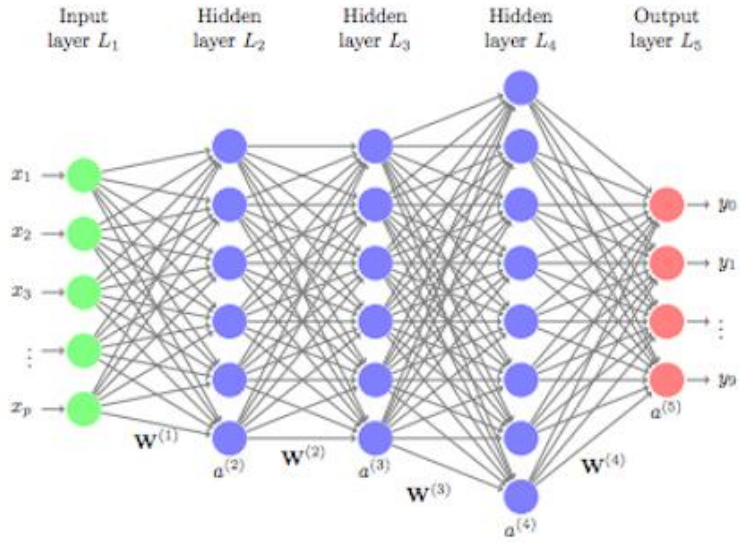
01

02

03

딥러닝 ?

- 신경망(Nueral Network)을 더 Deep하게 구성한 것
- 여기서 'Deep 하다'는 Hidden Layer의 계층수가 많아 졌다는 의미



01

02

03

- Hidden Layer의 계층수가 더 많기 때문에 일반 Neural Network 보다 더욱 일반적, 추상적으로 학습 가능

- 알파고도 딥러닝으로 구현된 모델

› 그렇다면 무조건 Deep 할수록,
즉 Hidden Layer의 계층수가 많을 수록
좋을까?

- 그렇다면 무조건 Deep 할수록,
  즉 Hidden Layer의 계층수가 많을 수록
  좋을까?

X

- 어떤 경우는 Backpropagation(오류역전파)가
  제대로 이루어지지 않기 때문
- 분석 대상에 따라서 깊이는 다르게, 적절하게
- 적정 수준은 실험을 통해 알아내야 함

딥러닝 모델

- CNN(Convoultion Neural Network)
- RNN(Recurrent Neural Network)
- GAN(Generative Adversarial Network)

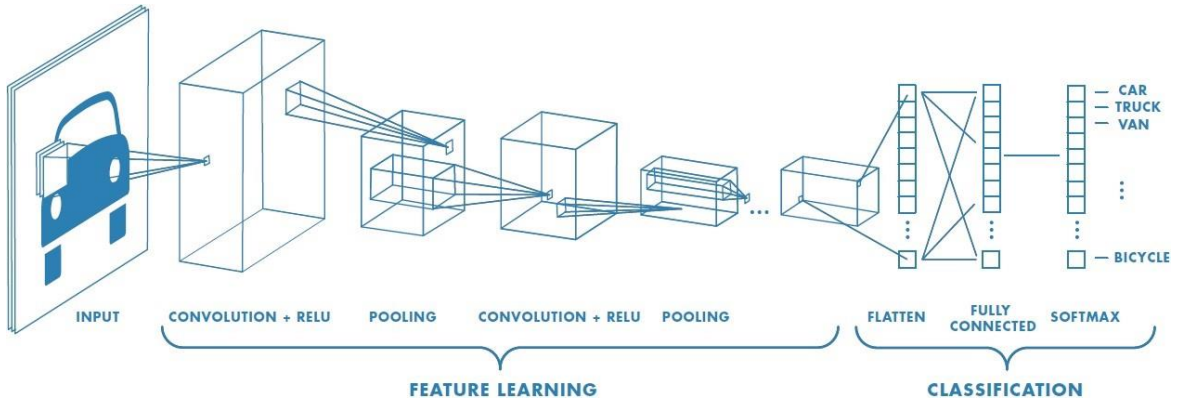
01

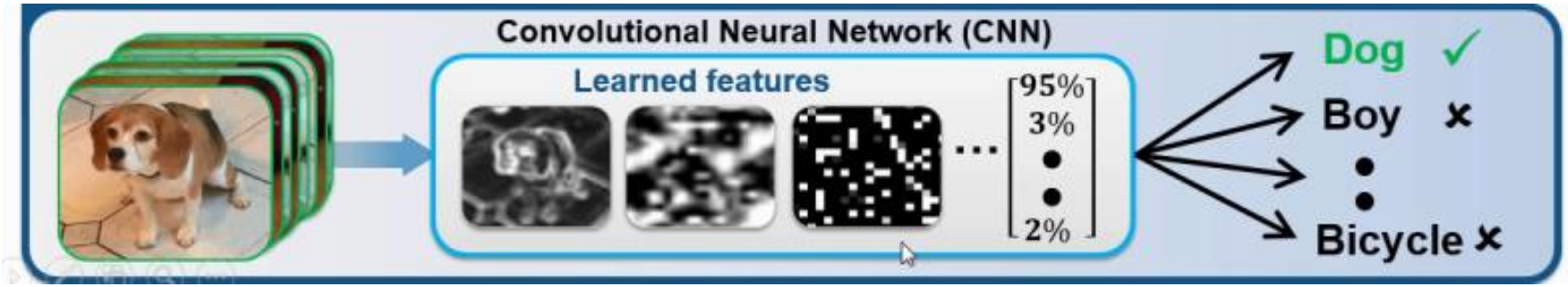
02

03

CNN

- 주로 이미지 분류에 사용
- Hidden Layer가 Convolution , Pooling 으로 이루어짐





- 각 이미지에 대한 특징을 학습
- 새로운 이미지 데이터가 들어오면 학습된 모델을 통해서 가장 확률이 높은 class를 예측

RNN

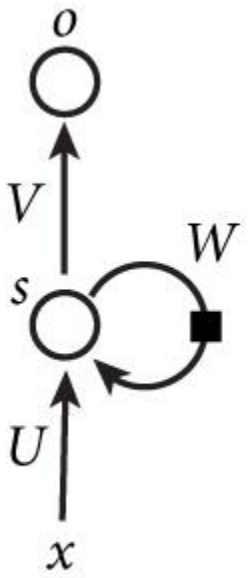
- 주로 학습 데이터가 Sequencial Data 일 때 사용
- Sequncial Data의 가장 대표적인 '언어'는 앞의 단어, 뒤의 단어와의 관계 등에서 문장의 '의미'를 파악해야 함
- RNN이 언어(자연어) 처리 분야에서 가장 뛰어난 모델
- 챗봇, 인공지능 스피커에 필수

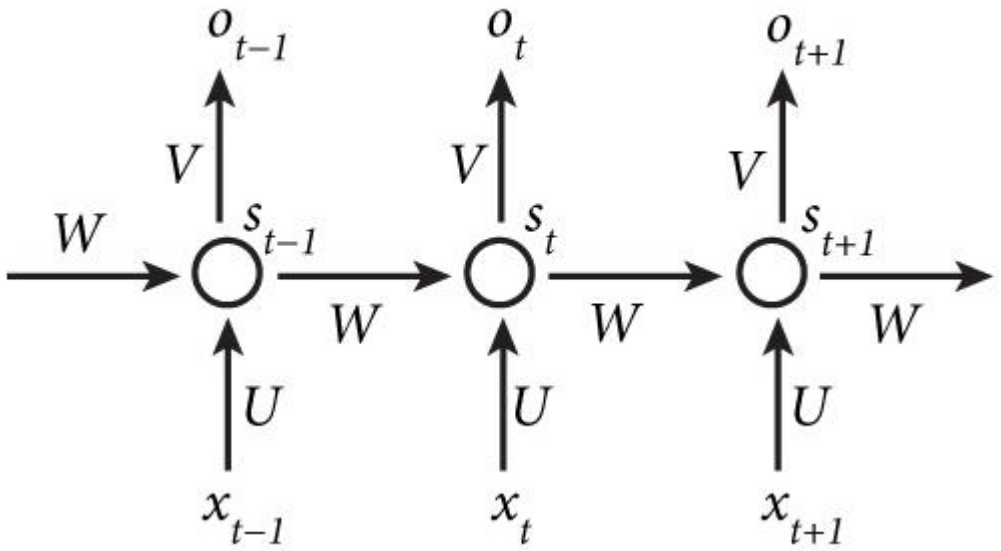
01

02

03

- Sequencial Data에 최적화 된 이유는 RNN(Recurrent Neural Network) 이라는 이름에서 처럼 순환적으로 구성되었기 때문임

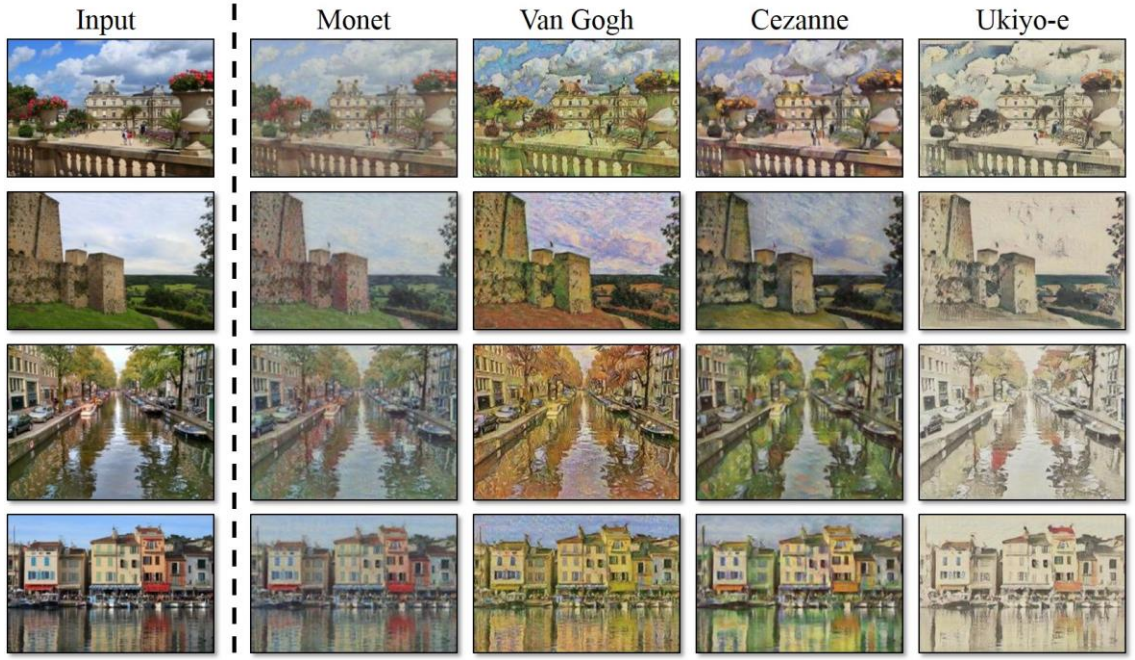




- ‘SSIDO’ 라는 단어를 훈련시킨다고 하면
  X(t-1) 에는 ‘S’ , O(t-1) 에는 ‘S’
  X(t) 에는 ‘S’ , O(t) 에는 ‘I’
  X(t+1) 에는 ‘I’ , O(t+1) 에는 ‘D’
  …
  Ex) 검색 할 때 자동완성

GAN

- 주로 흉내를 낼 때 쓰임

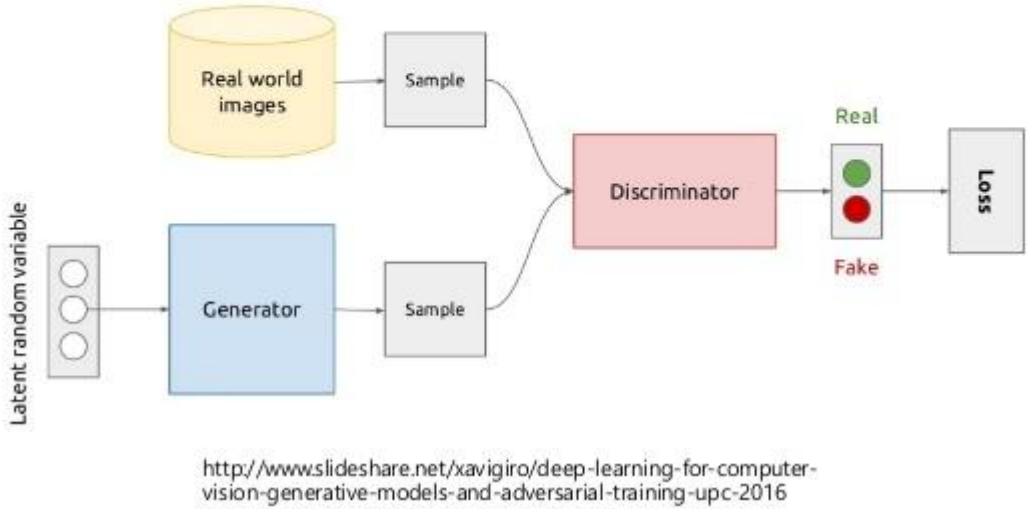


01

02

03

Adversarial Learning



http://www.slideshare.net/xavigiro/deep-learning-for-computer-vision-generative-models-and-adversarial-training-upc-2016

› 쉽게 설명하자면 <경찰 - 위조지폐범> 관계

01

02

03

딥러닝 = [노력,시간,돈]

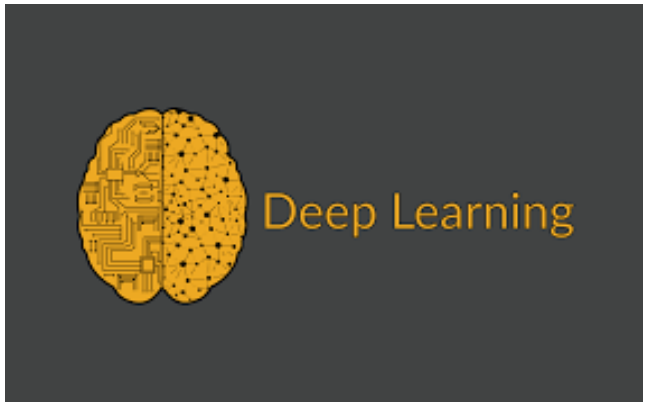
- 하드웨어가 발전 했다고는 하지만
  여전히 딥러닝 분야는 노력, 시간, 돈이 필요
- CPU로 그럴듯한 모델을 훈련시키려면
  최소 몇 시간은 걸림
- 결과가 마음에 안들면 다시 훈련

- 병렬처리가 가능한 GPU 사용
- 그래픽카드를 직접 구매해서 사용하거나 AWS에서 딥러닝 서버를 대여해서 훈련

01

02

03

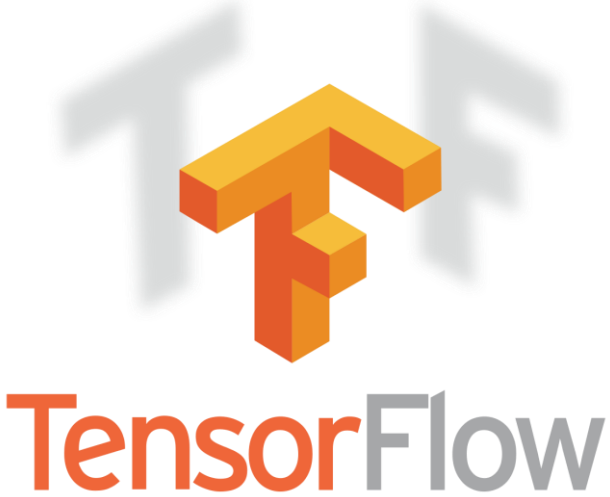


- 딥러닝 주력 언어는 Python
- 다른 언어로도 딥러닝 구현은 가능

01

02

03

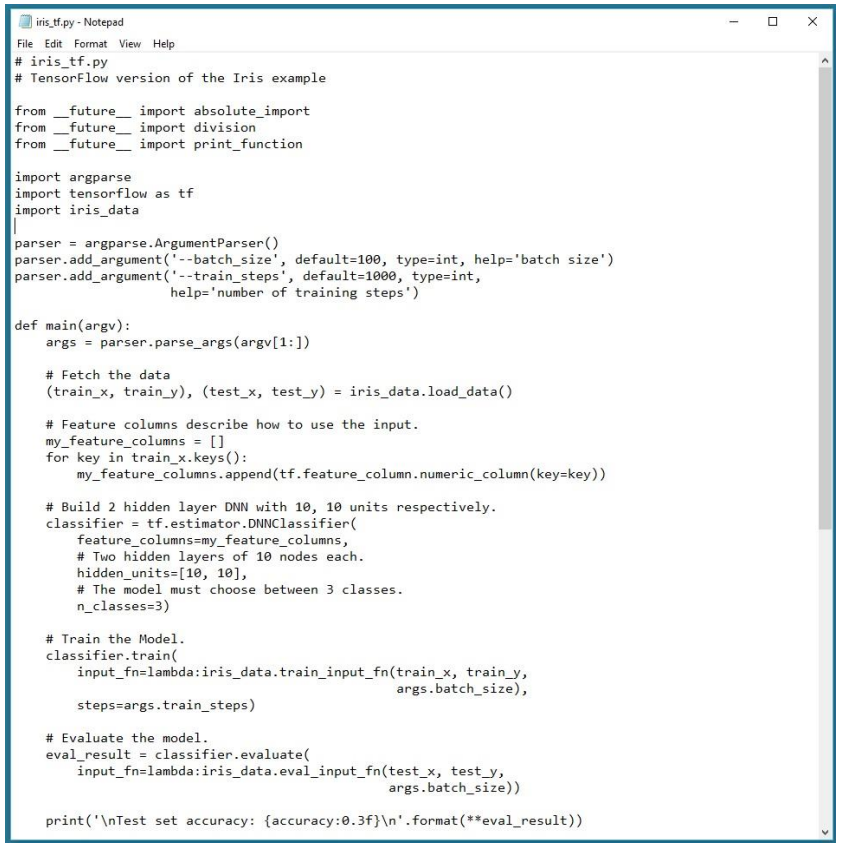


- 기존에 가장 많이 쓰이던 라이브러리는 Tensorflow
- 최근 Tensorflow 보다 더 간편하고 직관적인 Pytorch도 인기 상승 중

01

02

03



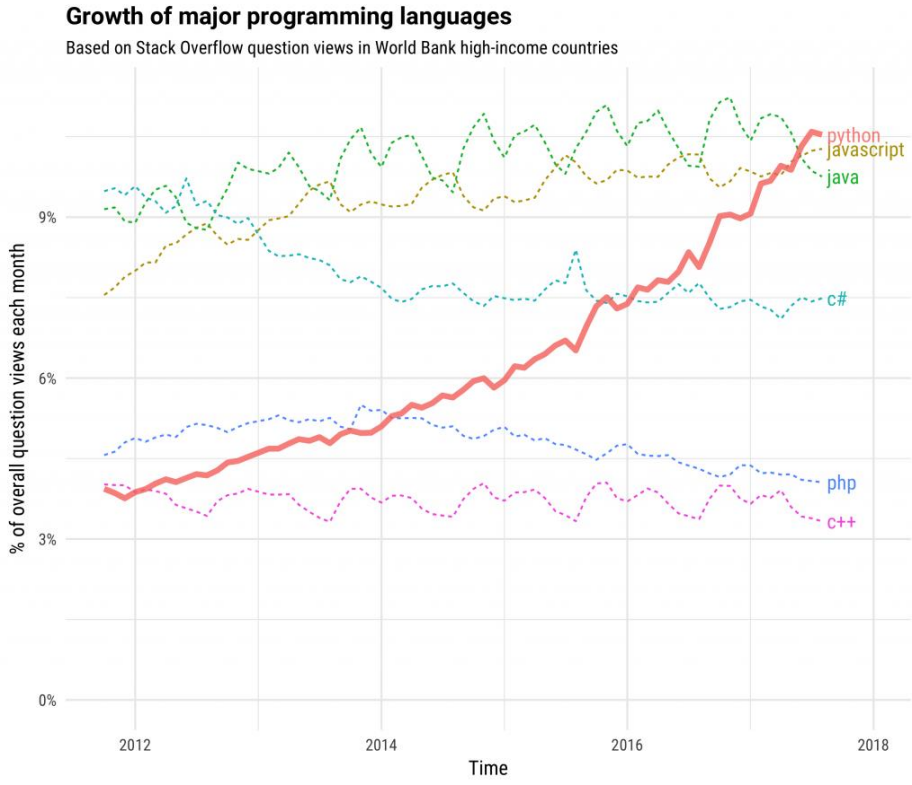
› 가장 간단한 편에 속하는 Tensorflow 로 구현한 딥러닝 코드

01

02

03

결론 : Python 열심히!



질문

Q & A