



MACHINE LEARNING & DEEP LEARNING

머신러닝과 딥러닝

2018.11.14 박정현

Daeyeon Jo

[강의 이력]

- (주) 한국인공지능아카데미 전임 강사 (삼성전자, 대한전기협회 등 B2B 머신러닝/딥러닝/데이터분석 강의)
- 엔코아 프로젝트 멘토링 및 교육 (데이터 사이언티스트 집중 양성과정 1기, 분석 프로젝트 운영 및 멘토링 진행)
- Shepherd International Education (SIE, 국제교육기관) 교원 (초/중등 학생 대상 파이썬 프로그래밍 교육)
- 서울대학교 언어교육원 한국어교육센터 교원 (해외 정부 장학생 대상 파이썬 프로그래밍 및 머신러닝 교육)
- 서울대학교 도시데이터사이언스 연구소 보조 강사 (머신러닝, 추천시스템, 웹크롤링, 자연어 분석 등)
- 서울대학교 석박사 머신러닝 & 데이터분석 프로젝트 멘토링 및 교육 (Python, R, SPSS, STATA, NetLogo 등)
- 서울대학교 4차산업혁명아카데미 인공지능에이전트 과정 부리더
- Tableau(데이터시각화 도구) 데이터 분석 및 시각화 강의 (기업 컨설팅 연구단체 대상)
- 서울특별시 창조경제타운 애플리케이션 기획 멘토

[경력]

- 사단법인 한국인공지능연구소 연구원
- 서울대학교 영상의학과 연구원
- 실리콘밸리 벤처캐피탈 페녹스 코리아 팀장
- 스타트업 오피스비긴 대표
- 스타트업 코스모엔젤스 공동대표
- 스타트업 라이트엔슬림 팀장
- 마이크임팩트 경영지원팀 인사담당

[교육]

- 서울대학교 경영학과 전공
- 서울대학교 공과대학 4차산업혁명아카데미 인공지능에이전트 전공
- KAIST 인공지능 및 기계학습 과정 (KCS470, K-MOOC)
- 응용 SW 엔지니어링 개발자 양성과정 (에이콘 아카데미)
- 빅데이터 연합 연구단체 BOAZ Advanced member

- Part1 쉽게 풀어서 살펴보는 인공지능
- Part2 머신러닝의 분류
- Part3 머신러닝과 딥러닝의 차이
- Part4 머신러닝과 딥러닝의 활용사례
- Part5 데이터관련 추천학습자료

PART1 쉽게 풀어서 살펴보는 인공지능

인공지능이란?

Artificial Intelligence

Artificial intelligence (AI), sometimes called machine intelligence, is **intelligence demonstrated by machines**, in contrast to the natural intelligence displayed by humans and other animals.

In computer science AI research is defined as the study of “**intelligent agents**” : any device that **perceives its environment** and **takes actions** that **maximize its chance of successfully achieving its goals**.

Perceives its environment
자신을 둘러싼 환경을 지각/인식하고



Take actions
적절한 행동들을 취하여



Maximize chance of achieving goals
목표를 성취할 가능성을
최대화 하는 기계(가 가진 지능)

사실은 마케팅 용어에 가깝게 쓰이는 “인공지능”

우리 회사는 인공지능을 적용해 ~~~를 하는 회사입니다.



우리 회사는

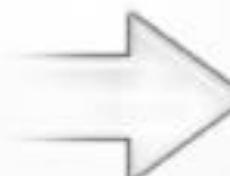
- 1) ~~~ 데이터를 활용하여,
 - 2) ~~~ 모델을 만들고,
 - 3) 이 모델을 적용한 ~~~ 기능을 만드는
- 회사입니다.

“인공지능”을 적용한다는 것은,

~~우리 회사는 인공지능을 적용해 ~~~를 하는 회사입니다.~~

우리 회사는

- 1) ~~~ 데이터를 활용하여,
- 2) ~~~ **모델**을 만들고,
- 3) 이 모델을 적용한 ~~~ **기능**을 만드는 회사입니다.



어떤 **모델**을 만들 것인가?

= 뛰어난 기술로부터 출발할 것인가?

어떤 **기능**을 만들 것인가?

= 고객의 Needs로부터 출발할 것인가?

→ 문제 해결을 위한 도구로서의 인공지능 기술 적용

모델(Model)이란?

구분	한국	미국	일본	영국	프랑스	독일
XS	85	95 - 90	S	0	40	44 - 46
		14	36			
S	90	90 - 95	M	1	42, 44	46
		15	38			
M	95	95 - 100	L	2	46, 48	48
		15, 5 - 16	40			
L	100	100 - 105	LL, XL	3	50, 52	50
		16,5	42			
XL	105	105 - 110	-	4	54, 56, 58	52
		17,5	44			
XXL	110	110 ~	-	5	60, 62	54
		-	46			

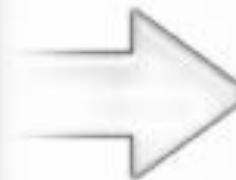
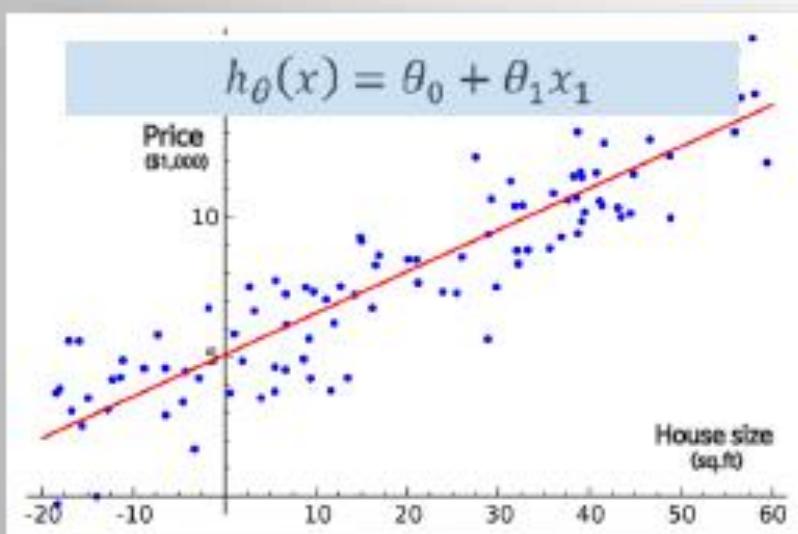
XS, S, M, L, XL, XXL 등의 **Size**는
각각의 **체형에 대한 모델**

내 몸에 맞는 **사이즈**를 찾는 과정
= 나의 체형을 가장 잘 표현하는 **모델**을 찾는 과정

내가 다루고 있는 **데이터를 가장 잘 설명하는 방법**을 찾는 과정
= 내 데이터에 맞는 **모델**을 찾는 과정 (= **Model fitting**)

Model의 종류
: 선형 함수, 비선형 함수, 확률 분포 함수,
Neural Network의 Layer architecture & Parameter set 등

모델(Model)이란?



$$y = a * x + b$$

= 모델 (Model)

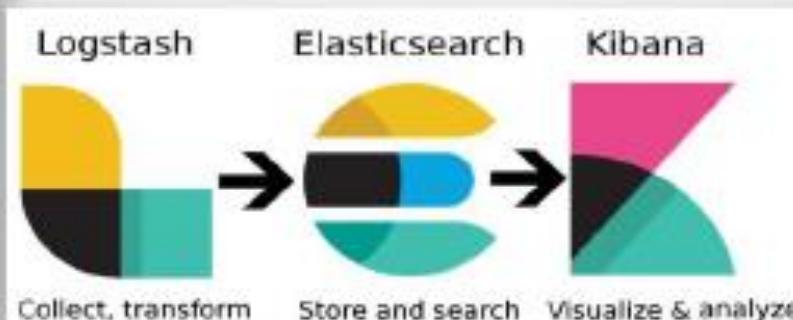
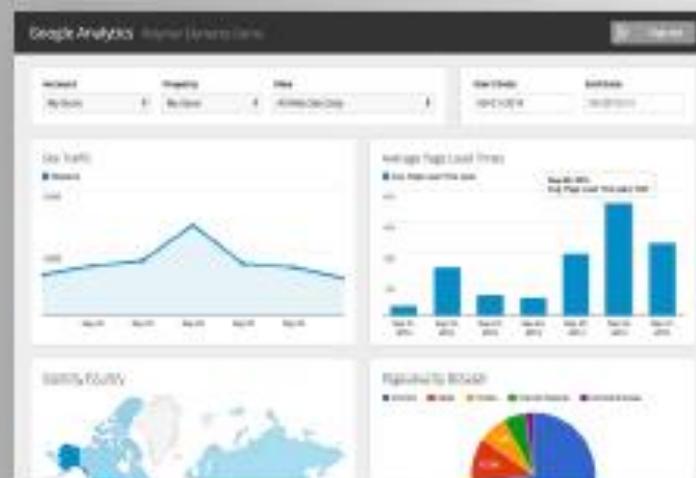
학습(Learning)이란?

내가 다루고 있는 **데이터를 가장 잘 설명하는 방법**을 찾는 과정
= 내 데이터에 맞는 **모델**을 찾는 과정 (= **Model fitting**)

데이터를 가장 잘 설명하는 모델을 어떻게 찾을 것인가?

- 1) 초기 모델(가설 모델)에 데이터를 넣는다.
- 2) 결과를 **평가**한다. (예측/분류의 정확도 등 = mean squared error, classification error, recall & precision, etc.)
- 3) 결과를 개선하기 위해 **모델을 수정**한다. (모델 내부의 **Parameter** 수정, 모델 **클래스**의 변경 등)

Various data collection - Owned data

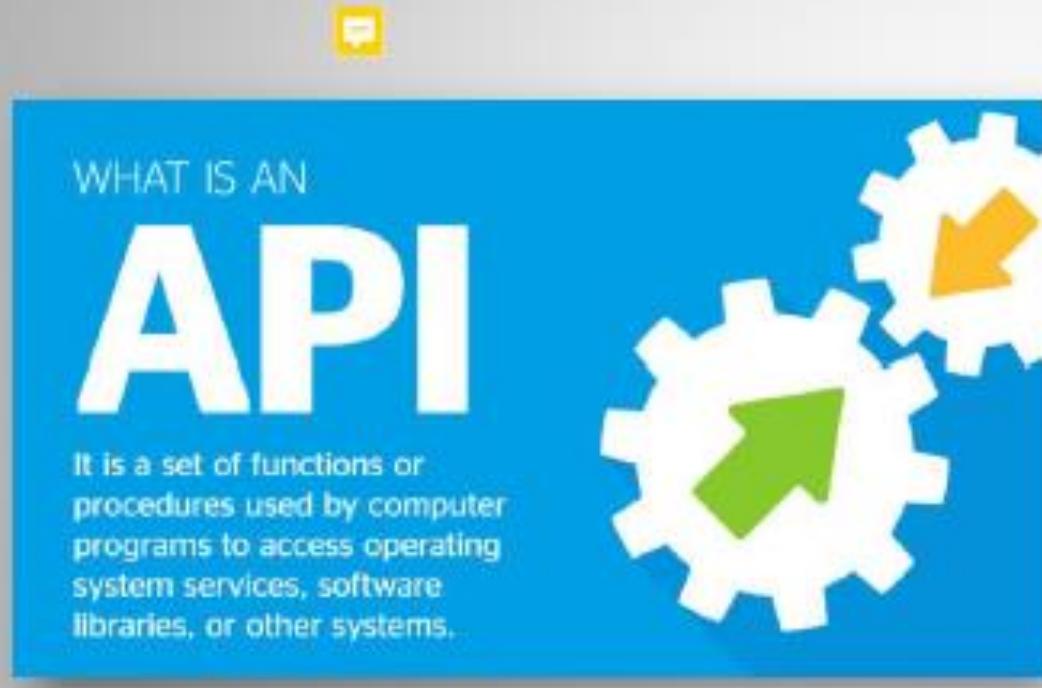


Various tools are available!

- Google Analytics (+ Facebook, Naver, etc)
- Elastic Stack (Logstash, Elasticsearch, Kibana) 
- Zeppelin

- * Elastic search 개요 @ <https://goo.gl/1x7Fq9>
 - * Elastic search 블로그 @ <https://goo.gl/KsojgK>
 - * Elasticsearch 시작하기 (한글 동영상) @ <https://goo.gl/TJnb7Z>
 - * Korea Elasticsearch User Group (Facebook group) @ <https://goo.gl/x131Pm>
- 본 교안은 11월 8일 강의를 위해 제작되었으며, 강의 외 배포/개시/공개를 금합니다.

Various data collection – Unowned data



Use APIs & Web scraper

- APIs (Twitter, Facebook, Instagram, etc)
- Bots (Web crawler, Web scraper)

Various data collection – Public data & Open data (APIs & files)



- 공공 데이터 포털 : <https://www.data.go.kr>
 - 국가통계 포털 : <http://kosis.kr>
 - MDIS (MicroData Integrated Service) : <https://mdis.kostat.go.kr>

머신러닝을 적용한 인공지능을 활용한다는 것은,

~~우리 회사는 인공지능을 적용해 ~~~를 하는 회사입니다.~~

우리 회사는

- 
- 1) (각종 정형/반정형/비정형) 데이터를 활용하여,
 - 2-1) 데이터를 **가장 잘 설명할 수 있는 함수(알고리즘)**의 후보들을 정하고,
 - 2-2) 함수에 포함된 **Parameter**들의 **가장 좋은 값을 컴퓨터로 찾아내어**,
 - 3) 이 모델을 적용한 **~~~ 기능**을 만드는

회사입니다.

PART2 머신러닝의 분류

머신러닝이란?

- A field of artificial intelligence that gives computers **the ability to learn** from data, without being explicitly programmed.

Arthur Samuel, 1959

- A computer program is said to **learn from experience E** with respect to **some class of tasks T** and **performance measure P**, **if its performance at tasks in T, as measured by P, improves with experience E**.

만약 작업 T에 대해 기준 P로 측정한 성능이 경험 E로 인해 향상되었다면,
그 프로그램은 작업 T에 대해 기준 P의 관점에서 경험 E로부터 "배웠다"라고 말할 수 있다.

Tom M. Mitchell, 1997

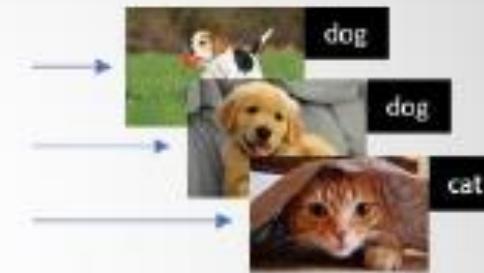


- 1) 어떠한 **과제**를 해결하는 과정에서
- 2) 특정한 **평가 기준**을 바탕으로
- 3) **학습**의 경험을 쌓아나가는 프로그램

머신러닝의 분류 Supervised, Unsupervised, Reinforcement



Supervised



Un-Supervised



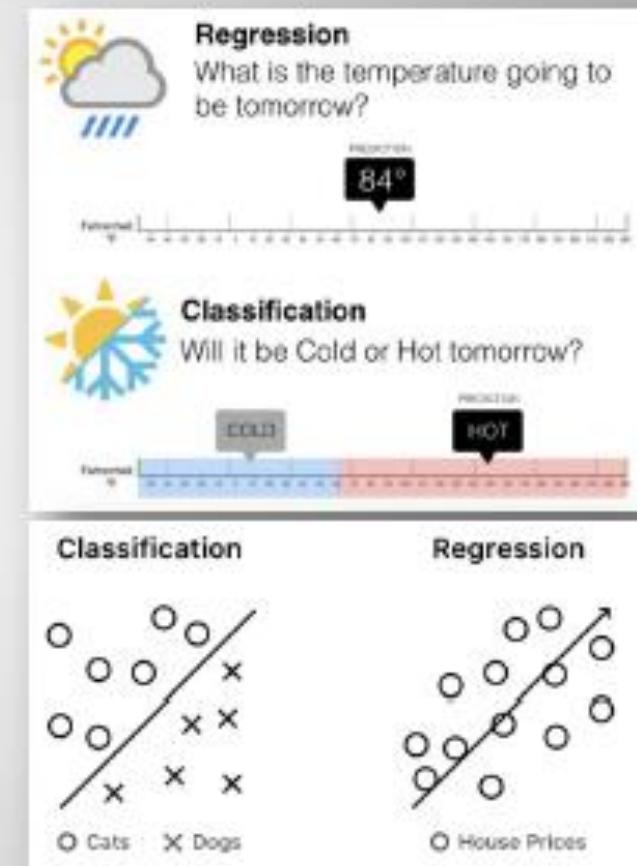
머신러닝의 분류 Supervised, Unsupervised, Reinforcement



1) Supervised learning (지도 학습)

- Input data에 대한 정답을 예측하기 위해 학습 (\rightarrow Function approximator)
- 데이터에 정답(Label, Target)이 존재함
- Output의 형태에 따라 회귀 분석과 분류 분석으로 나눌 수 있음
 - 회귀 Regression (Output이 실수 영역 전체에서 나타남)
 - 분류 Classification (Output이 class에 해당하는 불연속값으로 나타남)
- 대표 알고리즘
 - : Linear/Logistic regression, Decision tree, Bayesian classification, (Basic) Neural Network, Hidden Markov Model(HMM) 등

ex) 스팸 분류기, 주식 가격 예측, 유방암 진단, 이미지 인식 등
 (강아지와 고양이가 어떻게 생겼는지 알려주고 강아지와 고양이 사진들을 분류시킴)



머신러닝의 분류

Supervised, Unsupervised, Reinforcement

2) Unsupervised learning (비지도 학습)

- Input data 속에 숨어있는 규칙성을 찾기 위해 학습 (\rightarrow (shorter) Description)
- 데이터에 정답(Label, Target)이 존재하지 않음

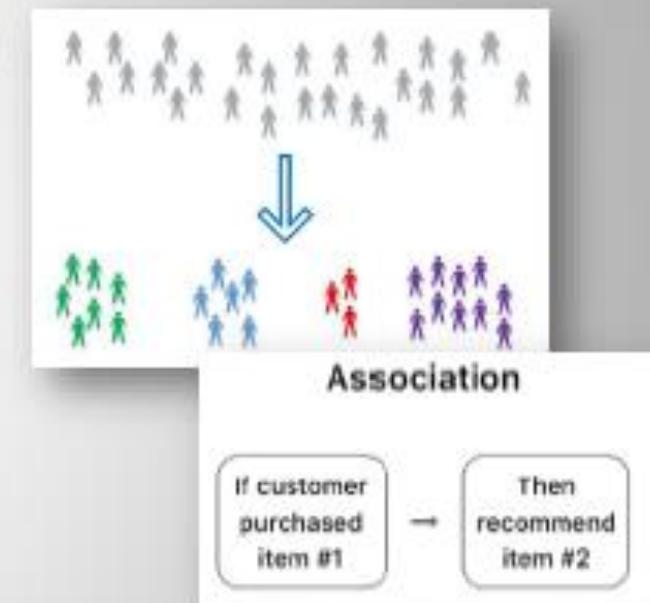
군집 분석 Clustering Algorithm

차원 축소 Dimensionality reduction (or Compression)

- 대표 알고리즘

: K-Means clustering, Nearest Neighbor Clustering, EM clustering, Principal component analysis (PCA), Linear Discriminant Analysis (LDA) 등

ex) 고객군 분류 (고객 세분화), 장바구니 분석(Association Rule), 추천 시스템 등
(강아지와 고양이가 무엇인지 알려주지 않고 여러 사진들을 보고 직접 비슷한 형태끼리 묶어보게 함)



머신러닝의 분류 Supervised, Unsupervised, Reinforcement

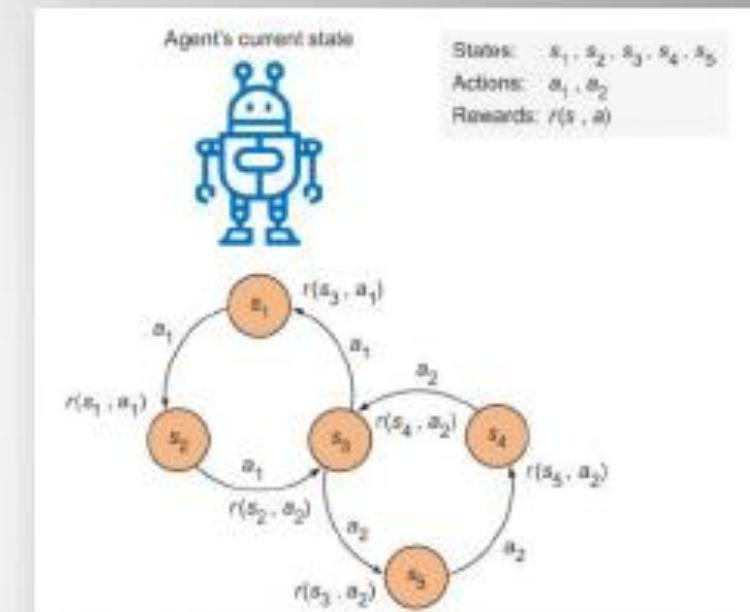
3) Reinforcement learning (강화 학습)



- Trial & Error 를 통한 학습 (\rightarrow Sequential decision making)
- 주위 환경과 자신의 행동(Decision) 사이의 반복적 상호작용을 바탕으로, 최종적으로 얻게 될 기대 보상을 최대화하기 위한 행동 선택 정책(Policy)을 학습
- 연속적인 단계마다 상태(State)를 인식하고, 각 상태에 대해 결정한 행동(Action)들의 집합에 대해, 환경으로부터 받는 보상(Reward)을 학습하여, 전체 행동에 대한 보상을 최대화하는 행동 선택 정책(Policy)을 찾는 알고리즘

- 대표 알고리즘
 - : Brute force, Monte Carlo methods, Markov Decision Processes, Q-learning, Deep Q-learning, Dynamic Programming 등

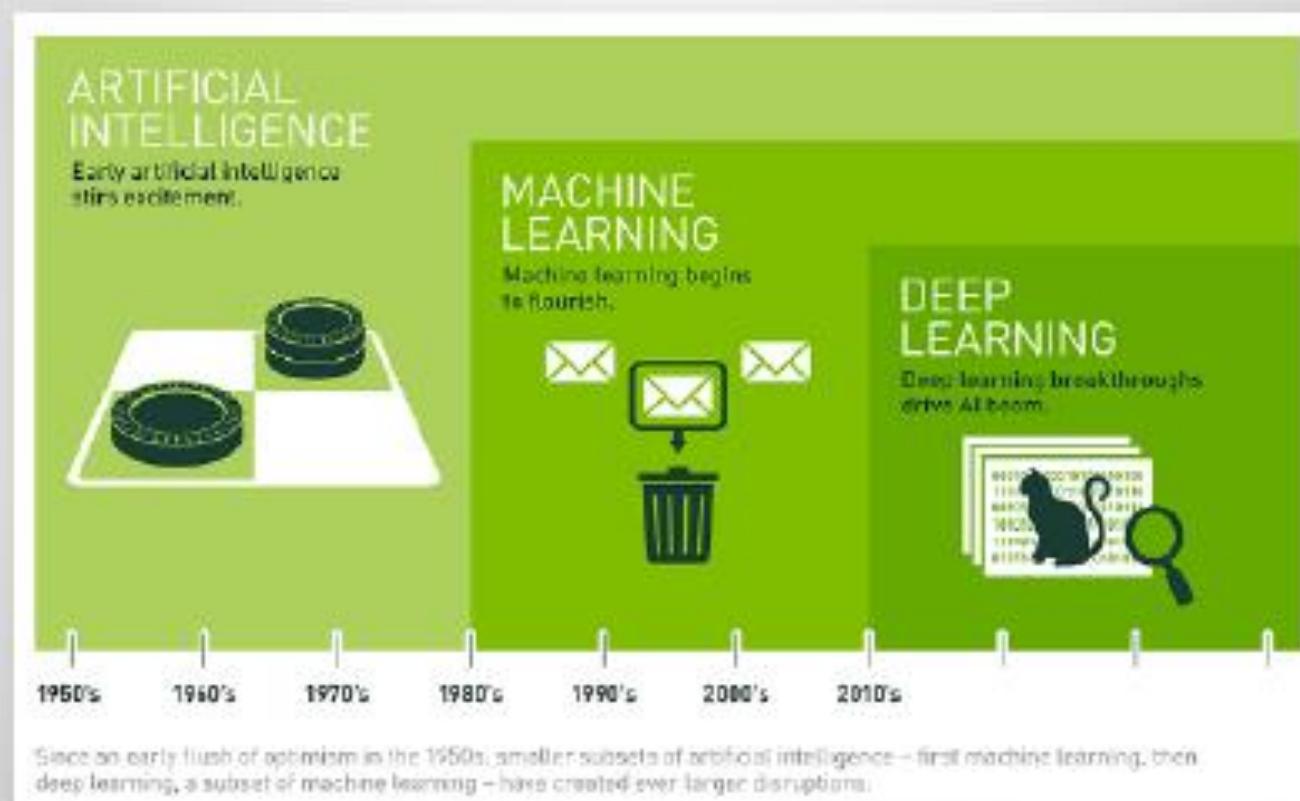
ex) 로봇 제어, 공정 최적화, Automated data augmentation 등



PART3 머신러닝과 딥러닝의 차이

머신러닝을 넘어 딥러닝으로

[전통적인 머신러닝]을 넘어 딥러닝으로

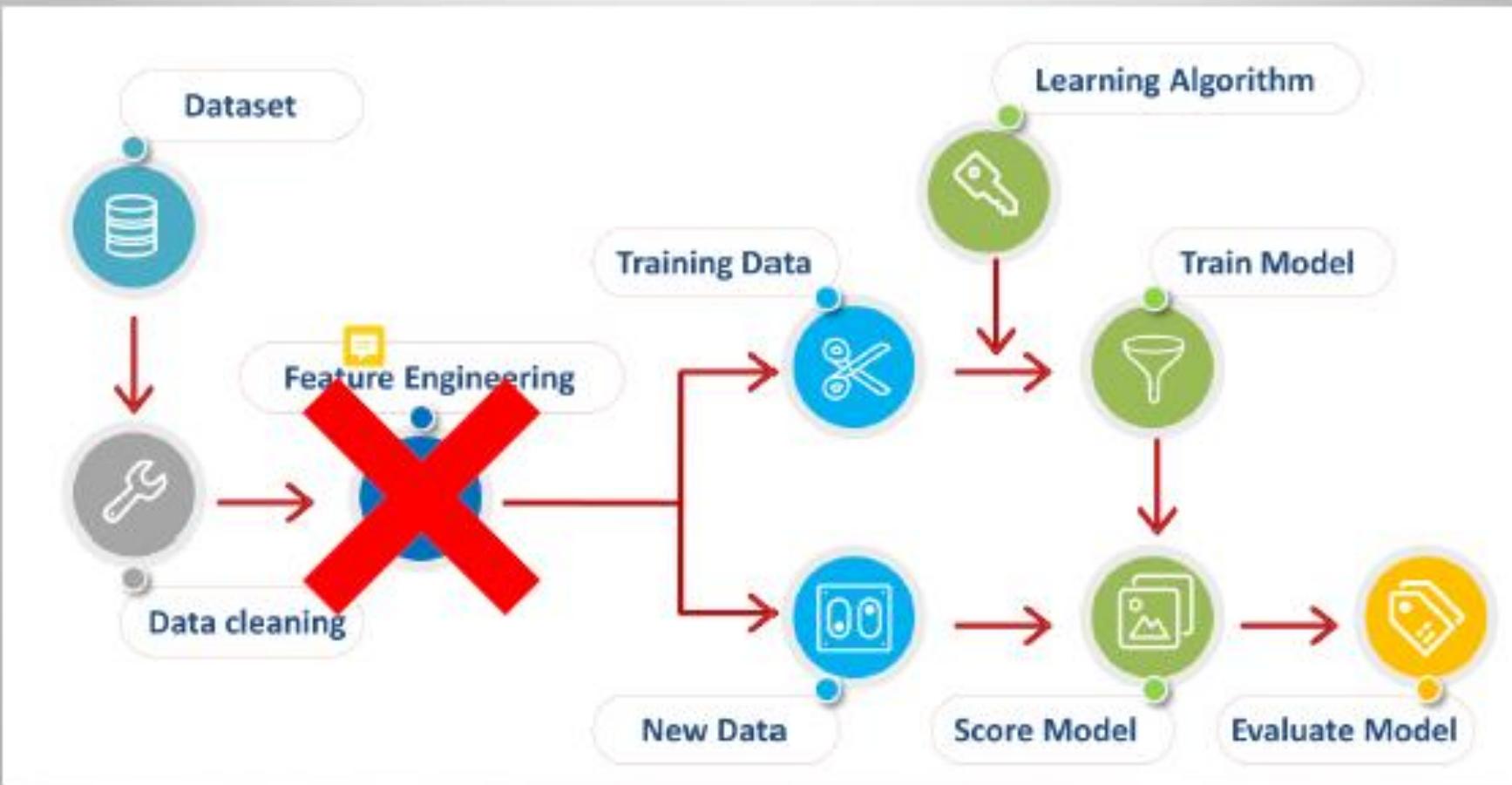


출처 : NVIDIA 공식블로그 @ <https://goo.gl/Dpfx2N>
(AI, ML, DL의 포함 관계이기도 합니다)

전통적인 머신러닝을 넘어 딥러닝으로



머신러닝과 딥러닝, 그 미묘한 차이



머신러닝과 딥러닝, 그 미묘한 차이

- 전통적인 머신러닝 : 사람이 직접 데이터의 중요한 부분들을 찾아 Feature로 정해주어야 함
- 딥러닝 : 신경망 내부에서 자체적으로 데이터의 중요한 Feature를 찾거나 구성함



* Feature = 예측을 수행하는데 사용되는 입력 변수 (ex. 주택 가격을 예측하는 모델의 feature : 우편번호, 택지 규모, 일자 조건 등)

* Feature engineering = 모델을 학습시키는데 유용할 특성이 무엇인지 판단하고 로그 파일 및 기타 소스의 원시 데이터를 해당 특성으로 변환하는 과정

머신러닝과 딥러닝, 그 미묘한 차이



아이의 손을 찍은 X-ray 사진을 보고,
아이의 나이를 알아맞추시오.

- 전통적인 머신러닝

- 길이를 재기 위한 뼈들의 위치를 찾는다.
- 뼈의 길이에 따라 나이를 예측하도록 모델을 훈련시킨다.

- 딥러닝 (End-to-end learning)

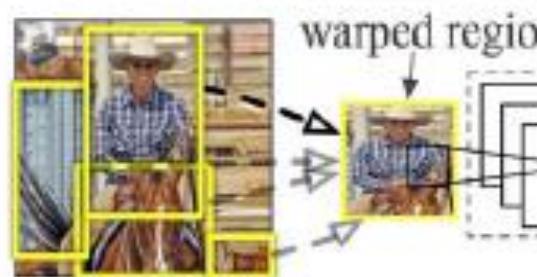
- (아주 많은) X-ray 이미지를 모델에게 준다.
- 모델이 이미지를 보고 아이의 나이를 예측하도록 학습한다.

PART4 활용사례

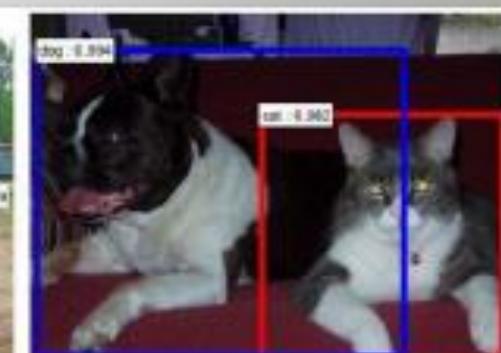
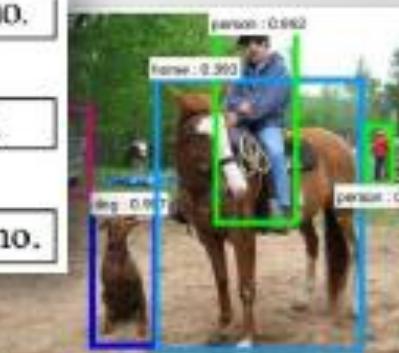
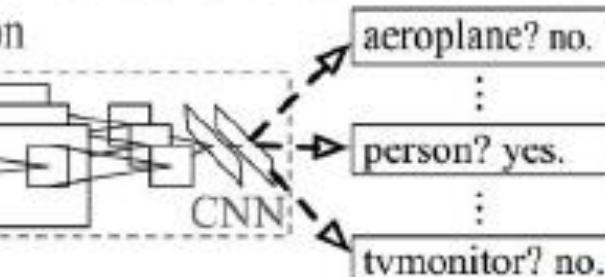
Object Detection (사물 검출)

: 이미지 속에 담긴 사물의 위치와 종류를 알아내는 기술

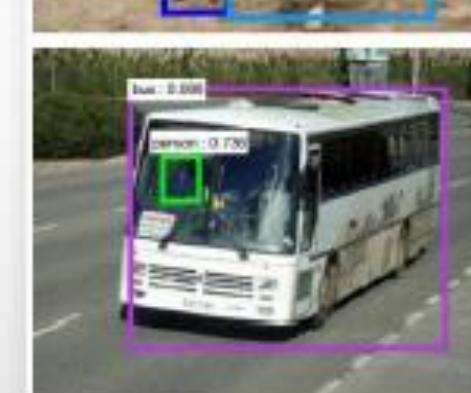
입력 이미지에서 후보 영역을 추출한 후 CNN을 적용하여 해당 영역에 무엇이 있는지 예측함



warped region

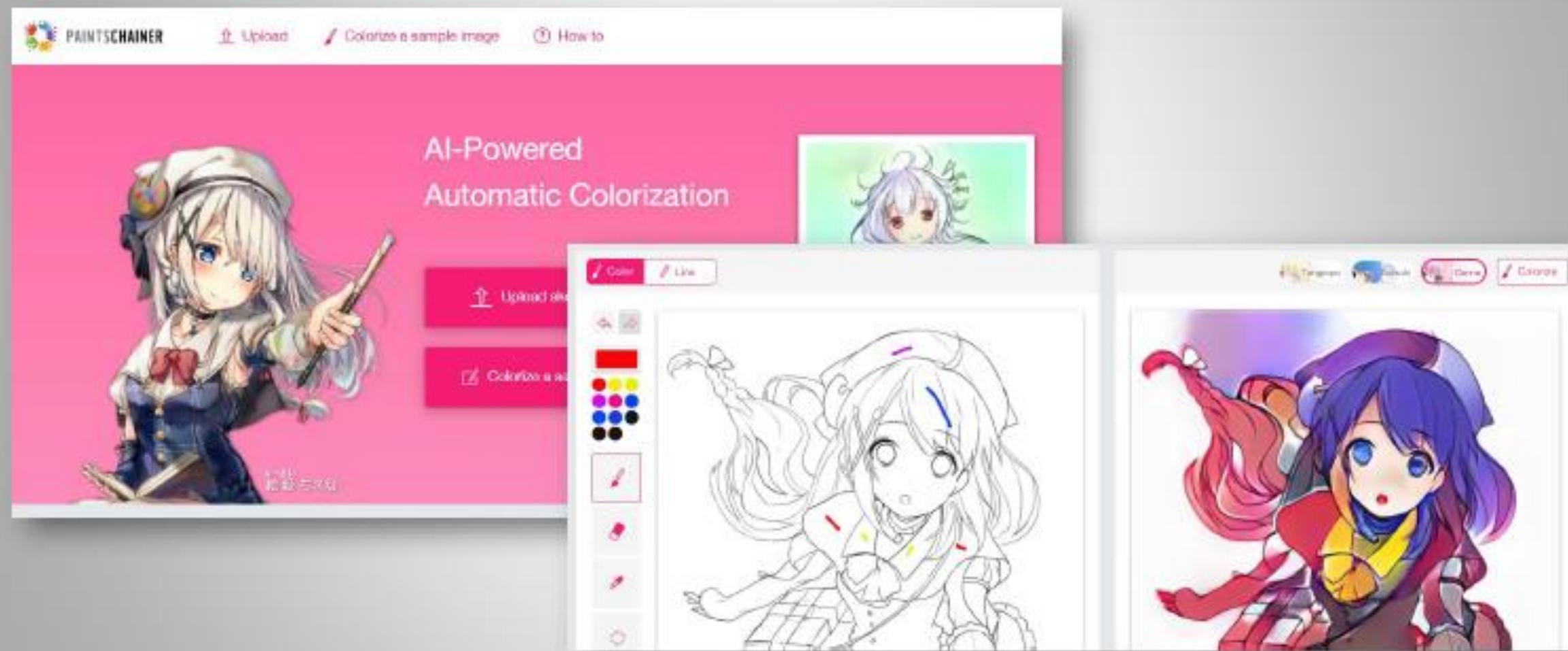


Faster R-CNN (Regions with Convolutional Neural Network)



딥러닝 스케치 채색 도구 PaintsChainer

: <https://goo.gl/xRWV00>

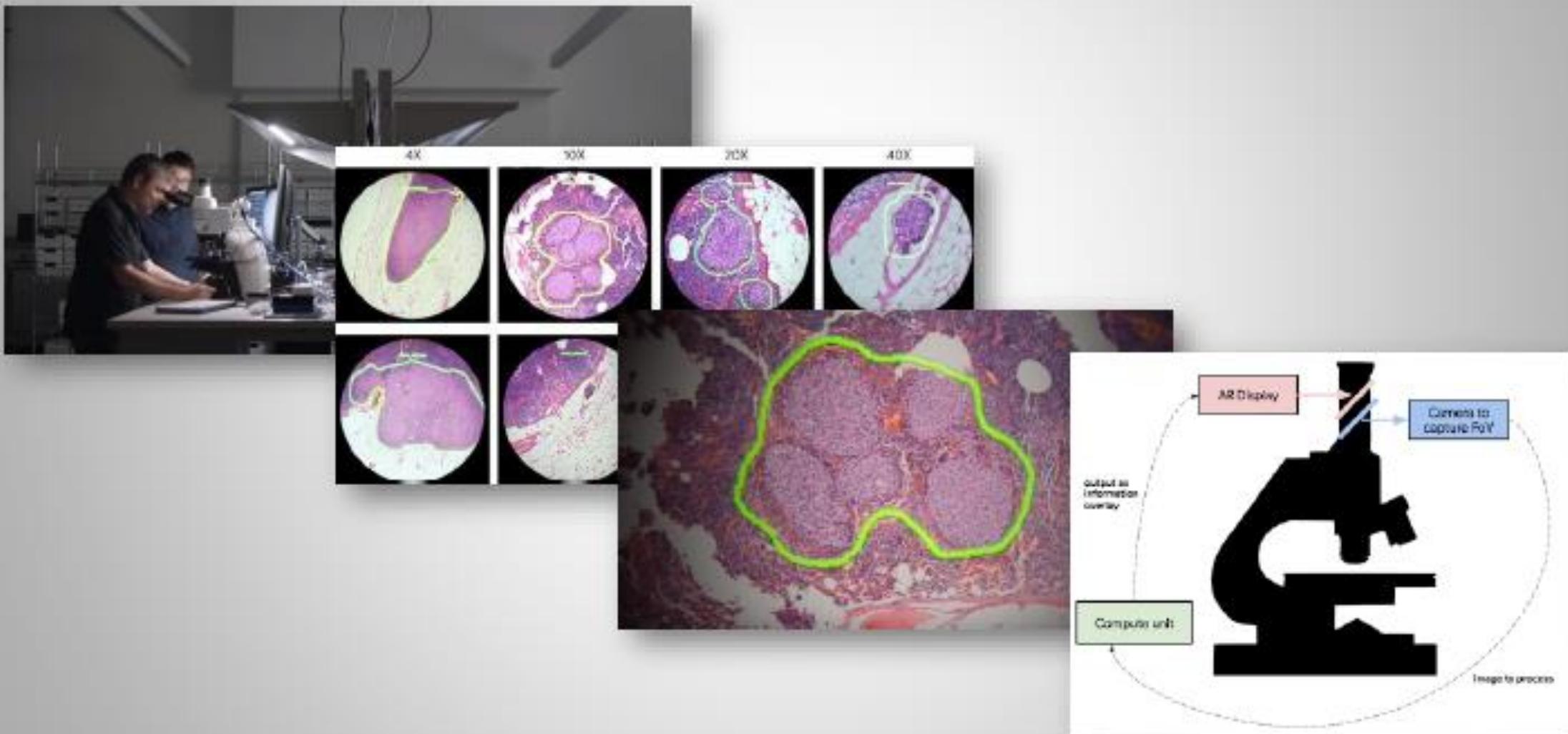


* MangaCraft (The current best semi-automatic manga colorization project on earth) @ <https://goo.gl/zcuxeu>

본 교안은 11월 8일 강의를 위해 제작되었으며, 강의 외 배포/개시/공개를 금합니다.

딥러닝 & AR 기반 암세포 감지 현미경 (Google, Prototype)

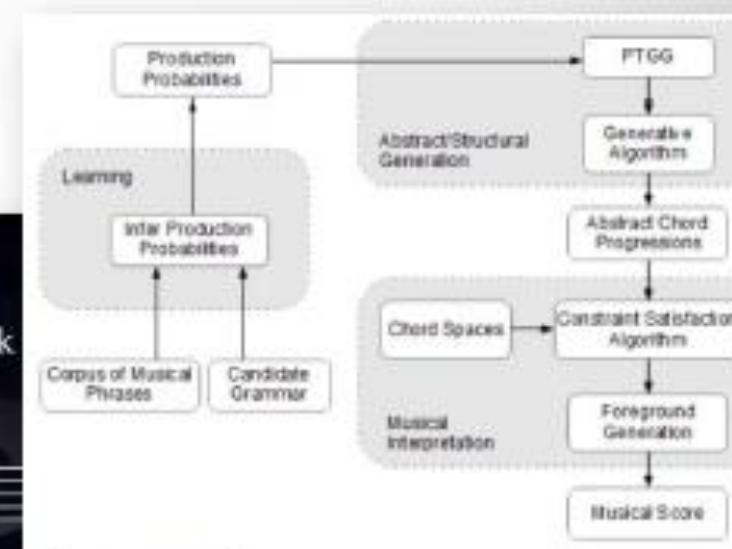
An Augmented Reality Microscope for Cancer Detection : <https://goo.gl/ZRqKe6>



딥러닝을 활용한 음악 작곡

: Kulitta – <http://donyaquick.com/kulitta/>

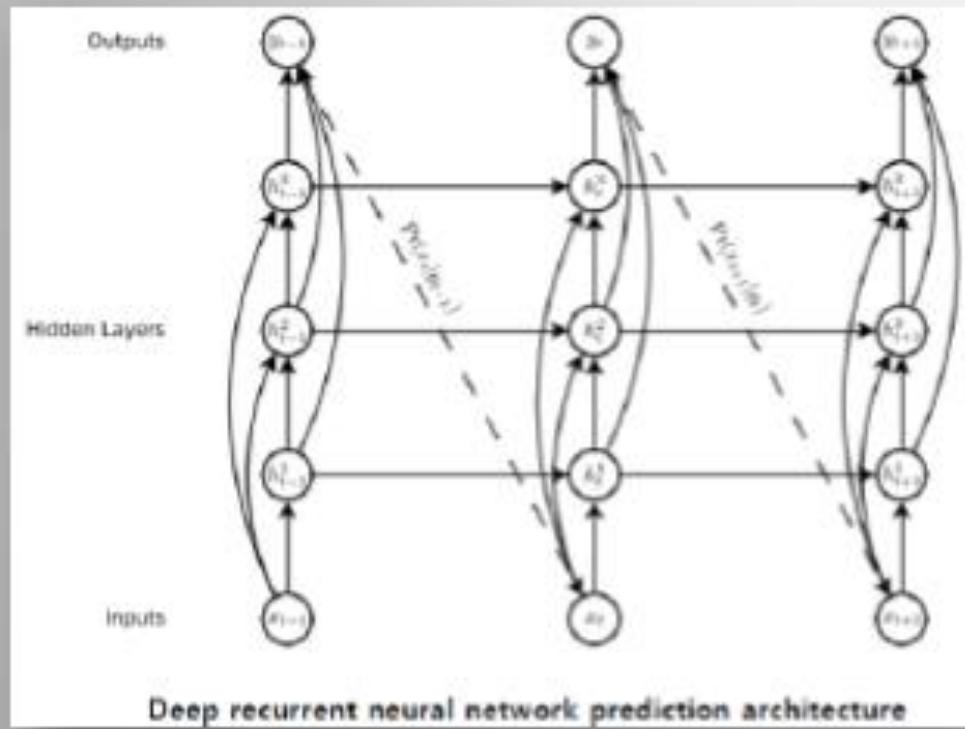
- 예일대 컴퓨터 사이언스 박사 Donya Quick이 개발한 작곡을 위한 인공지능 모델
- 딥러닝을 활용해 특정 장르의 음악을 스스로 학습하고 비슷한 특성을 가진 창작곡을 내놓는 방식
- Blind test 결과 '사람이 만든 음악'으로 평가를 다수 받음



딥러닝을 활용한 필기체 생성

: Generating Sequences With Recurrent Neural Networks @ <https://arxiv.org/pdf/1308.0850.pdf>

- RNN 의 발전된 형태인 LSTM-RNN 를 이용하여 손글씨를 만드는 기술을 개발 (토론토 대학)
- Demo @ <https://www.cs.toronto.edu/~graves/handwriting.html>



HANDWRITING GENERATION

Recurrent neural network handwriting
 Recurrent neural network handwriting
 Recurrent neural network handwriting

PART5 데이터관련 추천학습자료

무엇을 공부하면 좋을까요?

모두를 위한 머신러닝/딥러닝 강의

모두를 위한 머신러닝과 딥러닝의 강의

알파고와 이세돌의 경기를 보면서 이제 머신 러닝이 인간이 잘한다고 여겨진 직관과 의사 결정능력에서도 충분한 데이터가 있으면 어느정도 또는 우리보다 더 잘할수도 있다는 생각을 많이 하게 되었습니다. Andrew Ng 교수님이 말씀하신것 처럼 이런 시대에 머신 러닝을 잘 이해하고 잘 다룰수 있다면 그야말로 "Super Power"를 가지게 되는 것이 아닌가 생각합니다.

더 많은 분들이 머신 러닝과 딥러닝에 대해 더 이해하고 본인들의 문제를 이 멋진 도구를 이용해서 풀수 있게 하기위해 비디오 강의를 준비하였습니다. 더 나아가 이론에만 그치지 않고 최근 구글이 공개한 머신러닝을 위한 오픈소스인 TensorFlow를 이용해서 이론을 구현해 볼수 있도록 하였습니다.

수학이나 컴퓨터 공학적인 지식이 없이도 쉽게 볼수 있도록 만들려고 노력하였습니다.

<http://hunkim.github.io/ml/>

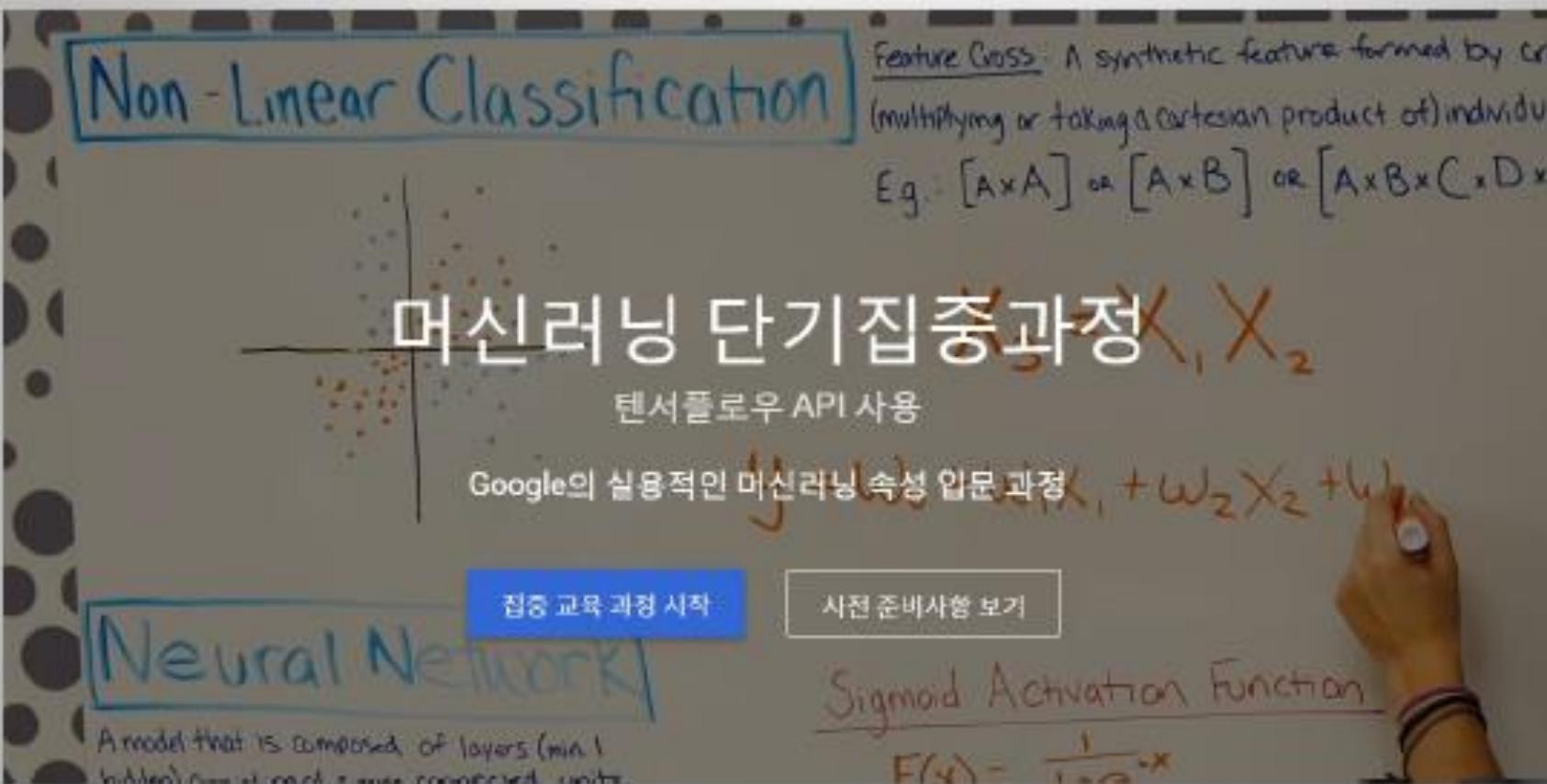
무엇을 공부하면 좋을까요?



The screenshot shows the 'Machine Learning' course page on Coursera. The left sidebar has links for 'Overview', 'Syllabus', 'FAQs', 'Creators', 'Ratings and Reviews', 'Enroll' (which says 'Starts Feb 05'), 'Apply for Financial Aid', and 'Training a team ? Try Coursera for Business'. The main content area shows the course title 'Machine Learning' and a brief description: 'About this course: Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it. Many...'. It also shows the course is 'Created by: Stanford University' with a Stanford University logo, and 'Taught by: Andrew Ng, Co-founder, Coursera; Adjunct Professor, Stanford University; formerly head of Baidu AI Group/Google Brain' with a circular profile picture of Andrew Ng.

<https://www.coursera.org/learn/machine-learning>

무엇을 공부하면 좋을까요?



<https://developers.google.com/machine-learning/crash-course/>

Recommended News letter & Facebook group list



DataCamp Weekly

Projects & Resources

[Understanding Blockchains by Coding One In R](#)

Implement a small blockchain in R and learn more about what a blockchain looks like and some of the core concepts behind it! The best way to understand what is happening under the hood is to build it yourself.

[datacamp.com](#)

[Data Science, the Environment and MOOCs \(with Roger Peng\)](#)

Air pollution, the environment, and data science: where do these intersect? Find out in Ep. 10 of the DataFramed Podcast.

[datacamp.com](#)

[Python Tuples Tutorial](#)

Learn about Python tuples: what they are, how to create them, when to use them, what operations you perform on them and various functions you should know.

[datacamp.com](#)

DataCamp weekly
@ <https://www.datacamp.com/home>

week, the newsletter
however, in place this
you to the
e full article. Also, the

Recommended News letter & Facebook group list

Kaggle
@ <https://www.kaggle.com/>

The screenshot shows the Kaggle homepage. At the top, there are two large, overlapping blue banners. The top banner features the word 'kaggle' in white, with 'Datasets' in a smaller font to its right. The bottom banner also features the word 'kaggle' in white. Below these banners, on the left, is a white sidebar containing a news snippet and a '5-Day SQL' course advertisement. The main content area on the right is titled 'Kaggle Cache' and lists 'New, intriguing, & overlooked datasets'. Below this, a numbered list of 10 datasets is provided. To the right of the list are four 'Learn Tracks' sections: 'Machine Learning' (with a 5-star icon), 'Python' (with a 3x3 grid icon), 'Data Visualization' (with a gear icon), and 'SQL' (with a database icon). At the bottom right is an 'R' icon with the text 'Learn the language designed for data analysis.'

kaggle Datasets

kaggle

Today we're excited BigQuery! We recently made it blazingly fast to query our training data.

5-Day SQL Scala

Need to learn or brush up on relational databases? On February 11, we're teaching SQL basics. [Learn it!](#)

Kaggle Cache

New, intriguing, & overlooked datasets

1. Visualizing Russian US Election Troll
2. 3 Million GitHub Repositories: A Big Cache
3. A Year in the International Wildlife Trade
4. A Very Chocolatey Analysis [\[link\]](#)
5. Austin Animal Shelter Outcomes [\[link\]](#)
6. Visualizing Connections in San Francisco
7. Big Data Analysis with Hacker News
8. notMNIST - MNIST, But Harder [\[link\]](#)
9. Ramen Lovers Just Love Eating [\[link\]](#)
10. Exploring Trendy YouTube Videos [\[link\]](#)

Learn Tracks

Get started building your machine learning skills or add a new area of expertise. Our tracks are suitable for beginners (with a bit of Python knowledge) and data scientists looking to expand their machine learning toolkit with more advanced topics.

Machine Learning
Get started with the fundamentals.

Python
Conquer data manipulation challenges.

Data Visualization
Make insightful and beautiful graphics.

SQL
Work with databases and Google BigQuery.

R
Learn the language designed for data analysis.

[Learn More Data Science](#)

Recommended News letter & Facebook group list

<https://www.facebook.com/groups/>

Python Korea
R Korea – KRSG
TensorFlow KR
케라스 코리아 (Keras Korea)
통계분석연구회

싸이그래머
데이터그램
캐글뽀개기



인공지능이란? 지능의 규정 범위에 따른 구분

Artificial Narrow Intelligence (“weak AI”)

- AI that specializes in **one specific area**.
- **Good at performing a single task**, such as playing chess or Go, making purchase suggestions.
- **Works within a very limited context**, and can't take on tasks beyond its field



Artificial General Intelligence (“strong AI”)

- AI that is **as smart as a human** and can perform any intellectual task that a human being can be.
- **Can understand and reason its environment** as a human would.



Artificial Super Intelligence

- AI that becomes **much smarter than the best human brains** in practically every field, including scientific creativity, general wisdom and social skills, we've achieved Artificial Super Intelligence.

교육적 시사점

- 1. 딥러닝과 머신러닝을 교육에 응용하려고 할 때 어떤 데이터들을 수집 할 수 있으며 현재 존재하는 데이터들은 어떤 것들이 있을까?
- 2. 학습을 분석하는데 어떤 모델들을 개발해야 할 것인가?
- 3. 학습을 촉진하기 위해 어떤 기능들을 만들어 낼 것인가?
- 4. 인공지능을 적응적 학습체제개발과 학습자분석에 사용될할 수 있는가?