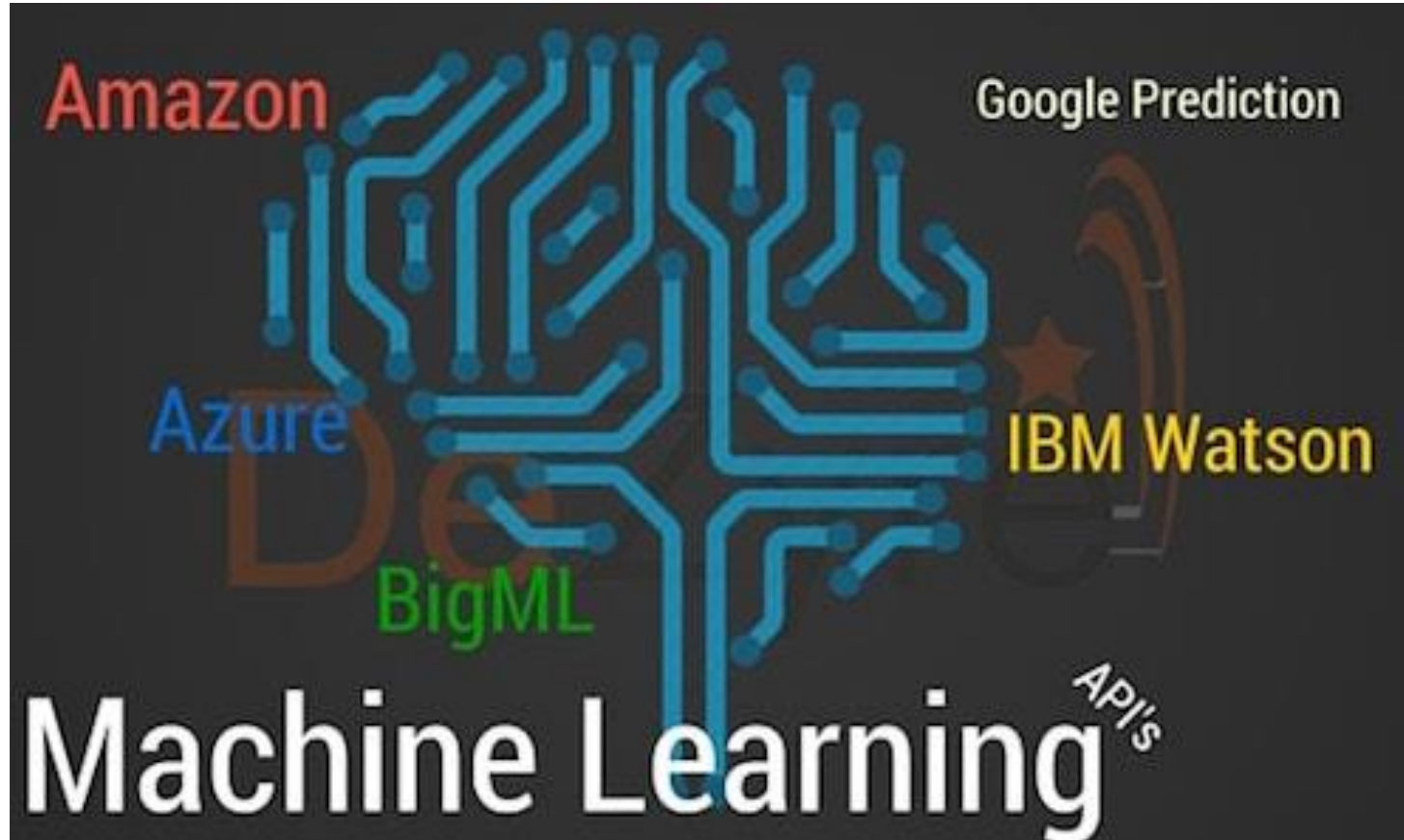


8장 머신 러닝: 인간의 학습 과정을 컴퓨터로 모방하기



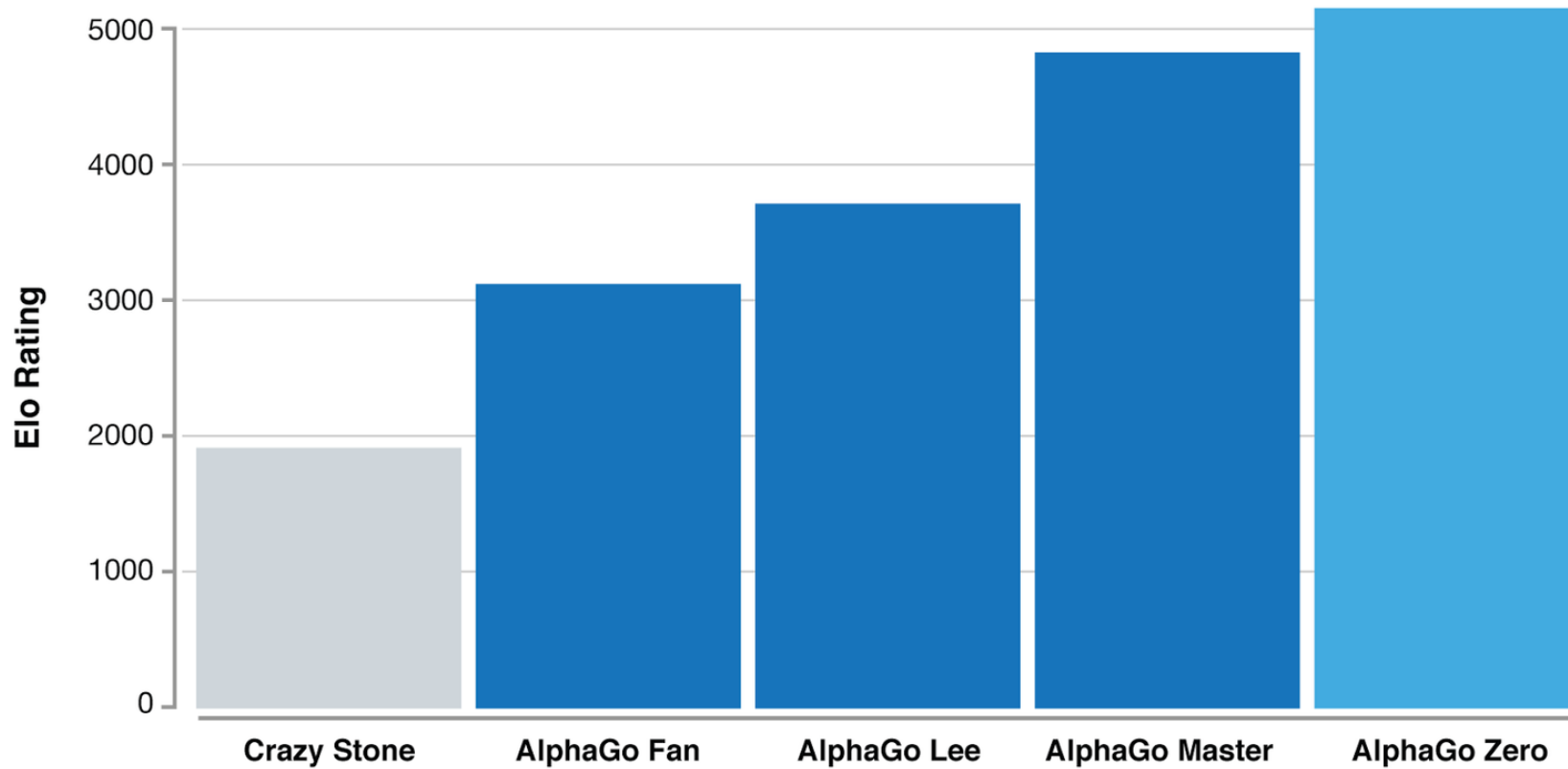
출처: <http://www.kdnuggets.com/2015/11/machine-learning-apis-data-science.html>

8장 머신 러닝: 적용 사례

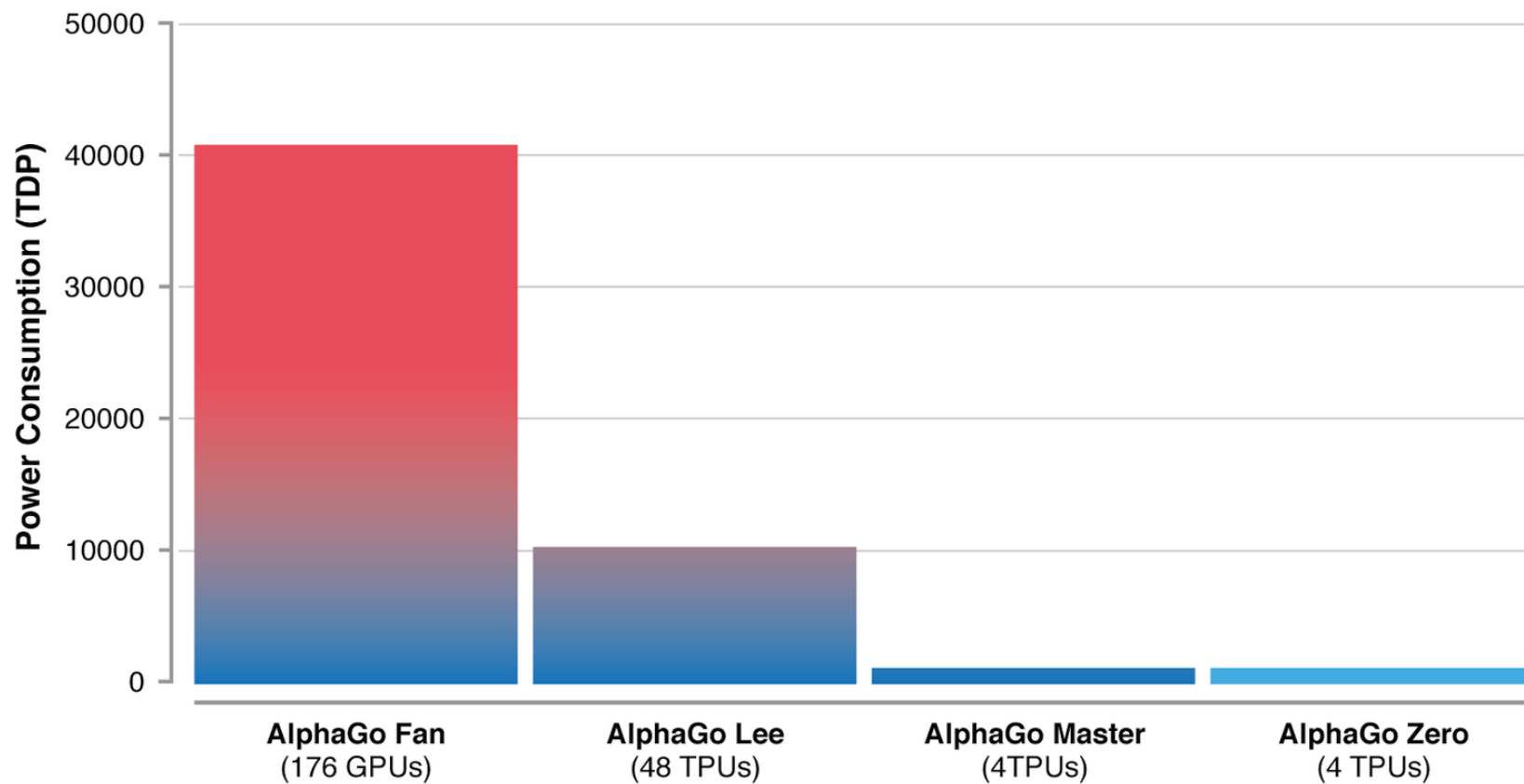


출처: <http://www.allchinatech.com/googles-alphago-beats-gos-world-champion-ke-jie-in-the-first-of-three-matches/>

결과



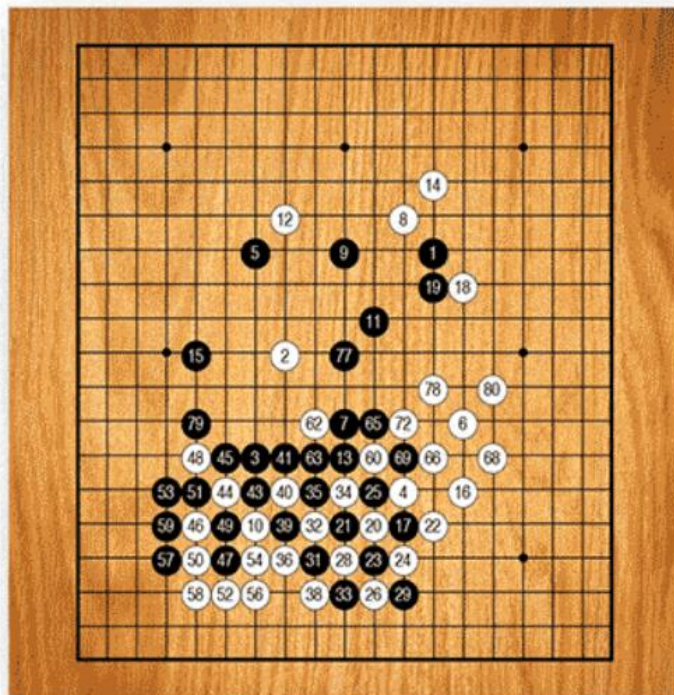
알파고



결과



알파고



3 hours

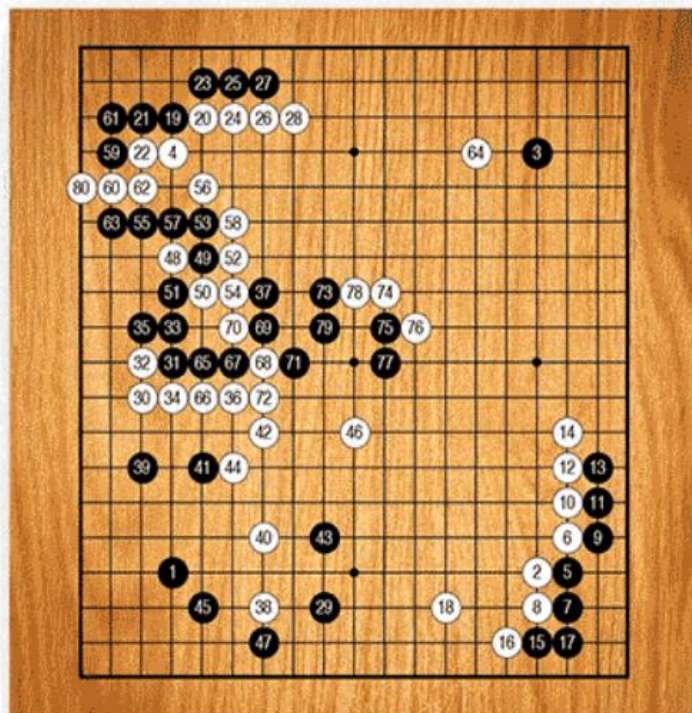
AlphaGo Zero plays like a human beginner, forgoing long term strategy to focus on greedily capturing as many stones as possible.

27 at 17 30 at 20 37 at 21 42 at 34 55 at 44 61 at 39
64 at 40 67 at 39 70 at 40 71 at 25 73 at 21 74 at 60
75 at 39 76 at 34

Captured Stones

4 of 4

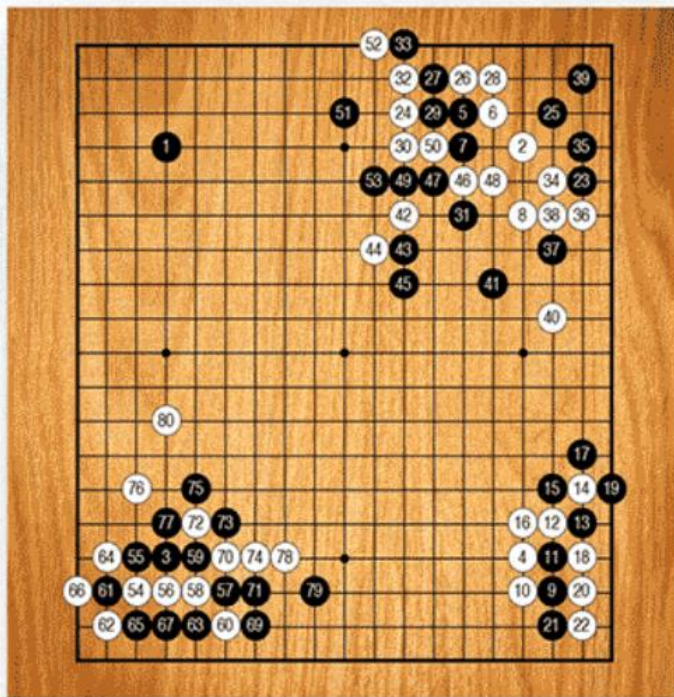
알파고



19 hours

AlphaGo Zero has learnt the fundamentals of more advanced Go strategies such as life-and-death, influence and territory.

알파고



68 at 61

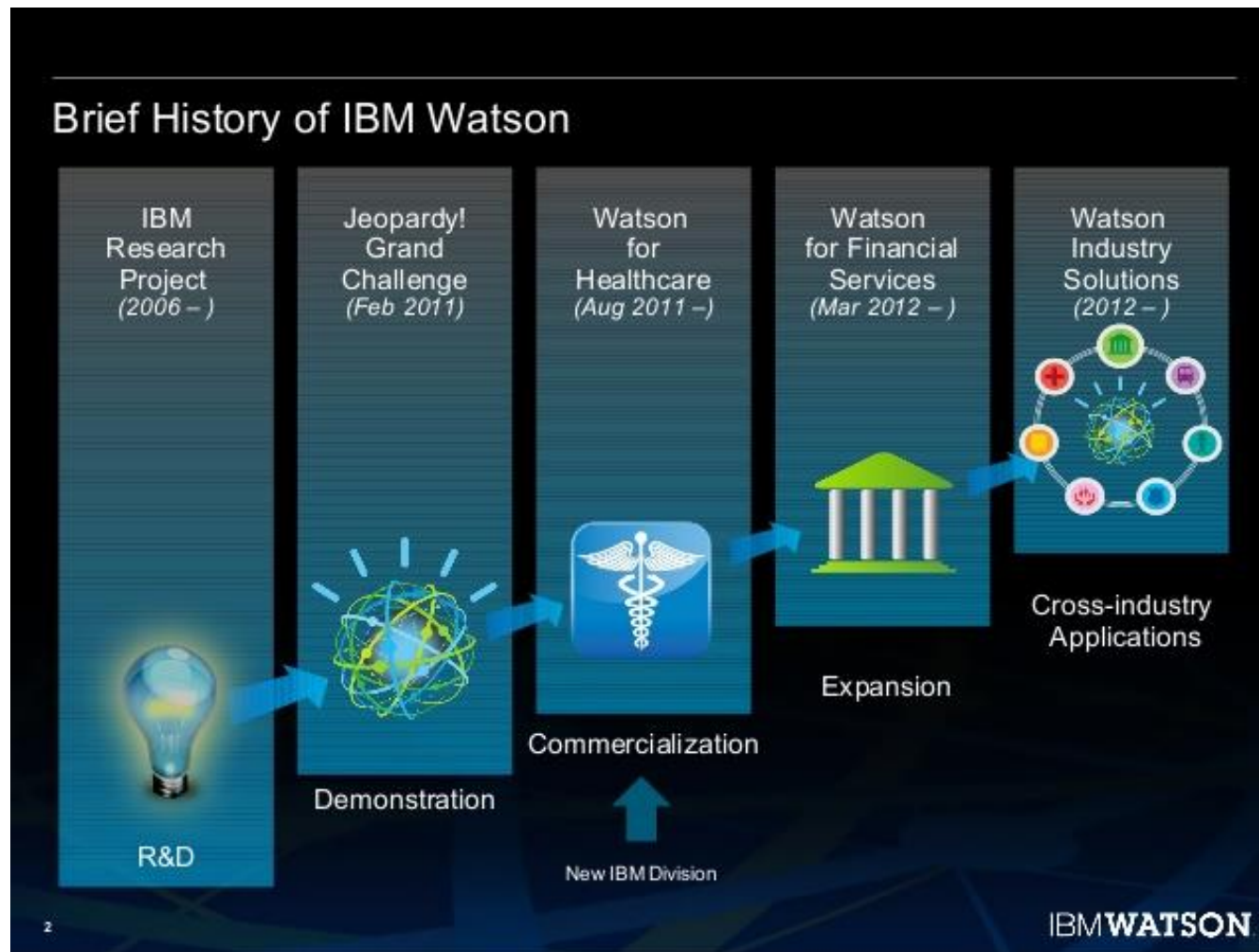
Captured Stones

70 hours

AlphaGo Zero plays at super-human level.
The game is disciplined and involves
multiple challenges across the board.

4 of 4

8장 머신 러닝: 적용 사례



출처: <https://krify.co/ibm-watson-empowering-mobile-app-developers-artificial-intelligence/>

8장 머신 러닝: 적용 사례

TIBCO | 10 Examples of Machine Learning

- Spam Detection
- Credit Card Fraud Detection
- Digit Recognition
- Speech Understanding
- Face Detection
- Shape Detection
- Product Recommendation
- Medical Diagnosis
- Stock Trading
- Customer Segmentation



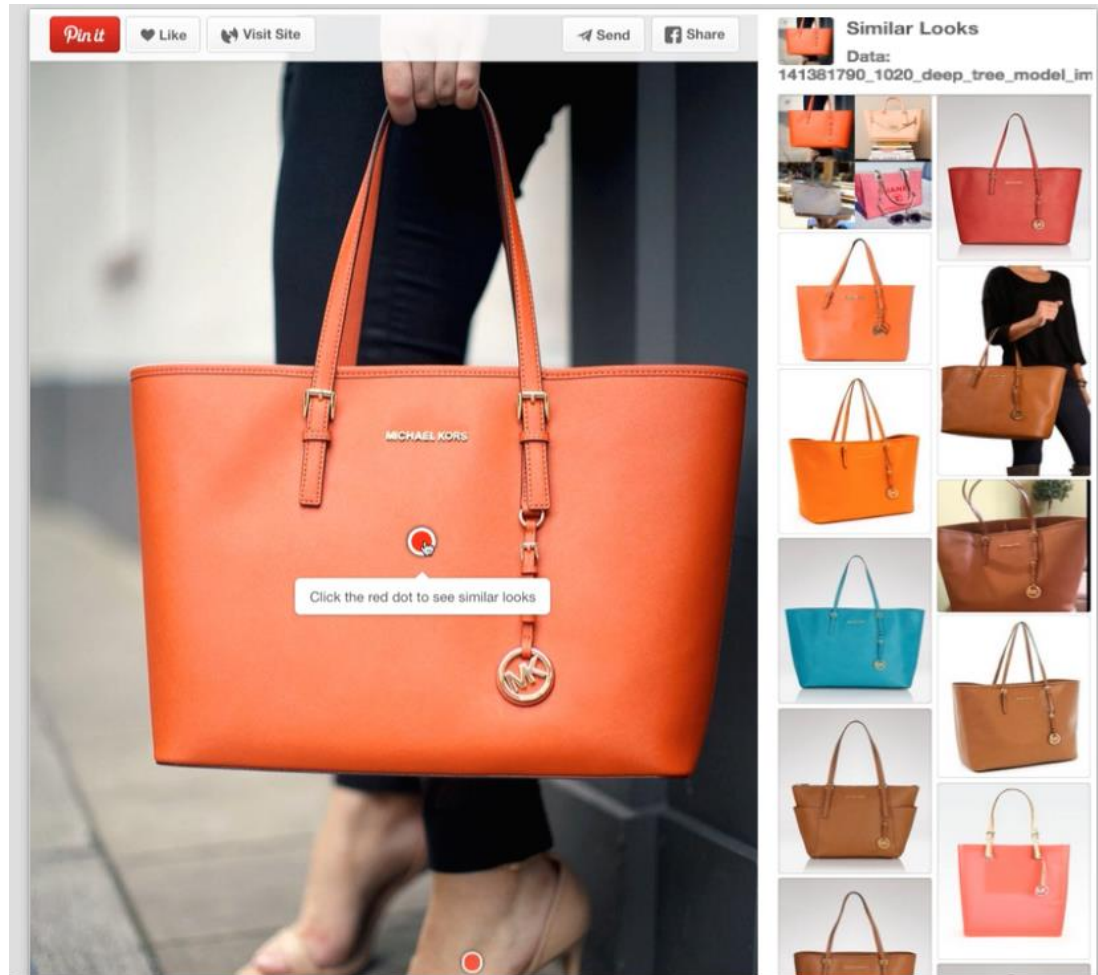
<http://machinelearningmastery.com/practical-machine-learning-problems/>

© Copyright 2000-2016 TIBCO Software Inc..

9

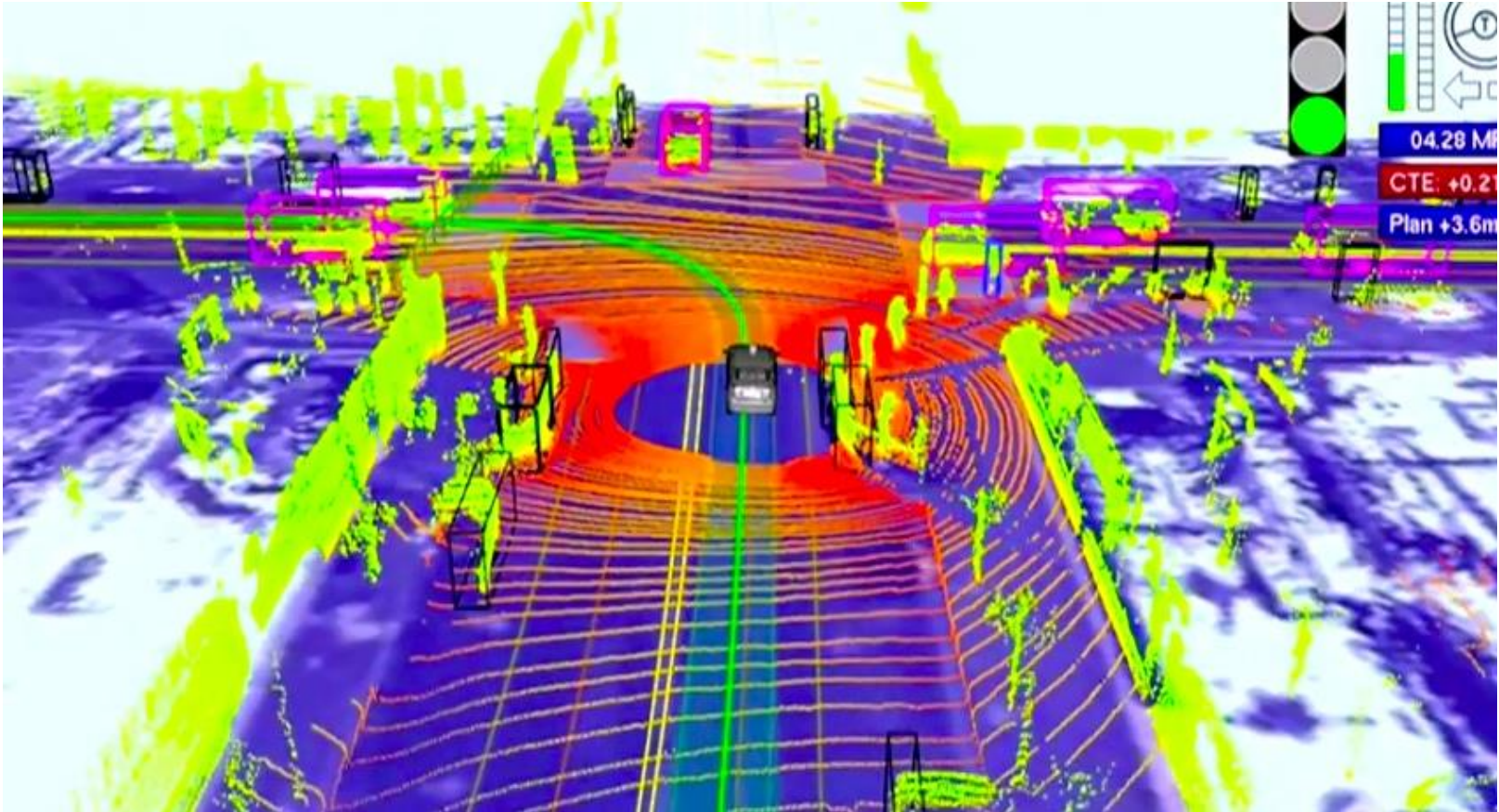
출처: <https://www.slideshare.net/KaiWaehner/how-to-apply-machine-learning-with-r-h20-apache-spark-mllib-or-pmml-to-real-time-streaming-analytics>

8장 머신 러닝: 적용 사례



Pinterest improves Related Pins with deep learning, plans **product recommendations** using object recognition

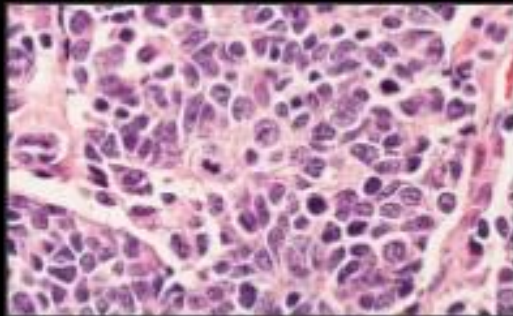
8장 머신 러닝: 적용 사례



<https://www.youtube.com/watch?v=51FS0t8hCSQ>

8장 머신 러닝: 적용 사례

DEEP LEARNING REVOLUTIONIZING MEDICAL RESEARCH



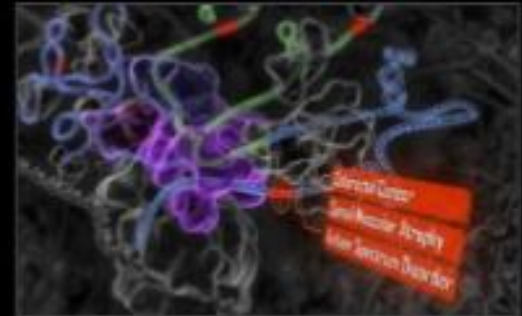
Detecting Mitosis in Breast Cancer Cells

— IDSIA



Predicting the Toxicity of New Drugs

— Johannes Kepler University



Understanding Gene Mutation to Prevent Disease

— University of Toronto

8장 머신 러닝: 인간의 학습 과정을 컴퓨터로 모방하기

- **학습 목표**
 - 머신 러닝의 기본 개념
 - 머신 러닝의 기술적 분류

8.1 머신 러닝의 기본 개념

- 기본적인 방법론으로는 다음과 같은 개념이 있음.
 - **교육 학습:** 교사의 지식을 형식 변환하여 기존 지식에 통합한다.
 - **연역 학습:** 기존 지식으로부터 구체적인 개념을 생성하여 새로운 지식으로 추가한다.
 - **귀납 학습:** 기존 지식 또는 새로운 교사 사례로부터 공통 개념을 추출하여 전체를 나타내는 지식으로 추가한다.
 - **강화 학습:** 환경과의 상호 작용에서 환경 적응 요인에 기초한 보상을 확인하면서 지식을 정돈한다.
 - **딥 러닝:** 주어진 환경에 적응하는 요인 자체를 자동으로 추출하고 그것을 최대화하는 쪽으로 지식을 정돈한다.
-
- 처음 3가지는 교사 사례가 있고 나머지 2가지는 직접적인 교사 사례가 없지만 학습 힌트는 필요할지 모름.

8.1.1 교육 학습

- **교육 학습(Learning by Being Told)**
 - 주어진 교사 사례를 그대로 기억하는 것.
 - 지식을 골라낼 때는 **교사 사례(Learning Examples)**와 완전히 같은 패턴으로 골라냄.
 - 말하자면, 학교 수업에서 학생이 선생님께서로부터 배운 대로 대답하는 것과 같음.

8.1.2 연역 학습

- 연역 학습(Deductive Learning)
 - 교사 사례를 그대로 기억할 뿐만 아니라 그것들을 조합하여 새로운 형태의 지식을 생성하므로 교사 사례와 다른 패턴도 처리할 수 있음.
 - 예를 들어, 3단 논법에 따라 (1)'개는 동물이다', (2)'동물은 움직인다'는 지식으로부터 (3)'개는 움직인다'고 지식을 유도할 수 있음.
 - 기존의 지식을 조합하여 새로운 지식을 유도하는 기법을 규칙 합성이라고 하며 증명 등에 응용 가능.
 - 단, 연역 학습에서는 기존 지식의 조합 범위에 머물러 교사 사례와 동등하거나 보다 상세화된 지식을 생성할 수는 있지만, 새로운 개념을 생성하는 것은 아님.

8.1.3 귀납 학습

- 귀납 학습(Inductive Learning)
 - 교사 사례를 보다 상위의 새로운 개념으로 집약하는 것이 가능하며 **개념 학습**에 상응.
 - 말하자면, 고등학교에 있어서 선생님이 말하고 있는 것은 결국 이런 것이구나 하고 학생들이 자신의 말로 이해하는 것.
 - 대표적인 기법으로는 복수의 교사 사례로부터 공통 요인을 추출하고 새로운 개념을 생성해 나가는 **버전 공간법**이 있음.
 - 이것은 **개념화 기법**으로, 고전적이기는 하지만 기본적인 개념이므로 나중에 자세히 설명함.

8.1.4 강화 학습

- 강화 학습(Reinforcement Learning)

- 교사 사례를 사용하지 않는 대신에 환경 적응에 따른 보상을 확인하면서 학습을 진행.
- 대표적인 기법으로는 Q 값이라고 부르는 평가값을 설정하여 이를 높여가는 '**Q 학습**'이 있음.
- 기본 원리는 매 time-step(t)마다 agent는 행동 a_t 를 선택하게 되고, 보상 r_t 를 받으며 새로운 상태 s_{t+1} 로 전이하고, Q 값이 갱신되는 과정이 반복됨.

$$Q(s_t, a_t) \leftarrow (1 - \alpha) \cdot \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} \right)$$

learned value

8.1.4 강화 학습

- α 가 1에 가까울수록 Q 값의 변화가 심한 학습이 되지만, 보통 학습 초기 단계에서는 α 를 크게, 최종 단계에서는 작게 설정.
- 강화 학습 분야는 최근 학습 연구의 중심에 있음.
- 더욱 발전한 발견 학습(Heuristic Learning)이라는 형태로 데이터 마이닝과 클러스터링에 사용되며 빅데이터 해석에도 응용되고 있음.

8.1.5 딥 러닝

- 딥 러닝(Deep Learning)
 - 근래에 가장 주목 받는 기법으로, **교사 사례도 보상도 없는 상태에서 주어진 환경에 적응.**
 - 예를 들어, 많은 사진 중에서 **고양이에 공통된 개념을 자동 추출하여 새롭게 주어진 사진이 고양이인지 아닌지 판단.**
 - 최근까지의 학습 기법에서는 **교사 신호가 필요하거나 교사가 없는 경우에도 무엇에 주목하고 학습하는가 하는 특징을 인간이 주지 않으면 안 되었음.**
 - **그러나 딥 러닝에서는 아무것도 주지 않아도 특징 자체를 추출하고 그 특징점에 따라 잡다한 데이터를 추상화 개념으로 정리하는 것이 가능함.**

8.1.5 딥 러닝

- 구글이 진행했던 고양이 인식은 고양이의 특징을 표현하는 데 필요한 요인을 인간이 부여한 것이 아니라, 딥러닝에 따라 자동으로 생성하여 새로운 영상 속의 동물을 고양이로 판단할 수 있었다는 점이 대단한 성과.



출처: <http://www.hardwarezone.com.my/feature-nvidia-gtc-2015-day-2-google-and-large-scale-deep-learning>

8.1.5 딥 러닝

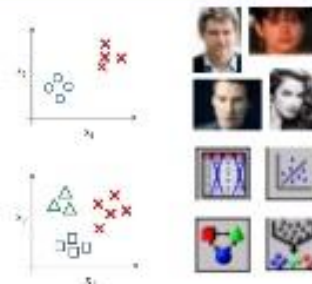
- 따라서 대량의 데이터를 정리하여 중요한 논점을 추출한다는 점에서 **빅데이터 해석에는 최적**.
- 추출된 특징 개념을 어떻게 부를지는 사람이 정하겠지만, 고양이 등과 같은 기존 개념에 머무르지 않고 **인간으로서는 생각해 낼 수 없었던 새로운 개념을 제시해 줄지도 모름**.
- 이로 인해 바야흐로 **발견 학습**에 다가섰다고 말할 수 있음.
- 단, '**주어진 데이터 또는 환경으로부터 완전히 기상천외한 아이디어가 나올까?**'라고 묻는다면, 그것은 아직 연구 단계에 있다고 생각됨.

8.2 머신 러닝의 기술적 분류

머신러닝의 세 가지 기본 학습 모드

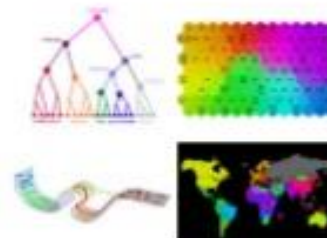
- **Supervised Learning (감독 학습, 지도 학습, 교사 학습)**

- 레이블(미리 정해놓은 정답) 달린 예제로 학습하기
- 예제가 매우 많은 경우 효과적인 학습이 가능하다
- 예) 분류(classification): 레이블이 이산적인(discrete) 경우.
- 예) 회귀(regression): 레이블이 연속적 값을 가지는 경우



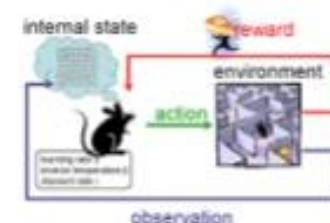
- **Unsupervised Learning (무감독 학습, 비지도 학습, 비교사 학습)**

- 데이터에 내재된 패턴, 특성, 구조를 학습을 통해 발견. 레이블은 고려하지 않음
- 학습 데이터는 개체에 대한 입력 속성만으로 구성됨 $D = \{(x)\}$
- 예: 차원 축소(dimension reduction), 군집화(clustering)



- **Reinforcement Learning (강화 학습)**

- 시스템의 동작이 적절성(right/wrong)에 대한 피드백이 있는 학습
- 소프트웨어 에이전트가, 환경(environment) 내에서 보상(rewards)이 최대화 되는 일련의 행동(action)을 수행하도록 학습하는 기법
- 환경의 상태, 에이전트의 행동, 상태 전이 규칙 및 보상, 관측 범위를 고려한 학습
- Action selection, planning, policy learning



8.2.1 머신 러닝의 기술적 주요 알고리즘

Machine Learning Algorithms *(sample)*

Unsupervised

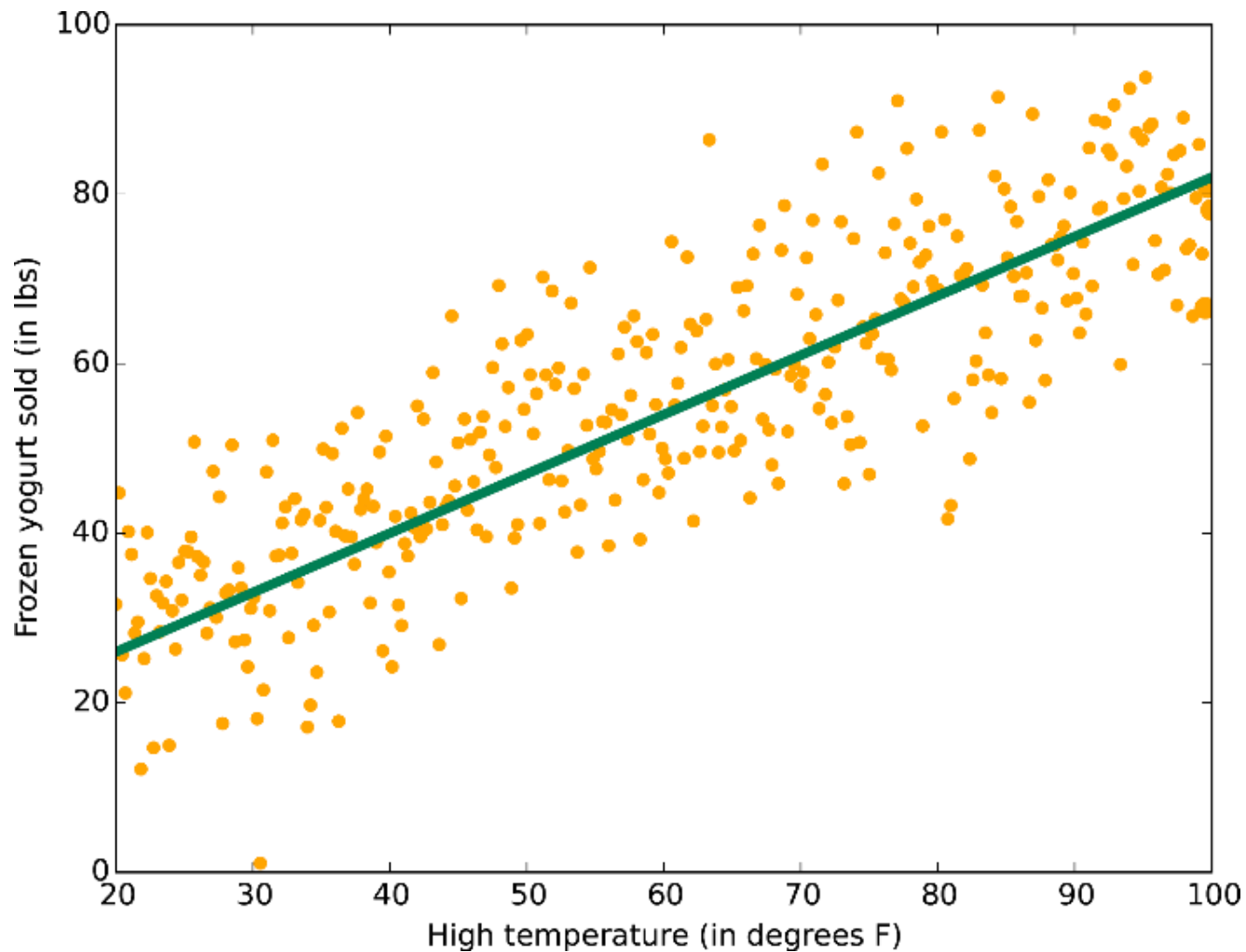
- Clustering & Dimensionality Reduction
 - SVD
 - PCA
 - K-means
- Association Analysis
 - Apriori
 - FP-Growth
- Hidden Markov Model

Supervised

- Regression
 - Linear
 - Polynomial
- Decision Trees
- Random Forests
- Classification
 - KNN
 - Trees
 - Logistic Regression
 - Naive-Bayes
 - SVM

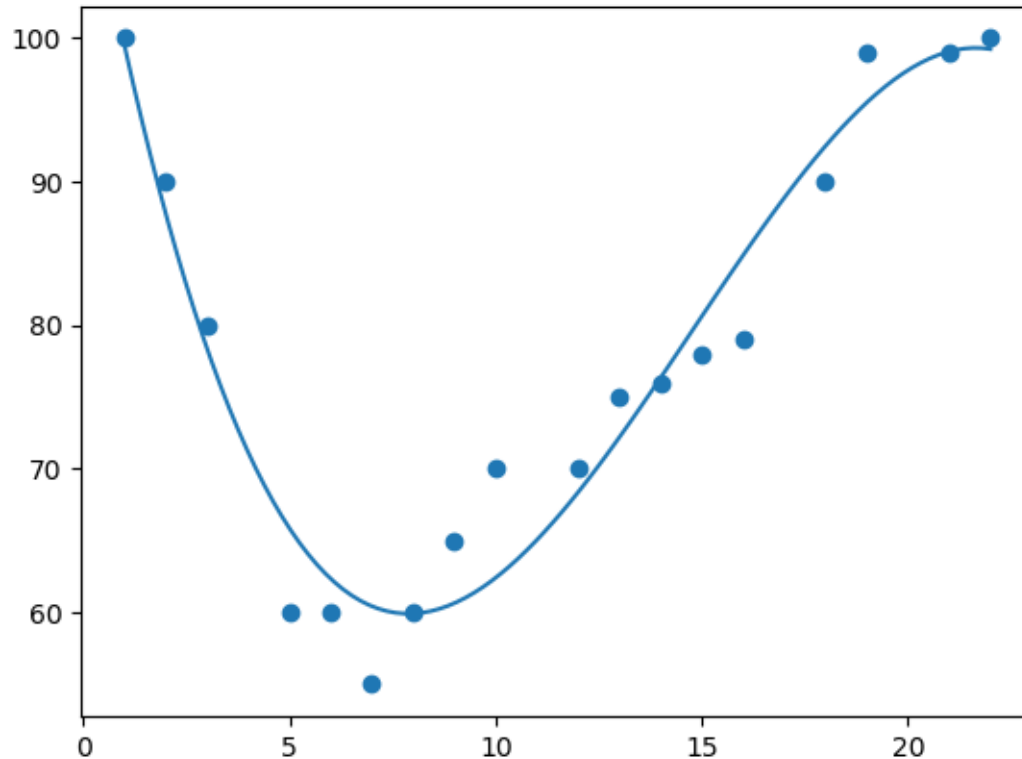
출처: <https://www.linkedin.com/pulse/machine-learning-primer-avinash-shenoy>

8.2.2 Linear Regression



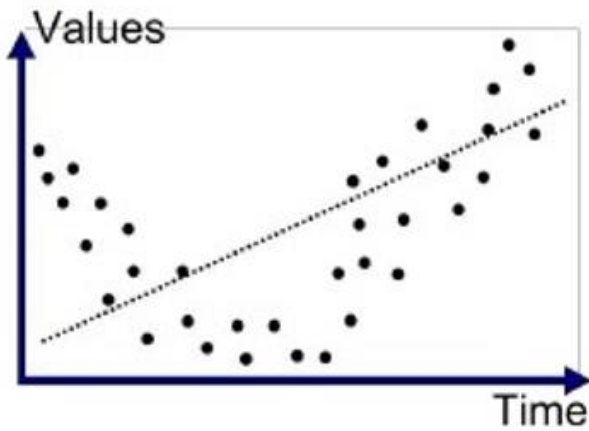
출처: <https://docs.microsoft.com/ko-kr/azure/machine-learning/machine-learning-algorithm-choice>

* Polynomial Regression

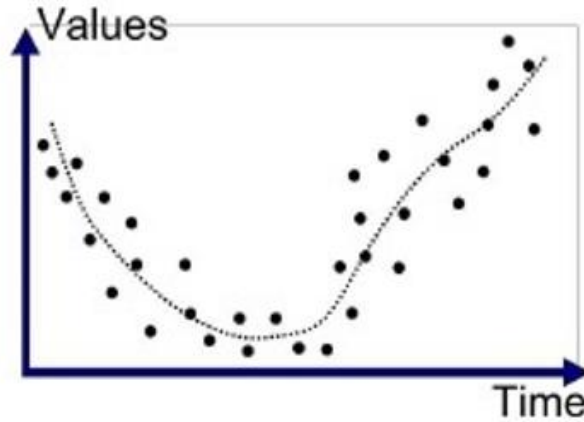


출처: https://www.w3schools.com/python/python_ml_polynomial_regression.asp

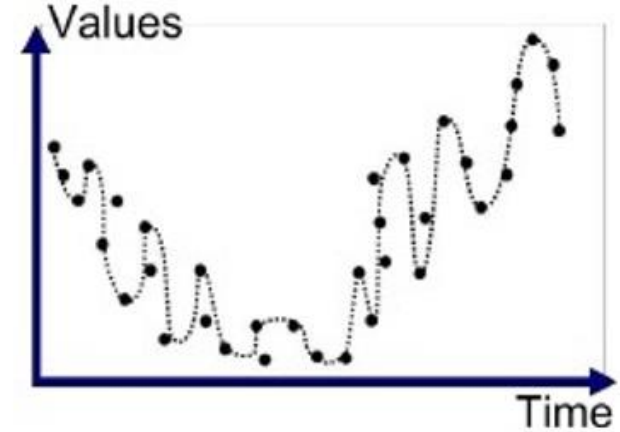
* Overfitting, Underfitting



Underfitted



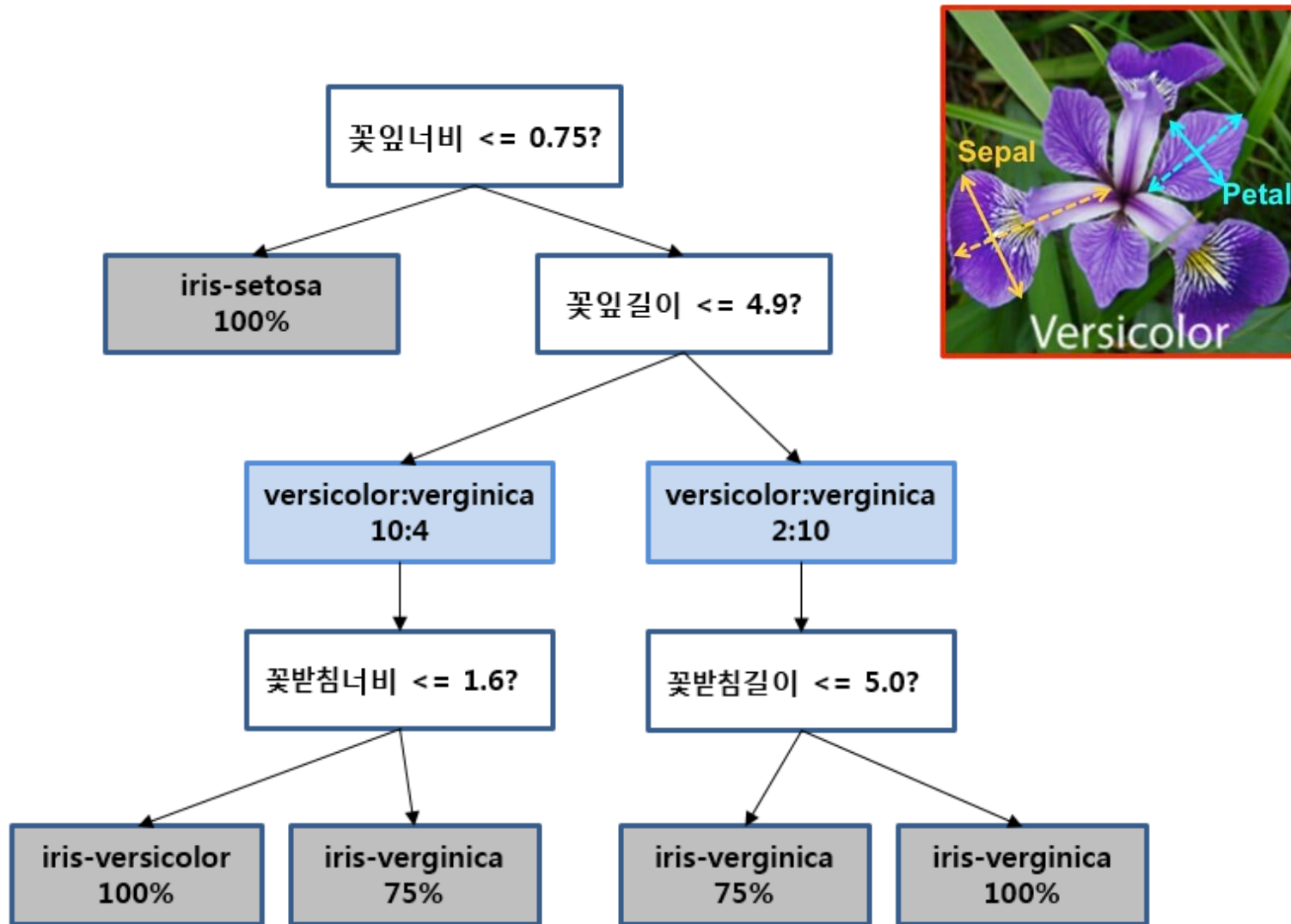
Good Fit/Robust



Overfitted

출처: <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>

8.2.3 Decision Tree



[illegible]




Supported By: In Collaboration With:

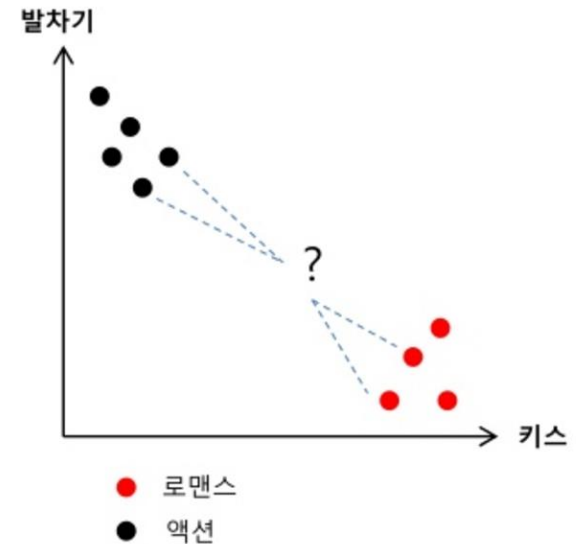
8.2.4 KNN(K-Nearest Neighbor)

kNN 예제

Press Esc to exit full screen

- 발차기와, 키스라는 항목으로, ? 대상의 유형을 유추

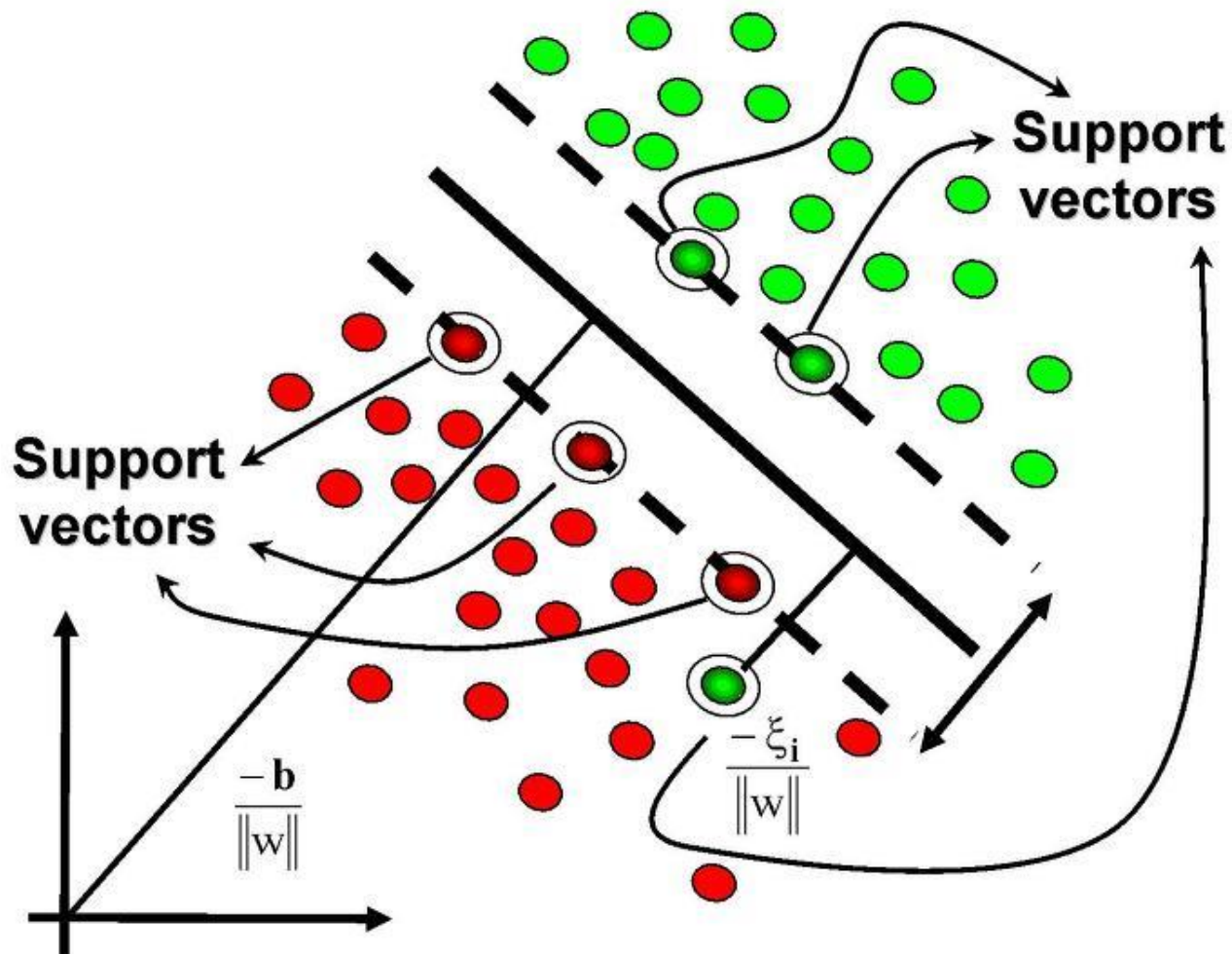
제목	발차기	키스	유형	?와의 거리
California Man	3	104	로맨스	20.5
He's Not Really into Dudes	2	100	로맨스	18.7
Beautiful Woman	1	81	로맨스	19.2
Kevin Longblade	101	10	액션	115.3
Robo Slayer 3000	99	5	액션	117.4
Amped II	98	2	액션	118.9
?	18	90	-(예측 대상)	



?(예측하고자 하는) 항목과 가장 거리가 가까운 항목의 "유형(라벨)"로 유출
→ 영화,쇼핑몰등 추천 시스템에 사용

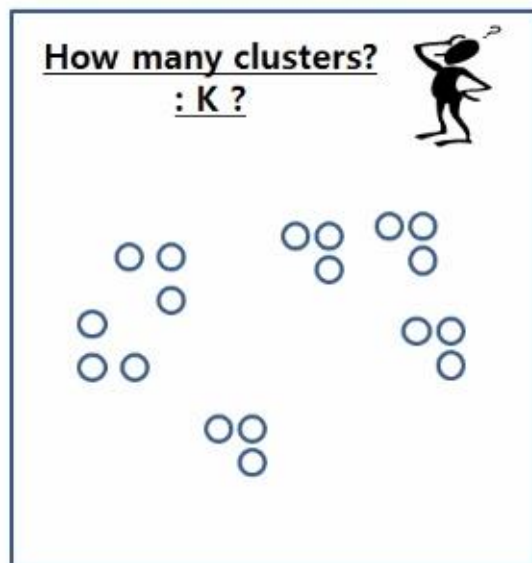
출처: <https://www.slideshare.net/Byungwook/1-knn>

8.2.5 SVM(Support Vector Machine)



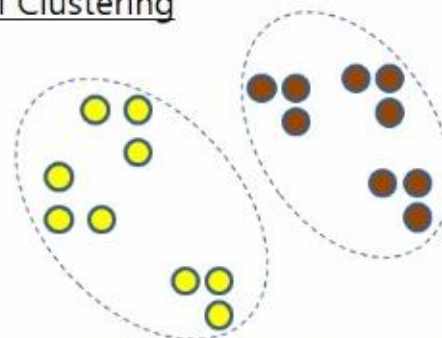
출처: <http://jangjy.tistory.com/96>

8.2.6 K-means clustering

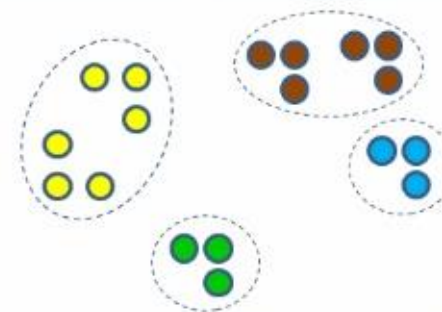


Partitional Clustering

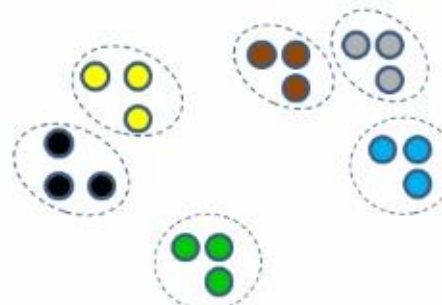
K = 2 ?



K = 4 ?



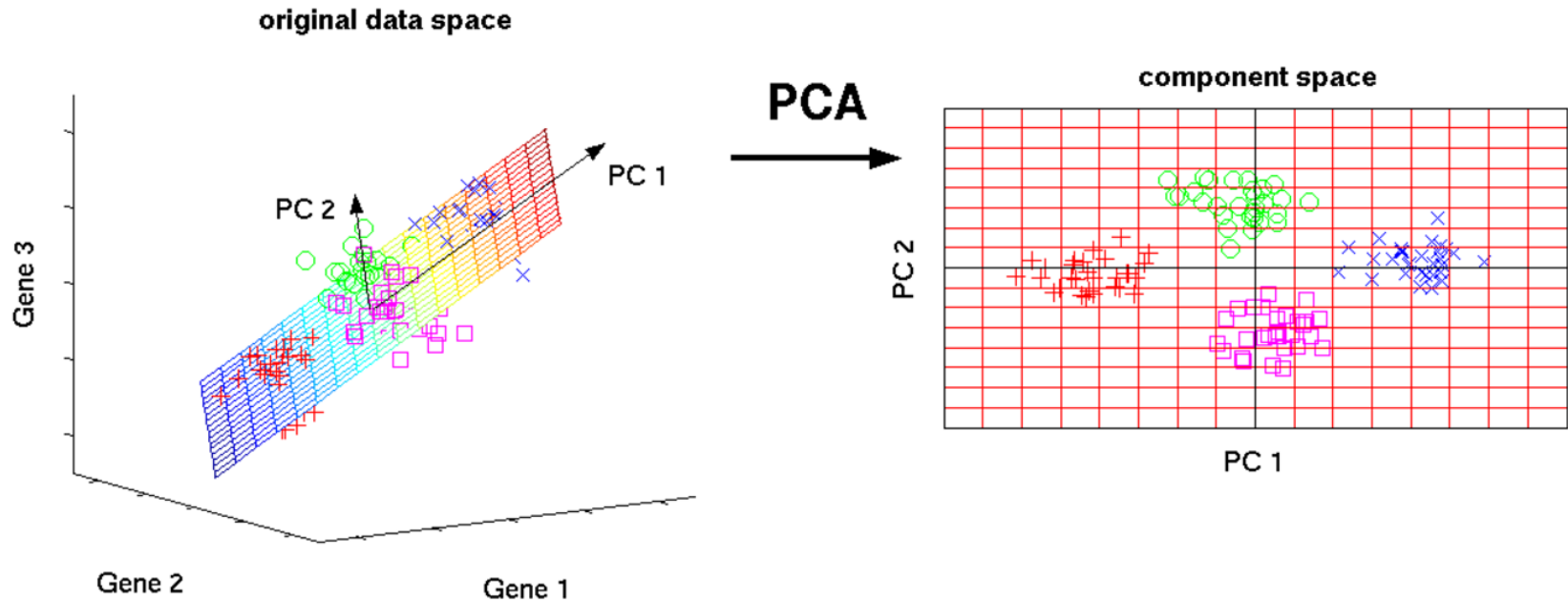
K = 6 ?



[Reference] Determining the number of clusters in a data set
⇒ https://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set

[R 분석과 프로그래밍] <http://rfriend.tistory.com>

8.2.7 PCA(Principal Component Analysis)



출처: <http://dokim.tistory.com/entry/PCA-Principal-Component-Analysis-주성분분석>

8.3 딥 러닝 관련 주요 이슈



출처: <http://twistedsifter.com/2016/03/puppy-or-bagel-meme-gallery/>

8.3 딥 러닝 관련 주요 이슈



chocolate cookie



[unknown]



fawn smooth Chihuahua



brown coated Chihuahua



baked blueberry muffin



baked muffin



white chihuahua



beige short coated puppy



fawn smooth Chihuahua



tan smooth Chihuahua puppy



blueberry muffin



blueberry cupcakes



fawn smooth Chihuahua



three smooth Chihuahua puppies

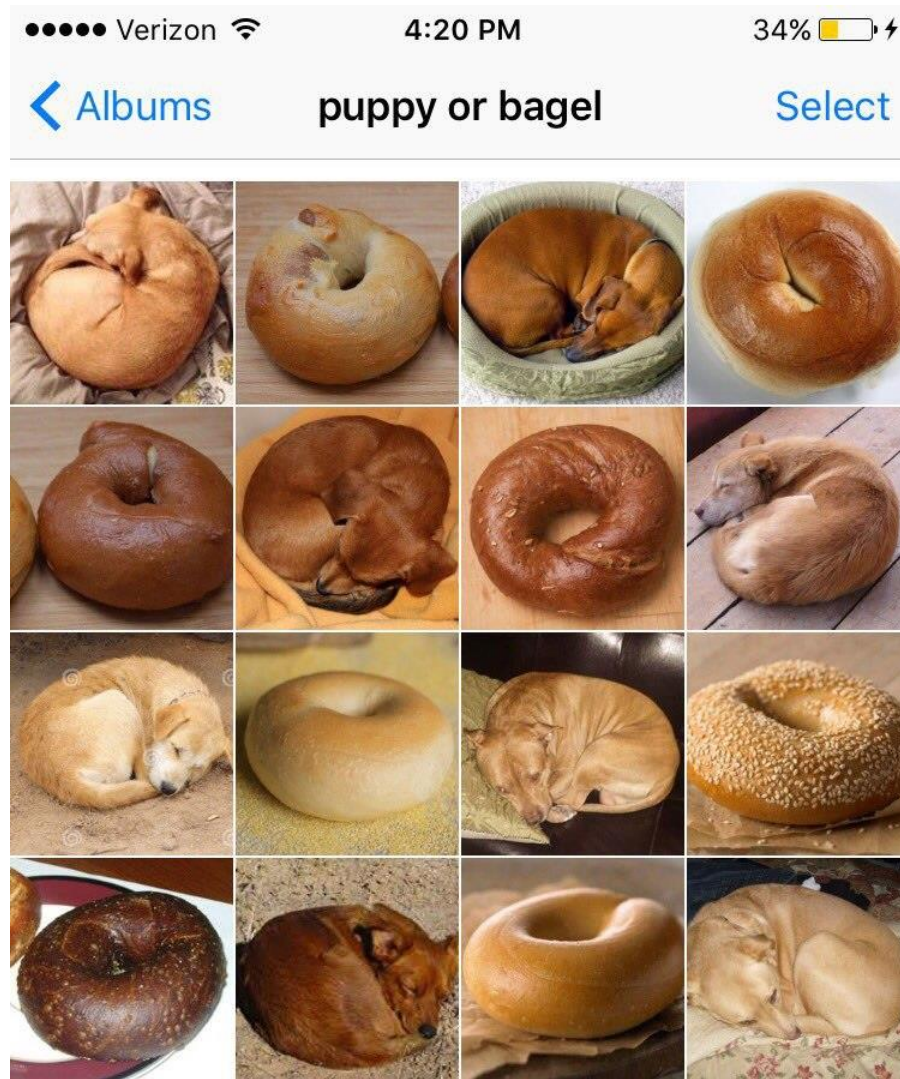


muffin



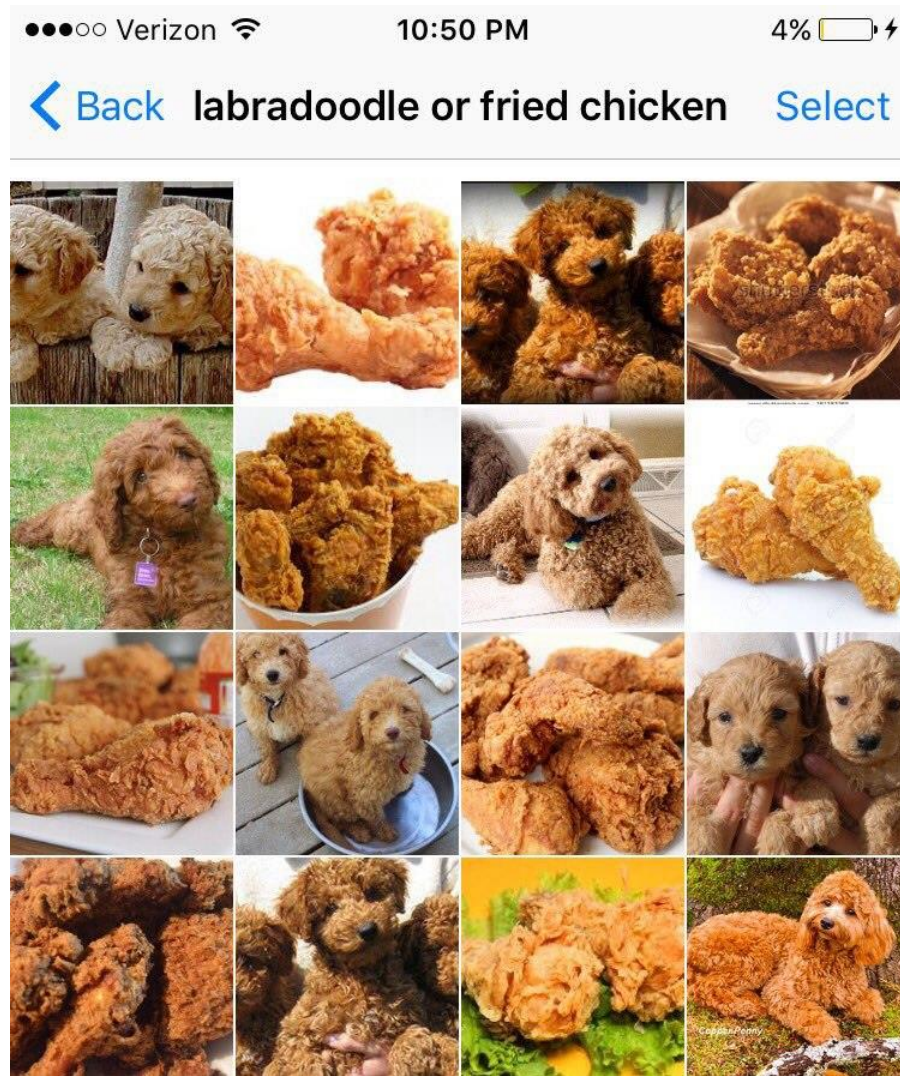
white and black muffin

8.3 딥 러닝 관련 주요 이슈



출처: <http://twistedsifter.com/2016/03/puppy-or-bagel-meme-gallery/>

8.3 딥 러닝 관련 주요 이슈



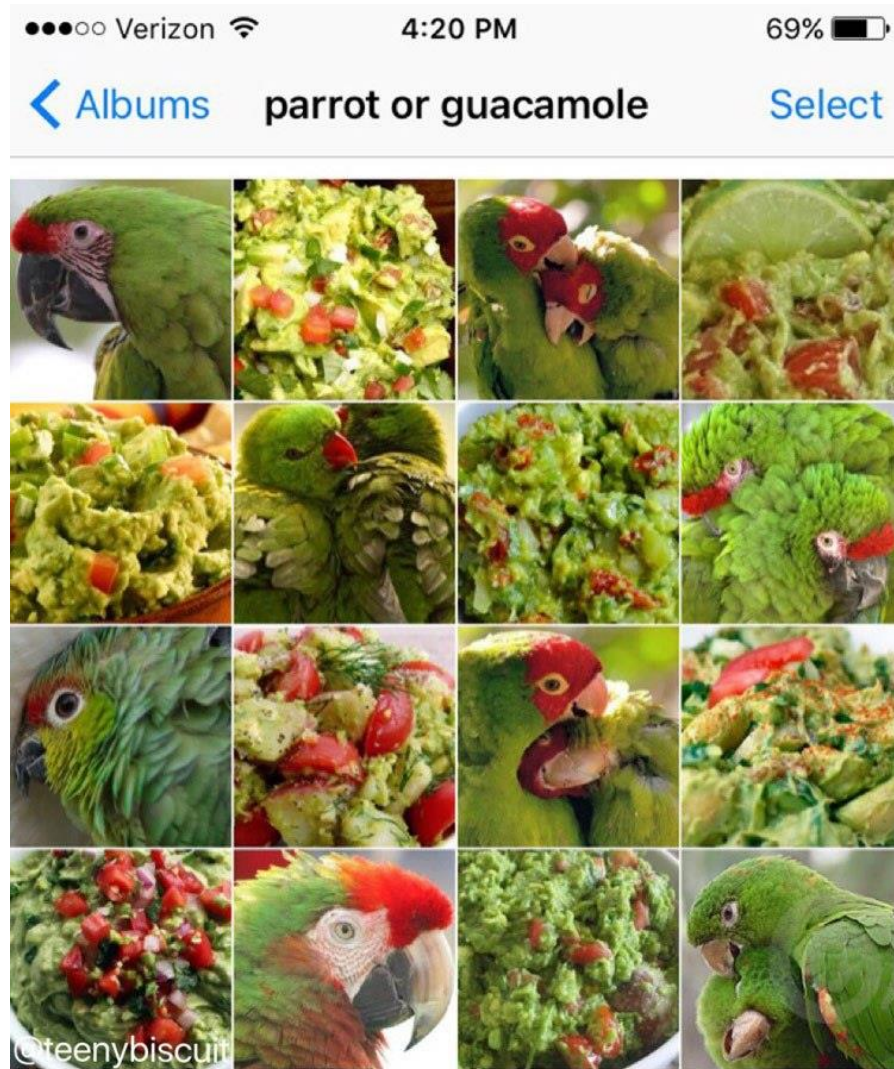
출처: <http://twistedifter.com/2016/03/puppy-or-bagel-meme-gallery/>

8.3 딥 러닝 관련 주요 이슈



출처: <http://twistedsifter.com/2016/03/puppy-or-bagel-meme-gallery/>

8.3 딥 러닝 관련 주요 이슈



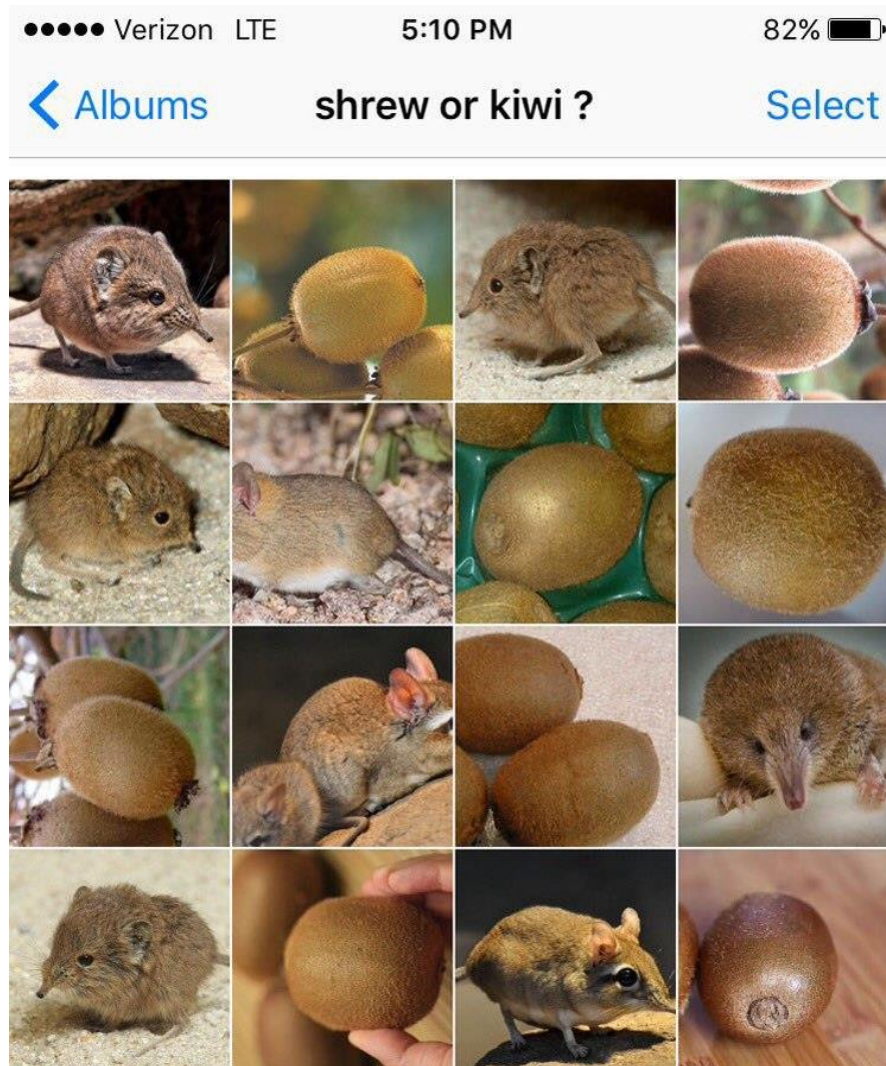
출처: <http://twistedsifter.com/2016/03/puppy-or-bagel-meme-gallery/>

8.3 딥 러닝 관련 주요 이슈



출처: <http://twistedrifter.com/2016/03/puppy-or-bagel-meme-gallery/>

8.3 딥 러닝 관련 주요 이슈



출처: <http://twistedifter.com/2016/03/puppy-or-bagel-meme-gallery/>

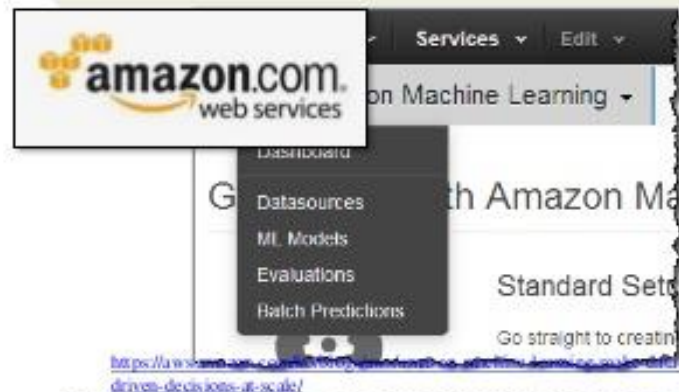
8.3 딥 러닝 관련 주요 이슈



8.4 머신 러닝 기능을 제공하는 상용 사이트

클라우드 기반 머신러닝 기술

클라우드(Cloud) 환경에서 빅데이터(Big Data)를 활용하는 머신러닝(Machine Learning) 기술이 지능형 서비스/솔루션의 핵심 포맷으로 부각되고 있다.



© 2016 SNU Biointelligence Laboratory, <http://bi.snu.ac.kr/>

70

8.5 머신 러닝 관련 사이트

- 상용 사이트 주소
- <https://console.bluemix.net/catalog/services/machine-learning/>
- <https://aws.amazon.com/machine-learning/>
- <https://azure.microsoft.com/ko-kr/services/machine-learning/>
- <https://cloud.google.com/products/machine-learning/>
- 학습 자료
- <https://www.slideshare.net/Byungwook/1-knn>
- <https://www.slideshare.net/TerryTaewoongUm/machine-learning-54531674>