



# 데이터 마이닝

구름

도시공학과 일반대학원

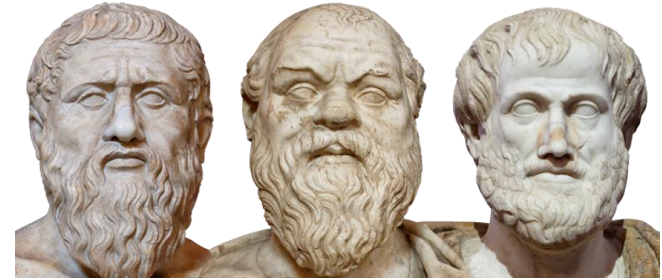
한양대학교

# 통계학 - Statistics

그리스

**Metaphysical**

Socrates, Platon, Aristoteles



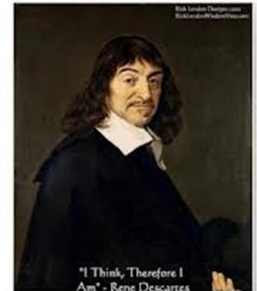
17세기

**Empiricism,  
Rationalism**

Francis Bacon, René Descartes



Francis Bacon  
(1561-1626)

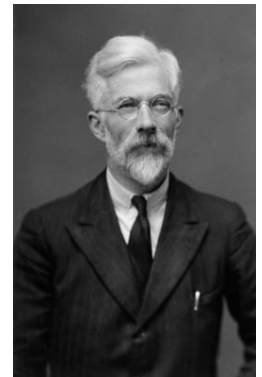
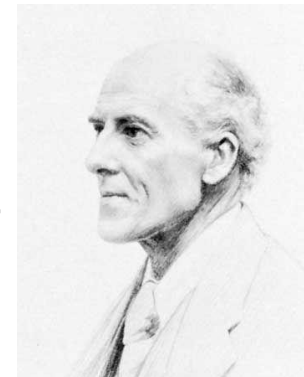


René Descartes  
(1596-1650)

19세기

**Statistics**

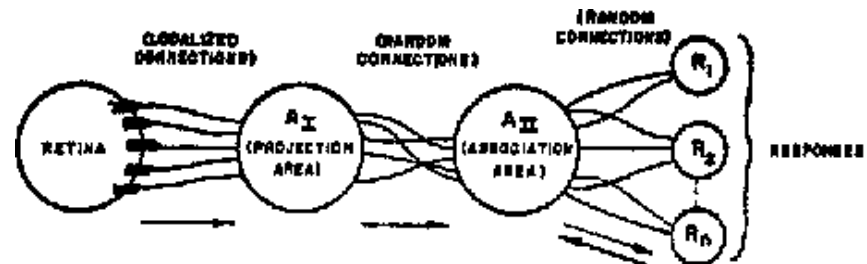
Karl Pearson, Ronald Aylmer Fisher



# 단층퍼셉트론

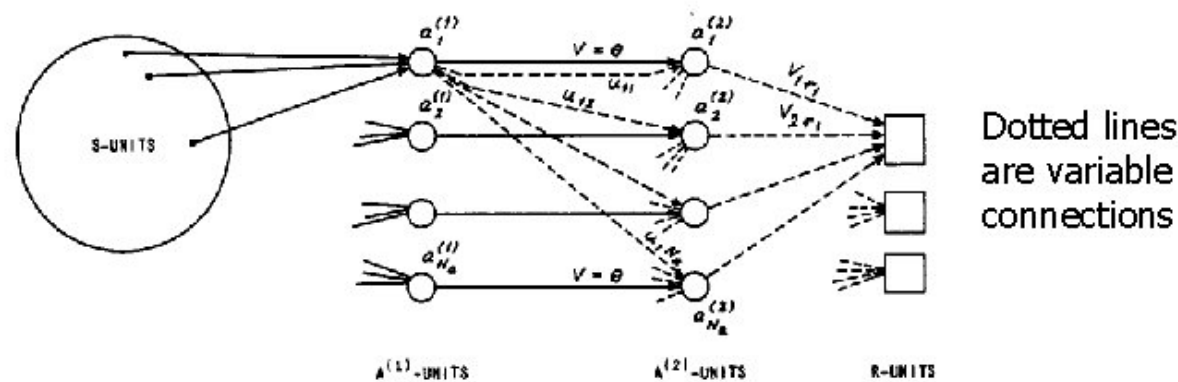
퍼셉트론은 1957년 코넬 항공 연구소(Cornell Aeronautical Lab)의 프랑크 로젠블라트(Frank Rosenblatt)에 의해 고안된 인공신경망이다. 로젠블라트에 의해 제안된 것은 가장 간단한 형태의 단층 퍼셉트론(single-layer perceptron)으로 입력 벡터를 두 부류로 구분하는 선형분류기이다.

Rosenblatt, Frank (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6,



# 다층퍼셉트론

1962년 로센블레트는 자신의 퍼셉트론 이론을 보완하고 정리해서 "신경 역학의 법칙(Principles of Neurodynamics)"이라는 책을 출간하여 다중 퍼셉트론 개념을 제시  
그러나 마땅한 수학적인 학습방법을 제시하지 못한 채 아이디어 수준에서 머무름



## 역전파 알고리즘 (back-propagation)

1986년 데이비드 러멜하트(David Rumelhart), 제프리 힌튼(Geoffrey Hinton), 로널드 윌리엄스(Ronald Williams)가 "Learning representations by back-propagating errors"라는 논문을 네이처지에 발표



**F. Rosenblatt**

1957

Perceptron  
learning rule



**A. Ivakhnenko**

1966

Group method of  
data handling



**S. Linnainmaa**

1970



**P. Werbos**

1982

Backpropagation



**D. Rumelhart**

1986

<https://towardsdatascience.com/towards-geometric-deep-learning-ii-the-perceptron-affair-fafa61b5c40a>

## Decision Tree (의사결정 나무)

### CHAID

Kass, G. V. (1980).

"An exploratory technique for investigating large quantities of categorical data". *Applied Statistics*. **29** (2): 119–127.

### CART

Breiman, Leo; Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984).

*Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

### ID3

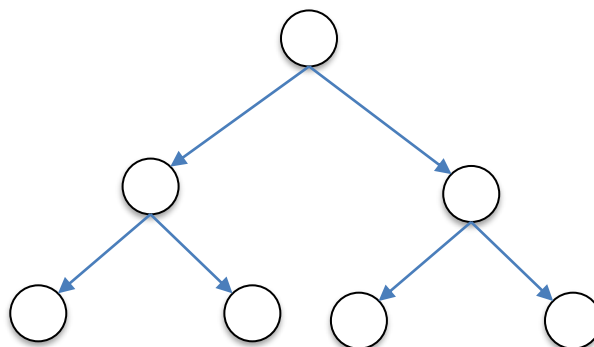
Quinlan, J. R. 1986. Induction of Decision Trees. *Mach. Learn.* 1, 1 (Mar. 1986), 81–106

### C4.5

Quinlan, J. R. C4.5: *Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.

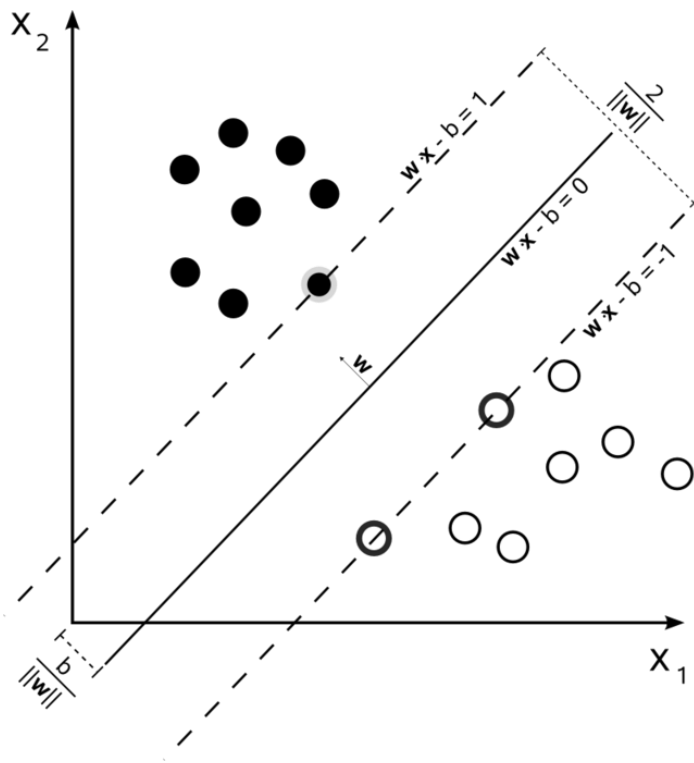


Quinlan, J. R.



# Support Vector Machine

Cortes, C. and Vapnik, V. (1995) Support-Vector Networks

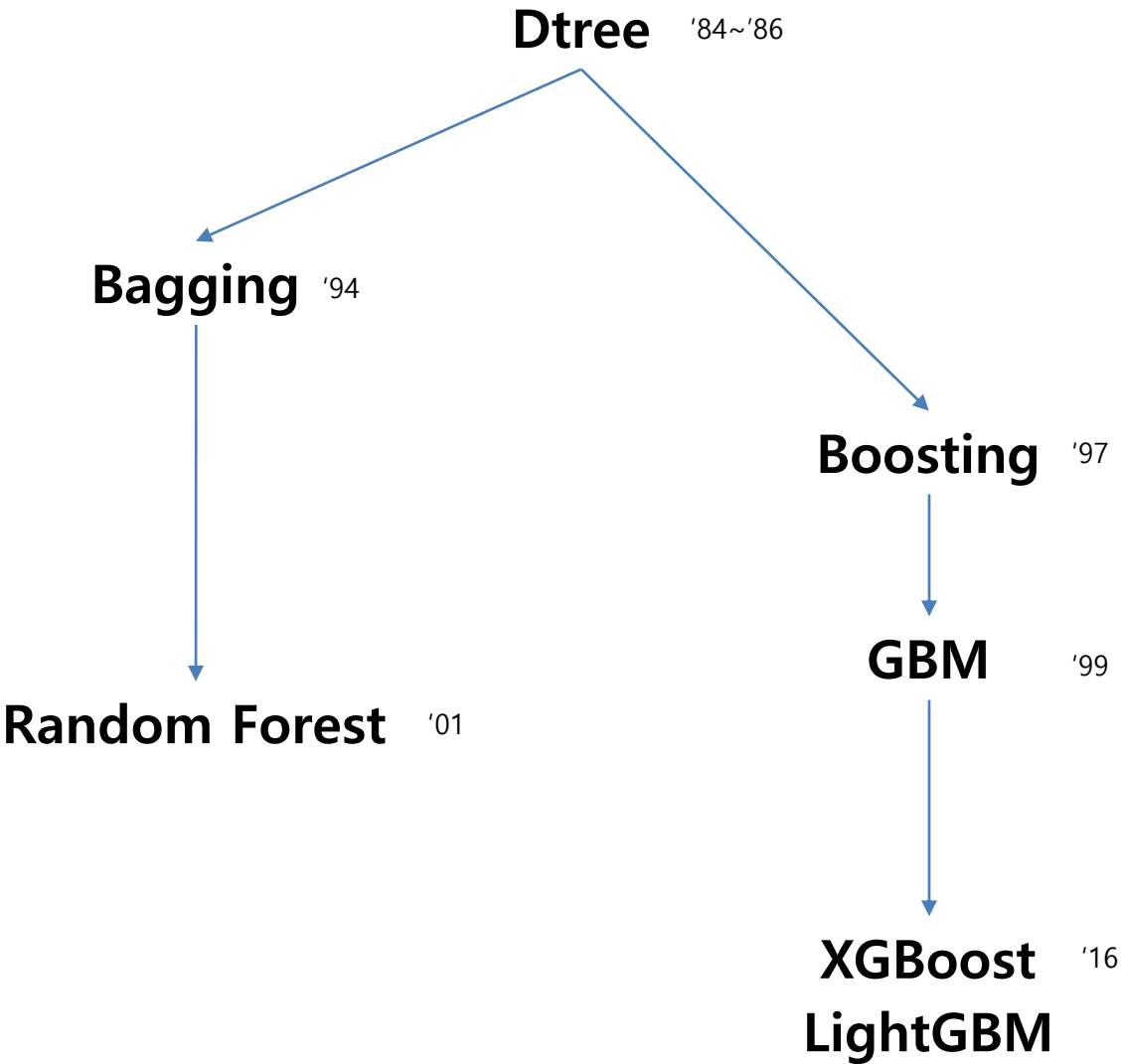


Vladimir Vapnik



Corinna Cortes

# 의사결정나무 알고리즘 확장





# Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?

**Manuel Fernández-Delgado**

MANUEL.FERNANDEZ.DELGADO@USC.ES

**Eva Cernadas**

EVA.CERNADAS@USC.ES

**Senén Barro**

SENEN.BARRO@USC.ES

*CITIUS: Centro de Investigación en TecnoloXías da Información da USC*

*University of Santiago de Compostela*

*Campus Vida, 15872, Santiago de Compostela, Spain*

**Dinani Amorim**

DINANIAMORIM@GMAIL.COM

*Departamento de Tecnologia e Ciências Sociais- DTCS*

*Universidade do Estado da Bahia*

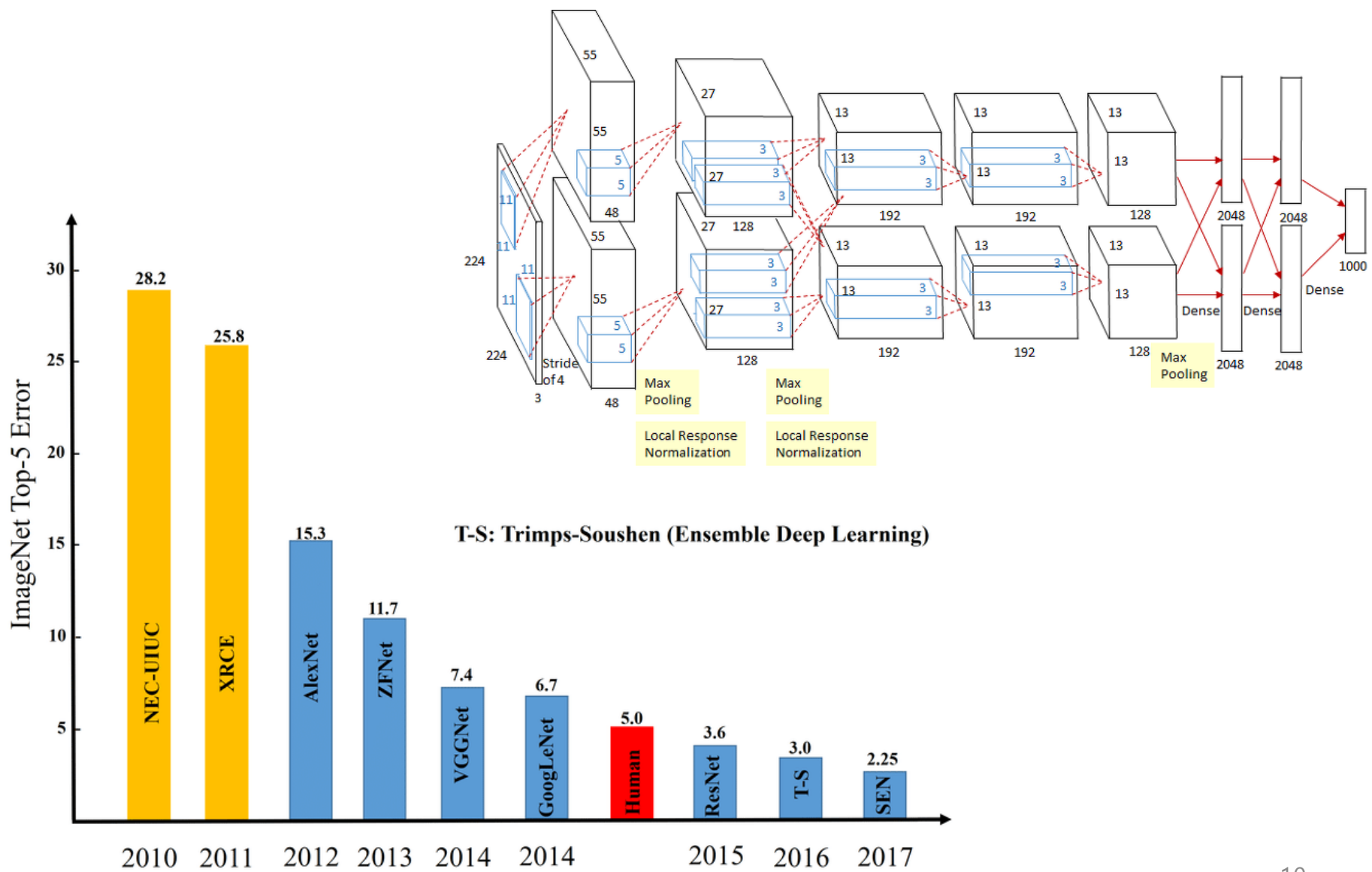
*Av. Edgard Chastinet S/N - São Geraldo - Juazeiro-BA, CEP: 48.305-680, Brasil*

**Editor:** Russ Greiner

## Abstract

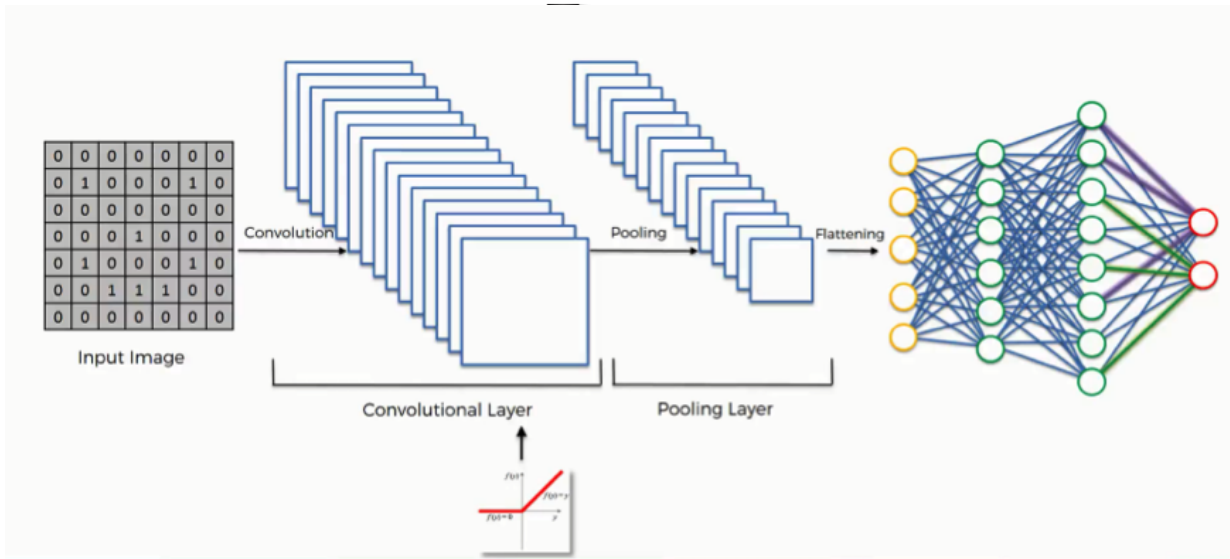
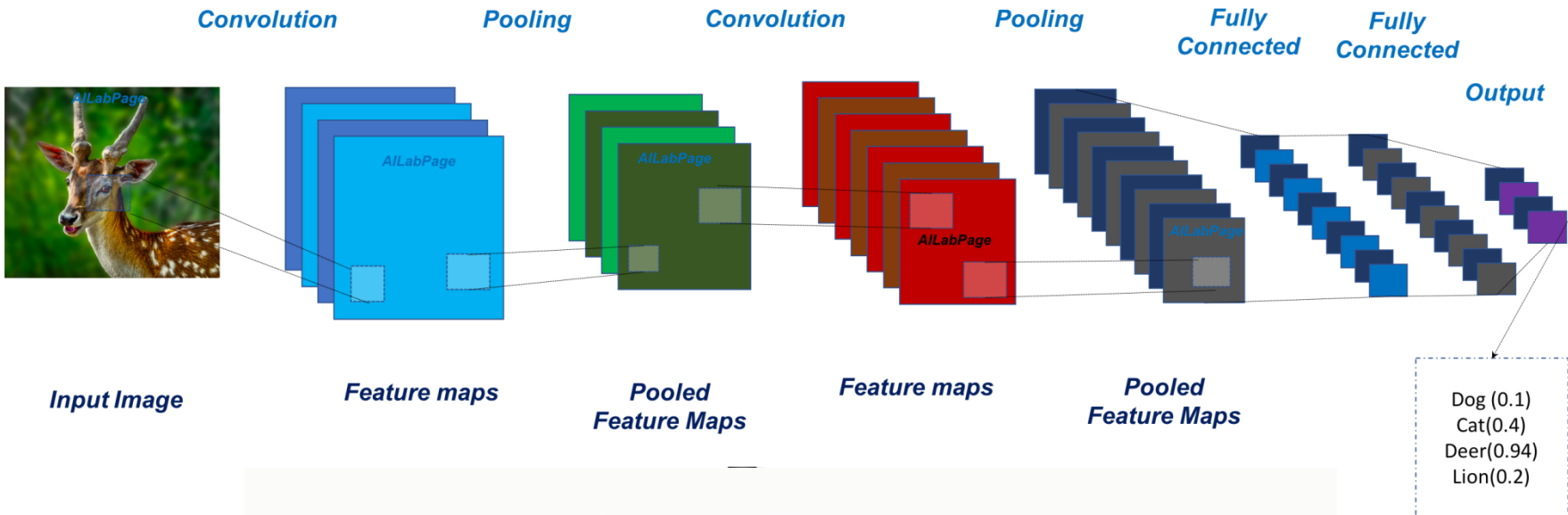
We evaluate **179 classifiers** arising from **17 families** (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest-neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods), implemented in Weka, R (with and without the caret package), C and Matlab, including all the relevant classifiers available today. We use **121 data sets**, which represent **the whole UCI data**

# AlexNet (2012) “ImageNet Classification with Deep Convilutional Neural Networks”

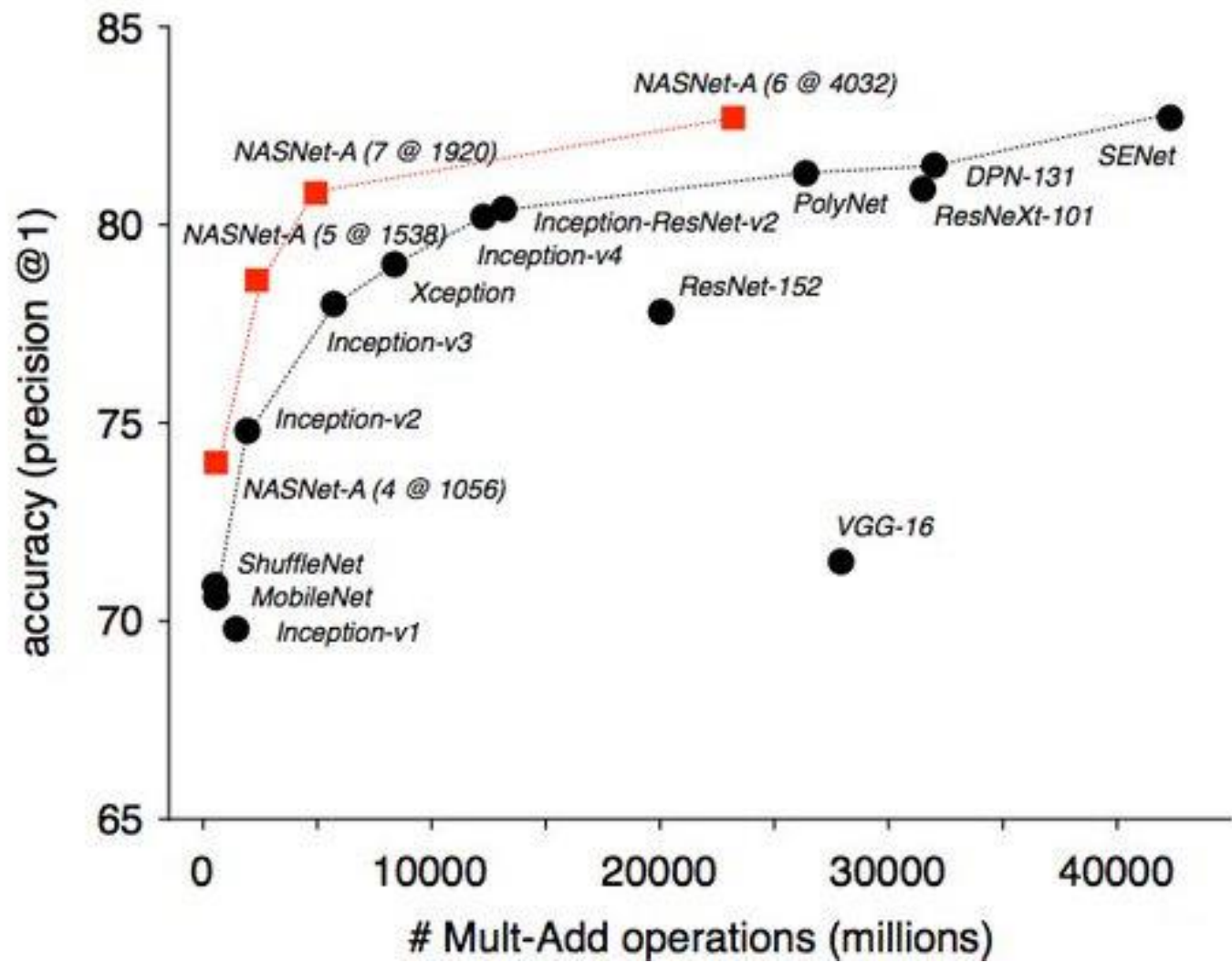


# 영상 인식과 딥러닝의 결합

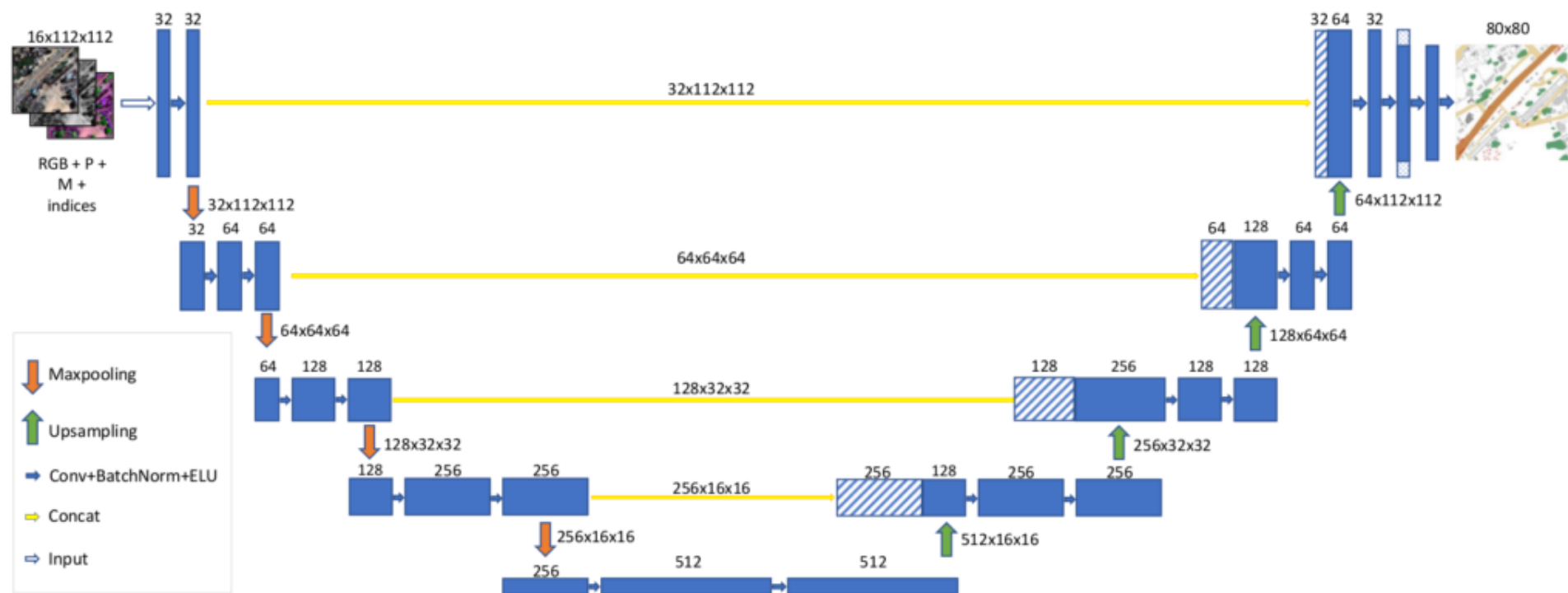
## Convolution Neural Network



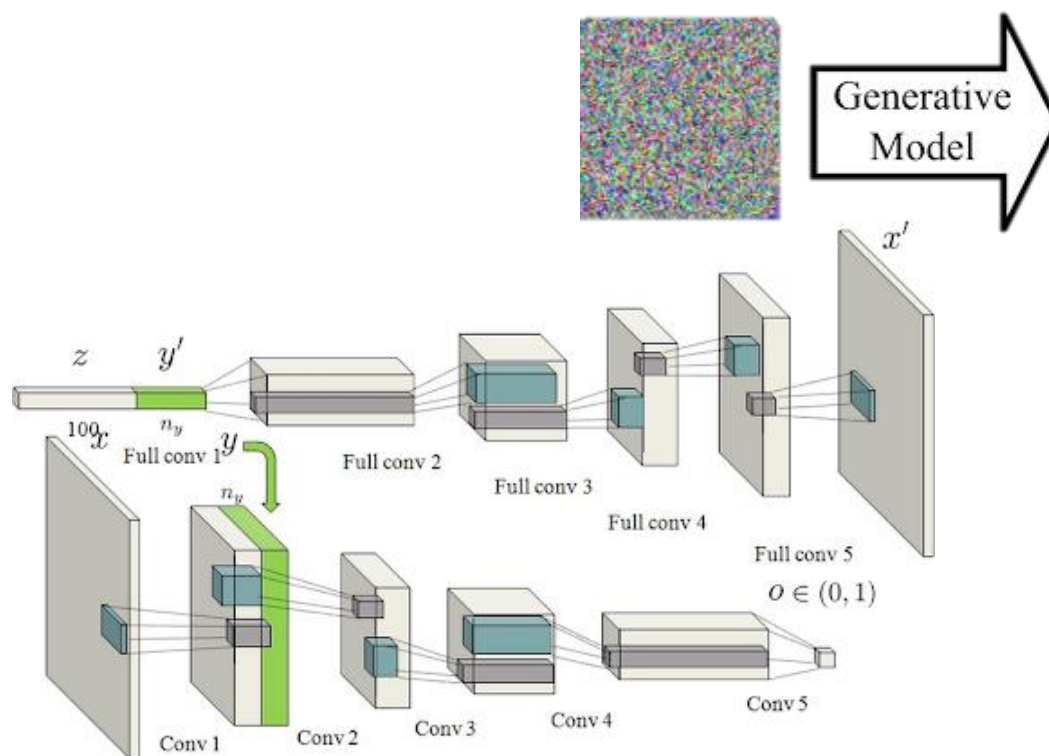
# 이미지 SOTA (State-Of-The-Art) 모델들



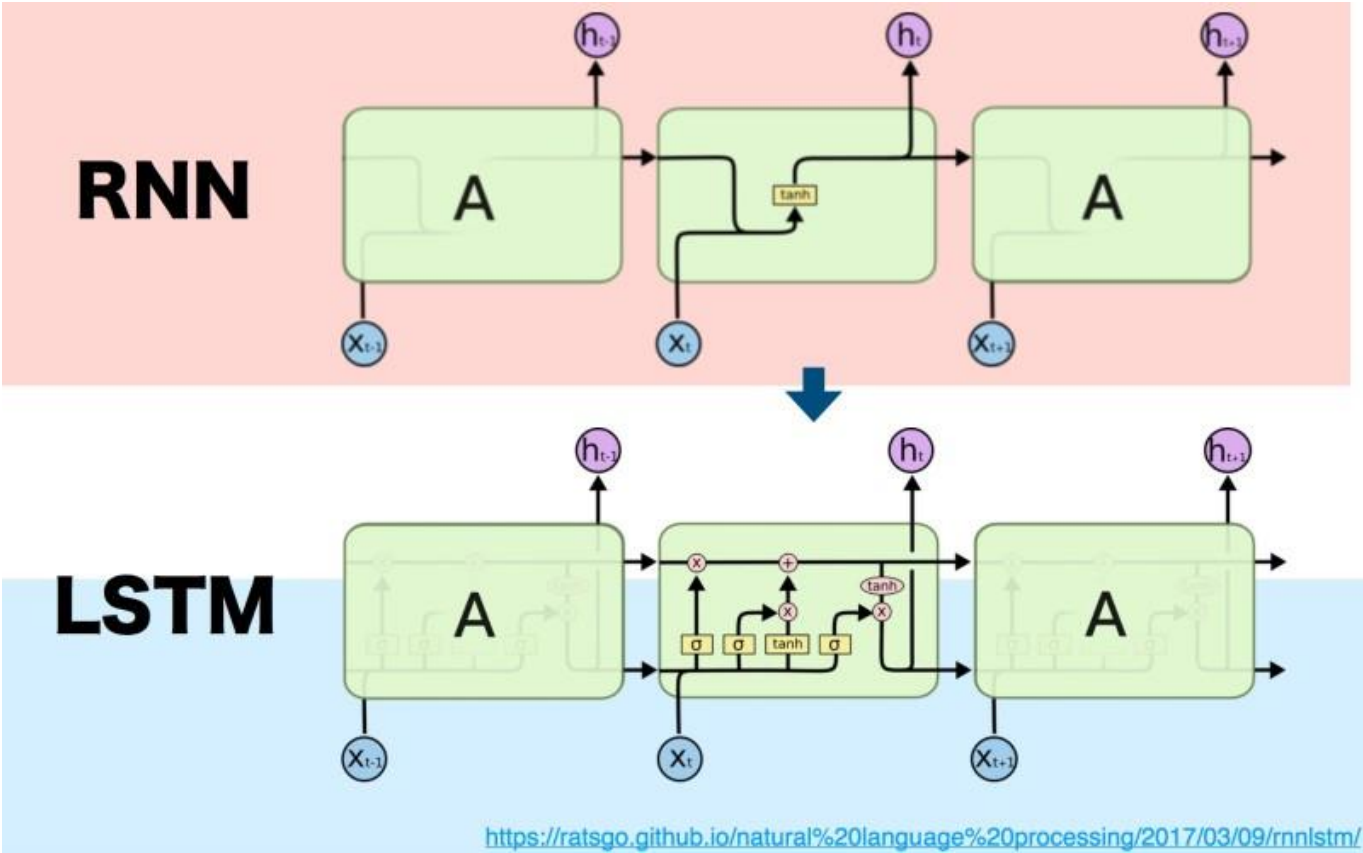
# Satellite Image Segmentation: a Workflow with U-Net



# 적대적 생성모델 GAN(Generative Adversarial Network)



# 시계열 모델

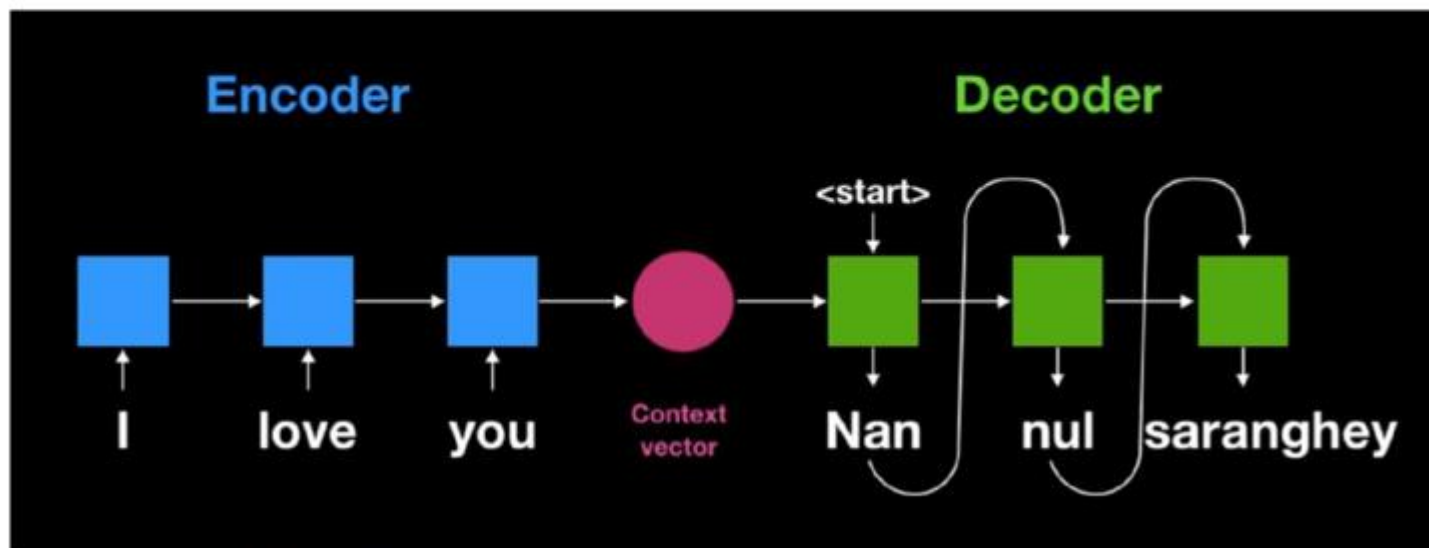




# 자연어처리 딥러닝 NLP(Natural Language Processing)

## 언어모델 : Sequence to Sequence

- Seq2Seq (Sequence to Sequence)
- Encoder layer: RNN 구조를 통해 Context vector 를 획득
- Decoder layer: 획득된 Context vector를 입력으로 출력을 예측



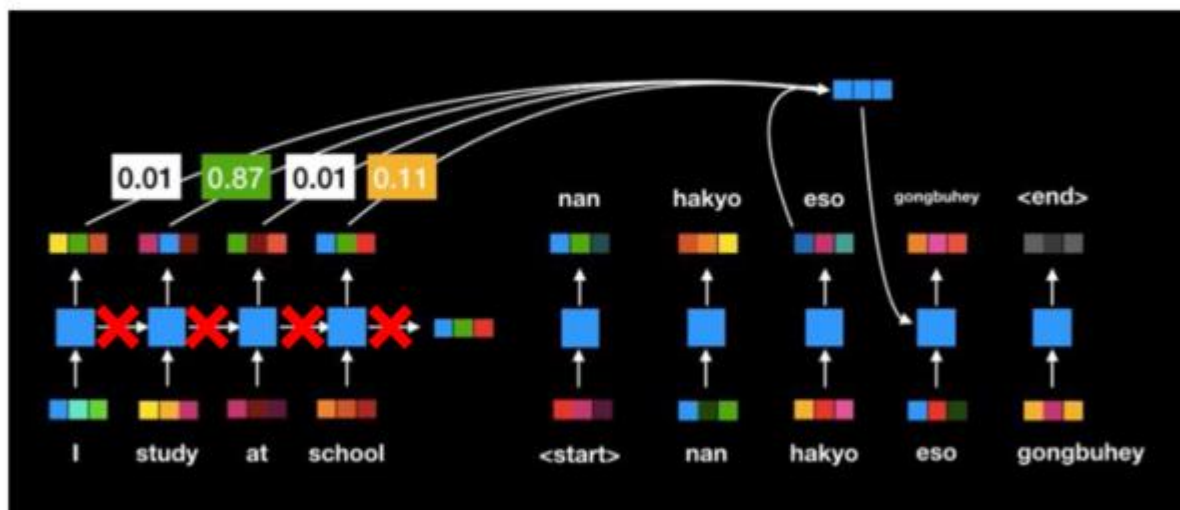
\* [딥러닝 기계번역] 시퀀스 투 시퀀스 + 어텐션 모델  
(<https://www.youtube.com/watch?v=WsQ1du2IMgl>)



# 자연어처리 딥러닝 NLP(Natural Language Processing)

## 언어모델 : Self-Attention

- 문맥에 따라 동적으로 할당되는 encode의 Attention weight로 인한 dynamic context vector를 획득
- 기존 Seq2Seq의 encoder, decoder 성능을 비약적으로 향상시킴
- 하지만, 여전히 RNN이 순차적으로 연산이 이뤄짐에 따라 연산 속도가 느림

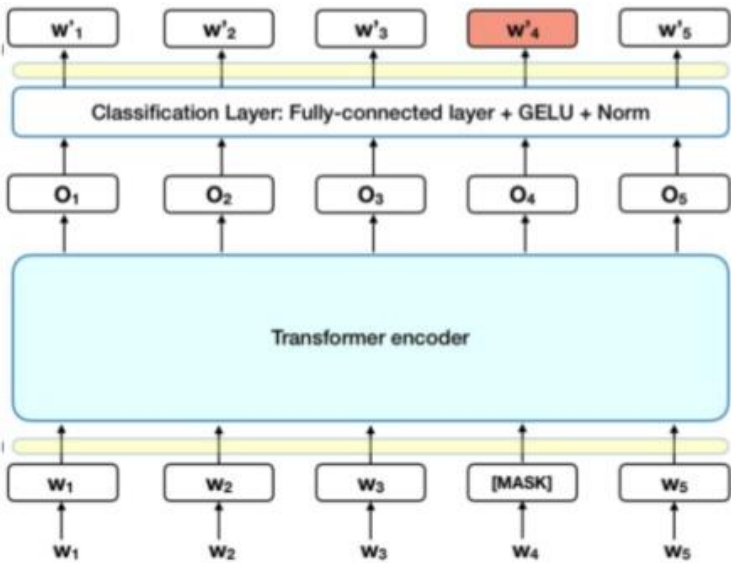


그냥 **RNN**을 없애는건 어떨까?

# 자연어처리와 딥러닝의 결합

## Google BERT

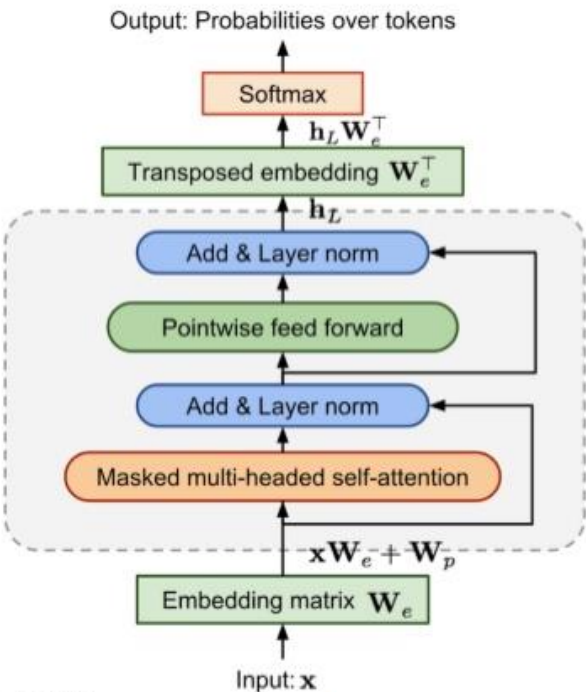
### Encoder



BERT(좌측)는 Language Understanding을 위해 인코더를 Pre-train하는 반면,  
GPT(우측)는 Language Modeling을 위해 디코더를 Pre-train 시키는 구조를 지니고 있음

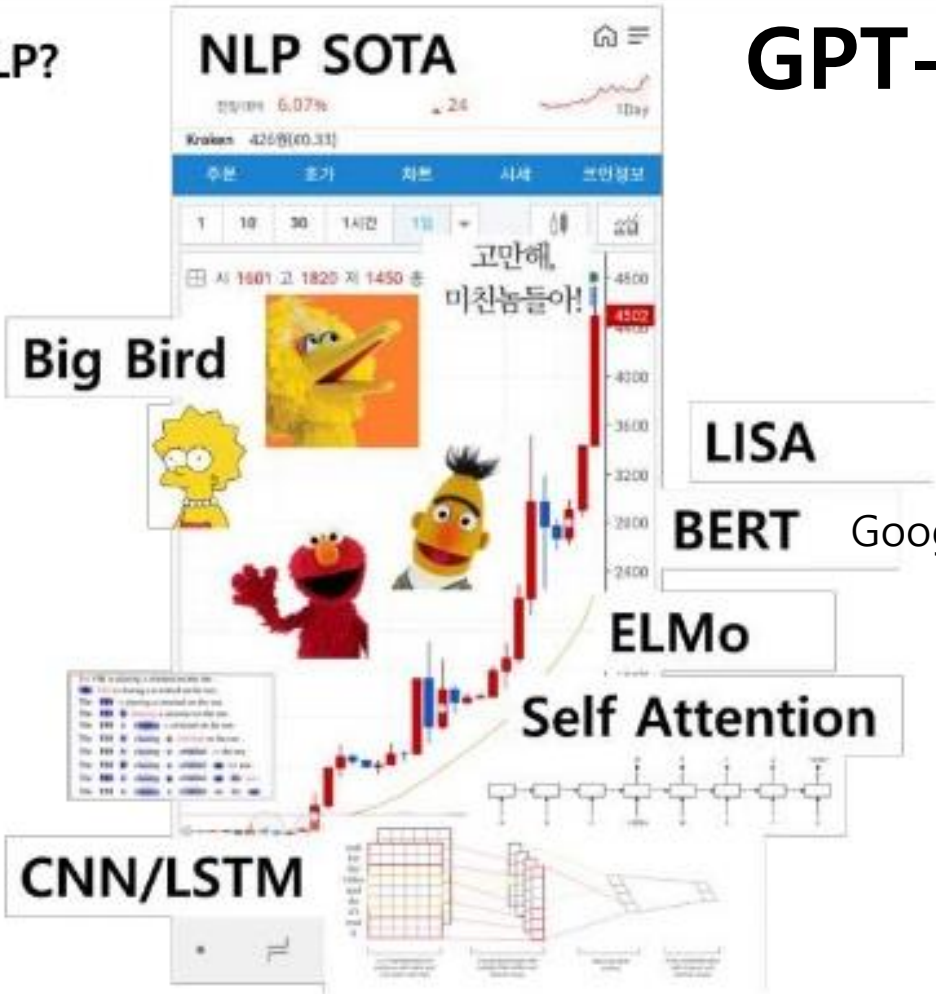
## OpenAI GPT

### Decoder



# 자연어처리 SOTA (State-Of-The-Art) 모델들

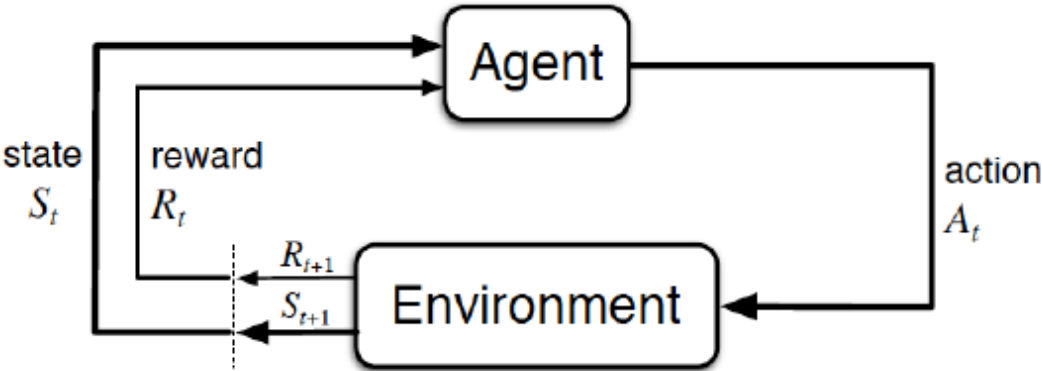
• ~2018년의 NLP?



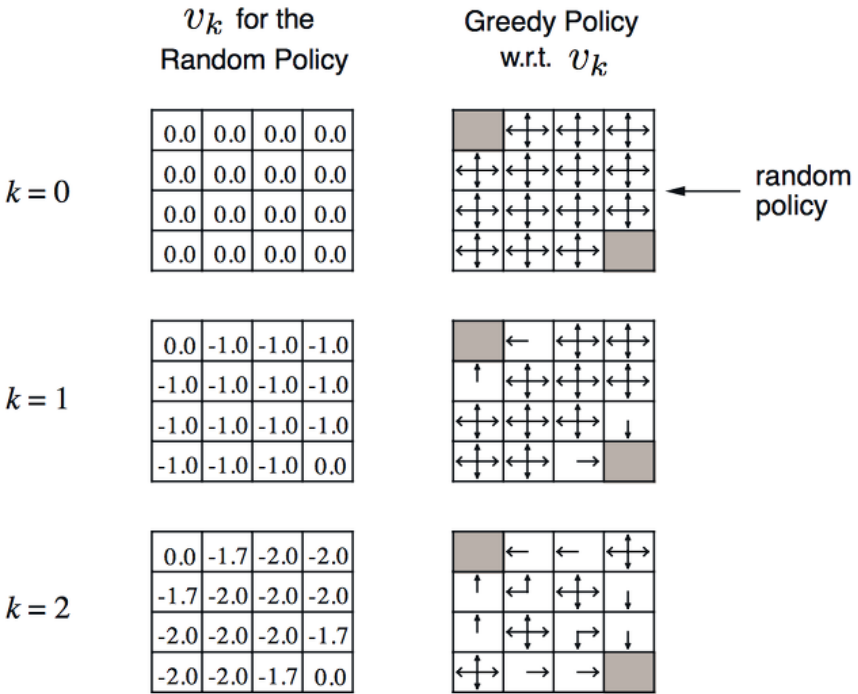
**GPT-3** Open AI

Google

# 강화학습(Reinforcement Learning)

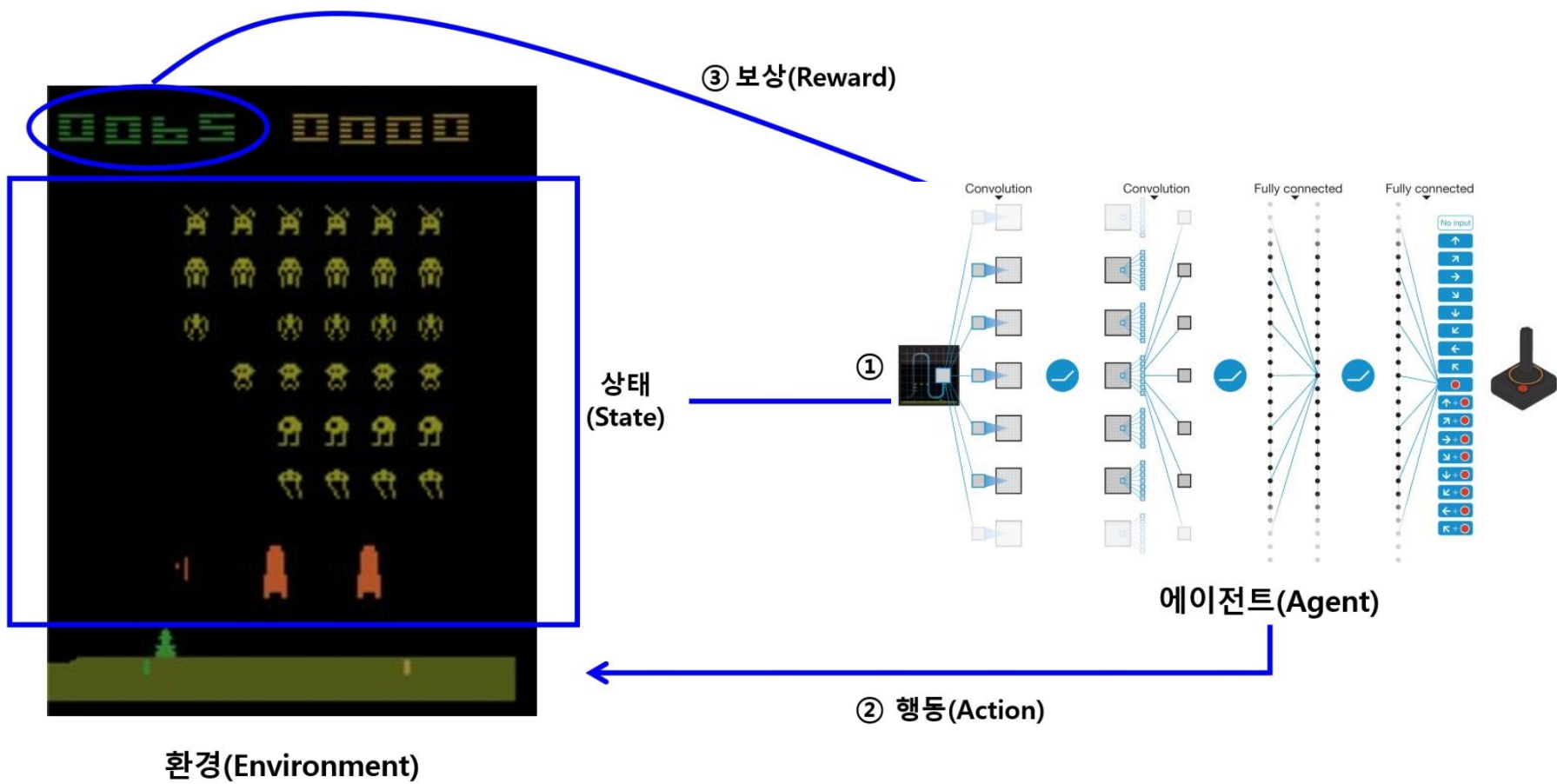


## Greedy Search



# Google DeepMind에서 2014년에 발표한 Deep Q-Network

## Atari 게임의 Q-network를 딥러닝 모델로 학습

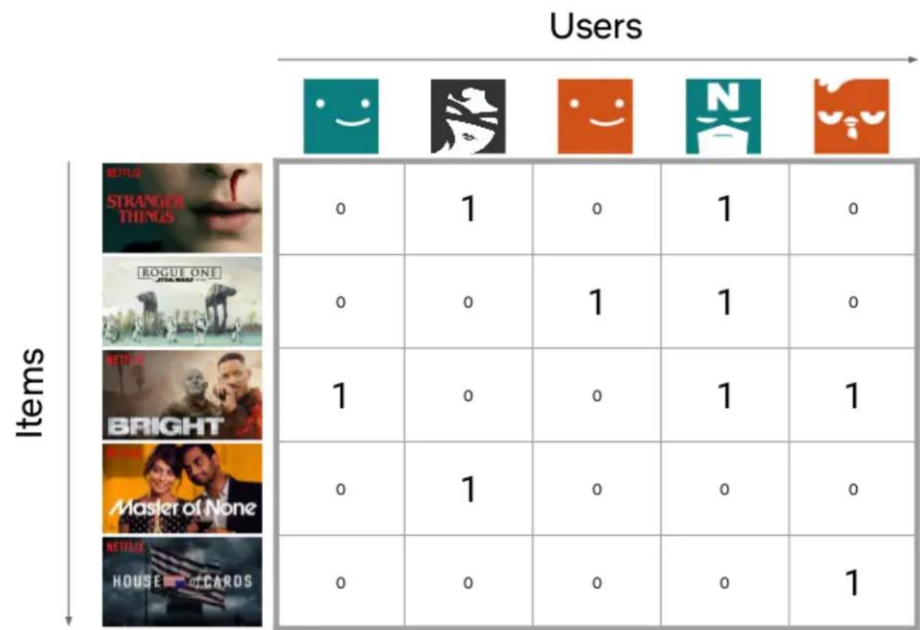




## Google DeepMind에서 2016년 AlphaGo 발표 – 이세돌과 대국 2017년 기존 바둑 기보를 학습하지 않은 alpha go zero를 발표



# Netflix 추천 시스템 (Collaborative Filtering)

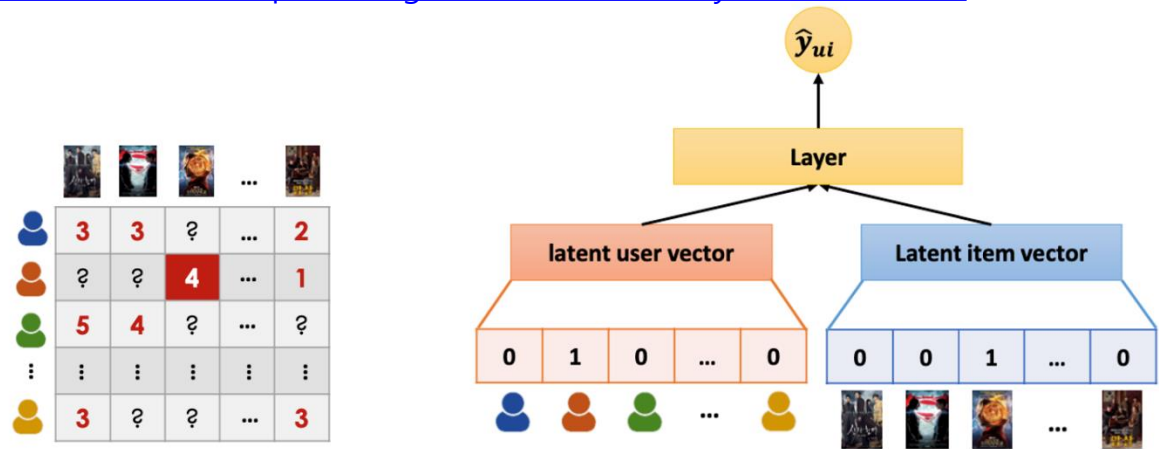


## A Matrix Factorization view

$$R \approx UV^T$$

$$\min_{u,v} \sum_{i,j \in R} (r_{ij} - u_i^T v_j) + \lambda (\|u_i\|^2 + \|v_j\|^2)$$

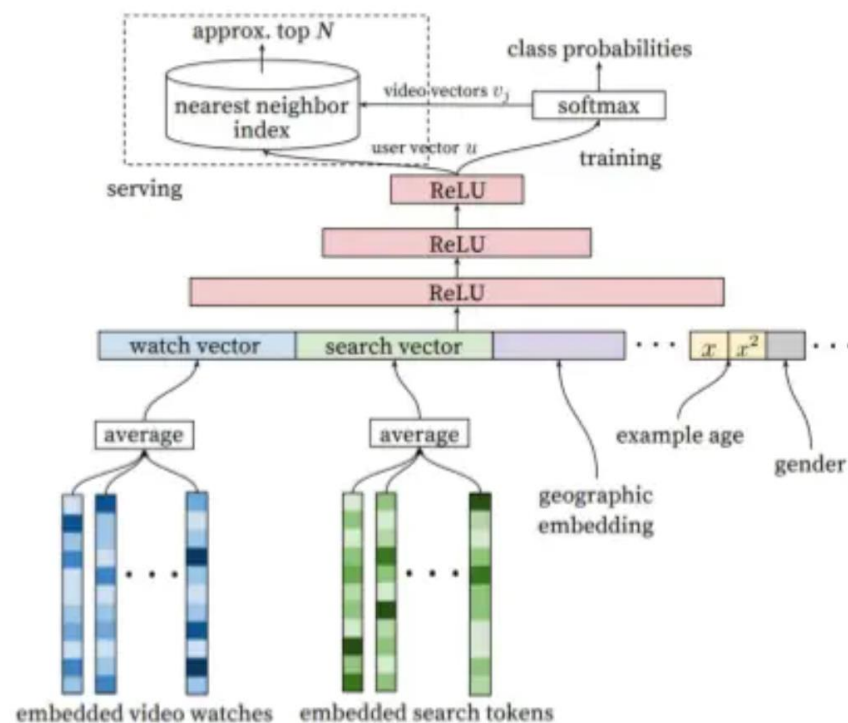
<https://www.slideshare.net/moustaki/deep-learning-for-recommender-systems-86752234>



# YouTube Recommendations

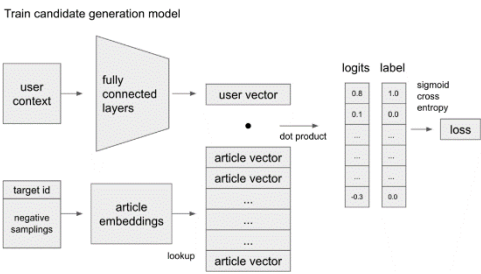
- Two stage ranker: candidate generation (shrinking set of items to rank) and ranking (classifying actual impressions)
- Two feed-forward, fully connected, networks with hundreds of features

[Covington et. al., 2016]





# 당근마켓 추천시스템



## B 사용자의 최근 본 글



생활/가공식품, 60,000 원  
노만코펜하겐 모르모르 그릇 판매 합니다.  
요리가 취미여서 한동안 서울 성동구 옥수동



디지털/가전, 26,000원  
브리츠 BZ-TWS5  
브리츠 블루투스 무선 이어폰입니다. 완전 새제품  
서울 강남구 압구정동



가구/인테리어, 60,000 원  
스피드락 철제선반  
1200\*400 선반6개짜리  
2개 1000\*400 선반5개  
서울 성동구 금호동1가



디지털/가전, 45,000원  
[미개봉] bsw 비엔나 에스프레소 머신  
bsw 비엔나 에스프레소  
서울 중구 약수동

## 서울 성동구 옥수동 지역 추천 글



가구/인테리어, 800,000 원  
수납장판매합니다  
휴대수제목공방 가구가있  
는품경에서구입한 대형수  
서울 서초구 잠원동



가구/인테리어, 10,000 원  
원목책장팔아요  
원목책장팔아요 책상 사  
이즈 120\*200\*20 신답  
서울 동대문구 답십리제1  
동



가구/인테리어, 30,000 원  
멋진 책상 팝니다(2개)  
2개 일괄 구매시 5만 5천  
원 입니다 면 너비 70cm  
서울 성동구 하왕십리동



가구/인테리어, 50,000 원  
리바트 티테이블  
리바트에서 작년 3월에  
구입했구요 상판은 생활  
서울 성동구 금호1가동

## A 사용자의 최근 본 글



유아동/유아도서, 15,000 원  
게스 아기옷 6m-9m 3가  
지 일괄  
일괄 15000원이고 따로  
서울 중구 동화동



유아동/유아도서, 5,000 원  
아기 우주복  
저희아기인 우주복 많이 안  
입게 되어 한번정도 입고  
서울 용산구 한강로동



유아동/유아도서, 3,000 원  
돌전후 아기 올인원  
깨끗해요  
서울 용산구 한강로동



유아동/유아도서, 3,000 원  
폴로 베레모 (돌전후)  
깨끗해요  
서울 용산구 한강로동

## 서울 성동구 옥수동 지역 추천 글



유아동/유아도서,  
콤비yt-180 보행기,점퍼  
루 팔아요  
콤피 사고서 2번사용한  
서울 성북구 돈암제2동



유아동/유아도서, 10,000 원  
킨더팜 유모차라이너 내  
놔요~  
첫째 2개월, 둘째 5개월  
서울 성동구 마장동



유아동/유아도서, 12,000 원  
아기 미끄럼틀  
쿠쿠토이즈 아기 미끄럼  
틀 팝니다. 첫 아기가 들  
서울 성동구 옥수동



유아동/유아도서, 20,000 원  
신생아아기침대 Angel  
Kid  
신생아아기침대 Angel  
서울 광진구 자양제3동