



인공지능의 탄생과 현재 그리고 미래

엑셈아카데미 DB기술연구소 소개 (edu@ex-em.com)



- 권건우 상무
 - MIS Ph.D.
 - 前 삼성 SDS DBA, TA
 - 現 엑셈 아카데미 상무
 - 現 DB 기술연구소 소장
 - DB Deep Internals 강의 (Oracle, PostgreSQL, MySQL, Aurora)
 - 엑셈아카데미 블록체인 인사이트 강의
 - 엑셈아카데미 인공지능 인사이트 강의
 - < Oracle, PostgreSQL, MySQL Core Architecture 1 > 책 집필
 - < Oracle, PostgreSQL, MySQL Core Architecture 2 > 책 집필



- 장서연 연구원
 - 現 엑셈 DB기술연구소 선임 연구원
 - 엑셈아카데미 DB Deep Internals 강의 (Oracle, PostgreSQL, MySQL, Aurora)
 - 엑셈아카데미 오로라 첫걸음 강의
 - 엑셈아카데미 블록체인 강의
 - 엑셈아카데미 왕초보 SQL 강의



- 이근오 팀장
 - 삼성카드/삼성생명/ING생명 차세대 PJT 참여
 - 現 DB 기술연구소 팀장
 - 엑셈아카데미 DB Deep Internals 강의 (Oracle, PostgreSQL, MySQL, Aurora)
 - 엑셈아카데미 블록체인 인터널 강의
 - 엑셈아카데미 인공지능 인터널 강의
 - < Oracle, PostgreSQL, MySQL Core Architecture 1 > 책 집필
 - < Oracle, PostgreSQL, MySQL Core Architecture 2 > 책 집필



- 이대덕 연구원
 - 現 DB 기술연구소 선임 연구원
 - 엑셈아카데미 DB Deep Internals 강의 (Oracle, PostgreSQL, MySQL, Aurora)
 - 엑셈아카데미 블록체인 인터널 강의
 - < Oracle, PostgreSQL, MySQL Core Architecture 1 > 책 집필
 - < Oracle, PostgreSQL, MySQL Core Architecture 2 > 책 집필



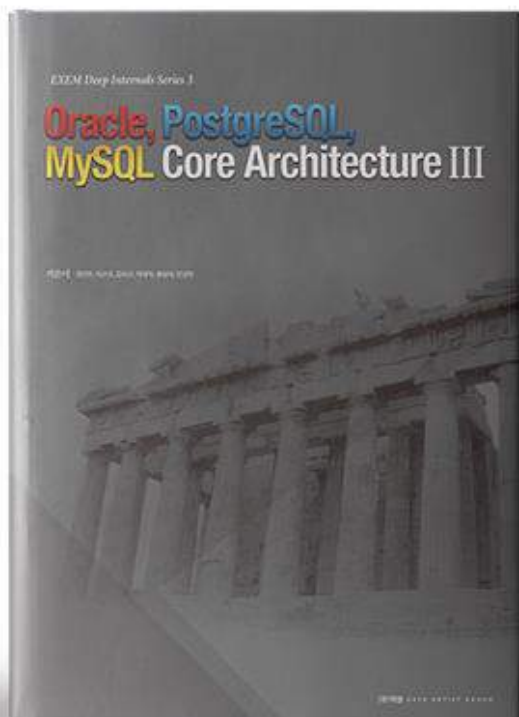
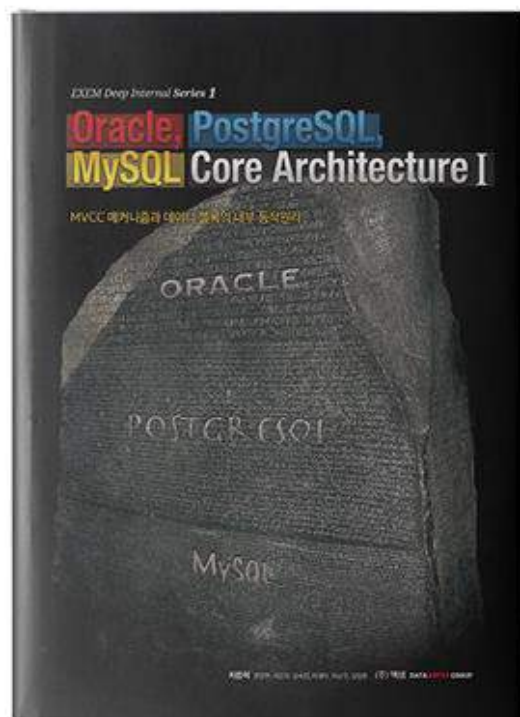
새로운 혁명의 시기

4차 산업혁명

'모든 것이 연결되고 보다
지능적인 사회로의 진화'

어떻게 생존할 것인가?





DBMS 비교 연구 20년



DBMS 내부 정밀 촬영



DBMS 천천히 돌려보기





Digital Forensics

PASSWORD



<http://m.photoviewer.naver.com/blog?listUrl=https%3A%2F%2Fm.blog.naver.com%2FPostView.nhn%3FblogId%3Dwiselaw70%26logNo%3D221053706952&imgId=8&host=https%3A%2F%2Fm.blog.naver.com%2Fphotoviewer&historyBack=true&blogId=wiselaw70&logNo=221053706952#main/8>

$$\Psi = \int e^{\frac{i}{\hbar} \int \left(\frac{R}{16\pi G} - \frac{1}{4} F^2 + \bar{\psi} i \not{D} \psi - \lambda \varphi \bar{\psi} \psi + |D\varphi|^2 - V(\varphi) \right)}$$

슈뢰딩거 파인만 아인슈타인 맥스웰-양-밀스 고바야시-마스카와
 노무라 베리 쿠로 디랙 우카와 하가

물리학의 모든 알려진 법칙들을 요약한 공식



2009/08/03



딥러닝과 수학의 연결 고리

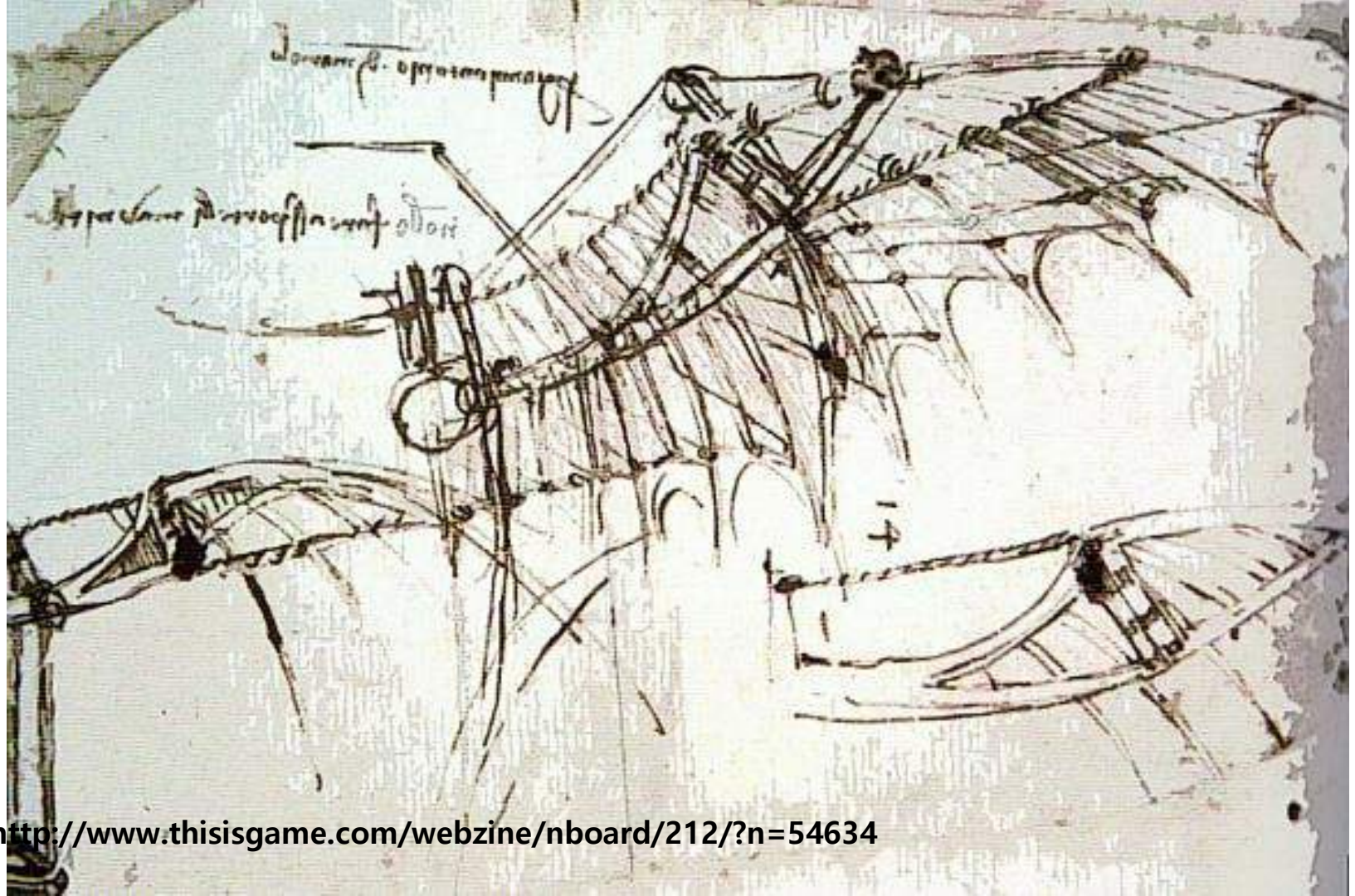
그러던 어느 날 민현석 박사님이 합성곱 연산(convolution operation)의 수학적인 직관에 대해 물어 보셨습니다. 사실 합성곱 연산은 신호처리 분야 뿐만 아니라 편미분방정식, 확률론 등 현대수학의 해석학(analysis) 분야에서 절대적인 위치를 차지하고 있습니다. 제가 박사과정 때 주로 연구했던 분야도 조화해석학(harmonic analysis)과 확률론이었습니다. 두 이론은 합성곱 연산과 푸리에 분석(Fourier analysis) 또는 웨이블릿 분석(wavelet analysis) 사이의 유기적인 관계를 이용해 시그널의 확률적인 패턴을 분석합니다. 민현석 박사님은 장시간에 걸쳐 딥러닝의 등장 배경과 CNN(convolutional neural networks)의 구조를 설명해 주셨고, 2012년도부터 이미지넷(ImageNet)[\[1\]](#) 등의 대회를 통해서 본격적으로 알려지게 된 딥러닝의 성능 발전에 대해서도 알려 주셨습니다.

한 동안 저는 딥러닝 분야에 수학자가 크게 기여할 수 있는 일이 많지 않다고 생각했습니다. **데이터 기반(data-driven)이면서 모델 비의존적(model-free)인 방법론을 추구하는 학문에서는 수학적인 접근보다 오히려 데이터를 잘 축적하기 위한 파이프라인(pipeline) 설계가 더 중요하다고 생각했기 때문입니다.** 저의 이러한 방어적(?) 태도에 대해 민현석 박사님은 다음과 같이 답변했습니다.

“지금으로선 나도 잘 모르겠다. 하지만 딥러닝의 가치가 커질수록 수학이 중요해질 날이 분명 올 것이다.” by 민현석박사님

A seagull is shown in flight, its wings spread wide, against a backdrop of a deep blue sky filled with soft, white clouds. The horizon line is visible in the distance, separating the sky from the ocean. The overall scene conveys a sense of freedom and aspiration.

하늘을 날고자 하는 인류의 꿈



출처 : <http://www.thisisgame.com/webzine/nboard/212/?n=54634>







생각하는 기계를 만들고자 하는 인류의 꿈

KEY MOMENTS IN DEEP-LEARNING HISTORY 1958-1986

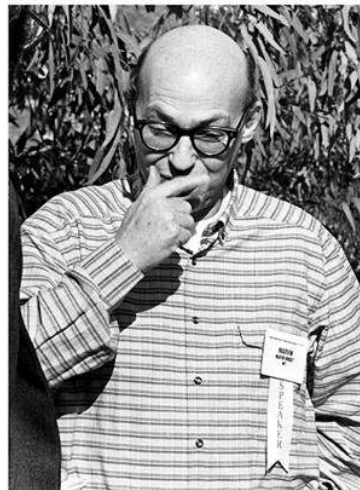
1958

Cornell psychologist Frank Rosenblatt unveils the Perceptron, a single-layer neural network on a room-size computer.



1969

AI giant Marvin Minsky of MIT cowrites a book casting doubt on the viability of neural networks. They fall out of favor.



1986

Neural nets pioneer Geoffrey Hinton and others find a way to train multilayer neural networks to correct mistakes. A flurry of activity ensues.

KEY MOMENTS IN DEEP-LEARNING HISTORY 1989-1997

1989

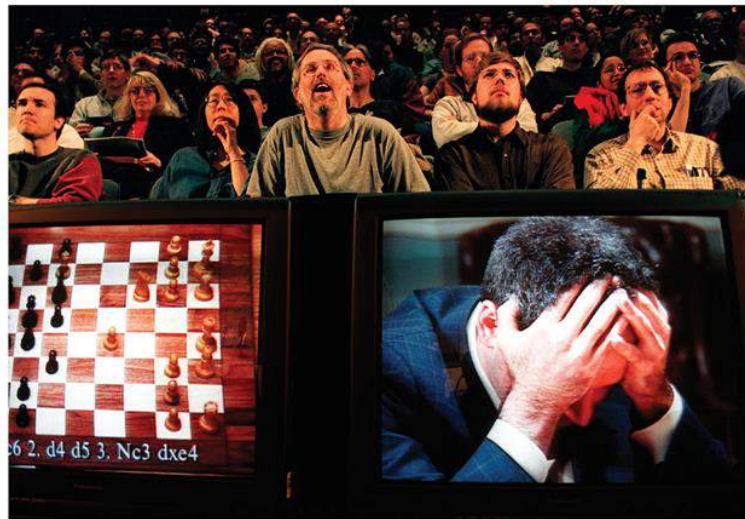
French researcher Yann LeCun, then at Bell Labs, begins foundational work on a type of neural net that becomes crucial for image recognition.

1991

German researchers Sepp Hochreiter and Jürgen Schmidhuber pioneer a neural net with memory features, which eventually proves superior for natural-language processing.

1997

IBM's Deep Blue beats **world champion Garry Kasparov** [right] in chess using traditional AI techniques.



KEY MOMENTS IN DEEP-LEARNING HISTORY 1990's-2011

Mid-1990s

Neural nets fall into disfavor again, eclipsed by other machine-learning techniques.

2007

Fei-Fei Li founds ImageNet and begins assembling a database of 14 million labeled images that can be used for machine-learning research. →



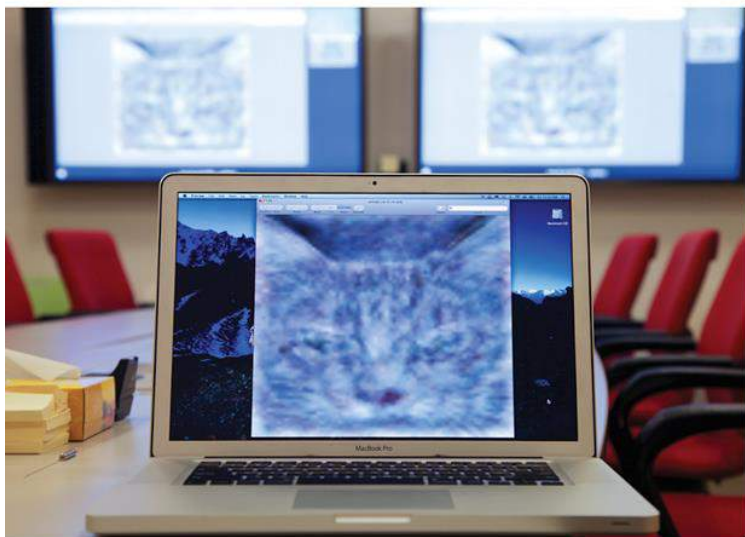
2011

Microsoft introduces neural nets into its speech-recognition features.

2011

IBM's Watson beats two champions at Jeopardy using traditional AI techniques.

KEY MOMENTS IN DEEP-LEARNING HISTORY 2012-2013



2012

JUNE

Google Brain publishes the “cat experiment.” A neural net, shown 10 million unlabeled YouTube images, has trained itself to recognize cats.



AUGUST

Google introduces neural nets into its speech-recognition features.

OCTOBER

A neural net designed by two of Hinton’s students wins the annual ImageNet contest by a wide margin.

2013

MAY

Google improves photo search using neural nets.

KEY MOMENTS IN DEEP-LEARNING HISTORY 2014-2016

2014

JANUARY

Google acquires DeepMind, a startup specializing in combining deep learning and reinforcement learning, for \$600 million.

2015

DECEMBER

A team from Microsoft, using neural nets, outperforms a human on the ImageNet challenge.

2016

MARCH

DeepMind's AlphaGo, using deep learning, defeats world champion **Lee Sedol** in the Chinese game of go, four games to one.



LEE JIN-MAN—AP PHOTO

MASTER 알고리즘

머신러닝은 우리의 미래를
어떻게 바꾸는가

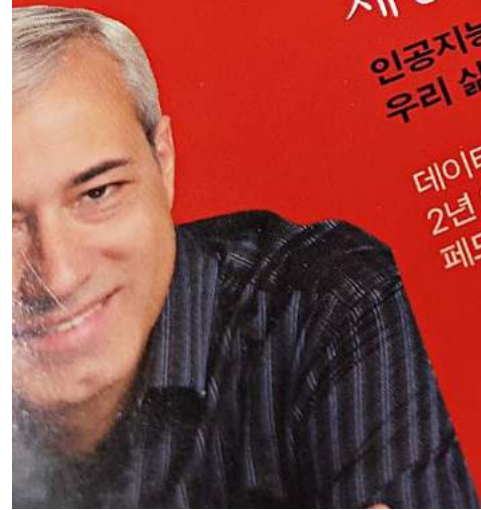
페드로 도밍고스 지음 | 강형진 옮김 | 최승진 감수

완벽한 마스터 알고리즘이 탄생하는 순간,
세상 모든 것이 재편될 것이다!

인공지능과 무인자동차, HCI, 클라우드컴퓨팅, 사물인터넷까지
우리 삶을 변화시킬 가장 혁신적인 기술, 머신러닝의 모든 것

데이터과학 분야의 최고 영예인 SIGKDD 혁신상을
2년 연속 수상한 세계 최고의 머신러닝 전문가
페드로 도밍고스가 전하는 머신러닝의 현재와 미래

아마존·기술
컴퓨터·기술
1위



비즈니스북스

인공지능/머신러닝

우도

기호주의자

- 규칙의 모음
- 귀저기를 사는 고객은 맥주도 사려한다.
- 역연역/의사 결정 트리
- 지식 공학과

연결주의자

- 기호주의는 순차적
연결주의는 동시적
- 퍼셉트론 / 신경망
- 볼츠만 기계 / 역전파
- 시그모이드/오토엔코더
- CNN / RNN

진화주의자

- 자연선택&유전이론
- 적합성 함수 /선발번식
- 유전 알고리즘
- 신경망 과 진화의 결합
- 볼드윈 효과

EM

로지스틱회귀

베이즈주의자

- 사전확률&사후확률
- 베이즈 정리
- 빈도주의 vs 베이즈
- 은닉마코프법(HMM)
- 마코프연쇄몬테카를로(MCMC)

p-value

유추주의자

- K-NN / SVM
- 유비추론/심층 유추법
- 협력 필터링
- 차원의 저주
- RISE (기호주의와
유추주의의 결합)

비지도학습

- 군집화
- K-means
- 주성분분석(PCA)
- 보상/처벌/강화학습
- 청킹 / 연관짓기

상관관계

특성

마스터 알고리즘



SYMBOLIC ARTIFICIAL INTELLIGENCE

meaning, definition, explanation...

창안자는 카네기멜론대학 교수며 1978년 노벨상을 수상한 허버트 사이먼박사다

사이먼은 "모든 인간의 의사결정에는 비이성적인 결정이 큰 역할을 한다"는 다소 이상한 법칙이 관류하는 것에 큰 흥미를 가지고 있었다. 그리고 그것은 몇개의 대수법칙으로 모델링하기 힘들다는 사실을 알았고, 이런 복잡한 시스템의 해석에 가장 적합한 도구가 바로 컴퓨터라는 사실을 방공시스템의 훈련프로그램 개발과정에서 깨달았다.

사이먼 교수는 러셀의 '수학의 원리'(Principia Mathematica)에 있는 몇가지 정리를 증명해 보이는 '논리 이론가'라는 프로그램을 만든다. 이것은 비록 조잡한 수준이긴 해도 인간의 고유영역에 인공지능이 처음내린 발이었다. 이후 사이먼은 '만능해결사'라는 프로그램으로 모든 문제를 해결하려는 왕성한 의욕을 보였지만 그것이 너무나도 힘든 과제라는 사실을 깨닫게 된다.

사이먼의 제자인 파이겐바움

은 유전학으로 노벨상을 수상한 레더버그의 도움으로 유기물의 분자구조를 분석해내는 '덴드럴'(DENDRAL)이란 전문가시스템을 완성한다. 덴드럴은 능숙한 화학자의 솜씨만큼이나, 그러나 그 속도는 십수배나 빠르게 분자구조를 밝혔다. 이제 덴드럴을 사용하는 것은 일상적인 일이 되어버렸다.

이 성공에 고무받아 스탠퍼드대학 의대생인 쇼틀리프는

이 대학 의사들의 도움을 받아서 항생제 선택에 관한 조언을 해주는 '마이신'(MYCIN)이란 의학용 전문가시스템을 개발해 낸다. 그 결과 의사의 오진율을 훨씬 밑도는 놀라운 결과를 보여주었다. 그는 더

Computer-Based
Medical Consultations:
MYCIN

Edward Hance Shortliffe



ELSEVIER COMPUTER SCIENCE LIBRARY

RULE037

IF: 1) THE IDENTITY OF THE ORGANISM IS NOT KNOWN
WITH CERTAINTY, AND
2) THE STAIN OF THE ORGANISM IS GRAMNEG, AND
3) THE MORPHOLOGY OF THE ORGANISM IS ROD, AND
4) THE AEROBICITY OF THE ORGANISM IS AEROBIC
THEN: THERE IS STRONGLY SUGGESTIVE EVIDENCE (.8)
THAT THE CLASS OF THE ORGANISM IS
ENTEROBACTERIACEAE

RULE145

IF: 1) THE THERAPY UNDER CONSIDERATION IS ONE OF:
CEPHALOTHIN CLINDAMYCIN ERYTHROMYCIN
LINCOMYCIN VANCOMYCIN, AND
2) MENINGITIS IS AN INFECTIOUS DISEASE DIAGNOSIS
FOR THE PATIENT
THEN: IT IS DEFINITE (1) THE THE THERAPY UNDER
CONSIDERATION IS NOT A POTENTIAL THERAPY FOR
USE AGAINST THE ORGANISM

MYCIN EXPERT SYSTEM

www.nipunjасwal.com

How Helpful Are Expert Systems In Medical ?

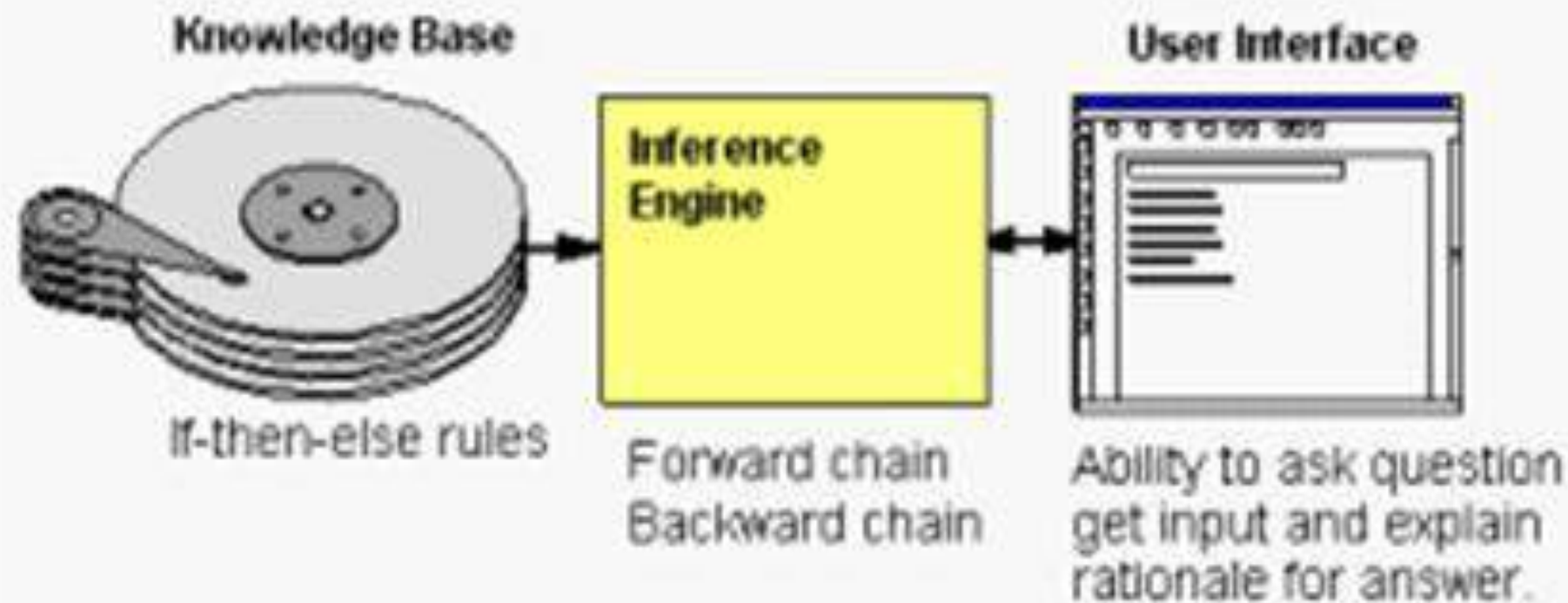
20, Apr 2013

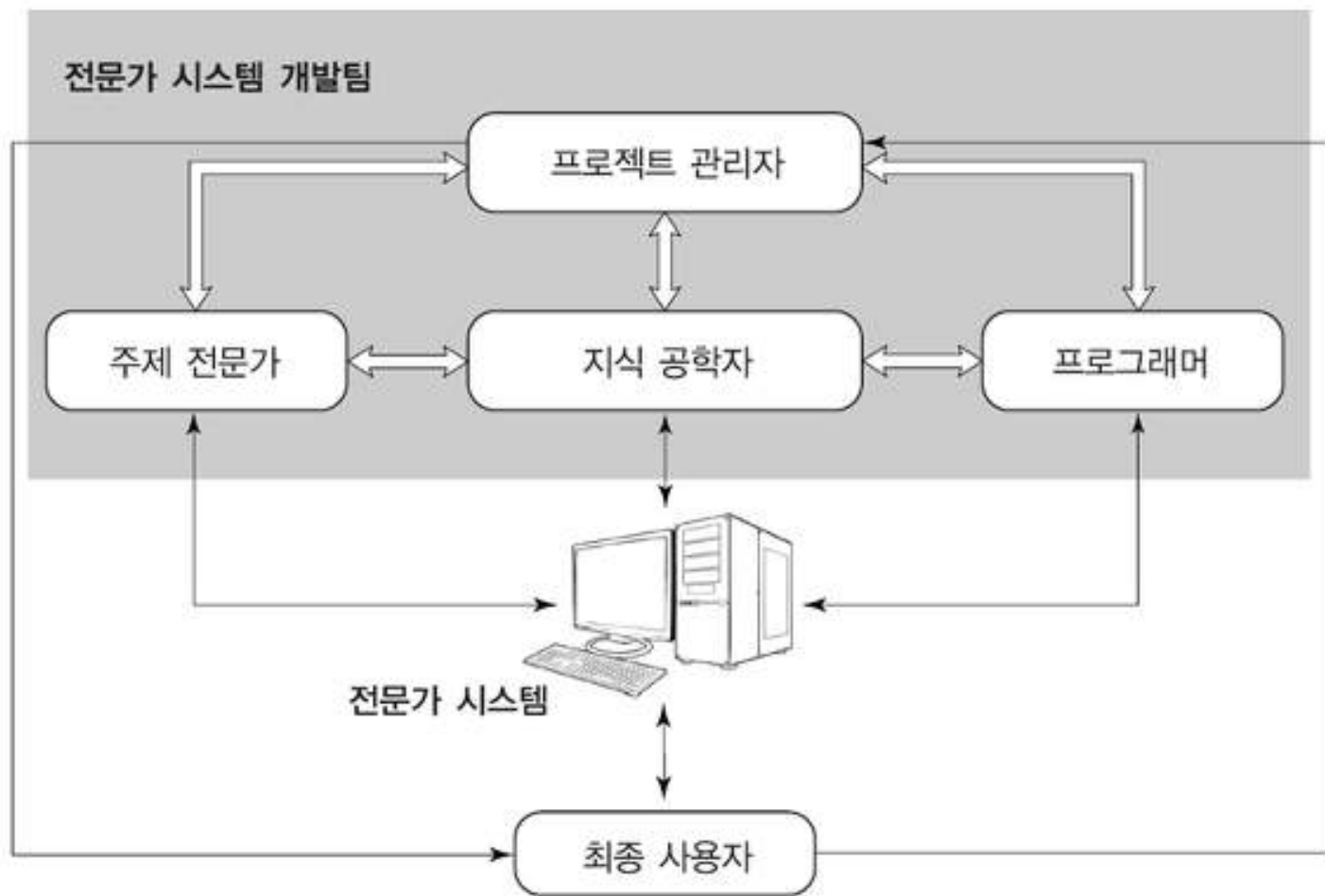
PRESENTED BY NIPUN JASWAL



MYCIN was an early expert system that used artificial intelligence to identify bacteria causing severe infections, such as bacteremia and meningitis, and to recommend antibiotics, with the dosage adjusted for patient's body weight — the name derived from the antibiotics themselves, as many antibiotics have the suffix "-mycin". The Mycin system was also used for the diagnosis of blood clotting diseases.

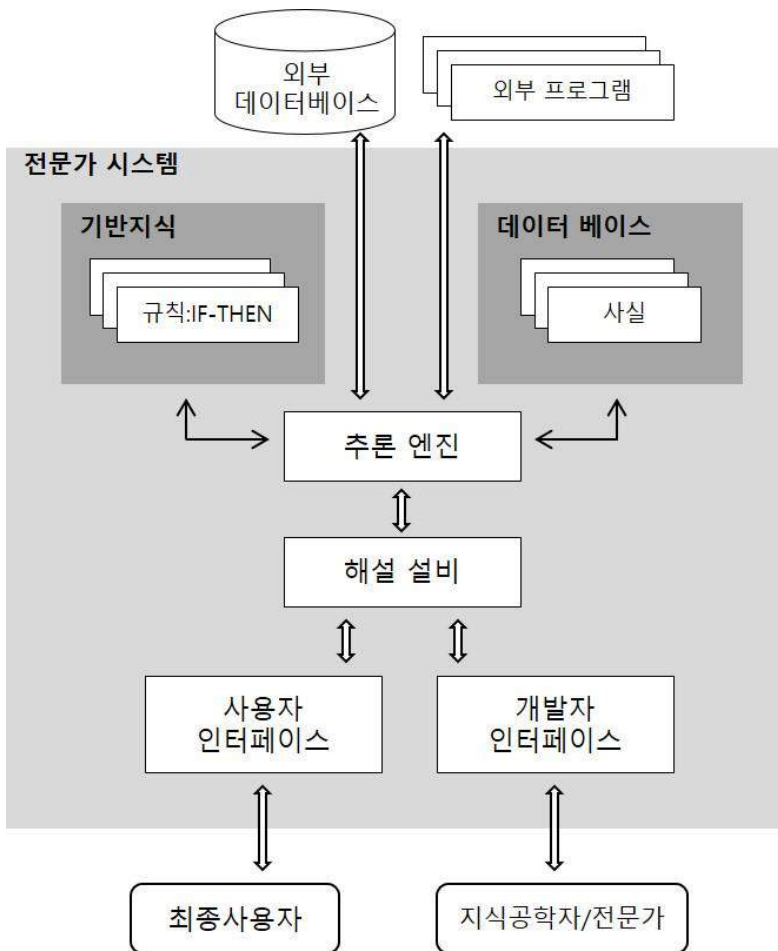
MYCIN was never actually used in practice. This wasn't because of any weakness in its performance. As mentioned, in tests it outperformed members of the Stanford medical school faculty. Some observers raised ethical and legal issues related to the use of computers in medicine — if a program gives the wrong diagnosis or recommends the wrong therapy, who should be held responsible? However, the greatest problem, and the reason that MYCIN was not used in routine practice, was the state of technologies for system integration, especially at the time it was developed.





[그림 2-1] 전문가 시스템 개발팀의 주요 구성원들

규칙기반 전문가 시스템



지식 표현 기법 :

IF <전건1>

AND <전건2>

OR <전건3>

...

THEN <후건1>

<후건2>

...

규칙기반 전문가 시스템의 장단점

- 자연스러운 지식 표현

- 통일된 구조

- 지식과 과정의 분리

메타지식 : 지식에 관한 지식으로 정의

- 메타지식은 전문가 시스템 내에 있는 분야 지식을 사용하고 제어하는데 필요한 지식이다.

<그림> 규칙기반 전문가 시스템의 구조

<http://www.jangun.com/study/IntroductionAI.html>

명인은 사라져도 전문가시스템이 남아

전문가시스템은 초기의 위세와 기대에도 불구하고 인간 전문가의 자리를 완전히 차지하는 데는 실패하였다. 그 이유는 아직도 프로그래밍화 하지 못하는 다양한 경우와 상식들이 입력될 수 없기 때문이다.

특히 특정한 개인이 책임을 져야 하는 기업적인 체계에서는 불리했다. 예를 들어 의학 진단에 완전히 '전문가시스템을 사용한다고 가정할 때, 그로부터 발생하는 의료사고는 누가 책임을 질 것인가. 지식을 제공한 의사인가 아니면 그것을 컴퓨터화한 기사(Knowledge Engineer)인가. 이를 구분하는 것이 어렵기 때문에 상당 기간 동안 전문가의 보조수단에 머물러 있을 수밖에 없을 것이다.

이와는 달리 최근 전문가시스템은 인간의 지혜를 보존, 정리하는 도구로서 훌륭히 사용되고 있다. 따라서 현재 많은 전문가시스템의 범위는 인간 상식을 추월하려는 노력보다는 특수한 좁은 분야, 예를 들면 PCB(인쇄회로기판) 검사라든지, 최적의 부품조립을 위한 조연자로서 역할을 수행한다.

명인은 사라져도 전문가시스템이 남아

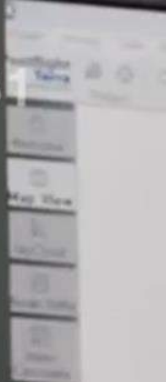
넓게 본다면 보다 지능화된 대형 데이터베이스(DB)라고도 할 수 있을 것이다. 1980년대 코닥사의 주물 시스템 고장진단 시스템, NASA(미 항공우주국)의 액체산소탱크 관리 시스템, 우체국의 재고관리시스템 등에 사용되고 있다. 그리고 IBM에서는 컴퓨터 회사답게 부품조립 등의 전문적인 작업에 1백개 정도의 전문가시스템을 활용하고 있다. 컴퍼스(COMPASS)란 전문가시스템은 미국에 있는 46개국 전화교환기의 오류를 진단해준다. 일본 NKK 제철은 스케플랜(Scheplan)이란 작업공정 조정용 전문가시스템의 도입으로 연간 70억원 이상의 경비를 절감했다.

미국 GM 자동차회사에는 진동분석계의 명인인 찰리 앵블이 은퇴하면서 자신의 모든 기술을 물려준 컴퓨터 후배 찰리시스템이 있다.

그리고 미국 최대의 식품회사인 캠벨 수프회사의 살균소독 전문가인 알도 시미노는 정년퇴직을 앞두고 컴퓨터 학자인 스미스와 함께 7개월에 걸쳐서 2백여개의 경험법칙을 담은 전문가시스템을 완성했다. 비록 찰리와 시미노보다는 덜 미더웠지만 회사측에서는 대만족이었다. 이제 호랑이는 죽어서 가죽을 남기고 명인은 사라져도 전문가시스템을 남기게 되었다.







명인은 사라져도 전문가시스템이 남아

4차산업혁명 시대에는

명인은 사라져도 인공지능 학습 모델이 남아

컴퓨터와 대화하다

● 인공지능 연구의 재도약 - 1980년대

- ◆ 추론과 탐색이라는 단순한 규칙의 한계 인식
- ◆ 2차 AI 붐을 주도한 것은 “지식”
 - 의사 = 병에 관한 많은 지식
 - 지식베이스

● 이런 방향에 영감을 준 1차 AI 붐 시절의 결과

◆ 1964년의 ELIZA

- 대화형 시스템
 - 지식은 존재하지 않고 입력 데이터에서 텍스트를 추출하여 주어진 규칙에 따라 대응하게 함
 - 지적인 대화라고 할 수 없음

◆ 전문가 시스템

- 1970년대 초 스탠포드 대학의 마이신(MYCIN)
 - 질문에 답한 내용을 바탕으로 감염질환을 판단해 항생제 처방을 내리는 시스템

◆ 1960년대 에드워드 파이겐바움(Edward Albert Feigenbaum)

- 미지의 유기 화합물을 특정하는 전문가 시스템 DENDRAL 개발
 - **Dendritic Algorithm**
 - Heuristic-DENDRAL: performance
 - Meta-DENDRAL: learning
 - MYCIN, MOLGEN, MACSYMA, PROSPECTOR, XCON 등도 이 DENDRAL의 후손



Edward Feigenbaum
The Father of Expert System

DENDRAL 개발 당시
스탠포드 대학 교수,
이후 공군 연구소
1994. Turing 상 수상

전문가 시스템의 한계

- 지식의 수집
 - ◆ 전문가로부터 지식을 청취하는 일의 복잡도
 - ◆ 축적된 지식의 규모가 커질 경우 지식들 사이의 모순 발생
- 지식의 범위
 - ◆ 한정된 범위 내의 지식은 축적이 용이
 - ◆ 범위가 넓어지면 지식을 기술(description)하는 것 자체가 어려움
- 용어의 모호함
 - ◆ 자연어 자체의 모호함
 - ◆ 지식 도메인(domain)에 의존적인 의미
- 가장 어려운 지식은 상식

가장 어려운 지식은

상식

상식적 질문

식탁 옆에 의자는 몇 개? 4개
 못이 몇 개나 있나? 가위 등이 걸린 걸 보면 10개 이상
 식탁이 있나? 안 보이지만 분명히 있다



온톨로지(ontology)를 이용한 지식 표현

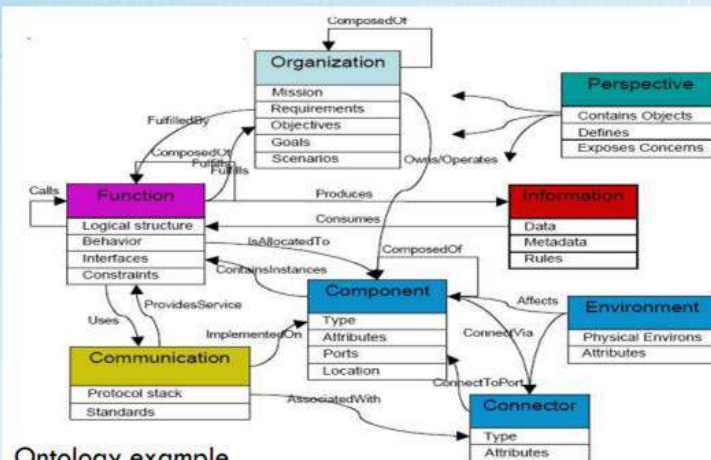
- 온톨로지(ontology)에 대한 미조구치 리이치로(溝口理一郎) 교수의 정의
 - Towards Ontology Engineering, Technical Report AI-TR-96-1, I.S.I.R., Osaka University
 - 철학적 의미는 "존재론 Theory of Existence"
 - 인공지능에서는 "개념화의 명시적 표현 An explicit representation of conceptualization"
 - 지식베이스 커뮤니티에서는 "인공 시스템을 구축하는 데에 필요한 어휘/개념의 이론"
 - A Theory of vocabulary/concepts used in building artificial system"



溝口理一郎 교수(JAIST)

지식 표현의 문제

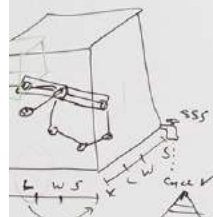
- 추이율(推移律, transitive law)의 성립 여부
 - Is-a
 - 강영민 is-a 사람 ^ 사람 is-a 동물 → 강영민 is-a 동물
 - 추이율 적용 가능
 - Part-of
 - 강영민 part-of 한국인 ^ 한국인 part-of 지구인
 - 강영민 part-of 지구인
 - 추이율 적용 가능
 - 큰머리 part-of 강영민 ^ 강영민 part-of 한국인
 - 큰머리 part-of 한국인
 - 추이율 적용할 수 없음



Ontology example

SS1 Biology DB

1. Taxons
OE



"Frame"

4th Force
to "Paris" City

(P,T,T₂)

(LL-)

(LLL)

"P" "F"

(-LL)

(- - - - -)
" - - - - -"
(LLL)



CyC Corp

Capacitación y Consultoría Corporativa

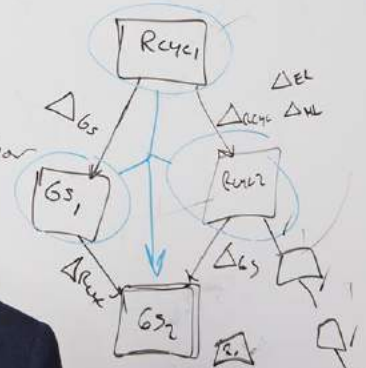
- Salient Assertions

- how to indicate e.g. type Primary function is salient?
- many Suggestions, etc.

- ds.
- contrast

- and dist. of labor
pages

MIND MEND



History of CYC

- Douglas B. Lenat and MCC (1984)
- GOFAI at a massive scale
- Construct “Common Sense” knowledge
 - 3 million rules of thumb (2002)
 - 300,000 terms



CYC: LOGICAL REASONING WITH THE WORLD'S LARGEST KNOWLEDGE BASE

그리고 2차 AI 붐의 끝

- **지식을 표현하여 입력하는 일**
 - ◆ 더 똑똑한 컴퓨터를 만들 수 있다
 - ◆ 아주 똑똑한 컴퓨터를 만들기 위해 지식을 얼마나 넣어야 하는가
 - 이것은 가능한 일일까
 - 온톨로지 연구의 심오함
 - 기계 번역의 어려움
 - 프레임 문제와 심볼 그라운딩 문제
 - 지식을 기술한다는 것은 불가능한 일이 아닌가
 - ◆ 입력할 지식의 방대함에 질려버린 사람들
- **2차 AI 붐의 끝. 또 겨울**
 - ◆ 인공지능은 금지어
 - ◆ 인공지능은 불가능한 것으로 취급



AI 연구에 장애가 된 사건들

1. 1966년 기계번역에 관한 ALPAC 보고서
"기계번역은 사람보다 비싸고 느리다"
2. 1969년 연결주의의 포기
Symbolic reasoning 쪽으로 연구 경향 이동
3. 1974년 Lighthill 보고서
영국회의의 질의에 대해 Lighthill 교수의 AI에 대한 부정적 의견
4. 70년대 DARPA 예산 삭감
미국방성에서 거의 무제한적 AI 지원을 하다가
1969년의 "mission-orient direct research"를 요구하는 법안의 통과로 예산 삭감
5. SUR debacle
DARPA가 CMU의 speech understanding research에 실망 (수백만불 예산 취소)
이후 많은 음성 인식 시스템이 이 CMU의 기술에 기반하여 성공
6. 1987년 LISP 기계 시장의 붕괴
7. 90년대 전문가 시스템 퇴조
8. 일본 5세대 컴퓨터 프로젝트가 흐지부지 종결
5세대 컴퓨터 프로젝트에 일본 정부가 8.5억 달러 투입했지만 목표 달성 실패

새로운 시대를 개척한 딥러닝

● 2012년 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)

◆ 토론토 대학의 슈퍼비전(SuperVision) 압승

Team name	Filename	Error (5 guesses)	Description
SuperVision	test-preds-141-146.2009-131-137-145-146.2011-145f.	0.15315	Using extra training data from ImageNet Fall 2011 release
SuperVision	test-preds-131-137-145-135-145f.txt	0.16422	Using only supplied training data
ISI	pred_FVs_wLACs_weighted.txt	0.26172	Weighted sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV, and CSIFT+FV, respectively.
ISI	pred_FVs_weighted.txt	0.26602	Weighted sum of scores from classifiers using each FV.
ISI	pred_FVs_summed.txt	0.26646	Naive sum of scores from classifiers using each FV.
ISI	pred_FVs_wLACs_summed.txt	0.26952	Naive sum of scores from each classifier with SIFT+FV, LBP+FV, GIST+FV, and CSIFT+FV, respectively.

Task 1
(분류)

무엇이 이런 차이를 가져 왔나?
모두들 기계학습...

SuperVision 이외의 기법들: 기계학습에 사용하는 특징의 설계를 사람이 수행
- 오차율을 낮추기 위해 다양한 특징을 설계하는 노력

SuperVision: 학습에 사용되는 데이터의 특징을 스스로 파악

“Deep Learning”



Geoffrey Hinton

딥러닝과 기타 기계학습의 차이

- 딥러닝(Deep learning)과 다른 기계학습의 차이
 - ◆ 데이터를 바탕으로 컴퓨터가 스스로 특징을 만들어 냄
- 인공지능의 주요 성과
 - ◆ 인공지능의 여명기에 몰려 있음
 - ◆ 이후의 성과는 어떤 측면에서 “minor change”
 - ◆ 딥러닝의 특징 표현 학습
 - 혁신적인 도약
 - 많은 사회적 논의가 거품 상태
 - 거대 인터넷 기업들을 중심으로 보이지 않는 움직임 활발

세계 인공지능 연구 투자의 가속화

● 글로벌 기업의 대규모 AI 투자

◆ Google

- 2013년 딥러닝의 1인자 제프리 힌튼(Geoffrey E. Hinton) 교수의 DNN 리서치 인수
- 2014년 4억 달러로 딥마인드 테크놀로지스 인수 (CEO: Demis Hassabis)
 - AlphaGo

◆ Facebook

- 2013년 인공지능 연구소 설립 (소장: Yann LeCun 교수)

◆ 바이두(百度)

- 2014년 Institute of Deep Learning(脑 脑学习研究院) 설립
- 3억 달러 투입,
- 소장: 앤드류 응(Andrew Ng) 교수 (기계학습 권위자)

◆ IBM

- Watson의 사업화
 - 10억 달러 (1조원) 투입
 - 2,000명 규모의 사업 부문 신설
 - 1억 달러 규모의 벤처 캐피탈 운용 - Watson 어플리케이션 벤처 회사에 투자

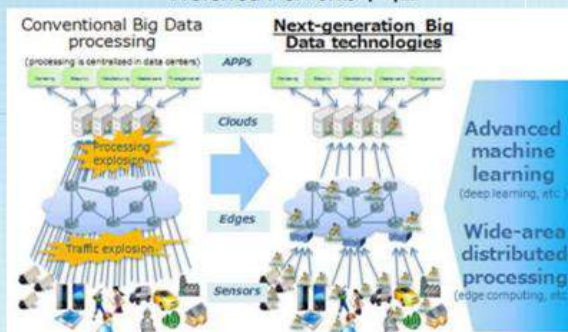
◆ Dwango (Dial-up Wide-Area Network Game Operation)

- 2014년 드완고 인공지능연구소 설립
- 소장: 야마카와 히로시(山川宏)

◆ PFI

- 2014년 빅데이터 관련 개발사 PFI(Preferred Infrastructure)가 설립한 Preferred Networks
 - 실세계 감지와 Deep-learning을 결합하여 인간 능력을 넘는 분석을 하고, IoT 기반으로 이러한 것들을 연결
- NTT가 2억엔 출자

Preferred Networks의 목표



Credit: NTT Laboratories

일 자리를 잃는 인간 (그리고 공포)

● 인간은 일자리를 잃을 것인가?

- ◆ 일시적으로 인간의 일자리를 빼앗을 가능성이 없지 않다
- ◆ 2014년 영국 딜로이트(Deloitte) 사의 전망
 - 영국의 일자리 가운데 35%가 20년 이내 로봇으로 대체
 - 일자리를 잃을 가능성은 연봉이 낮을수록 높음
- ◆ 옥스포드 대학의 연구
 - IT 기술로 20년 이내에 미국내 직업 50% 정도가 사라질 것으로 예측



● 영화에 비친 인간을 대신하는 컴퓨터

- ◆ 2014년 트랜센던스(Transcendence) – 인간을 컴퓨터에 업로드
- ◆ 2014년 허(Her) – 운영체제와의 사랑, 그리고 운영체제의 바람(이것은... 창의적이다)
- ◆ 2015년 이미테이션 게임(Imitation game) – 앨런 튜링(Alan Turing)의 일생
- ◆ 1968년 스탠리 큐브릭(Stanley Kubrick)의 2001년 스페이스 오딧세이(2001: Space Odyssey)
 - HAL 9000
 - 자유의지와 생존의지
- ◆ 1984년 터미네이터(Terminator) – 인공지능의 파괴적 능력에 대한 공포



인류 위기의 도래 - 기술적 특이점(Technological Singularity)

● 기계가 우리를 지배하게 되는 때는 언제인가?

◆ 특이점(singularity)

- 레이 커즈와일(Ray Kurzweil)의 주장
- 인공지능이 자신보다 나은 인공지능을 만들기 시작하는 시점
- 커즈와일은 2045년으로 예측

● 아이언 맨 일런 머스크(Elon Musk)

- ◆ “인공지능을 상당히 신중하게 취급할 필요가 있다. 결과적으로 악마를 호출하기 때문이다.”

● 빌 게이츠(Bill Gates)

- ◆ “나도 인공지능을 걱정하는 부류에 있는 한 사람”

● 윤리의 문제 부각

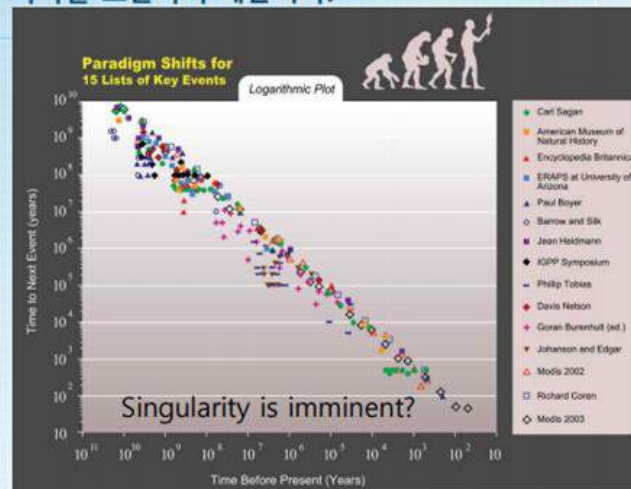
◆ Google

- DeepMind 인수하며 사내에 인공지능에 관한 윤리위원회 설치

◆ 일본 인공지능학회

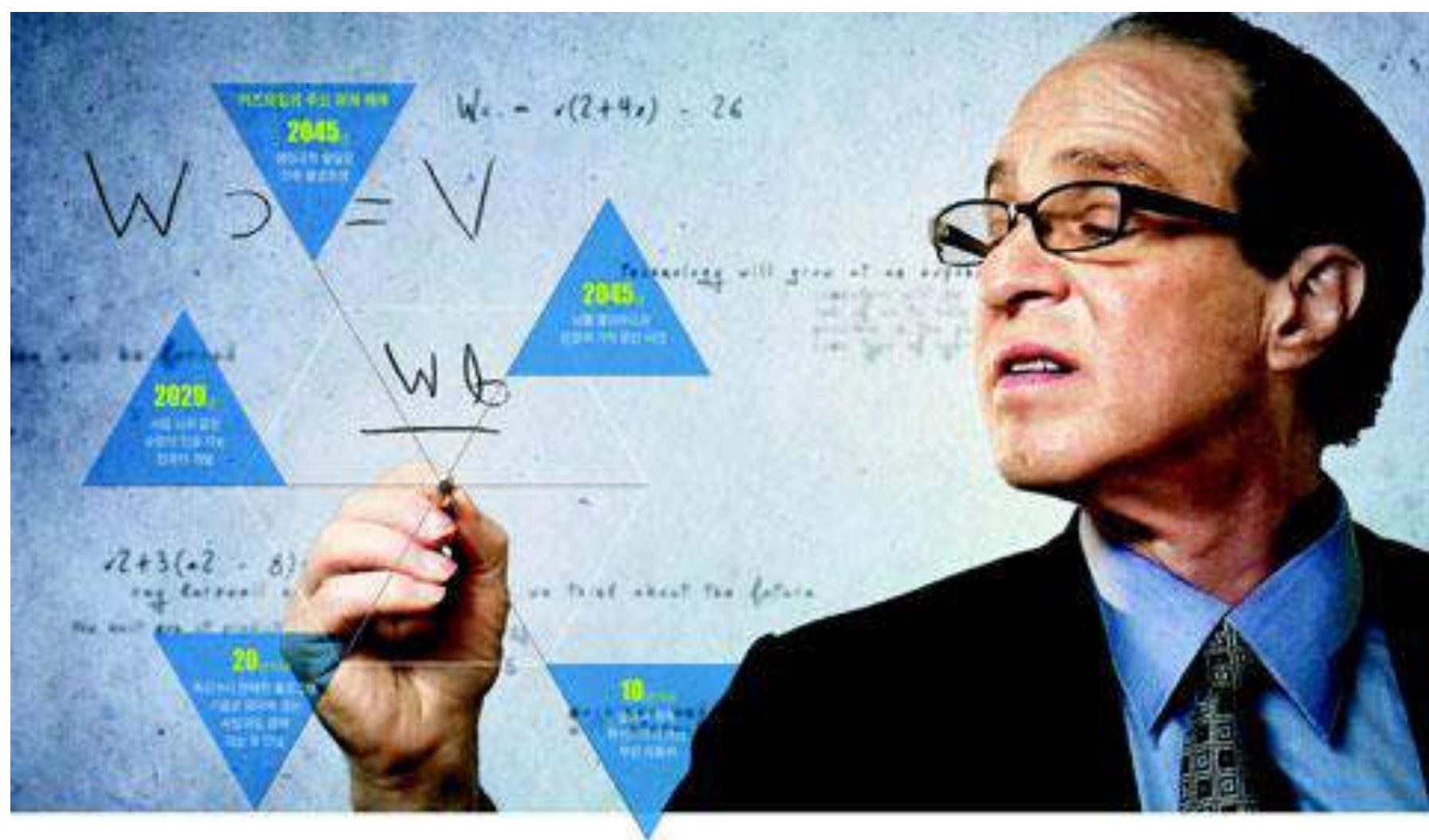
- 2014년 윤리위원회 신설

● 우리는 결국 멸망할 것인가?



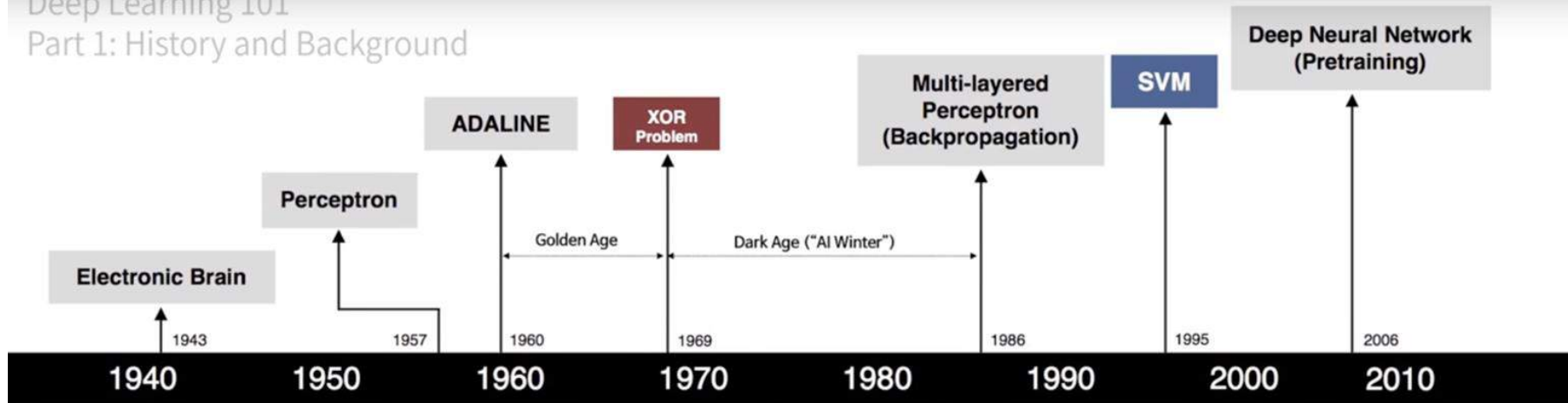
Ray Kurzweil – 2005 • The Singularity “It’s a future period during which the pace of technological change will be so rapid, its impact so deep, that human life will be irreversibly transformed.” o Kurzweil, “The Singularity is Near” pg. 7



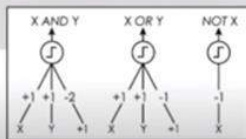


Deep Learning 101

Part 1: History and Background



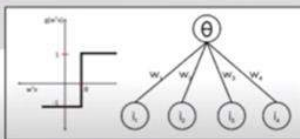
S. McCulloch - W. Pitts



- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



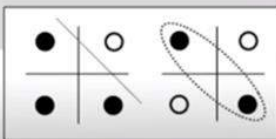
- Learnable Weights and Threshold



B. Widrow - M. Hoff



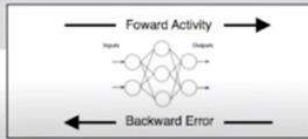
M. Minsky - S. Papert



- XOR Problem



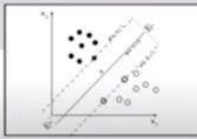
D. Rumelhart - G. Hinton - R. Williams



- Solution to nonlinearly separable problems
- E.g. computation, image processing and handwriting



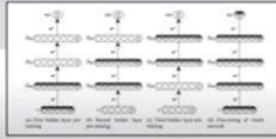
V. Vapnik - C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Non-linear Intervention



G. Hinton - S. Ruslan



- Hierarchical feature Learning

첫 번째로 아이디어 자체는 굉장히 오래된 아이디어거든요?

A



B



C

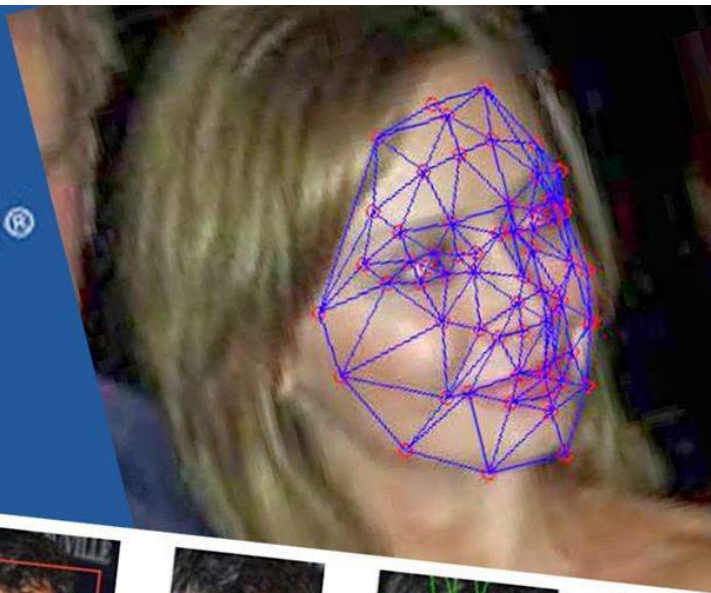


D



facebook

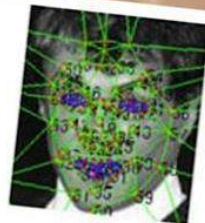
FACEBOOK
DEEPFACE



(a)



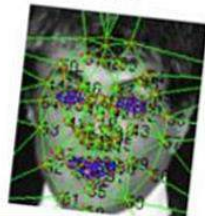
(b)

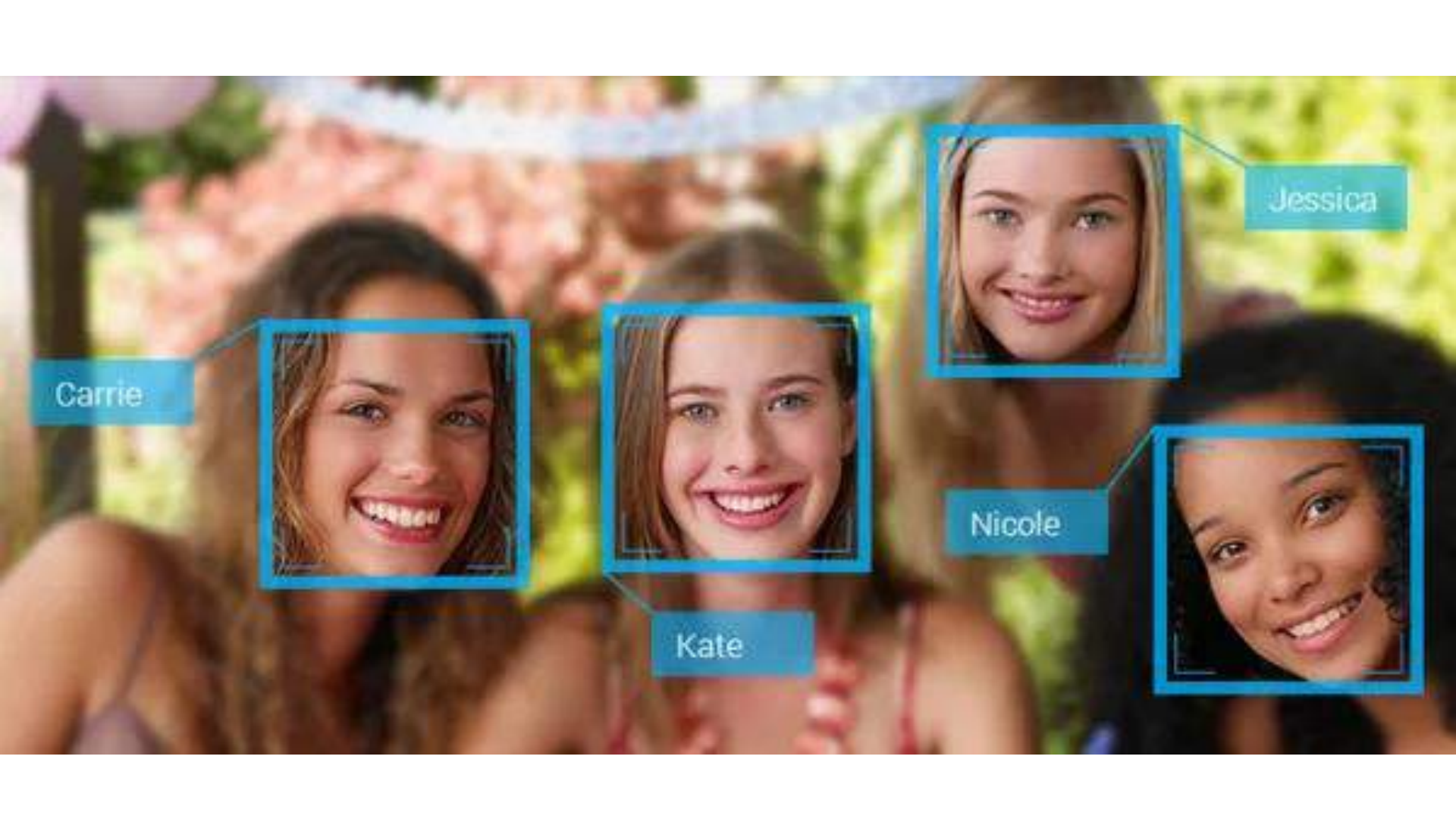


(c)



(d)





Carrie



Kate

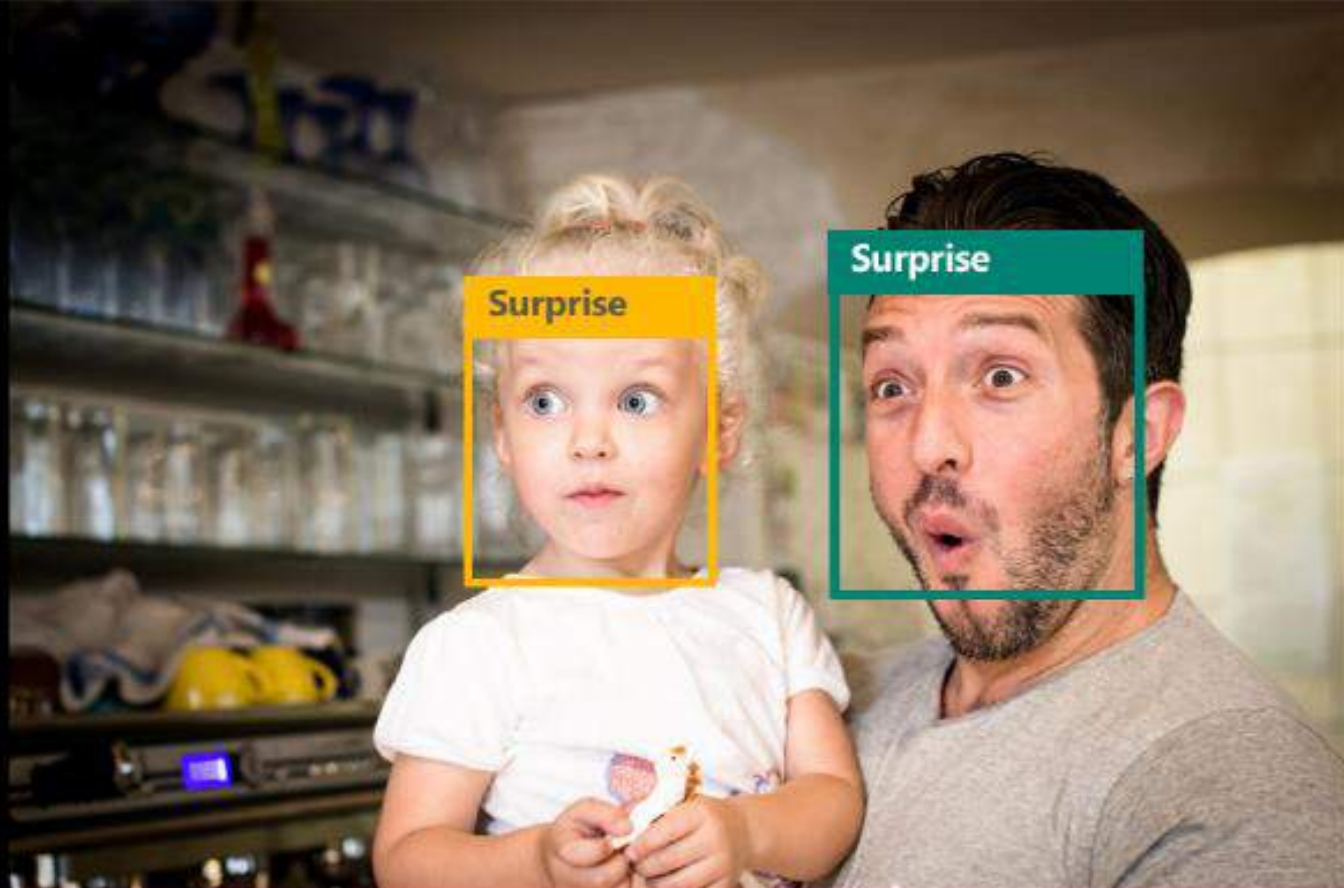


Jessica



Nicole





Neutral:
Happiness:
Surprise:
Sadness:



Anger:
Disgust:
Fear:
Contempt:



 Microsoft

Get started for free at projectoxford.ai



자동출입국심사 이용 방법



얼굴인식
Facial Recognition



출입국·외국인정책본부
KOREA IMMIGRATION SERVICE



ALL

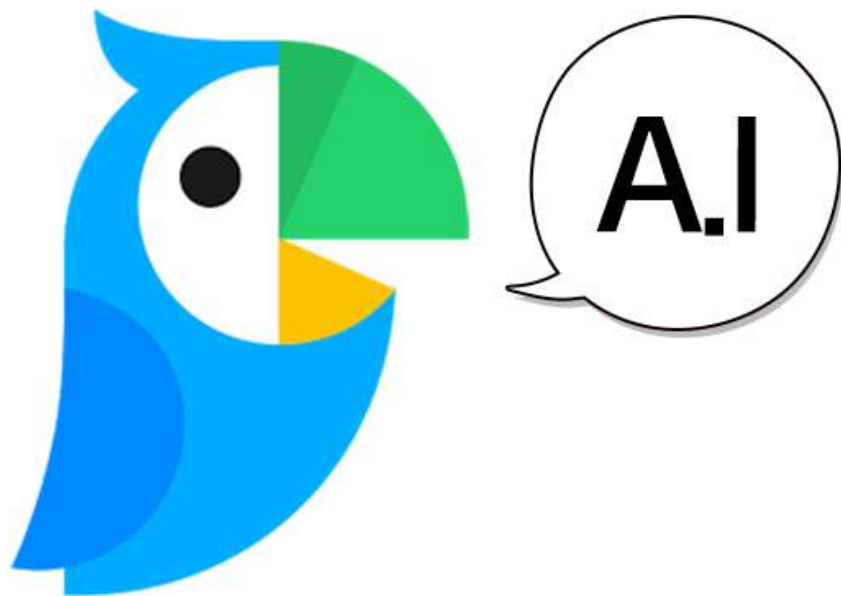
DEEP STYLE

THIN STYLE

DEEP DREAM

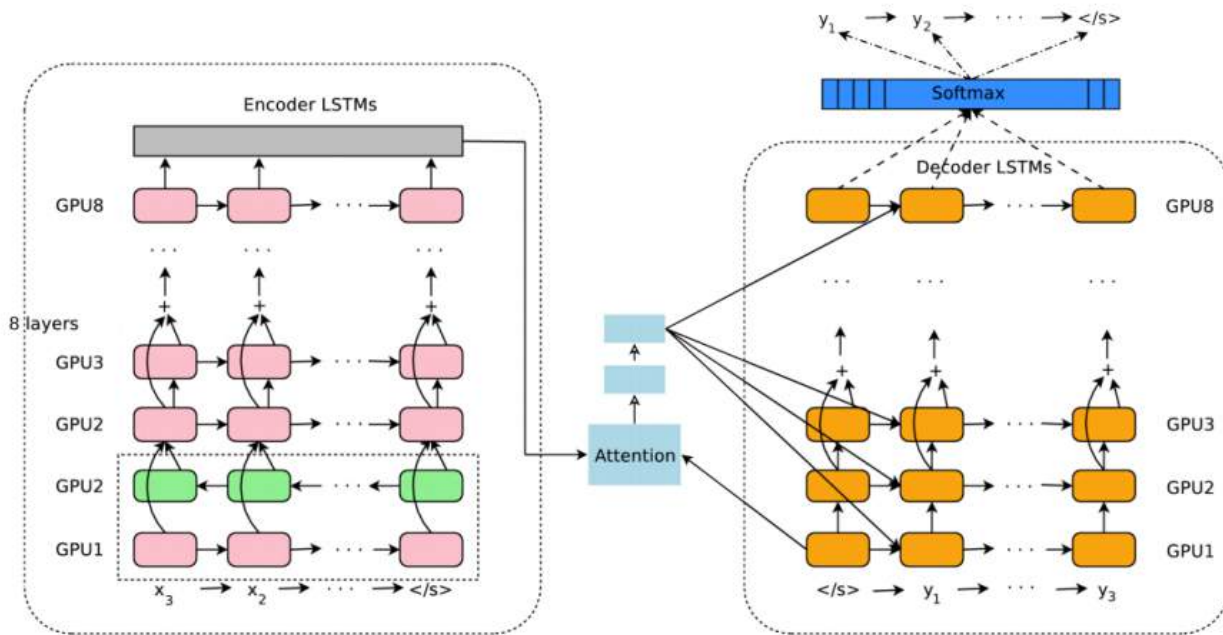


네이버 번역기 파파고,
인공지능으로 발전하다



구글 뉴럴 기계번역기(2016)

전체 구성은 Encoder-Decoder



Encoder와 Decoder는 각각 8층의 순환신경망으로 구성
순환신경망을 구성하는 기본 모듈로는 LSTM을 사용

Heroes of Deep Learning: Andrew Ng interviews Geoffrey Hinton

전체화면을 종료하려면 Esc 을(를) 누르세요.

beauty and instead of letting things
settle down just use one iteration in a

Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증
합니다.



9:35 / 39:45



구글 자율주행차 특징

유선형 외부

센서 감도 높이기 위한 디자인

인공지능

운행 중 발생할 수 있는
여러 시나리오 학습

엔진

전기 배터리

자료:구글 자율주행차 페이지

위치 파악하는 레이더

360도로 방향 인식, 지도·센서·카메라 이용해
차의 현재 위치 파악

핸들·페달 없는 내부

2인용 좌석, 출발과 정지버튼,
운행정보 보여주는 스크린



위험물 피하는 센서

소프트웨어·카메라·센서 이용해
보행자·자전거·자동차 등을 인지,
위험 피함

PIENSE

THINK

ΣΚΕΨΟΥ

DENKE

SMACHUŠ

PENSER

\$24,000

Who is Stoker?
\$ for each question on
the current situation
\$1,000

\$77,147

Who is Bram
Stoker?
\$ 17,973

\$21,600

Who is
BRAM STOKER?
\$5000



Siri





echo
amazon

체험해봅시다 1: 영어 필기체 합성

접속

- Alex Graves의 데모 페이지 접속
- <http://www.cs.toronto.edu/~graves/handwriting.html>

글자입력

- 영어로 100글자까지 'Text'란에 입력
- 되도록이면 소문자 위주로 입력하세요.

스타일

- 필기체 스타일 선택: Style
- 글자를 정자체로 쓸지 갈겨 쓸지 선택: bias (오른쪽으로 밀 수록 정자체에 가까움)

샘플

- 몇 개의 필기체를 생성할지 지정
- 각 샘플은 jpg 파일로 다운로드 가능

There's no place like home.

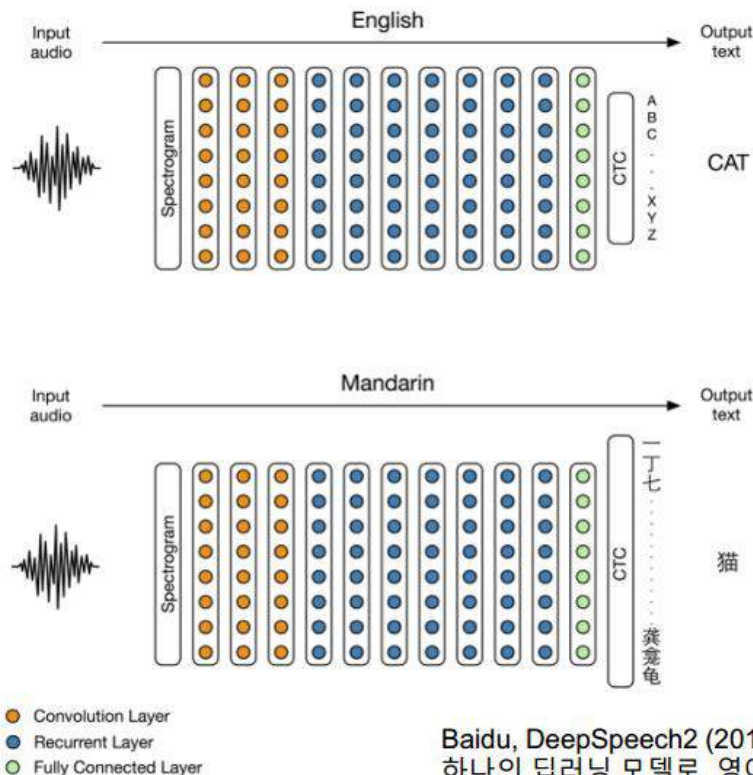
There's no place like home.
There's no place like home.

Let's count: 1, 2, 3

Let's count: 1, 2, 3
Let's count: 1, 2, 3

RNN 적용 대표 사례: 뉴럴 음성 인식기

RNN 모듈을 여러 층을 쌓아 복잡한 문제를 해결



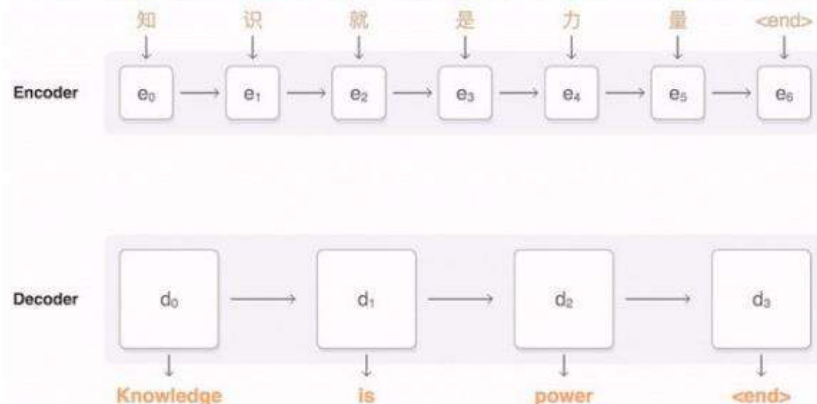
Baidu, DeepSpeech2 (2015)

하나의 딥러닝 모델로, 영어와 중국어를 모두 인식

RNN 적용 대표 사례: 뉴럴 기계번역

Neural Machine Translation (NMT)

입력에 대한 RNN과 출력에 대한 RNN을 연결하여 구성



Input sentence:	Translation (PBMT):	Translation (GNMT):	Translation (human):
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

<https://research.googleblog.com/2016/09/a-neural-network-for-machine.html>

RNN 적용 대표 사례: 지문을 보고 질문에 대해 답하기

Memory Network: 순환신경망에 외부 메모리를 붙인 모델

End-To-End Memory Network for bAbI Tasks

Story

daniel went to the office
john moved to the garden
john went back to the kitchen
daniel moved to the garden
mary went to the kitchen
daniel went to the bedroom
john went back to the hallway
sandra travelled to the garden
sandra travelled to the bedroom
daniel moved to the kitchen

Question ⓘ

is sandra in the bedroom?

Answer

Predict answer

Get new story

Text

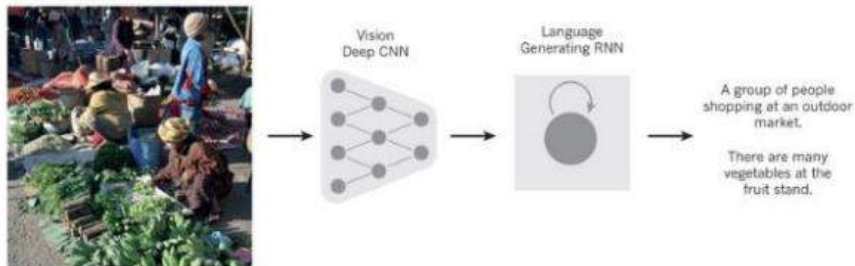
Mem 1

Mem 2

Mem 3

딥러닝 모델의 조합

CNN과 RNN을 조합하여 이미지의 캡션을 자동으로 생성



A woman is throwing a **frisbee** in a park.



A **dog** is standing on a hardwood floor.



A **stop** sign is on a road with a mountain in the background



A little **girl** sitting on a bed with a teddy bear.



A group of **people** sitting on a boat in the water.



A giraffe standing in a forest with **trees** in the background.

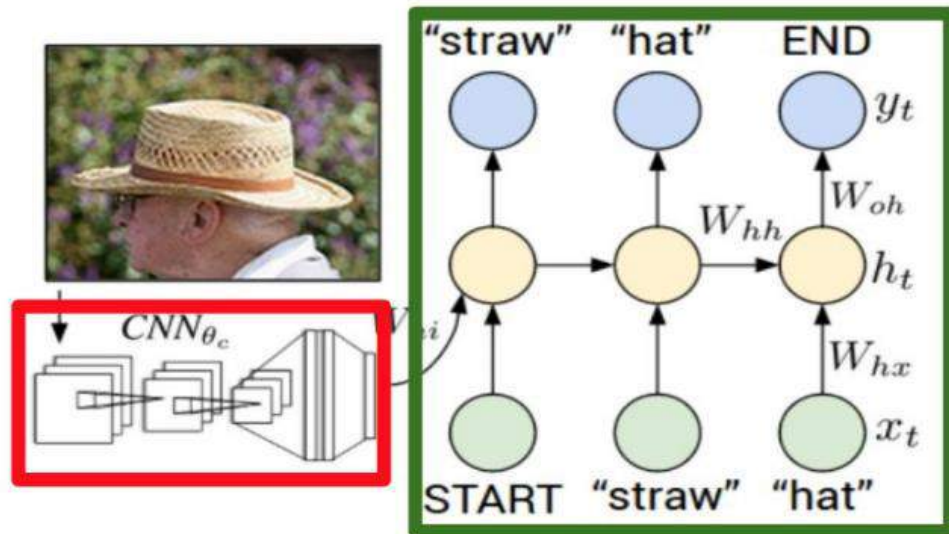
(Y. Lecun, Y. Bengio, and G. Hinton, 2015)

http://www.nature.com/nature/journal/v521/n7553/fig_tab/nature14539_F3.html

© 2017, SNU-CSE Biointelligence Lab., <http://bi.snu.ac.kr>

이미지 캡션 생성 모델

Recurrent Neural Network



Convolutional Neural Network

Google DeepMind


Challenge Match

8 - 15 March 2016



Go



A photograph of Sergey Brin, co-founder of Google, speaking into a microphone. He has dark, curly hair and a beard, and is wearing a dark shirt. The background is dark and out of focus.

구글의 공동 창업자 세르게이 브린 알파벳 CEO은 지난 4월 ‘창업자 서한(founder’s letter)’을 통해 “인공지능 기술의 부활은 내 생애에 일어난 가장 중요한 컴퓨터 기술 발전”이라며 “알파벳이 계속 딥러닝 기술의 선두 주자가 되려면 기술과 윤리적 진전을 같이 이뤄야 한다"고 말했다.

원문보기:

http://biz.chosun.com/site/data/html_dir/2018/06/04/2018060400011.html#csidxf42e2ca6131e7e8a21c1f42dede1bc6



브린 알파벳 CEO은 사진 인식(구글 포토), 사물 인식(자율주행차 웨이모), 하드웨어의 소리와 카메라 성능 개선, 음성 인식(구글 홈), 100여개 언어 번역(구글 번역), 10여개 언어로 10억여개 동영상의 캡션 달기(유튜브), 데이터 센터 효율 개선(구글 클라우드), 이메일 기능 지원(지메일), 당뇨병망병 진단 등 질병 연구(구글 라이스 사이언스), 새로운 우주 행성 발견(구글 브레인), 자동화 머신러닝(AutoML) 등 구글의 거의 모든 사업 분야에서 인공지능 기술이 핵심이 되고 있다고 밝혔다.

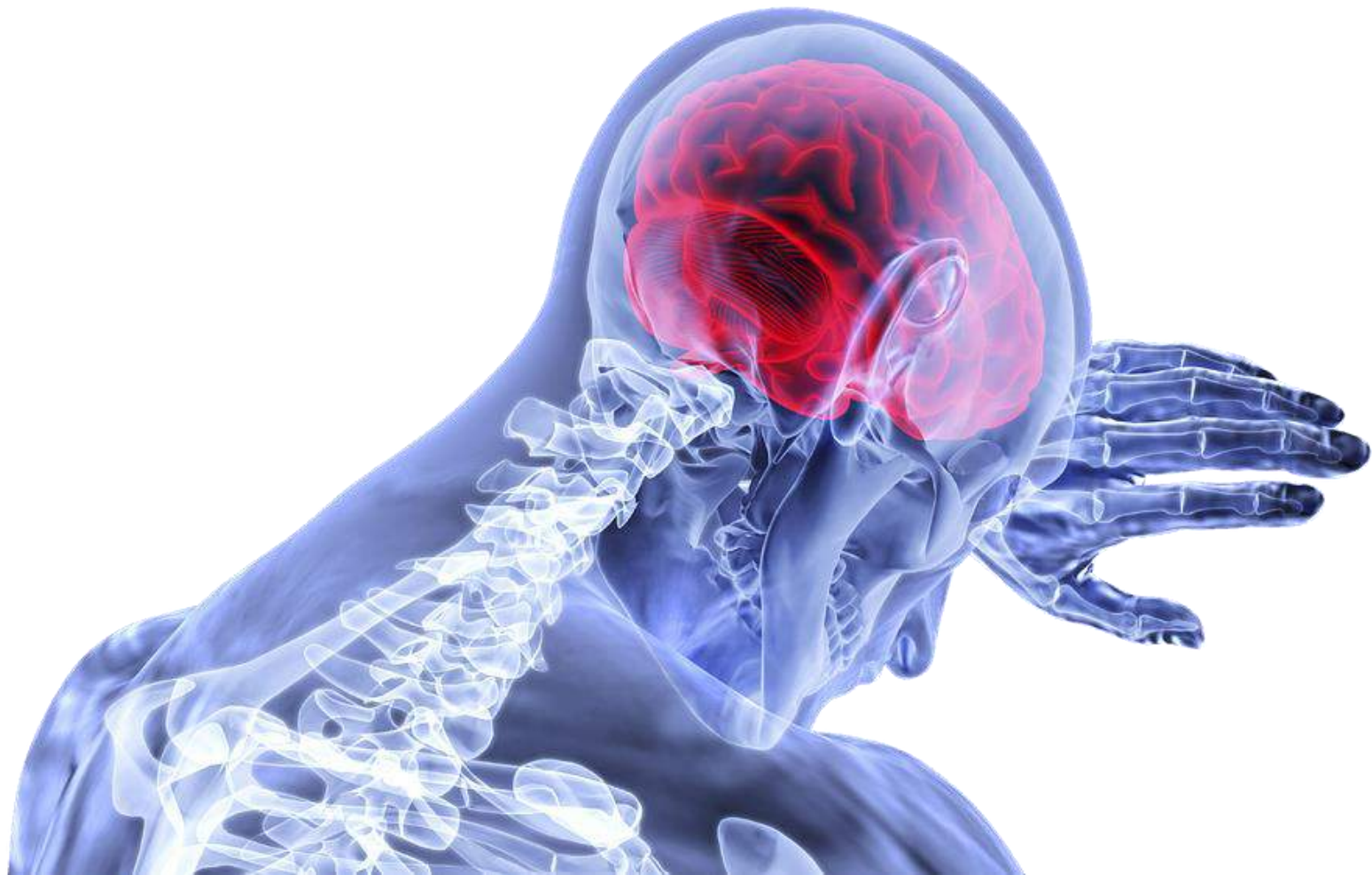
원문보기:

http://biz.chosun.com/site/data/html_dir/2018/06/04/2018060400011.html#csidx8fa03a4d21bdaa3a5347ac3bf23e083

The image shows a person's face, specifically their eyes and nose, in a close-up shot. The person is looking directly at the camera. The background is out of focus, showing a computer screen with lines of code, likely JavaScript or CSS, which are partially legible. The code includes various functions and variables, such as `function b(b)`, `c.VERSION`, and `c.RESET`. The overall tone of the image is technical and focused.

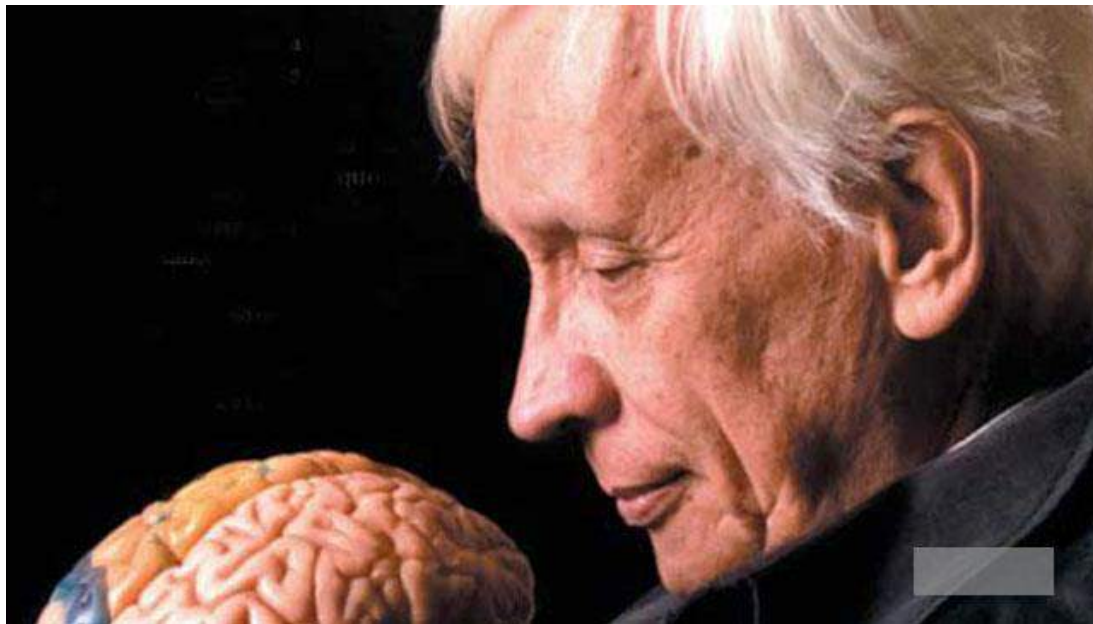
CONNECTIONIST AI

- Consequently:
- Connectionist is more flexible than symbolic AI.
- Connectionism is weak at doing logic.
- What symbolic AI does well, connectionism does badly, and the opposite.
- Hybrid systems combine the two, switching between them as appropriate.



Rodolfo R Llinás

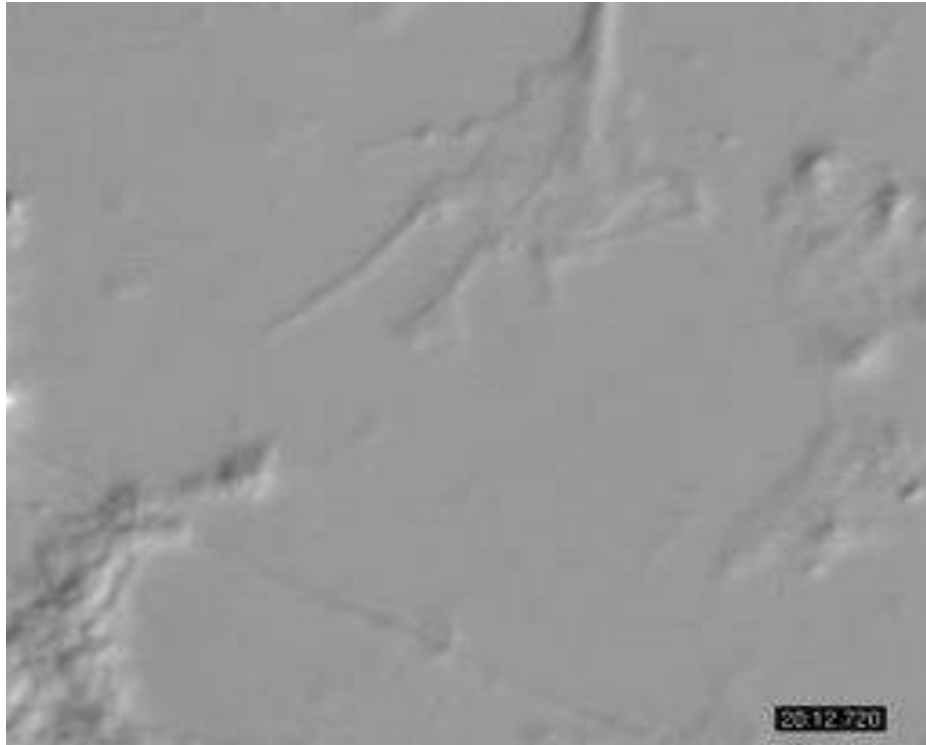
진화의 측면에서 보면 생각은 내면화된 운동이다.

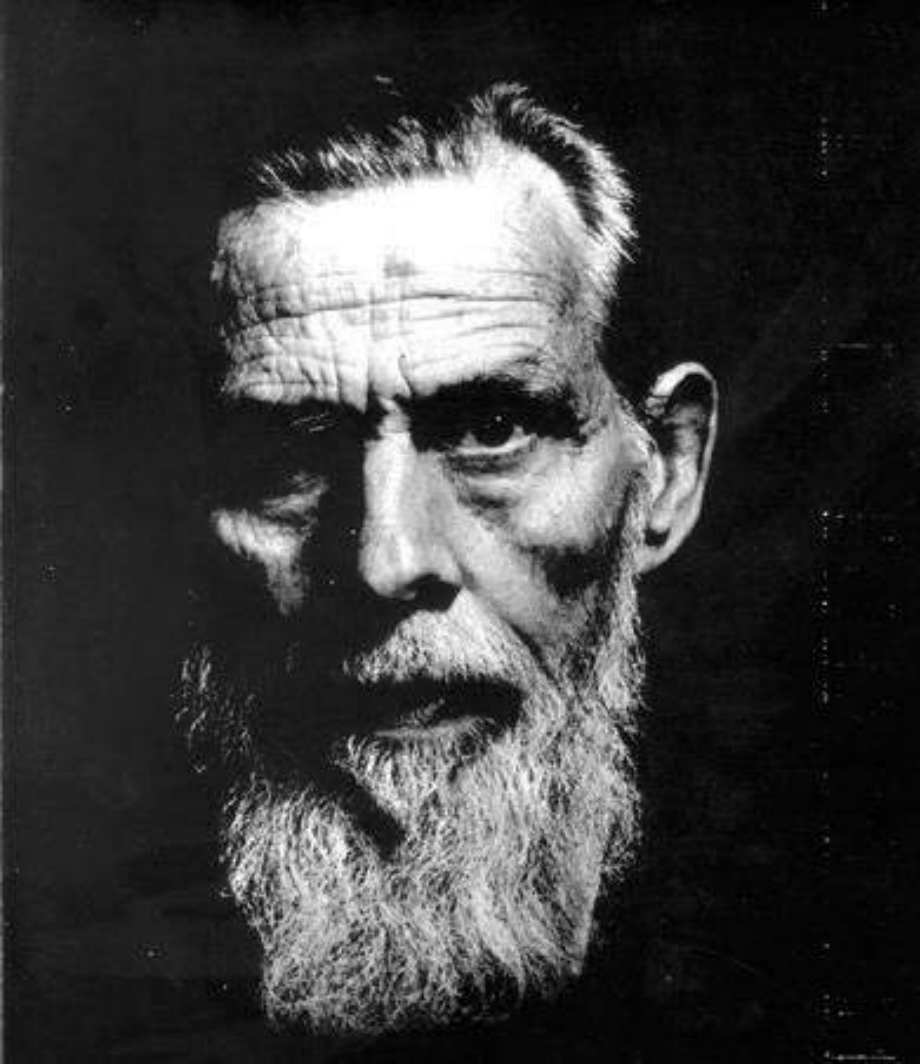


출처 : <https://www.arcoiris.com.co/2013/02/vivimos-para-trabajar-para-pensar-para-cambiar-eso-es-la-vida-rodolfo-llinas/>

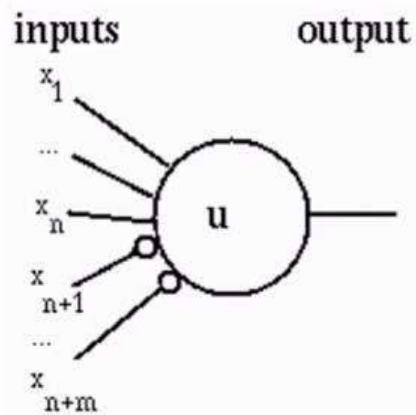
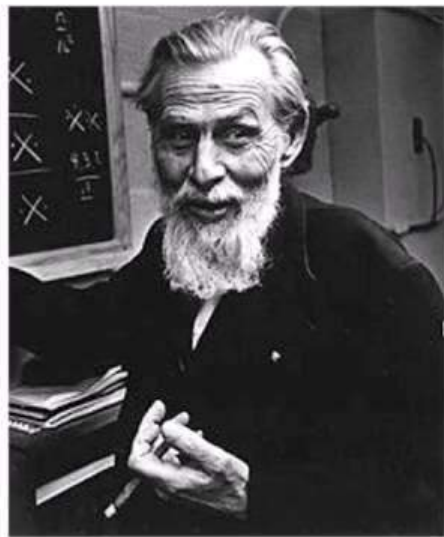
Robert S. McNeil

<http://www.neuralimages.org/>





Neurona de McCulloch – Pitts



Bulletin of Mathematical Biology Vol. 52, No. 1/2, pp. 99–115, 1990.
Printed in Great Britain.

0092-8240/90\$3.00+0.00
Pergamon Press plc
Society for Mathematical Biology

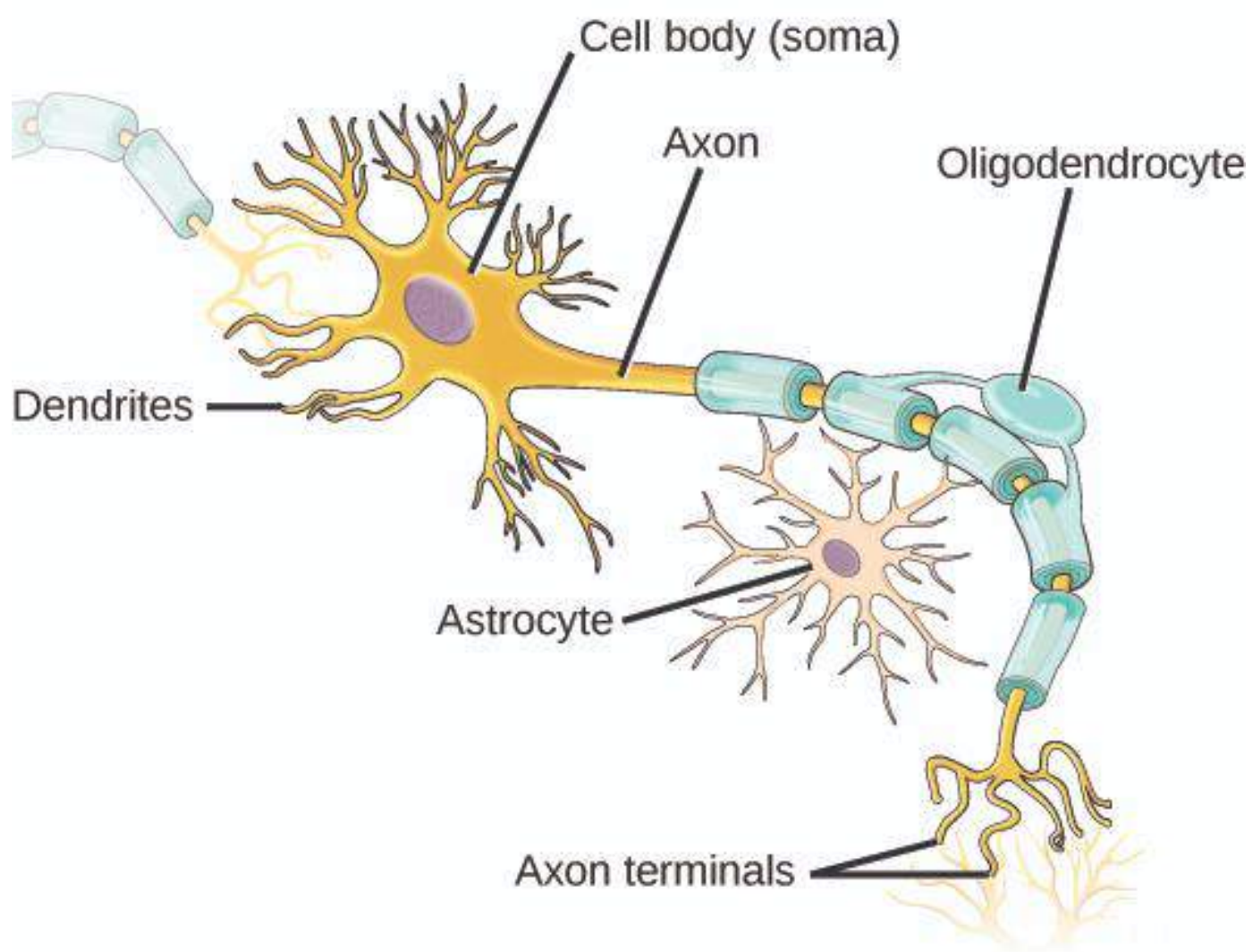
A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*

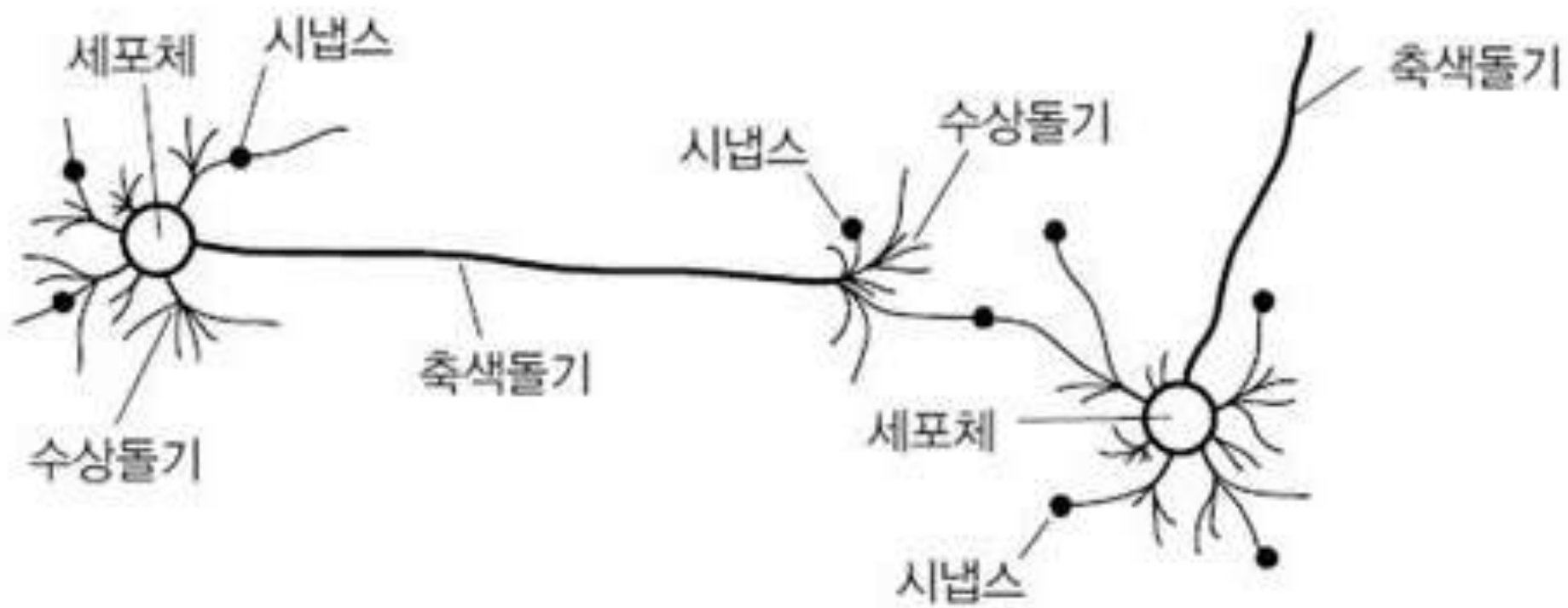
■ WARREN S. MCCULLOCH AND WALTER PITTS

University of Illinois, College of Medicine,
Department of Psychiatry at the Illinois Neuropsychiatric Institute,
University of Chicago, Chicago, U.S.A.

Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

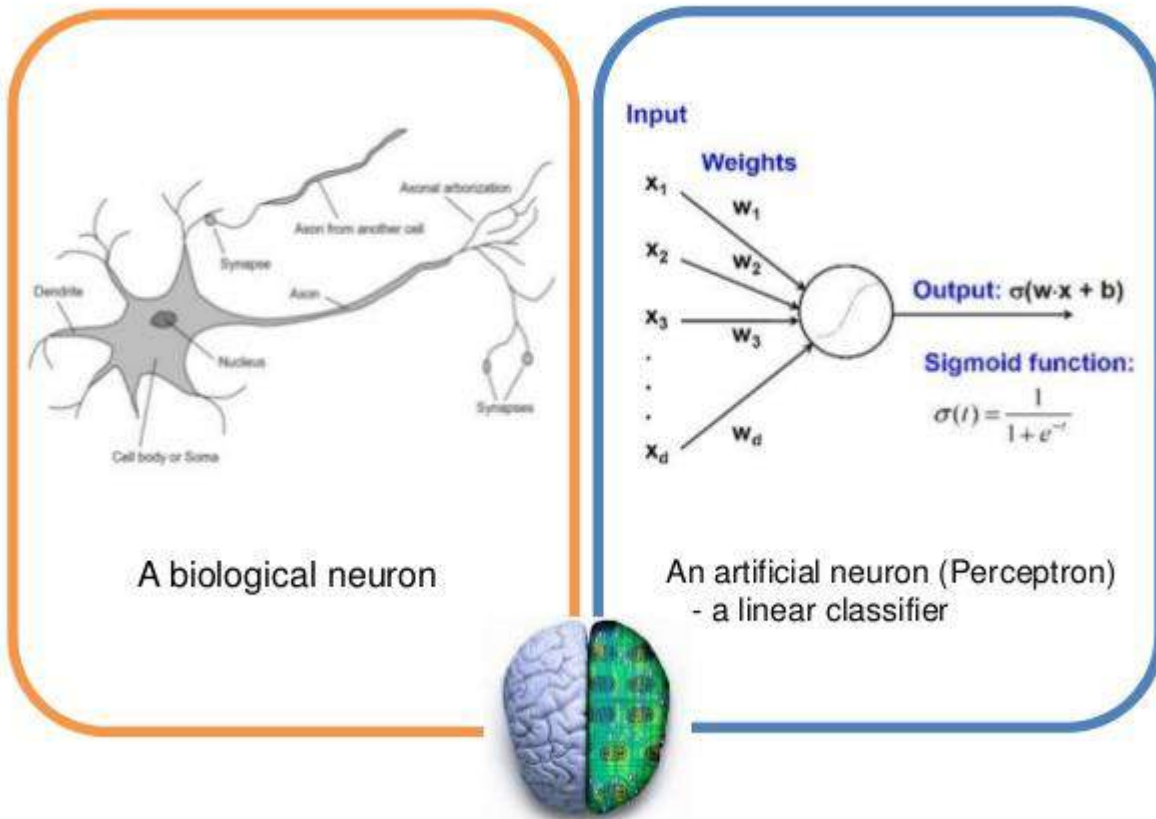
1. Introduction. Theoretical neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a soma and



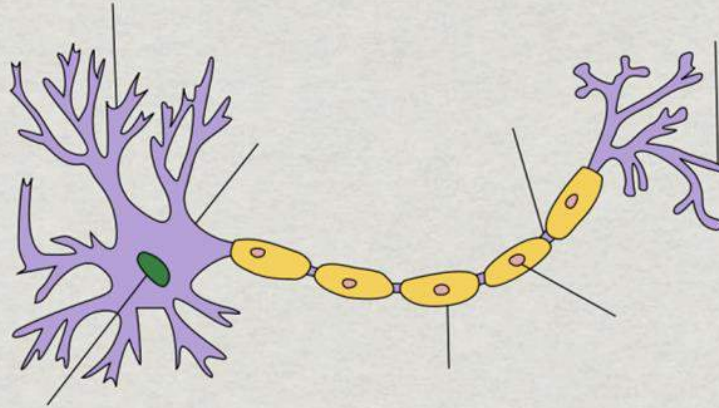


[그림 6-1] 생물학적인 신경망

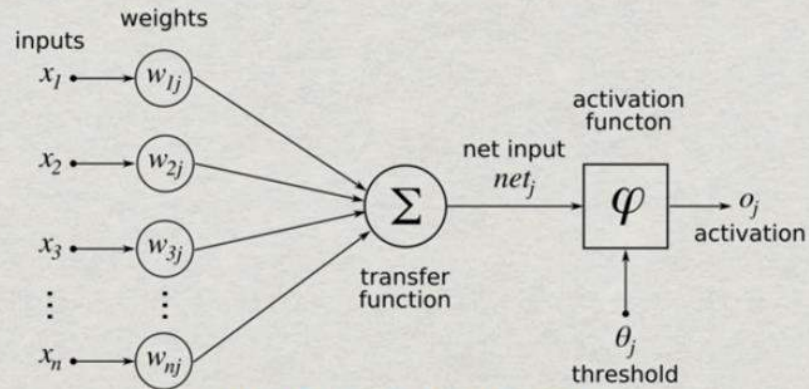
Biological neuron and Perceptrons



Neurons, Real and Artificial



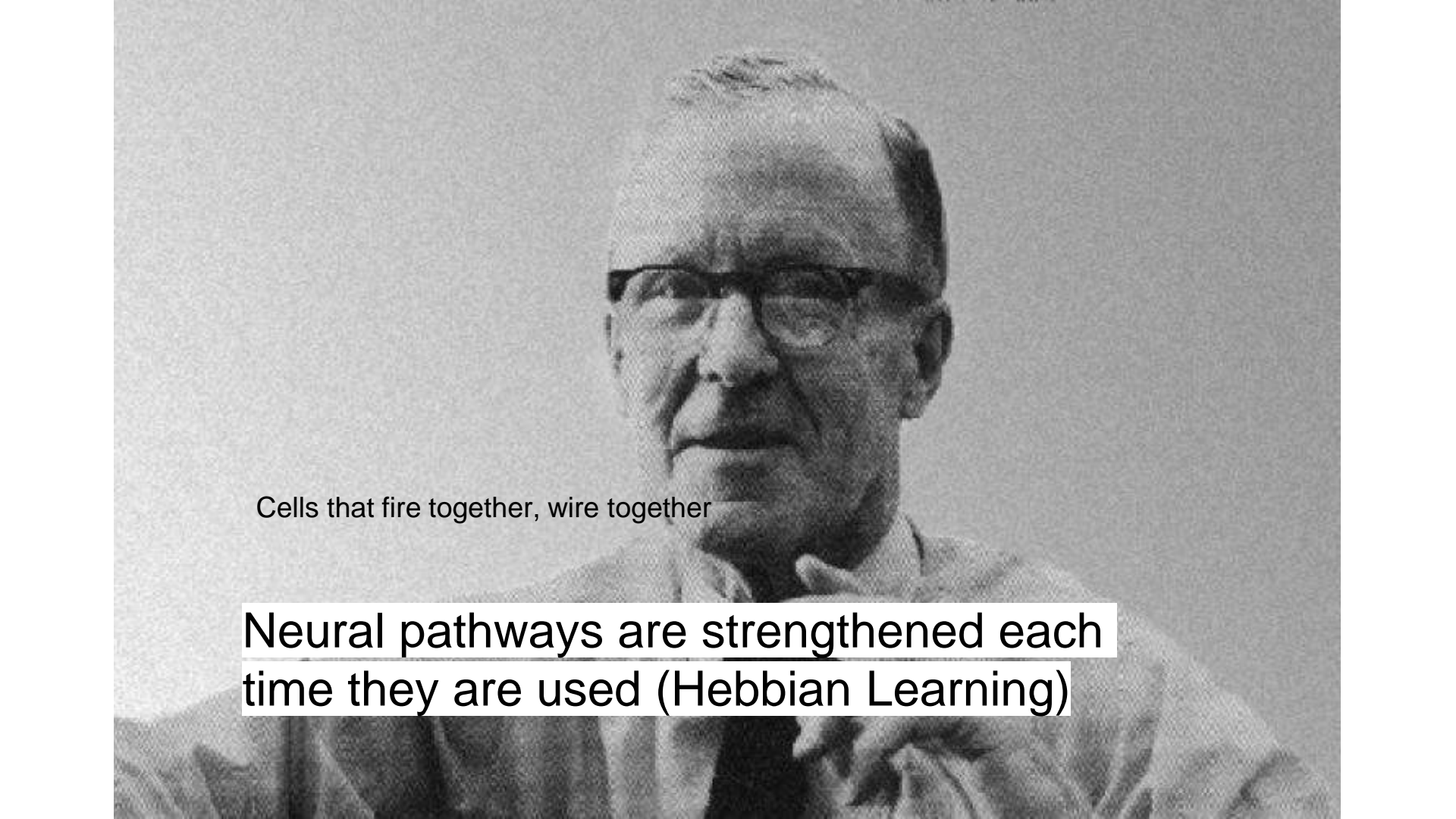
NEURON HAND-TUNED
QUASAR JAROSZ UNDER GFDL - WIKIMEDIA COMMONS





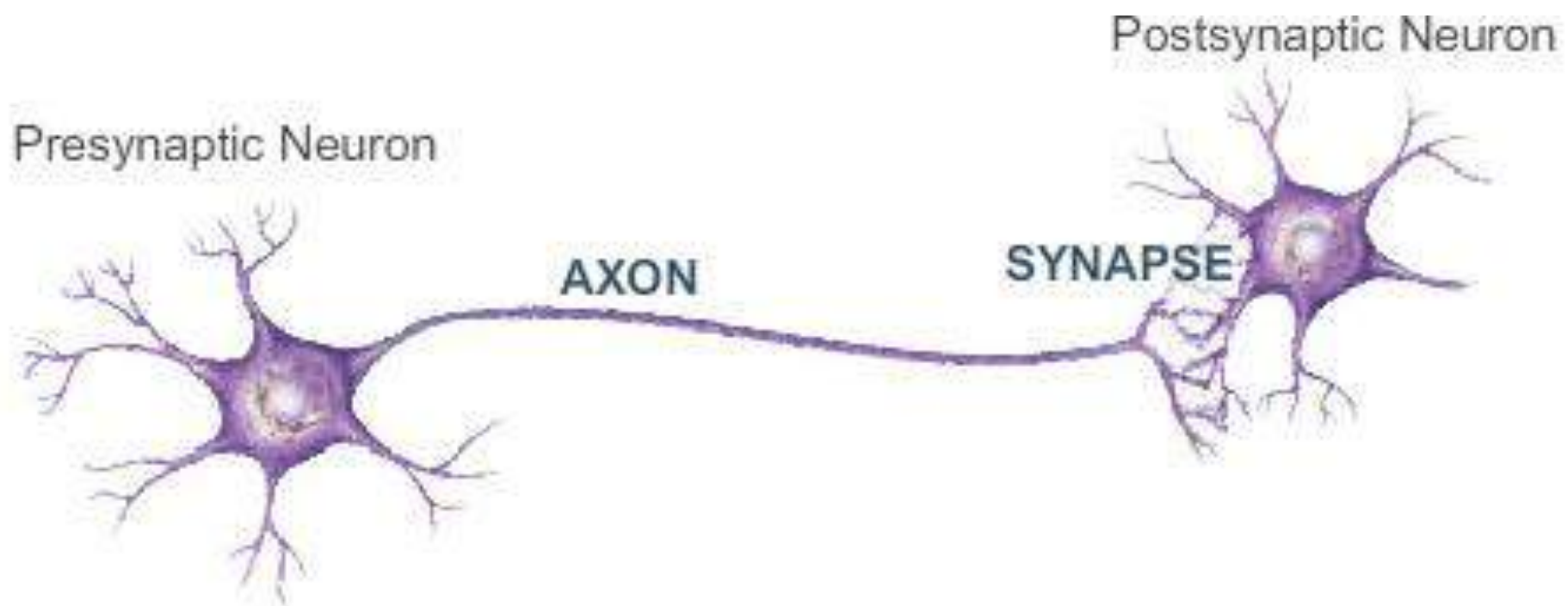


fineart
america



Cells that fire together, wire together

Neural pathways are strengthened each time they are used (Hebbian Learning)



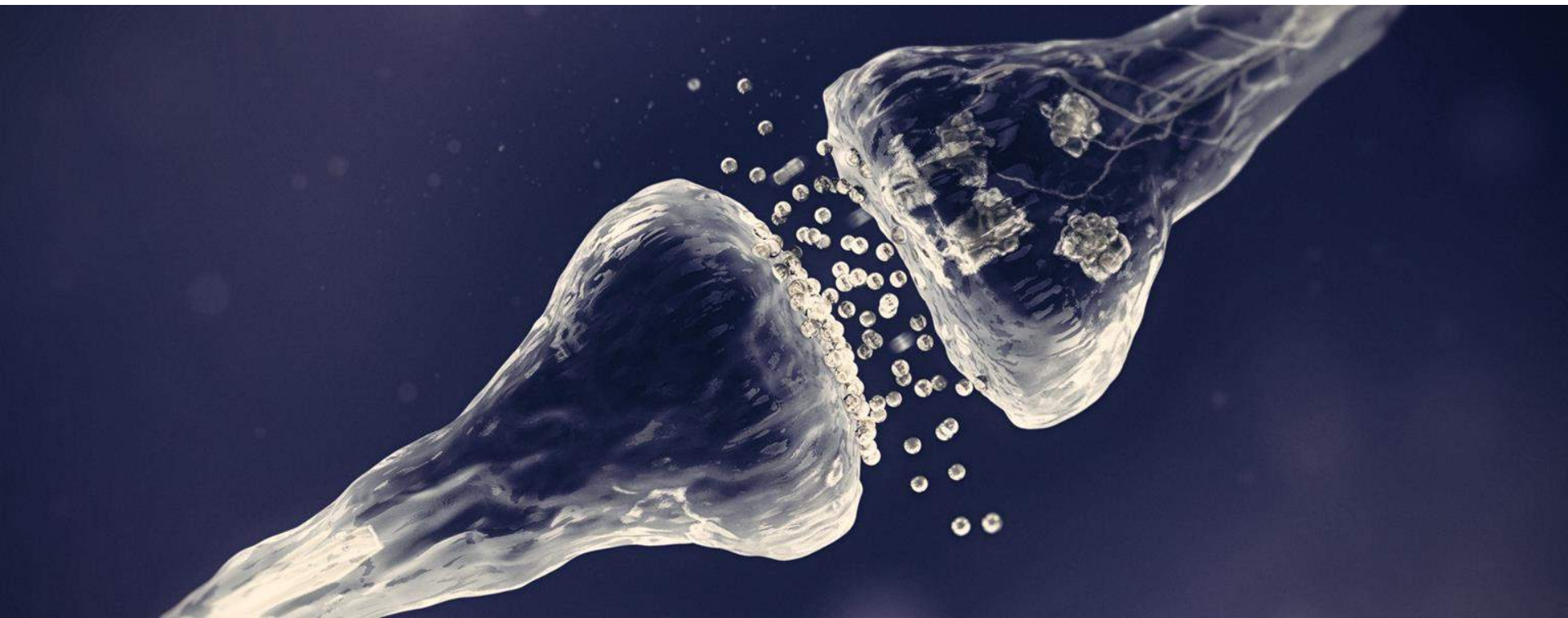


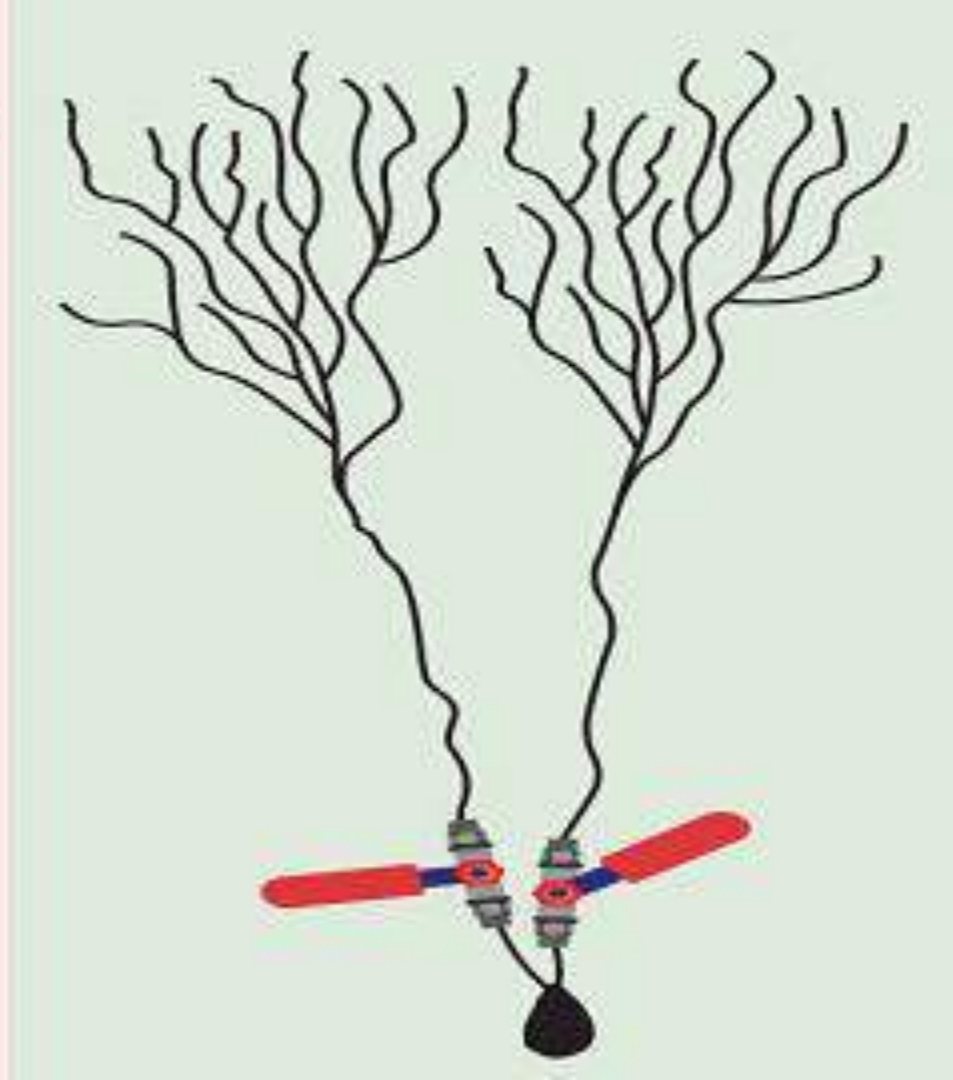
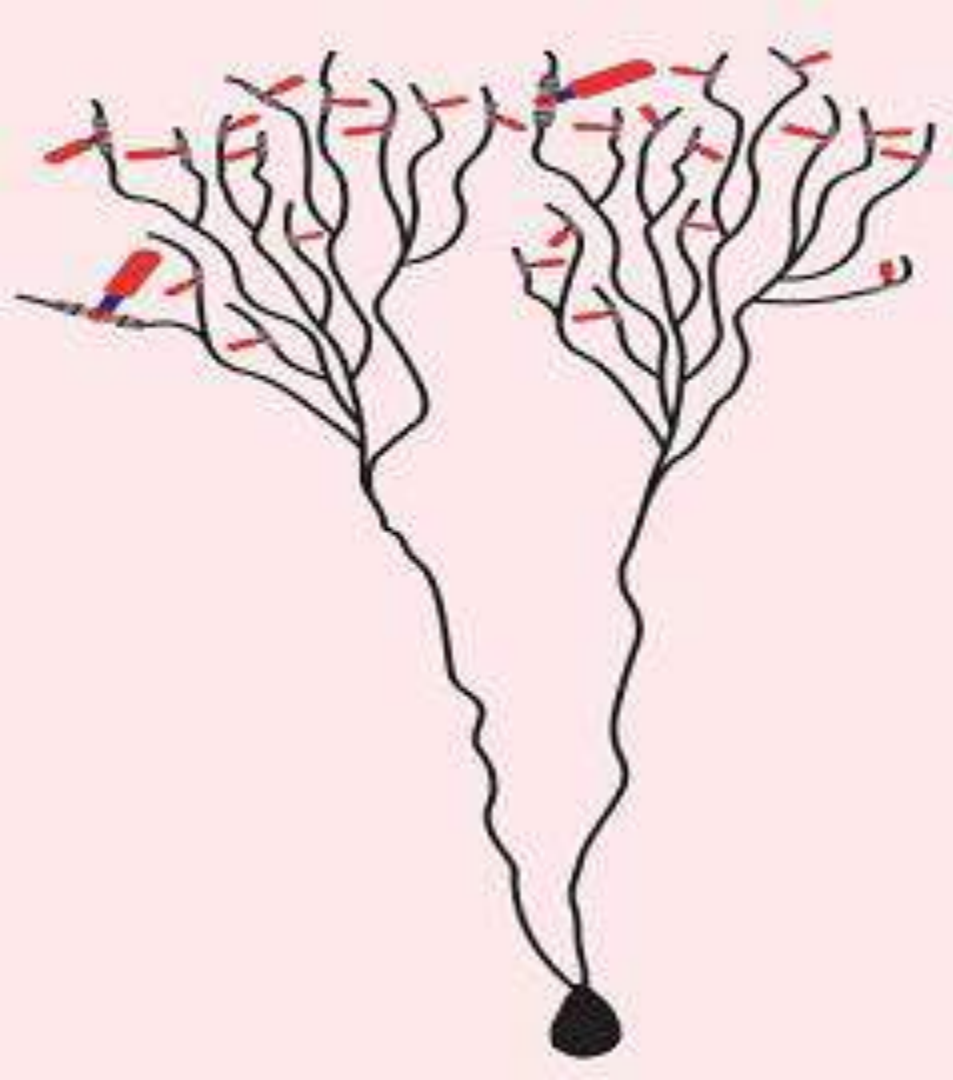
"When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing B, **growth process** or **metabolic change takes place** in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

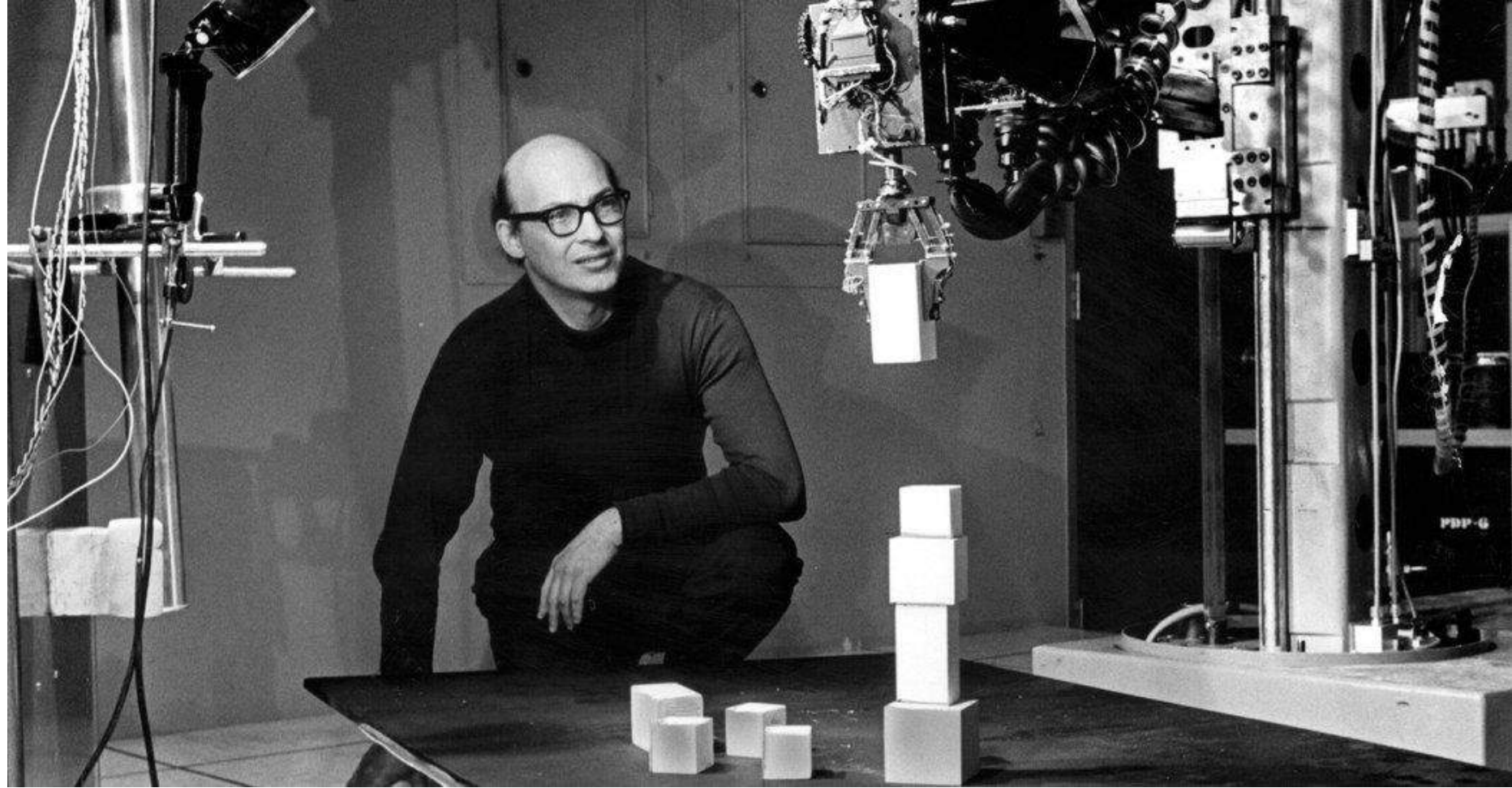
D. O. Hebb, *Organization of Behavior*, 1949

1949년 캐나다의 심리학자인 [Donald Hebb](#) 은 그의 저서인 'The Organization of Behavior' 에서 '헵의 시냅스'라고 알려진 시냅스의 연결강도 조정을 위한 생리학적 학습 규칙을 기술했는데 이 책은 1949년경의 신경심리학에 대해 폭넓고 깊이있게 기술하고 있다. 그는 이 책의 도입 부분에서 복잡한 두뇌 모델링에 대해 '커넥셔니즘(connectionism)'이란 말을 처음으로 사용하였다 "우리는 어떻게 학습하는가?"에 대한 비교적 간단한 학습 이론에 대하여 살펴보자. 중심적인 아이디어는 다음과 같다

두 개의 뉴런 A, B 가 서로 반복적이고 지속적으로 점화(firing)하여 어느 한쪽 또는 양쪽 모두에 어떤 변화를 야기한다면 상호간의 점화의 효율 (weight) 은 점점 커지게 된다. 헵의 학습 규칙은 나중에 개발된 다른 신경망 모델들의 학습 규칙의 토대가 된다 ([by 김대수](#) 1992)



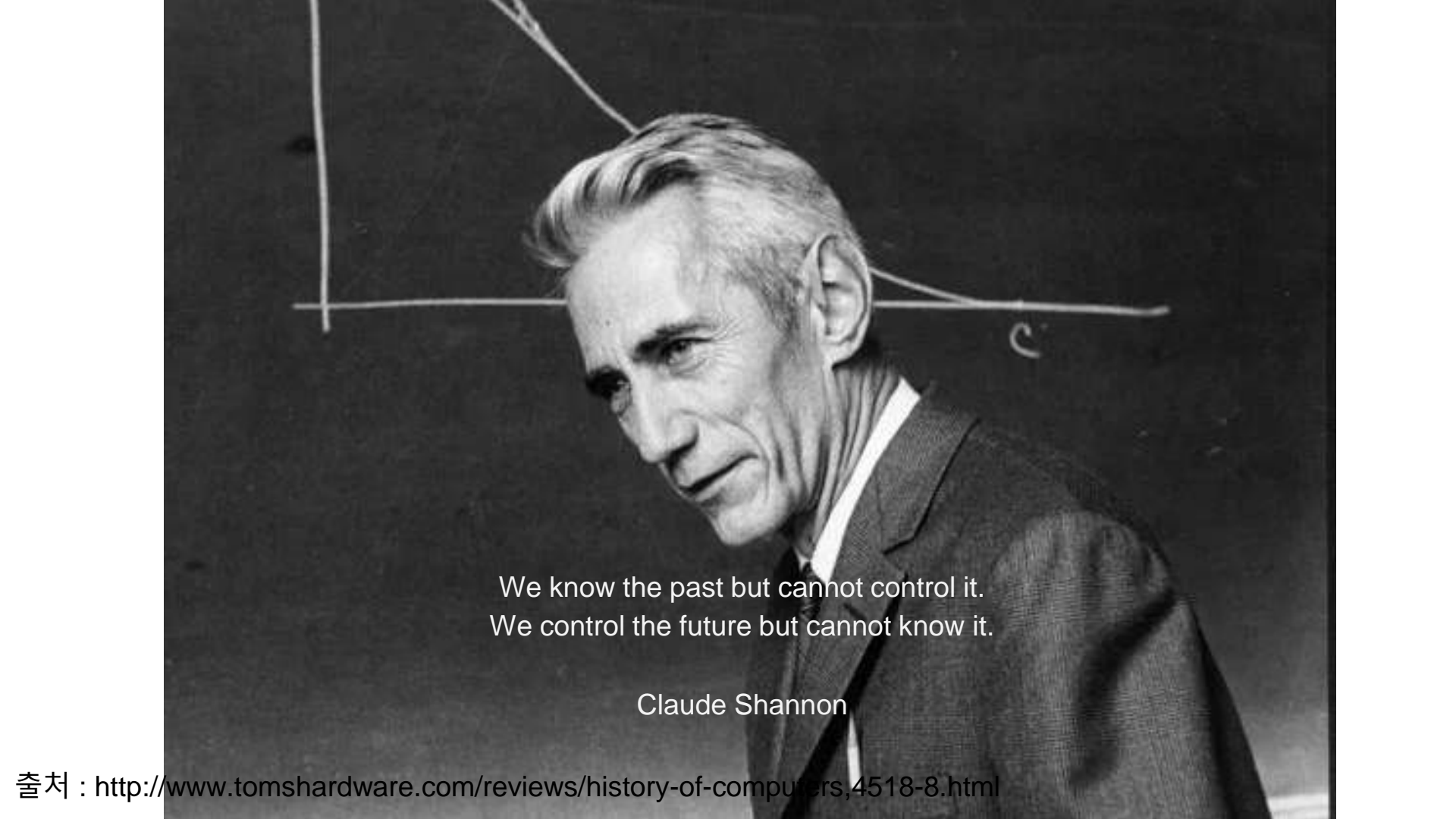






© Stanford News Service

출처 : <http://www.paginaq.it/2014/04/09/dalla-carta-schermo/>

A black and white portrait of Claude Shannon, an older man with light-colored hair, wearing a suit and tie. He is looking slightly to the left. The background is dark with some faint white lines, possibly from a chalkboard.

We know the past but cannot control it.
We control the future but cannot know it.

Claude Shannon

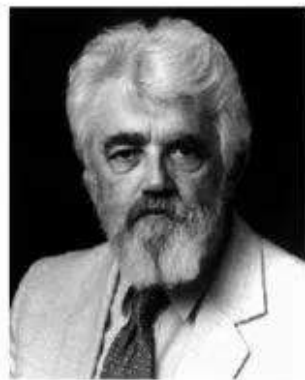


1956 Dartmouth AI Project



Five of the attendees of the 1956 Dartmouth Summer Research Project on AI reunited in 2006: Trenchard More, [John McCarthy](#), [Marvin Minsky](#), [Oliver Selfridge](#), and [Ray Solomonoff](#). Missing were: [Arthur Samuel](#), [Herbert Simon](#), [Allen Newell](#), [Nathaniel Rochester](#) and [Claude Shannon](#).

Dartmouth Conference: The Founding Fathers of AI



John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff

Alan Newell



Herbert Simon



Arthur Samuel



And three others...

Oliver Selfridge
(Pandemonium theory)

Nathaniel Rochester
(IBM, designed 701)

Trenchard More
(Natural Deduction)

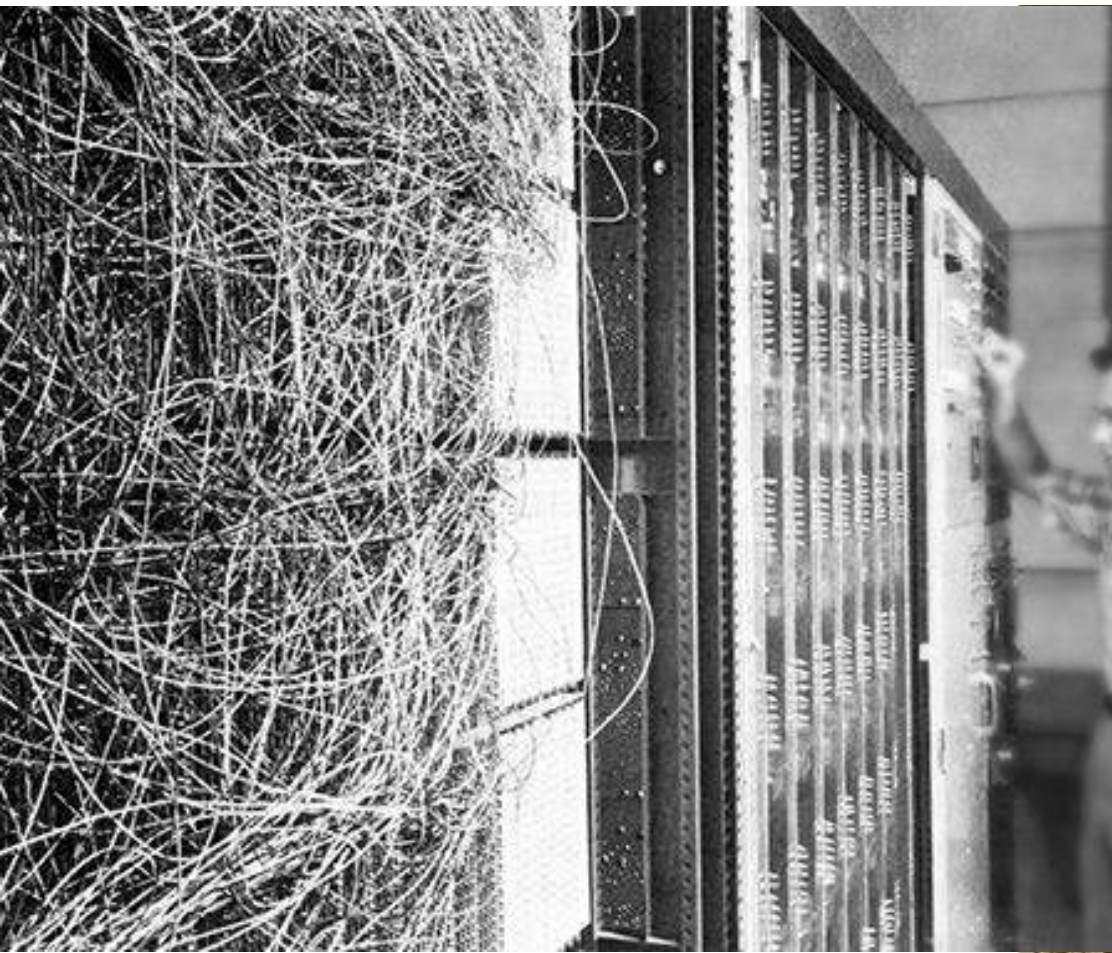
A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories
August 31, 1955

“We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in [Hanover, New Hampshire](#). (1956년 여름 뉴 햄프셔 하노버에 있는 다트머스대에서 두 달 동안 10명의 과학자가 모여 인공지능을 연구할 것을 제안합니다)”



출처 : <http://atreeofstories.blogspot.kr/2016/12/blog-post.html>

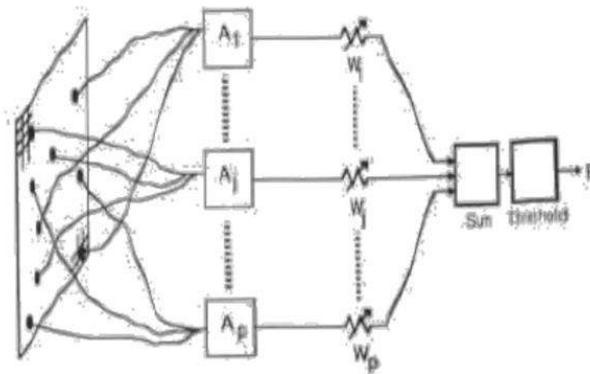


출처 : <https://blogs.umass.edu/comphon/2017/06/15/did-frank-rozenblatt-invent-deep-learning-in-1962/>

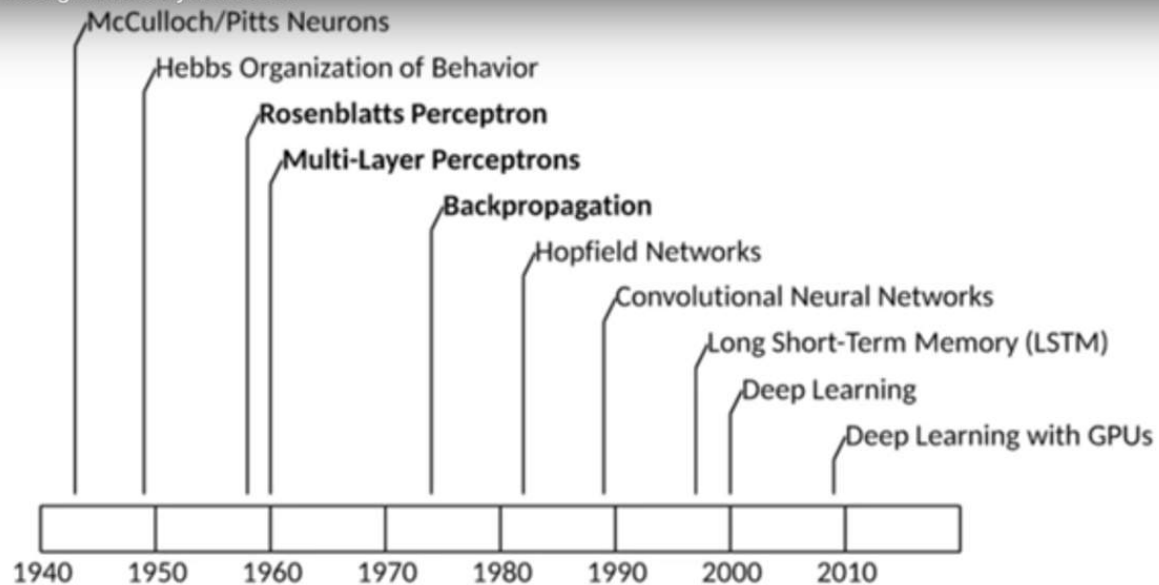
Rosenblatt, with the image sensor of the Mark I Perceptron (Source: Arvin Calspan Advanced Technology Center; Hecht-Nielsen, R. Neurocomputing (Reading, Mass.: Addison-Wesley, 1990).)



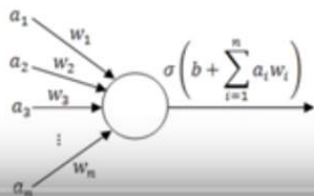
Frank Rosenblatt



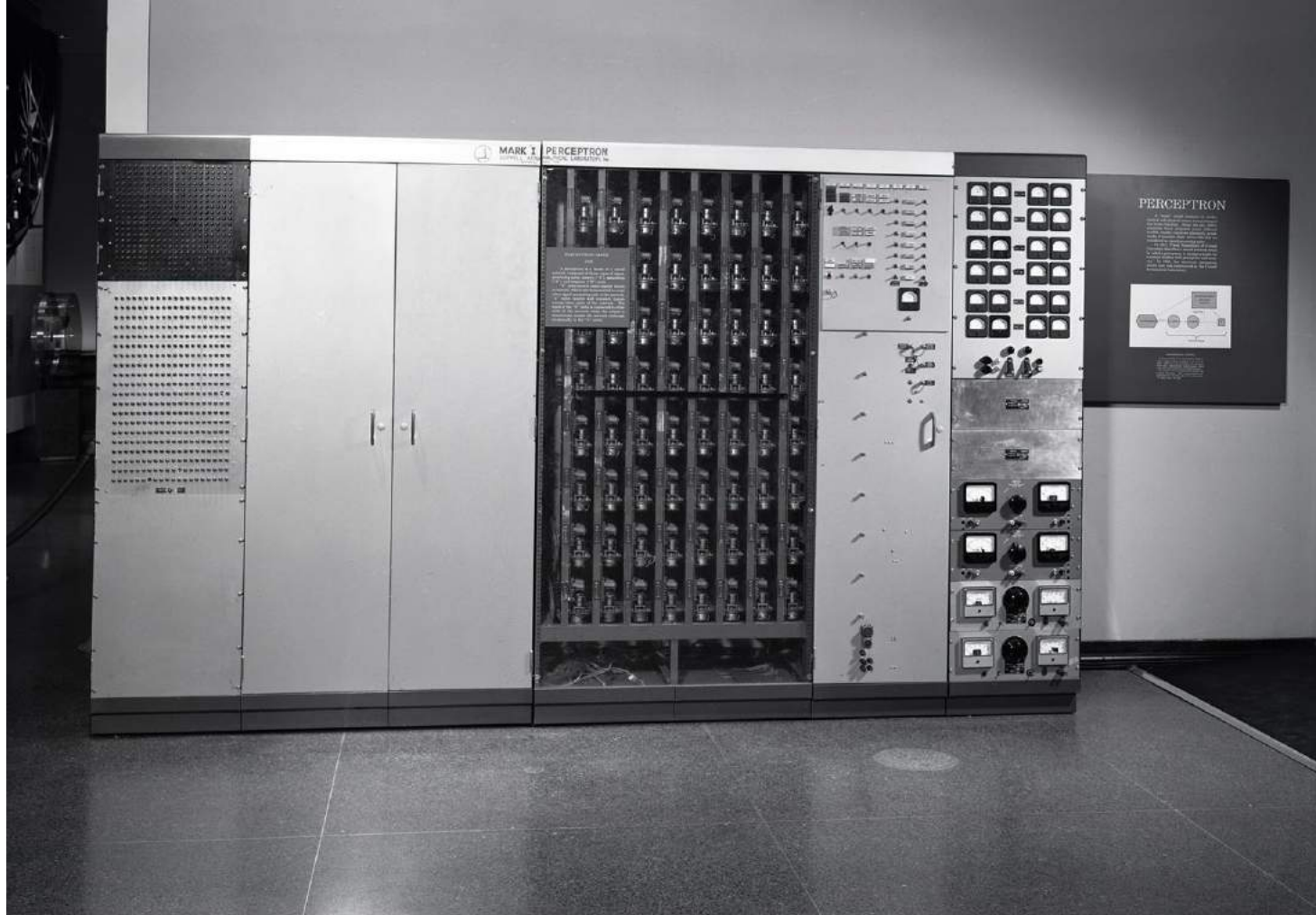
- Input : real-valued
- weights : learning
- Output : 0 or 1



Perceptron



"Mark 1 Perceptron"



출처 : <http://www.cs.cmu.edu/~cga/RobotMuseum/>

- [The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, by Frank Rosenblatt, 1958 \(PDF\)](#)

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN ¹

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

1. How is information about the physical world sensed, or detected, by the biological system?
2. In what form is information stored, or remembered?
3. How does information contained in storage, or in memory, influence recognition and behavior?

and the stored pattern. According to this hypothesis, if one understood the code or "wiring diagram" of the nervous system, one should, in principle, be able to discover exactly what an organism remembers by reconstructing the original sensory patterns from the "memory traces" which they have left, much as we might develop a photographic negative, or translate the pattern of electrical charges in the "memory" of a digital computer. This hypothesis is appealing in its simplicity and ready intelligibility, and a large family of theoretical brain

2563
CATALOGED BY ASTIA
AS AD No. _____

61-3-2
XEROX

PRINCIPLES OF NEURODYNAMICS

ASTIA
FOOTNOTES
MAR 21 1961

PERCEPTRONS AND
THE THEORY OF
BRAIN MECHANISMS

REPORT NO. 1196-G-8
15 MARCH 1961

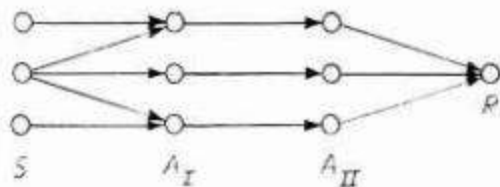


CORNELL AERONAUTICAL LABORATORY, INC.

CORNELL UNIVERSITY, BUFFALO 25, N. Y.

15. MULTI-LAYER PERCEPTRONS WITH FIXED PRETERMINAL NETWORKS

The perceptrons considered in Part II have all consisted of three "layers" of signal generating elements: a sensory layer, a single layer of association units, and a layer of R-units (containing only a single unit in the case of simple perceptrons). A perceptron with additional layers of A-units between S and R-units will be called a multi-layer system. Thus the network diagram:

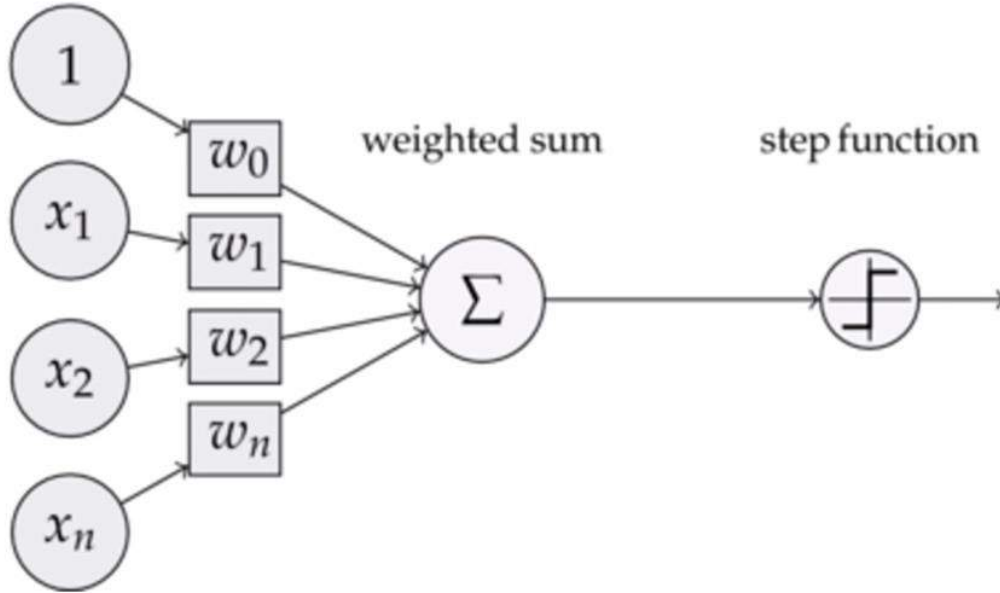


Perceptron Research from the 50's & 60's, clip



https://youtu.be/cNxadbrN_al

inputs weights



Perceptron (Input \rightarrow Output)

- Input values
 $(1, x_1, x_2, \dots, x_n)$
- are multiplied by their respective weights
 $(w_0, w_1, w_2, \dots, w_n)$
- This entire expression is then summed together inside the neuron
 $(w_0 + x_1 w_1 + x_2 w_2 + \dots + x_n w_n)$
- If the sum is larger than a specified threshold, the step function outputs a 1, otherwise it outputs a 0.

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo
of Computer Designed to
Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's \$2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of \$100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human be-

ings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

1958 New York Times...

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

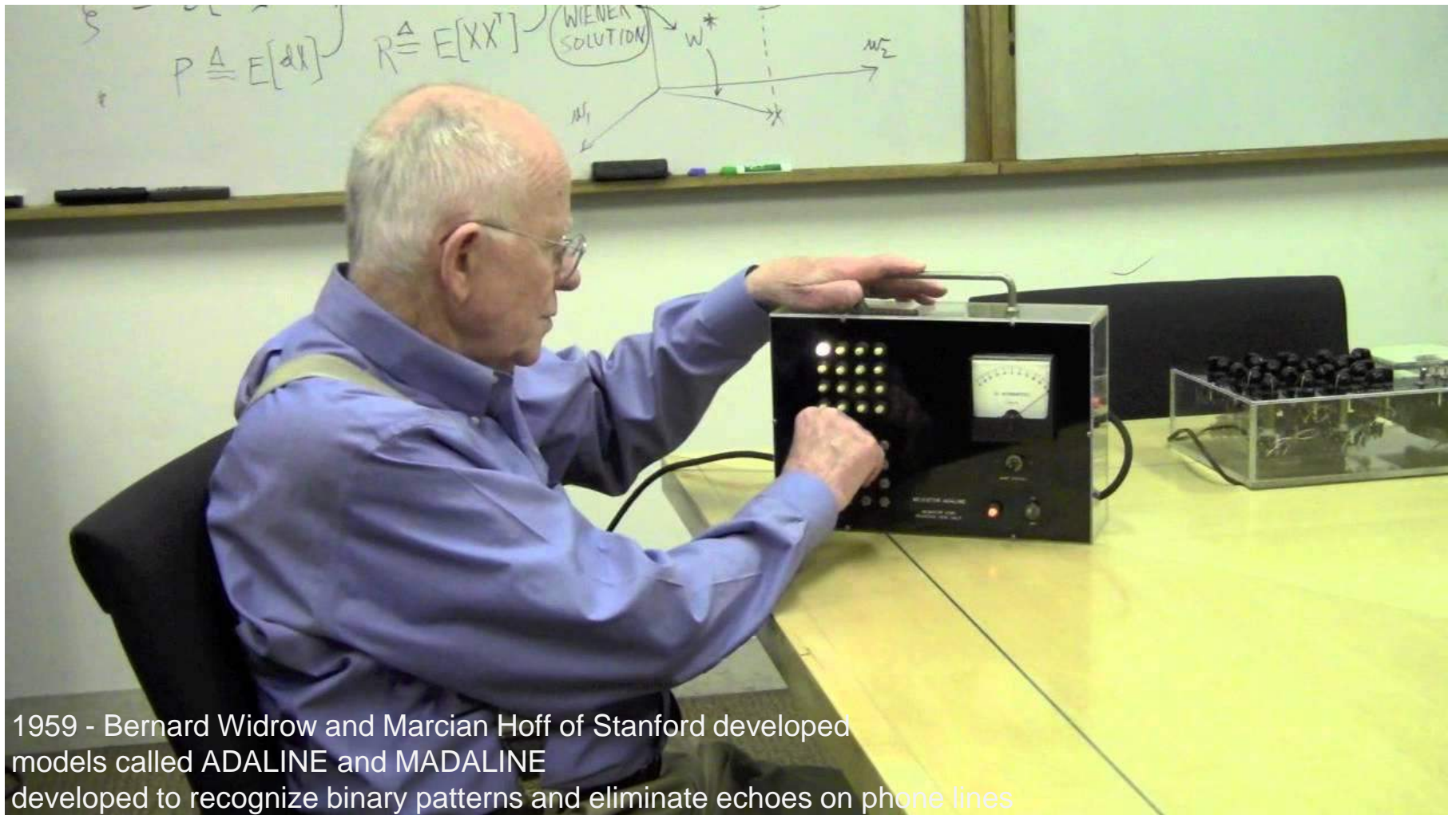
In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

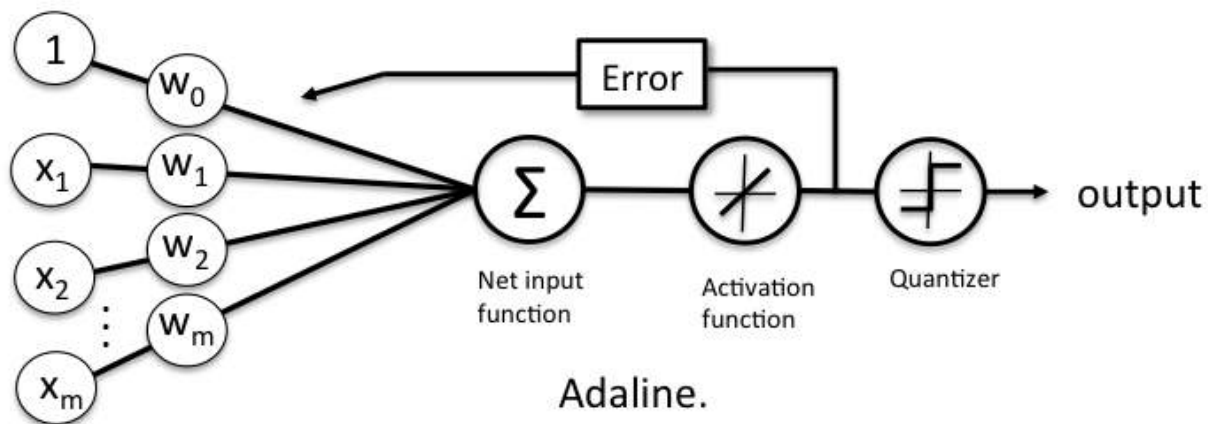
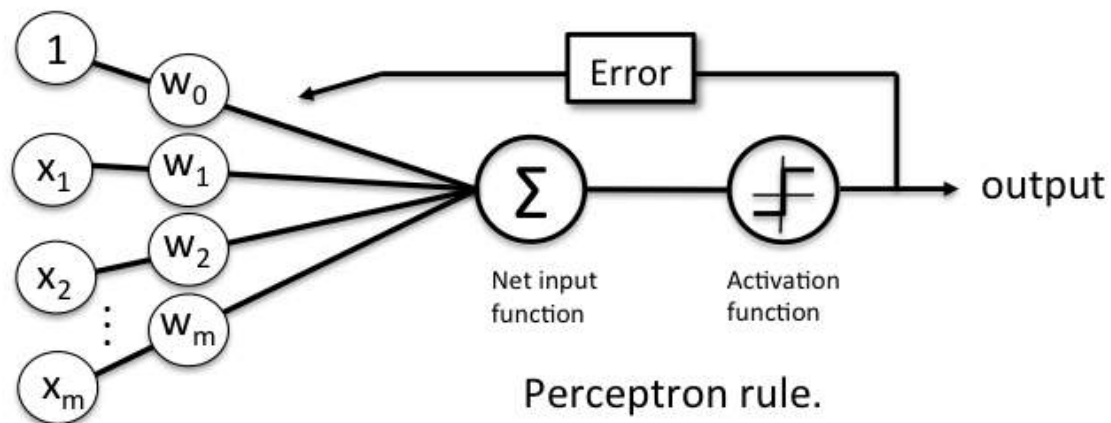
The Navy last week demonstrated the embryo of an electronic computer named the Perceptron which, when completed in about a year, is expected to be **the first non-living mechanism** able to "perceive, recognize and identify its surroundings without human training or control."

- NewYork Times 1958년



1959 - Bernard Widrow and Marcian Hoff of Stanford developed models called ADALINE and MADALINE developed to recognize binary patterns and eliminate echoes on phone lines

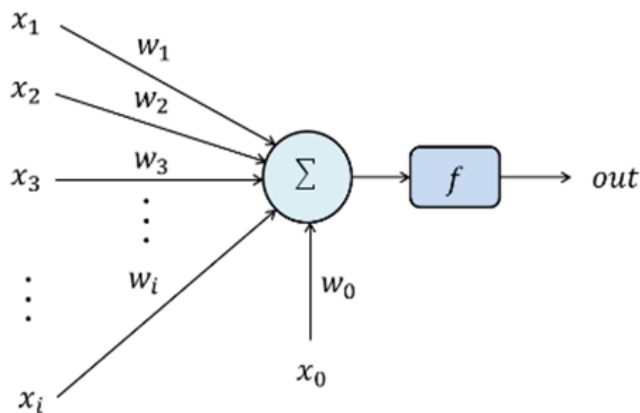






델타규칙이란?

- 단층신경망을 학습시키는 방법
- **Gradient Descent** 원리를 이용해 단층신경망을 통해 출력된 output과 실제 output의 평균오차제곱합인 **손실함수 MSE(Mean Square Error)**를 최소화하는 (w_1, w_2, θ)를 찾기 위한 학습



m개의 observations

- Input 변수: x_1, x_2, \dots, x_n
- 가중치 합: h_1, h_2, \dots, h_m
- Output 값: y_1, y_2, \dots, y_m

Output 출력 과정

1. (x_1, x_2, \dots, x_n) n개의 변수에 대한 m개의 observation의 가중치합 (h_1, h_2, \dots, h_m) 생성
2. 가중치 합을 활성화함수에 넣어 (y_1, y_2, \dots, y_m) m개의 output 생성



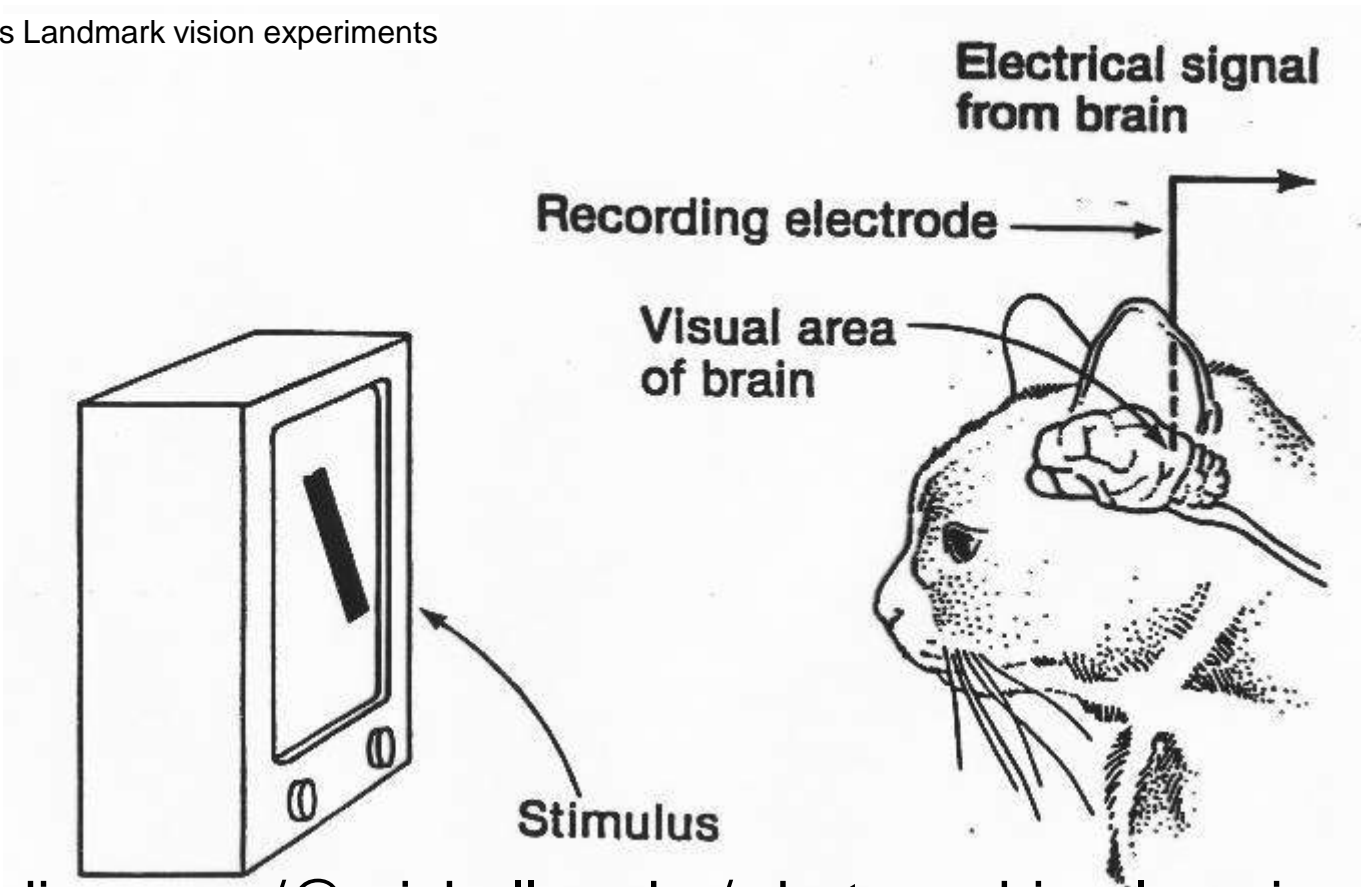


출처 : <https://twitter.com/i/web/status/955160651105619968>



출처 : <https://youtu.be/IOHayh06LJ4>

Hubel and Wiesel's Landmark vision experiments

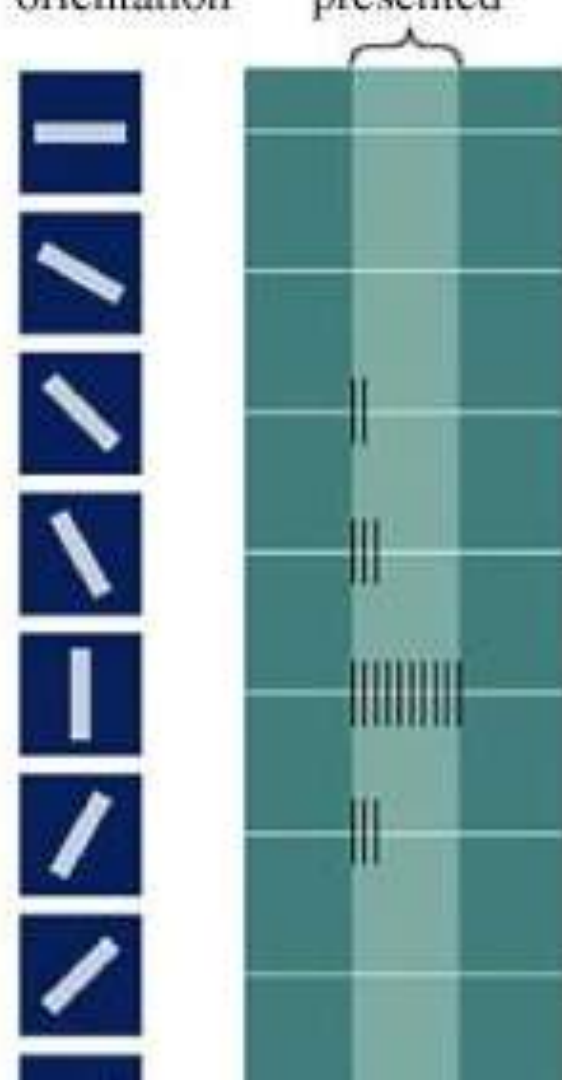
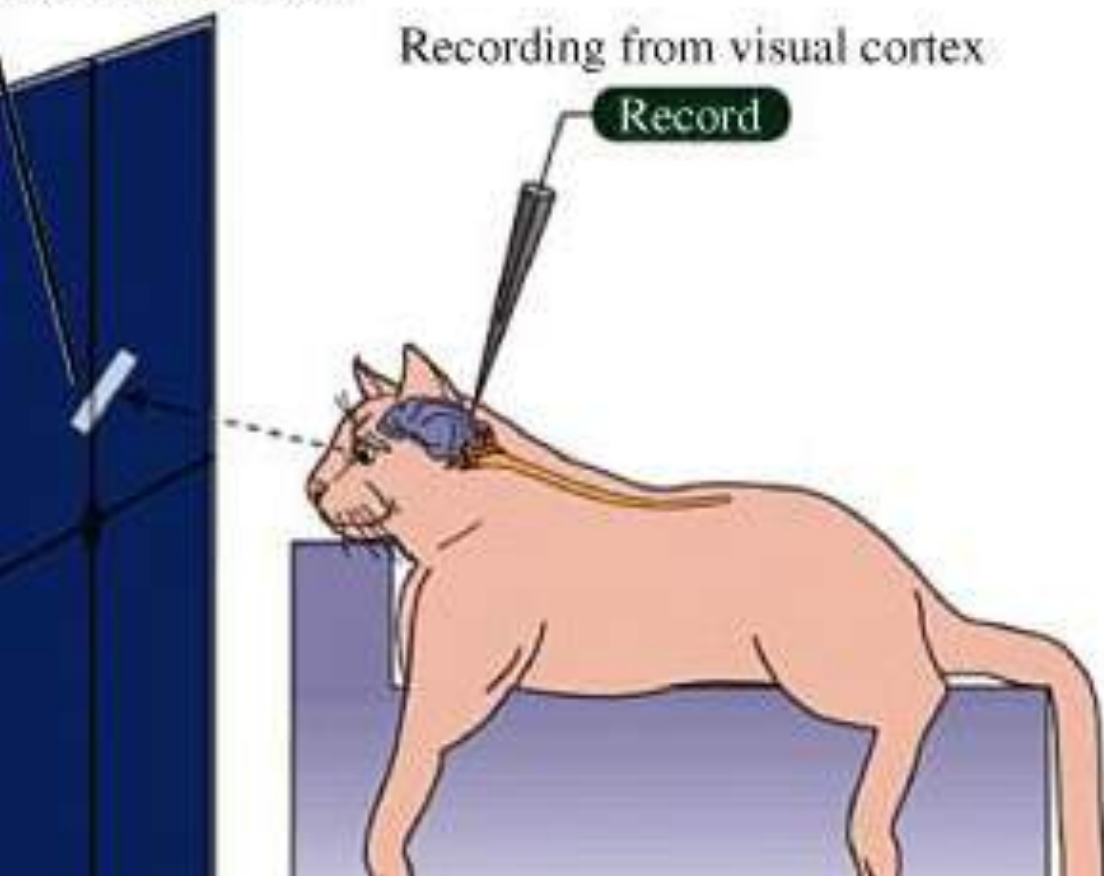


<https://medium.com/@michellewshu/what-machine-learning-borrowed-from-vision-science-469a20779b7a>

Light bar stimulus
projected on screen

Recording from visual cortex

Record

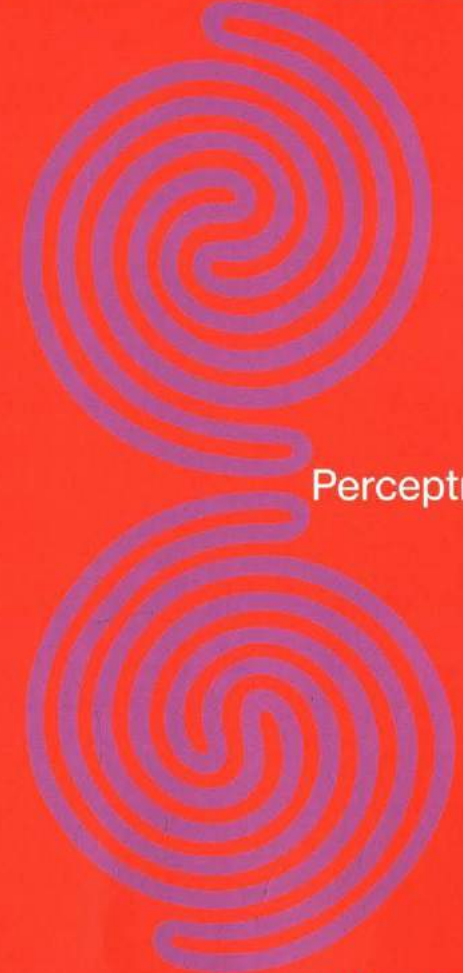




As a result, in 1969, **Minsky** co-authored with [Seymour Papert](#), *Perceptrons: An Introduction to Computational Geometry*. In this work they attacked the limitations of the perceptron. They showed that the perceptron could only solve linearly separable functions. Of particular interest was the fact that the perceptron still could not solve the XOR and NXOR functions. Likewise, Minsky and Papert stated that the style of research being done on the perceptron was doomed to failure because of these limitations. This was, of course, Minsky's equally ill-timed remark. As a result, very little research was done in the area until about the 1980's.

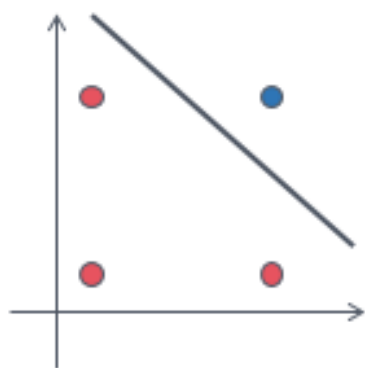
Reissue of the 1988 Expanded Edition with a new foreword by Léon Bottou

Marvin L. Minsky and Seymour A. Papert



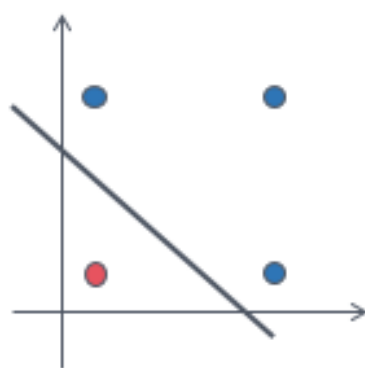
Perceptrons

An Introduction to Computational Geometry



AND

AND		
Input_A	Input_B	Output
0	0	0
0	1	0
1	0	0
1	1	1



OR

OR		
Input_A	Input_B	Output
0	0	0
0	1	1
1	0	1
1	1	1


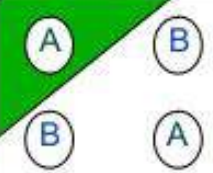


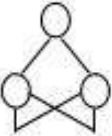
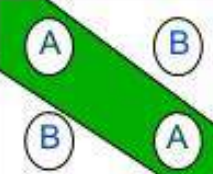
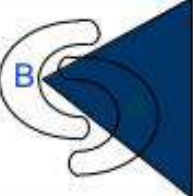

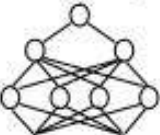
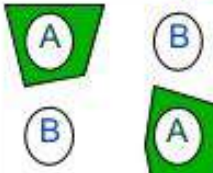




XOR

XOR		
Input_A	Input_B	Output
0	0	0
0	1	1
1	0	1
1	1	0

Different Non-Linearly Separable Problems

<http://www.zsolutions.com/light.htm>

Structure	Types of Decision Regions	Exclusive-OR Problem	Classes with Meshed regions	Most General Region Shapes
Single-Layer 	Half Plane Bounded By Hyperplane			
Two-Layer 	Convex Open Or Closed Regions			
Three-Layer 	Arbitrary (Complexity Limited by No. of Nodes)			

신은 가끔 같은 장소, 같은 시간에 천재 라이벌들을 보내 그들의 대결을 지켜보는 고약한 취미가 있는가합니다. 제갈량과 주유, 그랜트와 리 그리고 메리와 롬멜 등 천재들의 손에 땀을 쥐게 하는 대결은 인류 역사에 드물지 않습니다. 지금 소개하는 두 명의 천재의 대결을 신은 어떻게 보았을까요.



<민스키와 로젠블랫>

마빈 민스키와 프랭크 로젠블랫은 각각 1927년, 1928년 1년의 시간차를 두고 뉴욕에서 태어나 그곳의 명문인 브롱스 과학 고등학교(The Bronx School of Science) 동문으로 인연을 맺게 됩니다. 고등학교 졸업이후 민스키는 하버드와 프린스턴을 거치면서 수학을 공부하게 되며, 로젠블랫은 하버드에서 심리학을 공부하게 됩니다. 각각의 방향은 다르지만 결국 인공지능 정복을 위해 각자의 길을 걷기 시작한 셈이죠.

두각을 먼저 드러낸건 1살 나이가 많은 민스키였습니다. 그는 1956년 다트머스 회의에서 최초로 인공지능(Artificial Intelligence)이라는 용어를 사용하였으며, 존 매카시와 나다니엘 로체스터, 클로드 새너 등과 함께 약 한 달 동안 브레인 스토밍을 통해 그 개념을 확립하였습니다. 민스키가 주장하는 것은 간단히 설명하자면, 인간의 지식을 기호화 하고 그 기호간의 관계를 일일이 입력하여 학습을 시키면 컴퓨터는 인간과 비슷한 입력을 얻었을 때 할 수 있는 비슷한게 낼 것이라는 가정에서 출발한 것입니다.이를 기호주의라고 합니다. 이 기호주의를 계승한 가장 전형적인 AI 기법이 바로 전문가 시스템입니다.

예를 하나 들어보겠습니다. MYCIN이라는 의료처방 시스템이 그 중 가장 성공적인 Case 인데 의사들이 자신이 가지고 있는 의학 지식을 컴퓨터에 시키는 것입니다. 어떤 증상이 나타날 경우에는 어떤 소견을 내야 한다는 정보를 넣어 컴퓨터도 의사만큼 처방을 내릴 수 있게 되는 것이지요. 이런 기호주의는 인간의 모든 지식을 다 때려 넣으면 언젠간 기계도 사람처럼 생각할 수 있지 않겠느냐란 인식에서 출발했습니다.

기호주의로 독주를 하고 있던 민스키에게 새로운 도전자가 나타납니다. 바로 고등학교 동문이었던 로젠블랫입니다. 그는 당시 주류였던 기호주의 대신 연결주의를 선택합니다. 인간의 뇌는 수많은 신경망으로 이루어져있으며, 이것을 모티브로 삼아 컴퓨터도 신경망으로 학습을 시키고 추론을 하게 해야 한다는 주장이지요. 이른바 퍼셉트론입니다.

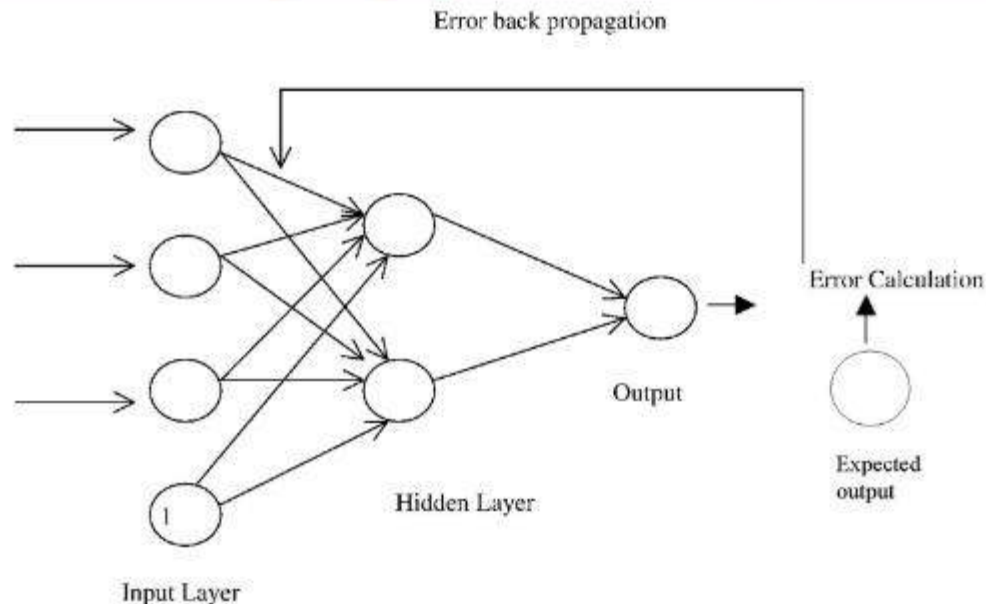


BRACE YOURSELF

**AI WINTER IS
COMING**



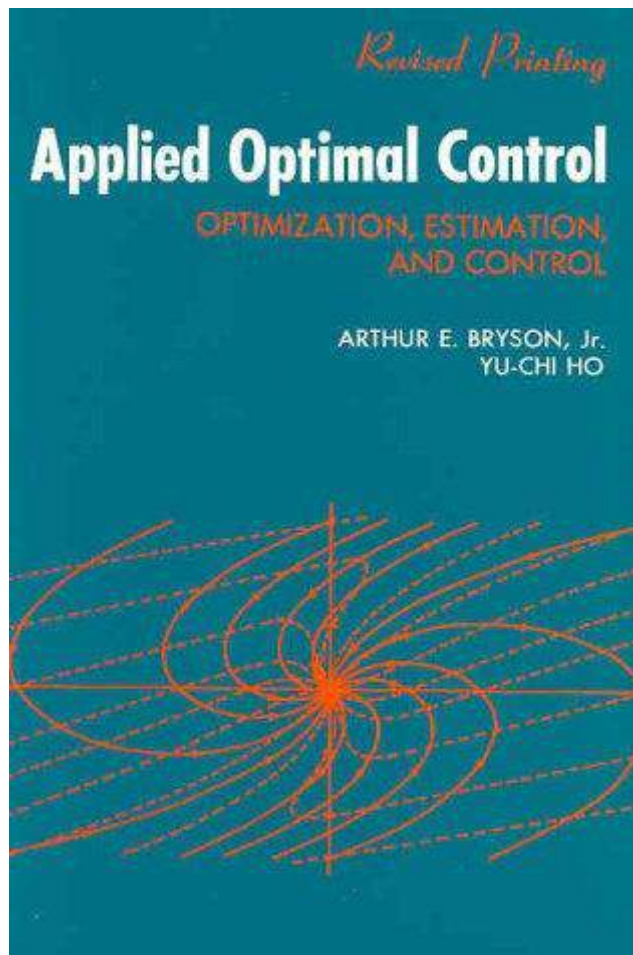
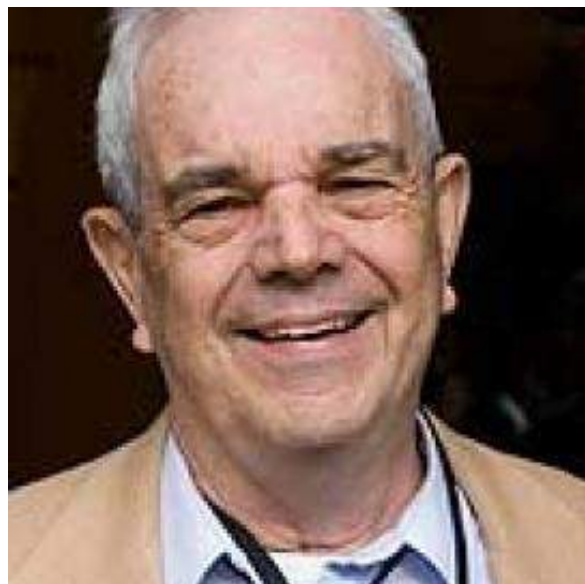
Backpropagation (1974/1986)



1974 Paul Werbos' invents Backpropagation algorithm for NN

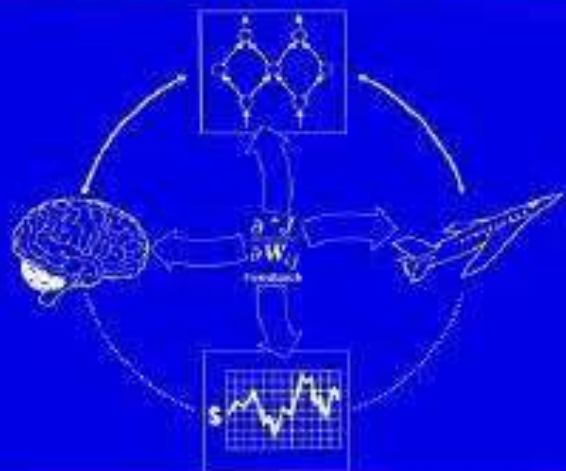
1986 Backdrop popularized by Rumelhart, Hinton, Williams

1990: Renewed Interest in NN's



THE ROOTS OF BACKPROPAGATION

From Ordered Derivatives
to Neural Networks
and Political Forecasting



PAUL JOHN WERBOS

A VISION OF THE FUTURE OF ARTIFICIAL INTELLIGENCE WITH NEW METHODS
FOR DESIGNING ROBUST, ADAPTIVE, AND CONTINUOUSLY LEARNING
SYSTEMS. HARVARD UNIVERSITY

- Back-propagation is a learning algorithm for multi-layer neural networks
- It was invented independently several times
 - Bryson and Ho [1969]
 - Werbos [1974]
 - Parker [1985]
 - **Rumelhart et al. [1986]**

Parallel Distributed Processing - Vol. 1

Foundations

David E. Rumelhart, James L. McClelland and the PDP Research Group

What makes people smarter than computers? These volumes by a pioneering neurocomputing.....

Zur Anzeige wird der QuickTime™
Dekompressor „TIFF (Unkomprimiert)“
benötigt.

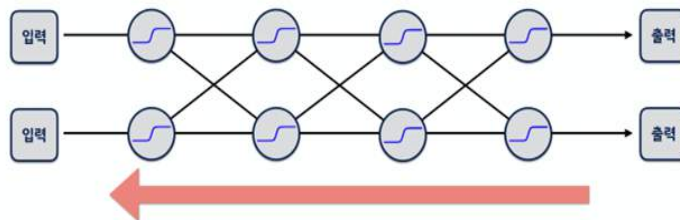
Zur Anzeige wird der
Dekompressor „TIFF (U
benötigt

Back propagation

“뒤로 전달”이라는 의미
사실상 미적분의 Chain Rule

뭐를 전달하는가?

현재 내가 틀린정도를 ‘미분(기울기)’ 한 거

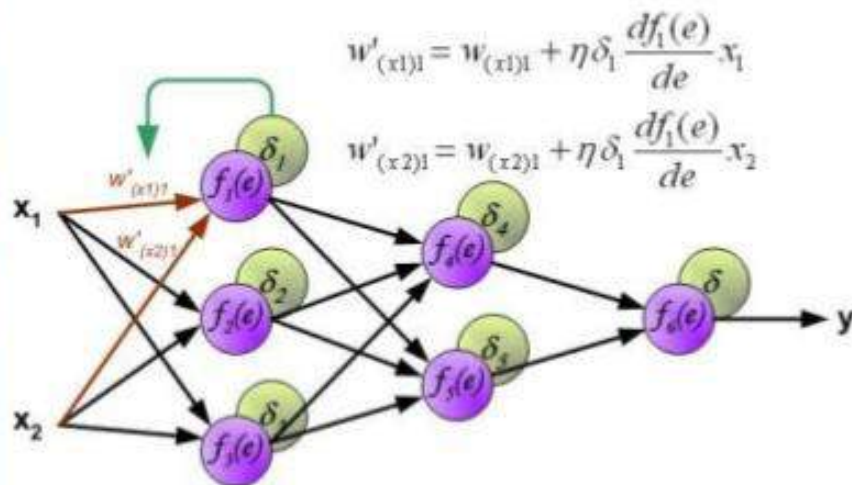


미분하고, 곱하고, 더하고를 역방향으로 반복하며 업데이트한다.

출처: 하용호님 발표

1974

- Paul Werbos's Ph.D. thesis describes the process of training neural nets through back-propagation.



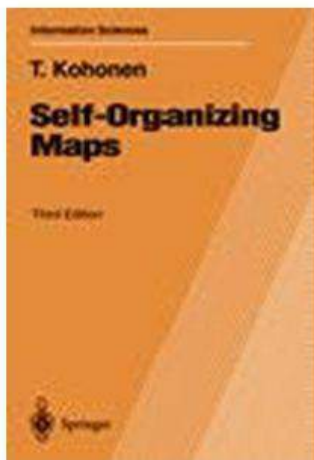
Self Organizing Maps



T. Kohonen

Dr. Eng., Emeritus Professor of the Academy of Finland

His research areas are the theory of self-organization, associative memories, neural networks, and pattern recognition, in which he has published over 300 research papers and four monography books.

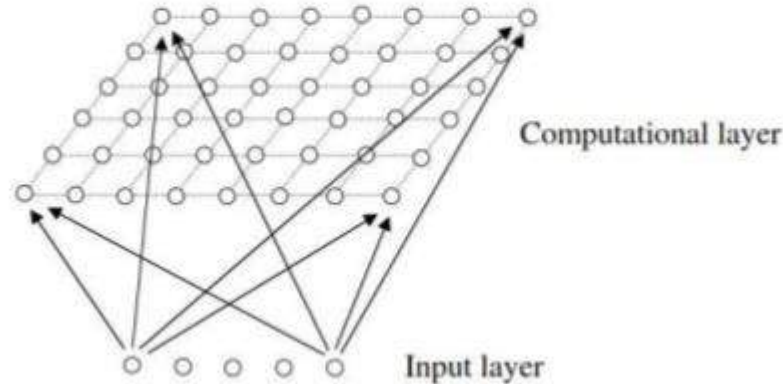


T. Kohonen (1995), Self-Organizing Maps.

Kohonen Networks



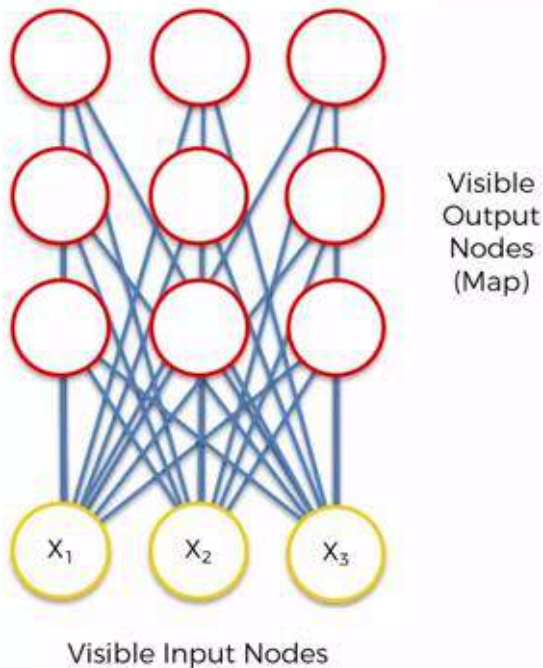
We shall concentrate on the particular kind of SOM known as a **Kohonen Network**. This SOM has a feed-forward structure with a single computational layer arranged in rows and columns. Each neuron is fully connected to all the source nodes in the input layer:



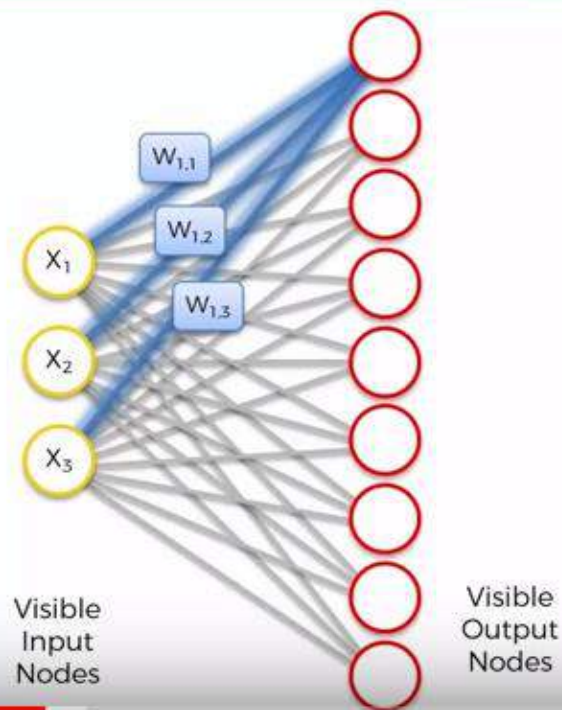
Clearly, a one dimensional map will just have a single row (or a single column) in the computational layer.



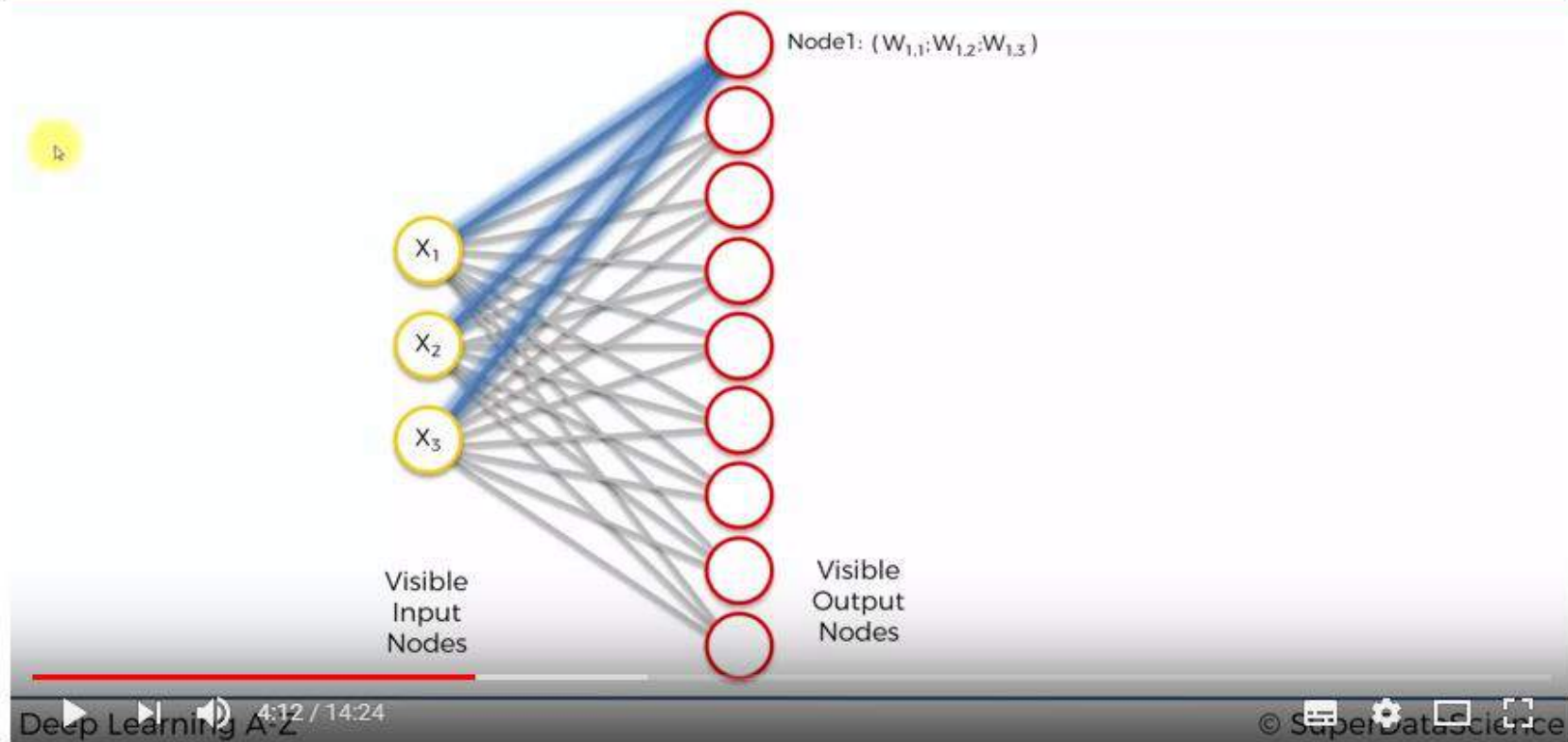
How Do SOMs Learn?



How Do SOMs Learn?

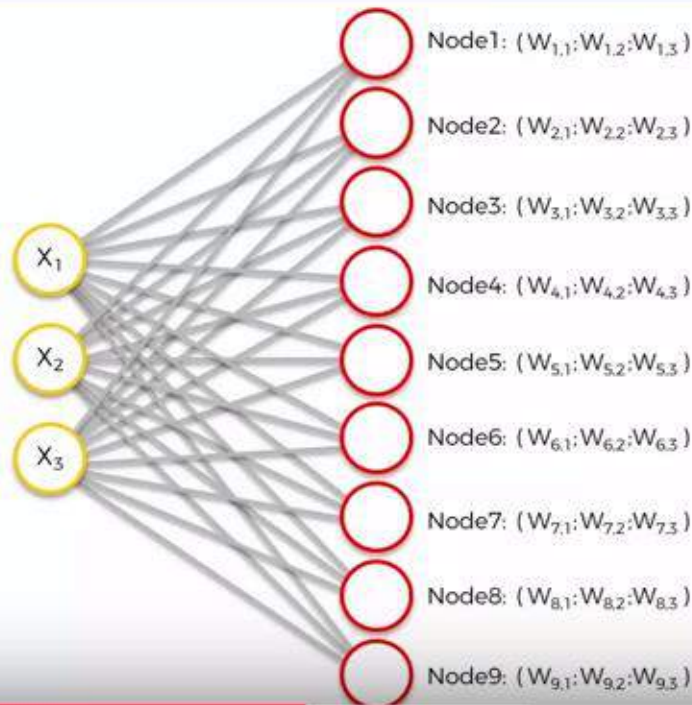


How Do SOMs Learn?



How do Self-Organizing Maps Learn? (Part 1)

How Do SOMs Learn?

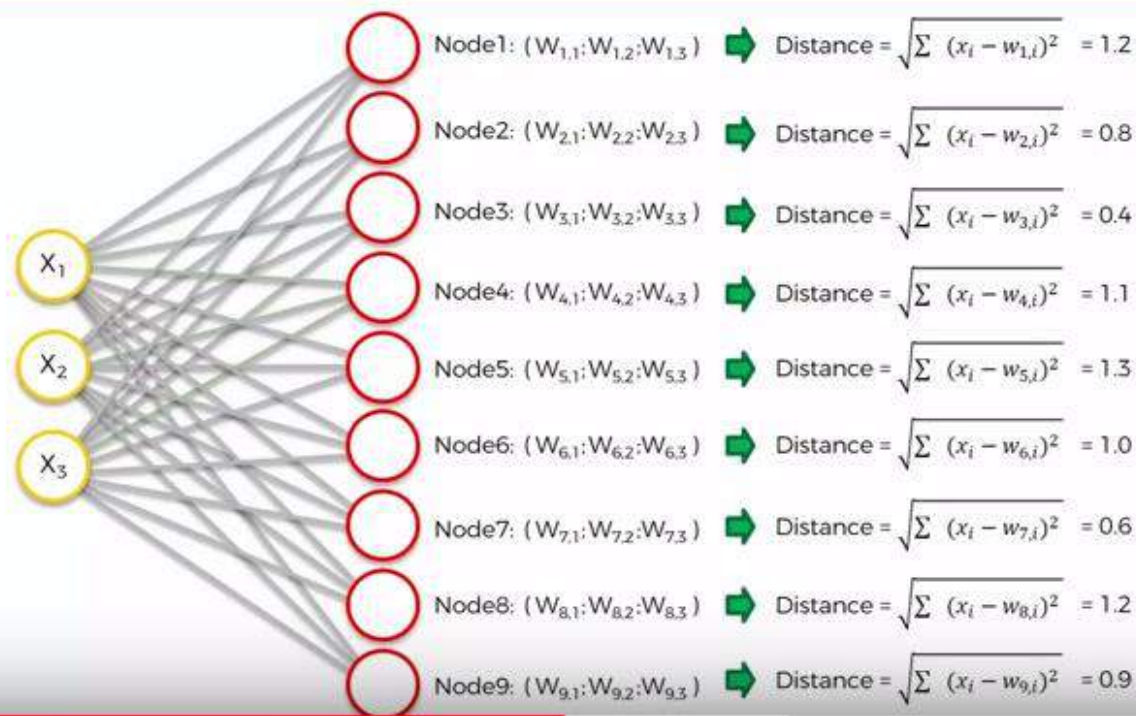


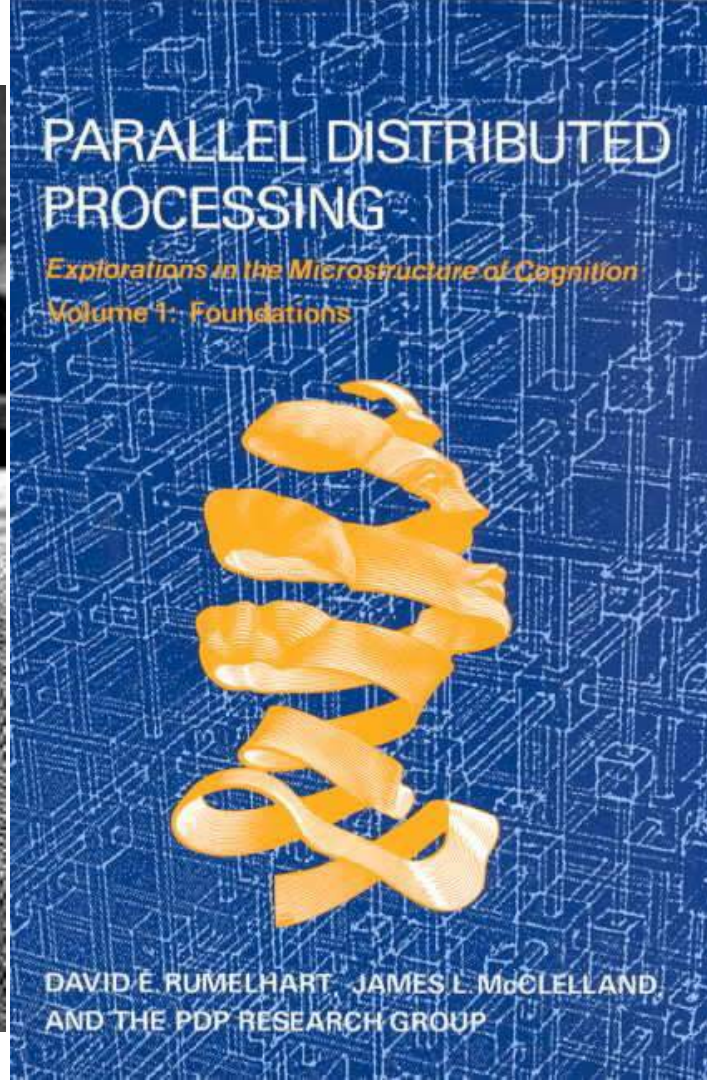
Deep Learning A-Z 6:03 / 14:24

© SuperDataScience

How do Self-Organizing Maps Learn? (Part 1)

How Do SOMs Learn?





james mcclelland



All the knowledge is in the
connections

— *David Rumelhart* —

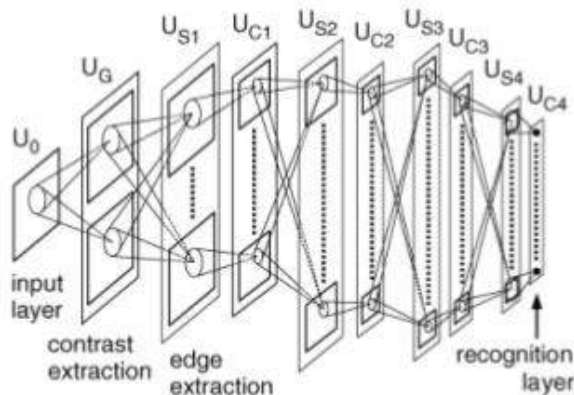
AZ QUOTES

Parallel distributed processing (PDP)

- Rumelhart & McClelland (1986) made a very strong case for the use of connectionist models, highlighting the qualities of **parallel** and *distributed* processing
- Parallel Processing
 - **Parallel** processing can be contrasted with **serial** processing
 - Processing several different pieces of information at the same time, rather than one after the other:
 - Example: face processing
- Distributed Processing
 - **Distributed** processing can be contrasted with **localist** processing
 - Representations of information are distributed across the whole neural network, rather than occupying specific locations
 - Example: Karl Lashley and the 'Engram' (location of memory)

Neocognitron

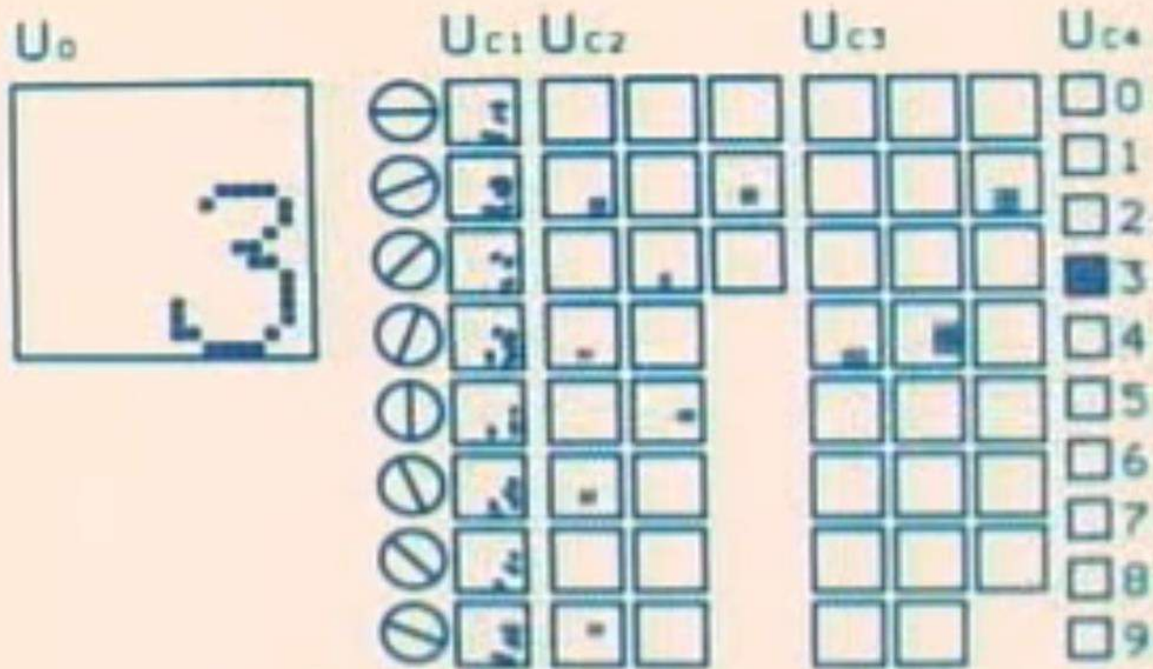
Fukushima (1980). Hierarchical multilayered neural network

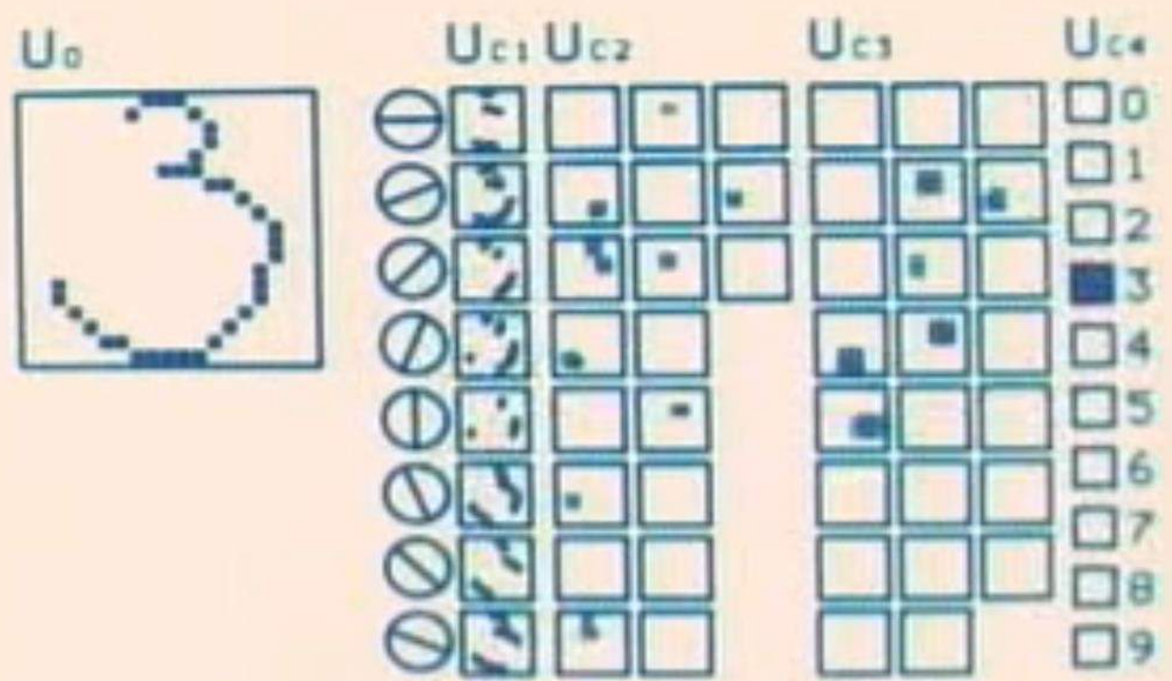


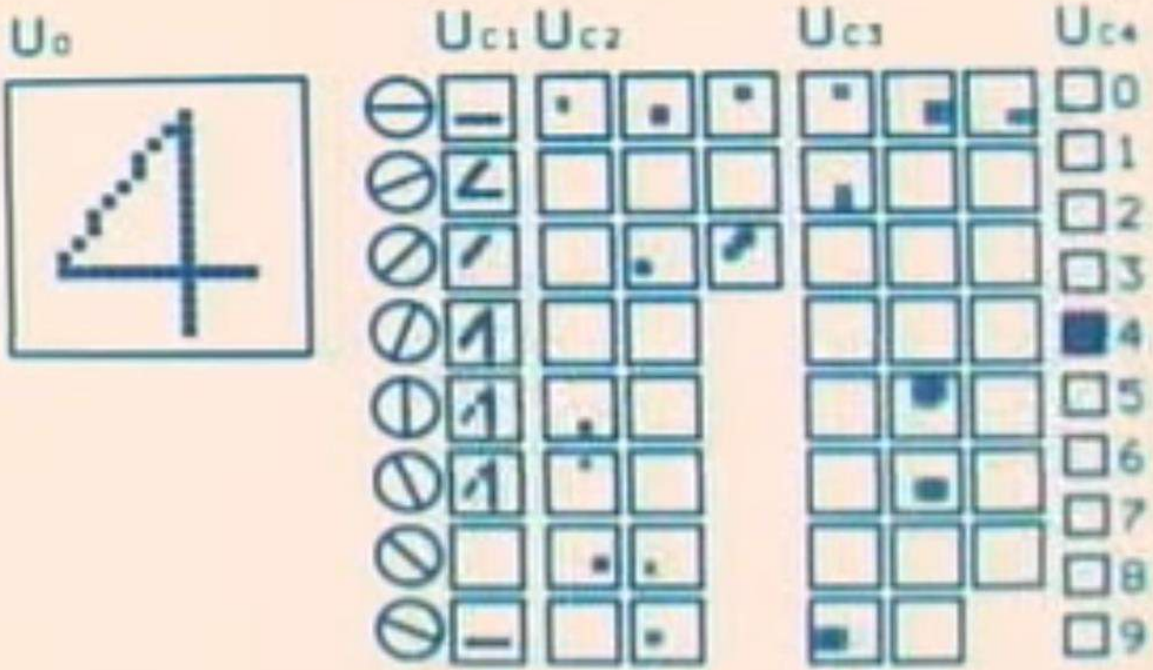
S-cells work as feature-extracting cells. They resemble simple cells of the primary visual cortex in their response.

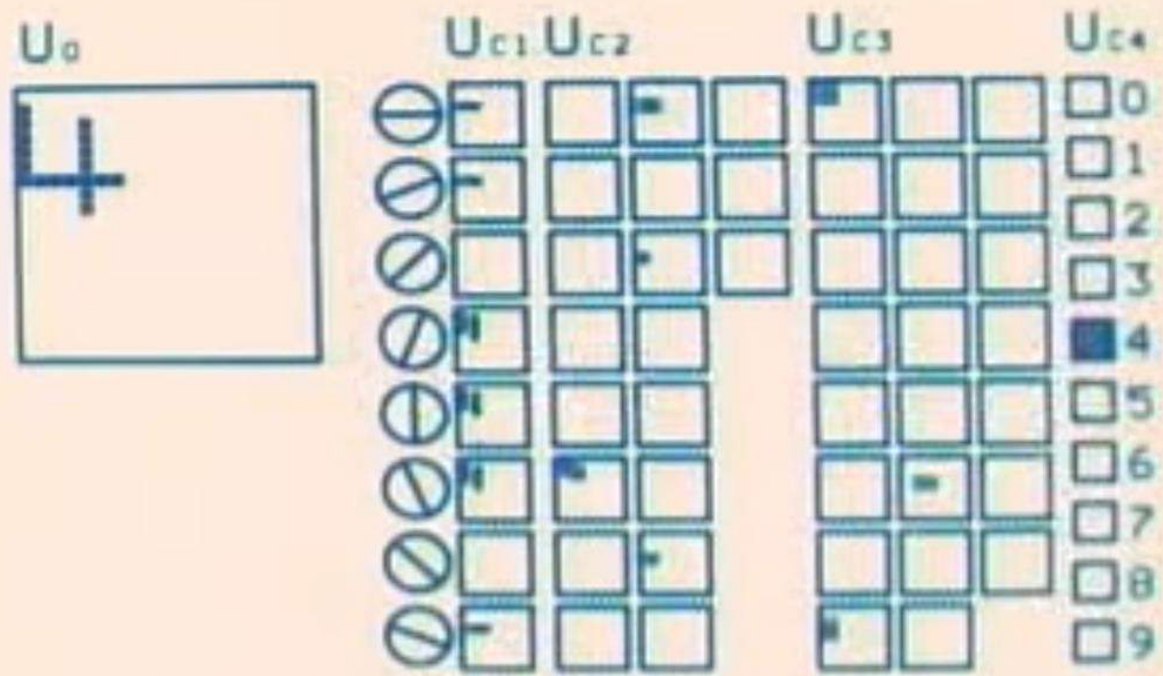
C-cells, which resembles complex cells in the visual cortex, are inserted in the network to allow for positional errors in the features of the stimulus. The input connections of C-cells, which come from S-cells of the preceding layer, are fixed and invariable. Each C-cell receives excitatory input connections from a group of S-cells that extract the same feature, but from slightly different positions. The C-cell responds if at least one of these S-cells yield an output.







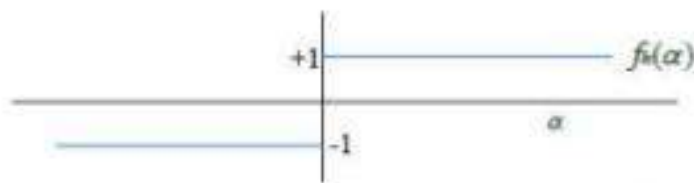




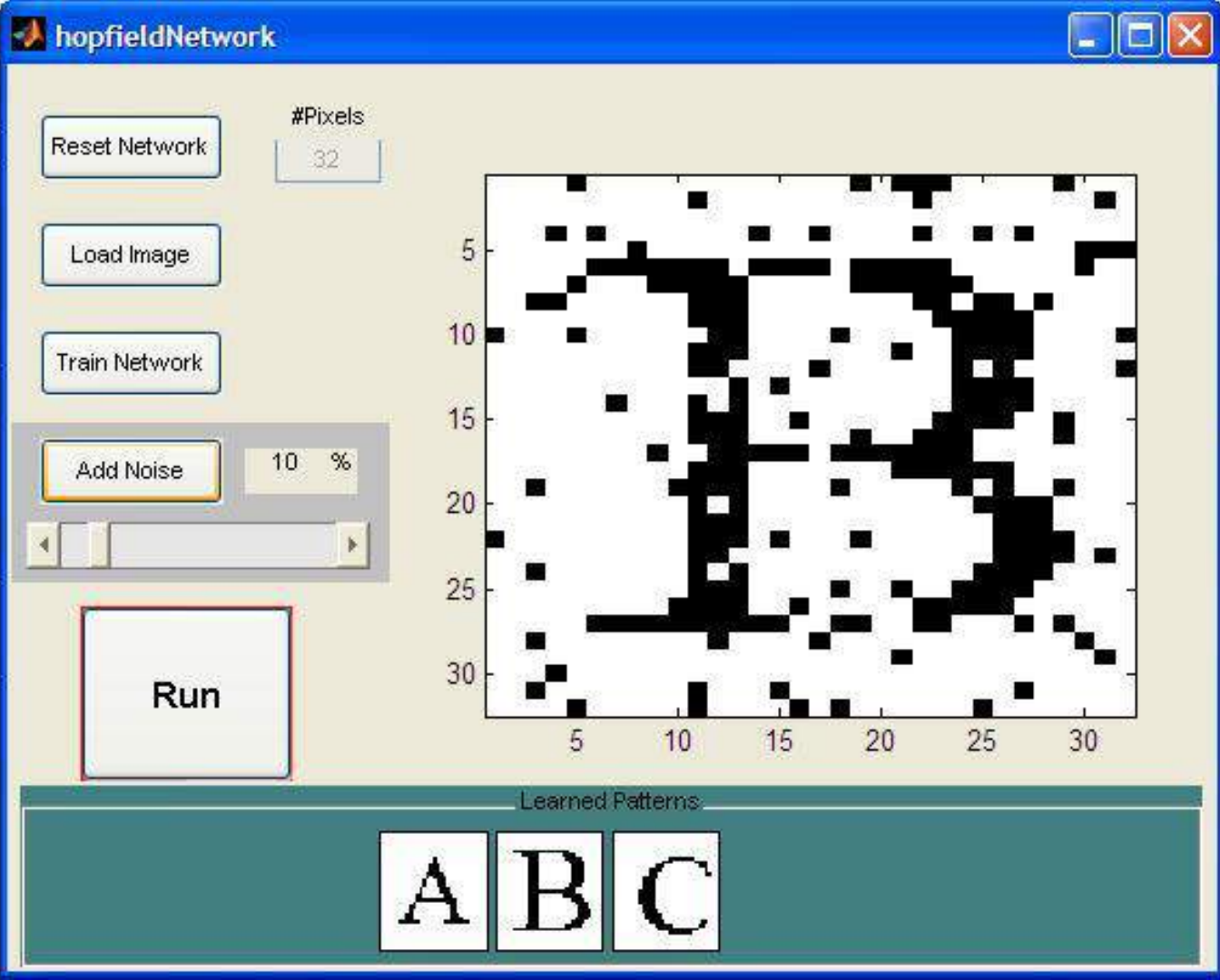


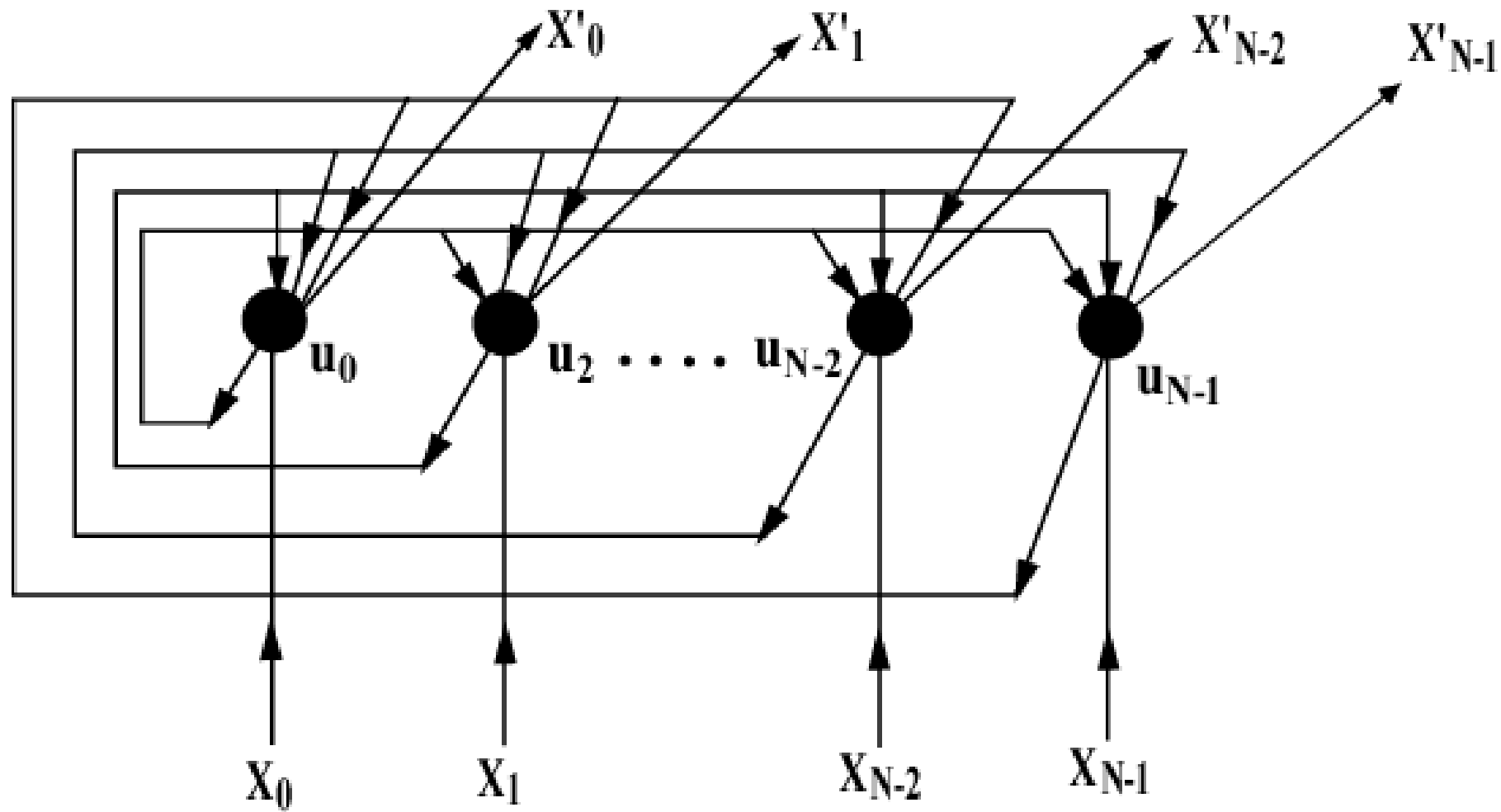
홉필드 네트워크 (HOPFIELD NETWORK)

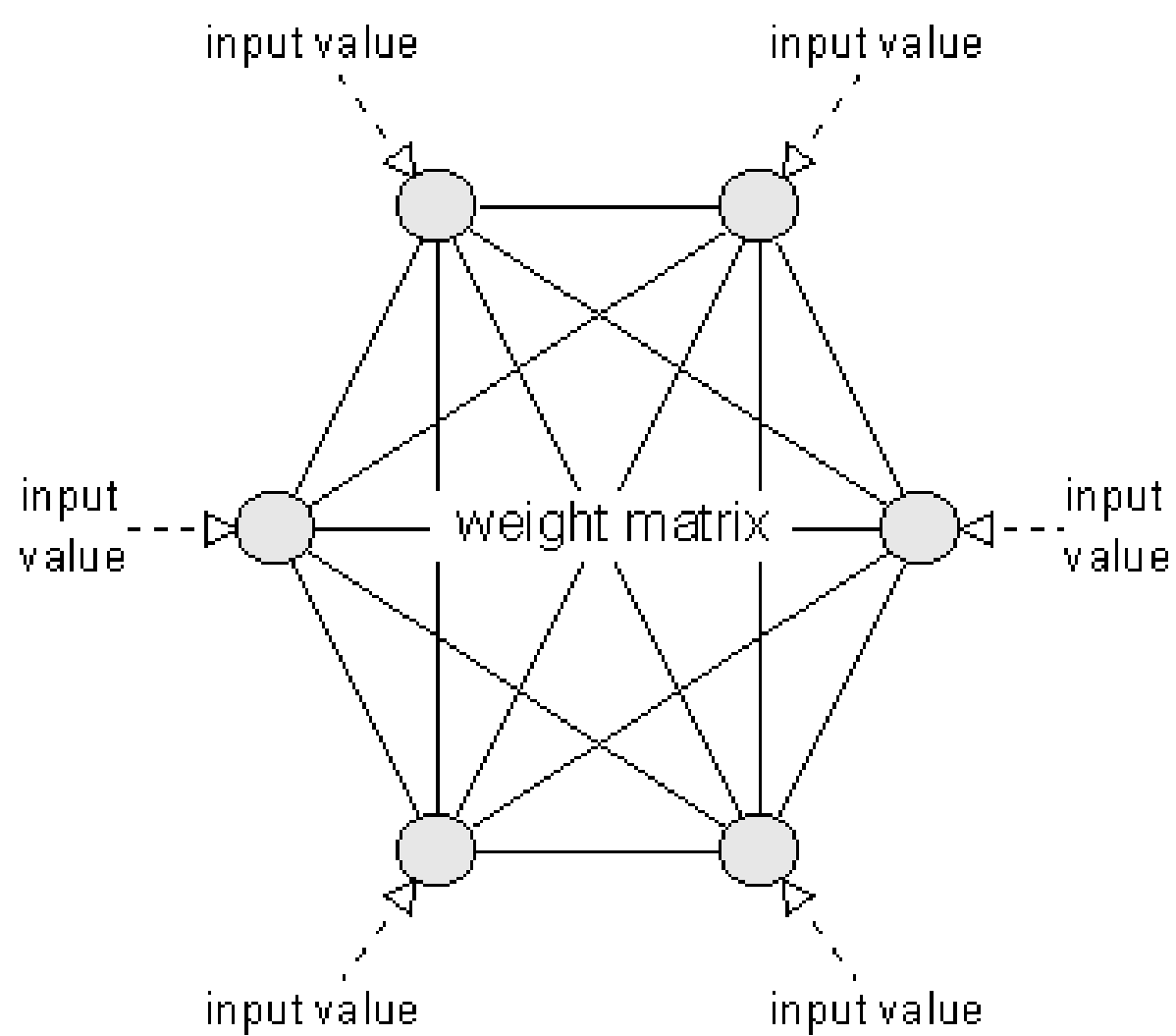
- 물리학적 스핀 모델로부터 Hopfield network 를 착안하였으며 에너지 개념을 신경망에 처음으로 도입
- 안정상태 도달을 위한 제약
 - 노드간 연결 가중치가 대칭적, ($w_{ij} = w_{ji}$)
 - 완전히 비동기적으로 (asynchronously) 동작할 때만 안정된 상태에 도달
- 입력값이 임계값을 넘으면 흥분하고 그 이하이면 억제됨:



$$E(t) = -\frac{1}{2} \sum_{i \neq j} w_{ij} o_i o_j(t) - \sum_i \theta o_i(t) (\theta: \text{임계값})$$











$$S = k \cdot \log W$$



LUDWIG
BOLTZMANN

1844-1906

HENRIETTE
BOLTZMANN
FEB. EDLE VON ARGENTIER
1854-1938

DR. PHIL. PAULA
BOLTZMANN

GEB. CHLARI

1891-1977

ARTHUR

BOLTZMANN

DIPL. ING. DR. PHIL. HOFER

1881-1952

LUDWIG

BOLTZMANN

1923-1943

SEINER MÄNNLICHEN NACHKOMME
GEFALLEN BEI STOLLEN

볼츠만머신 VS 홉필드 네트워크

○ 공통점

- 각 프로세싱 유닛은 +1, -1의 값만 가지며 대칭적 연결
- 각 유닛은 임의로 선택되며 한 순간에 하나의 유닛만 변화.
self-feedback이 없다.

● 차이점

- Stochastic neuron(볼츠만) vs Deterministic model(홉필드)
- unsupervised learning(볼츠만) vs Supervised learning of a probabilistic form(홉필드)
- 에너지가 증가하는 상태의 전이에 대해서도 허용하는 동작규칙 적용(볼츠만)하나 홉필드는 에너지 감소방향으로만 변화

1986

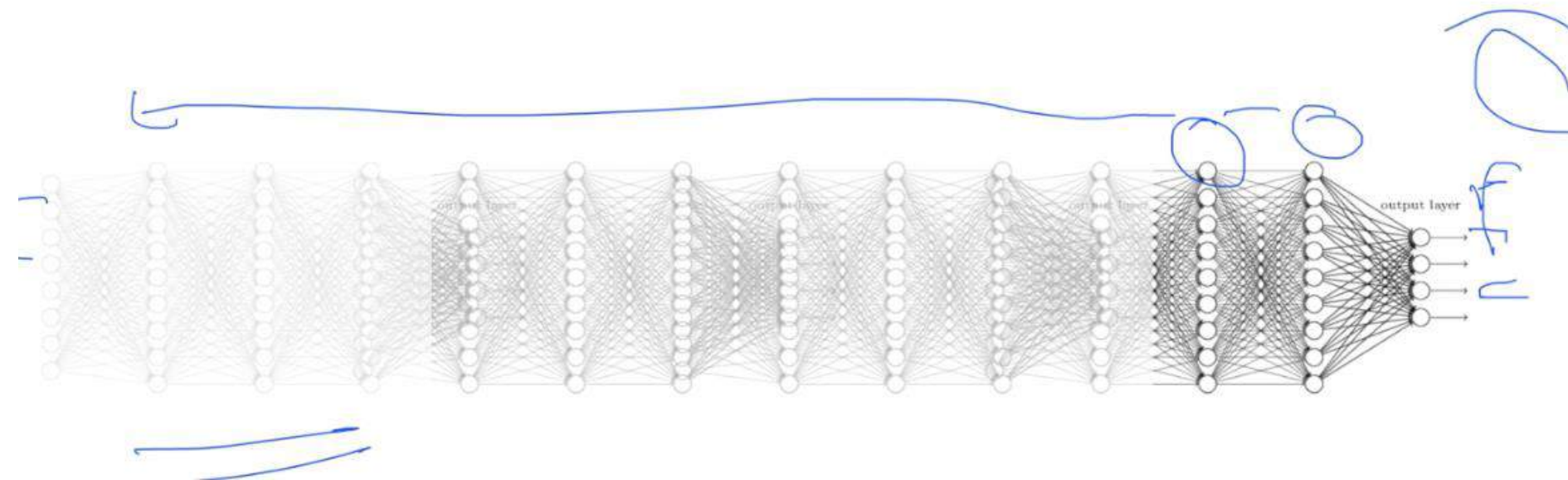


D.E. Rumelhart, G.E. Hinton, R.J. Williams

Learning representation by back-propagating errors. *Nature*, 323 (1986), pp. 533–536

- | | |
|--|--|
| <input type="checkbox"/> Solved learning problem | <input checked="" type="checkbox"/> Hard to train (non-convex, tricks) |
| <input type="checkbox"/> Biological system | <input checked="" type="checkbox"/> Hard to do theoretical analysis |
| <input type="checkbox"/> ... | <input checked="" type="checkbox"/> Small training sets ... |

Vanishing gradient (NN winter2: 1986-2006)







Resistance

[noisy_3 and 6](#)

[noisy_2](#) (anim)

[noisy_4](#) (anim)

Multiple

Character

[various stills](#)

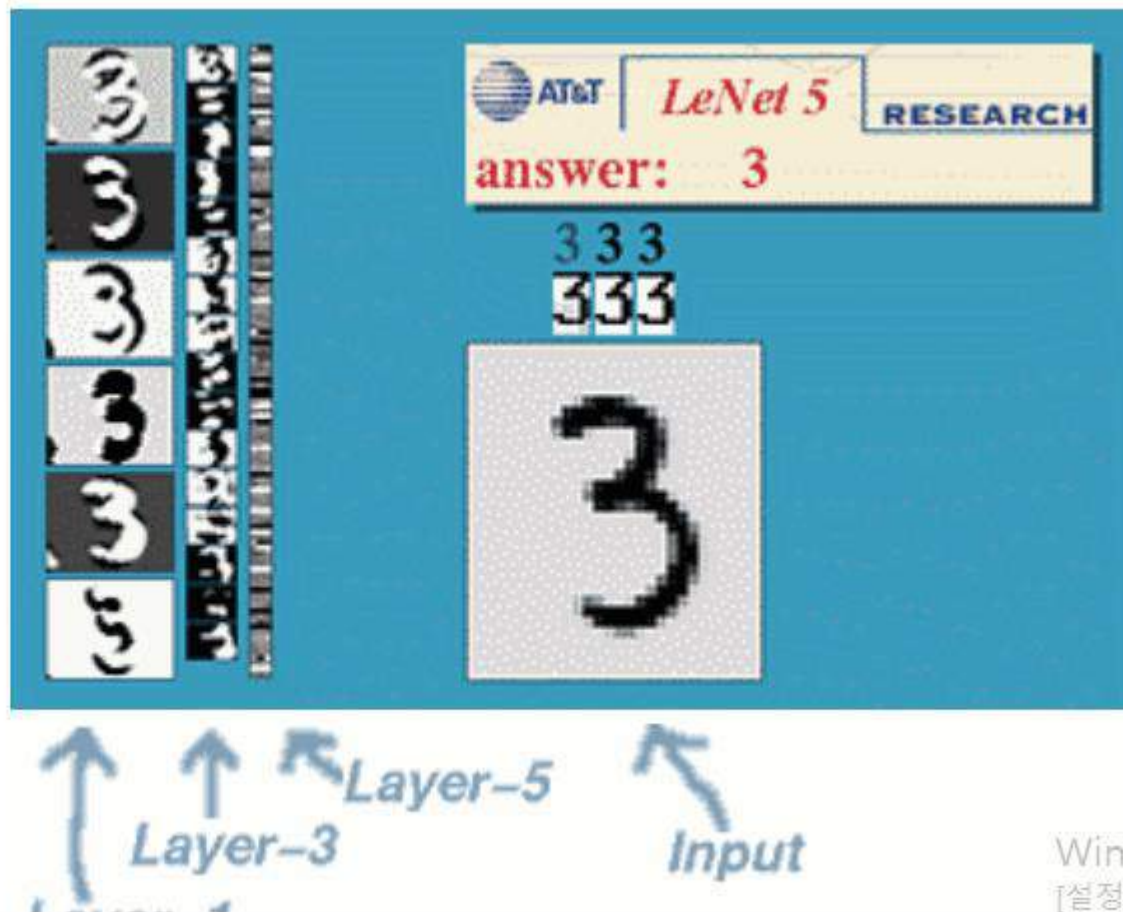
[dancing_00](#) (anim)

[dancing_384](#)

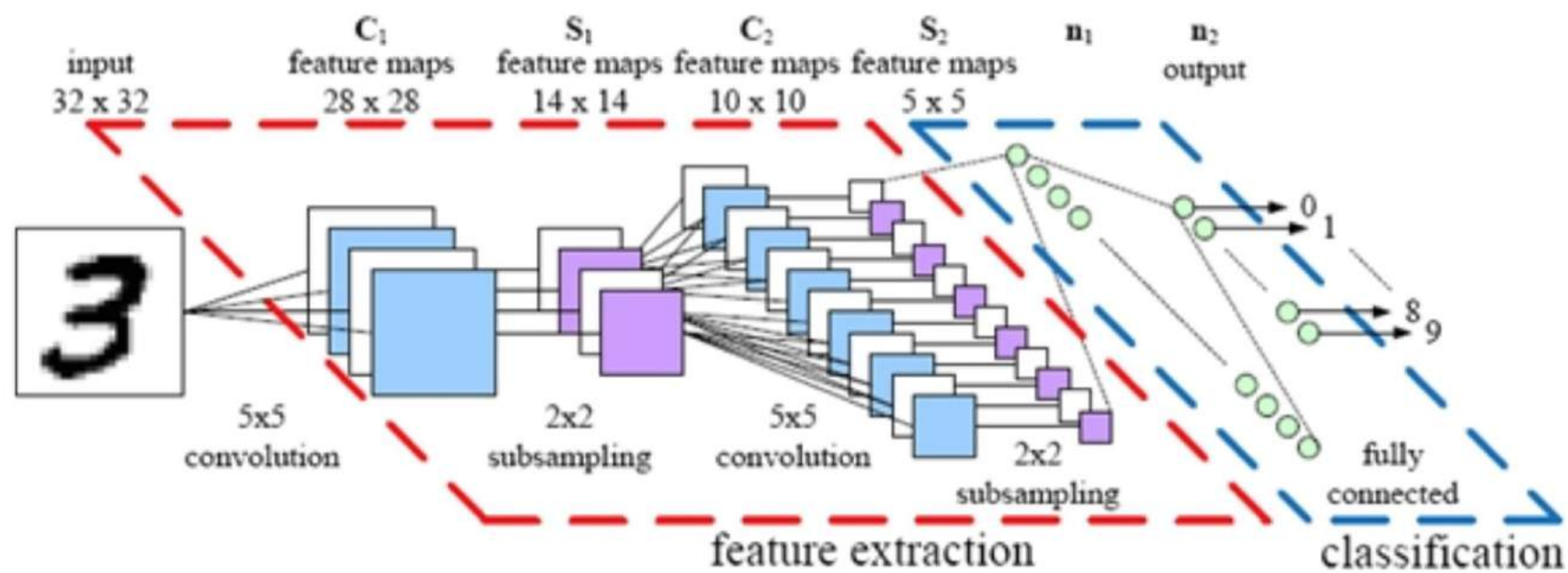
(anim)

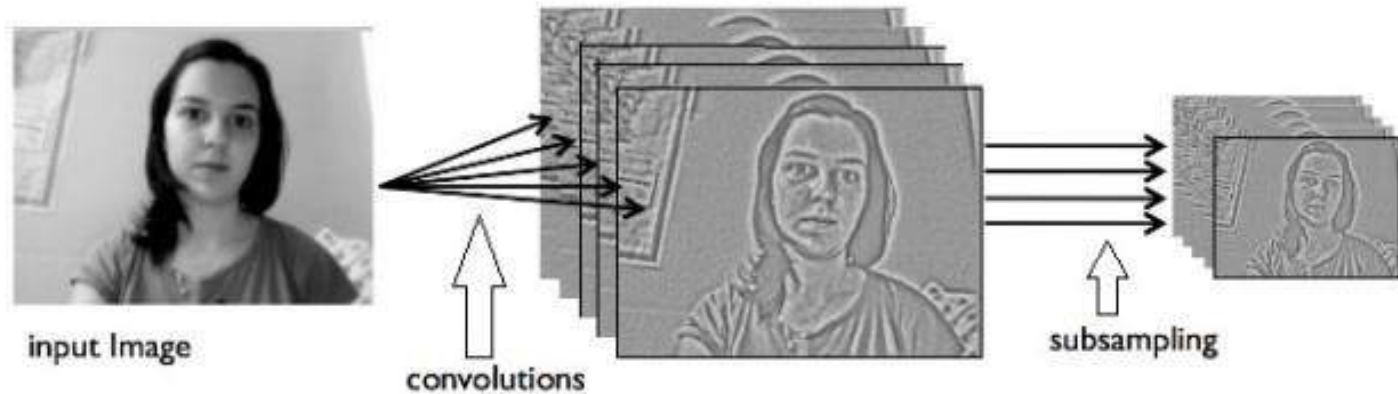
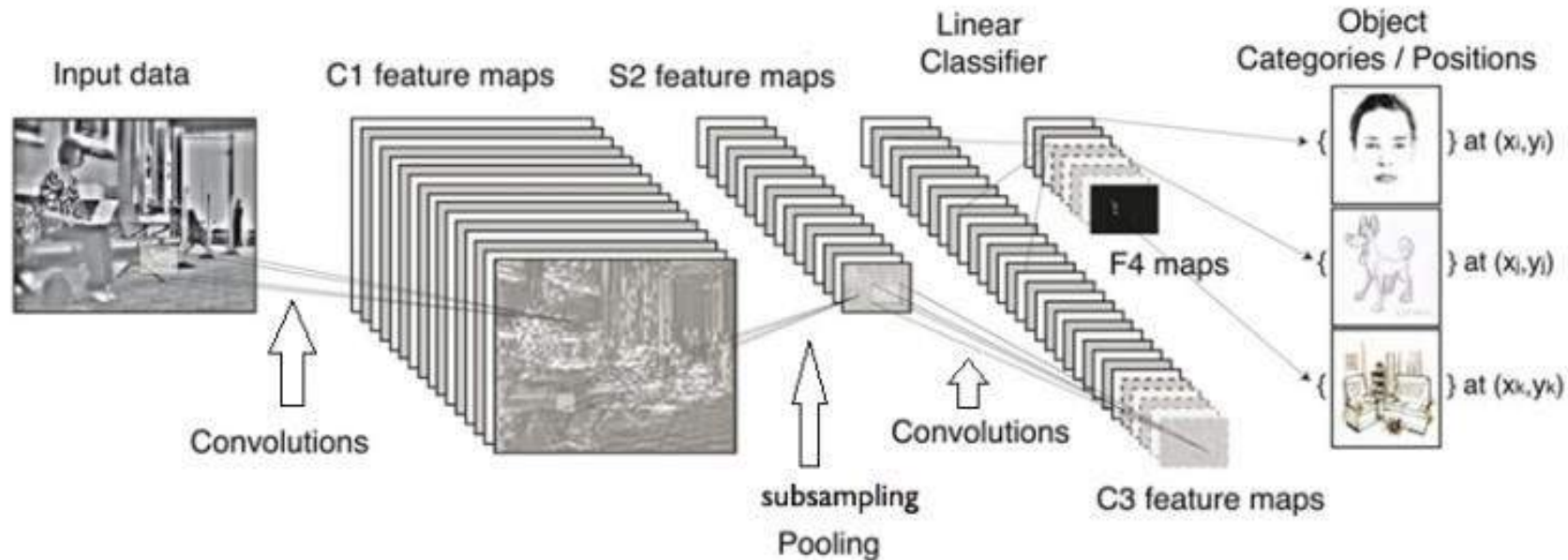
Complex cases

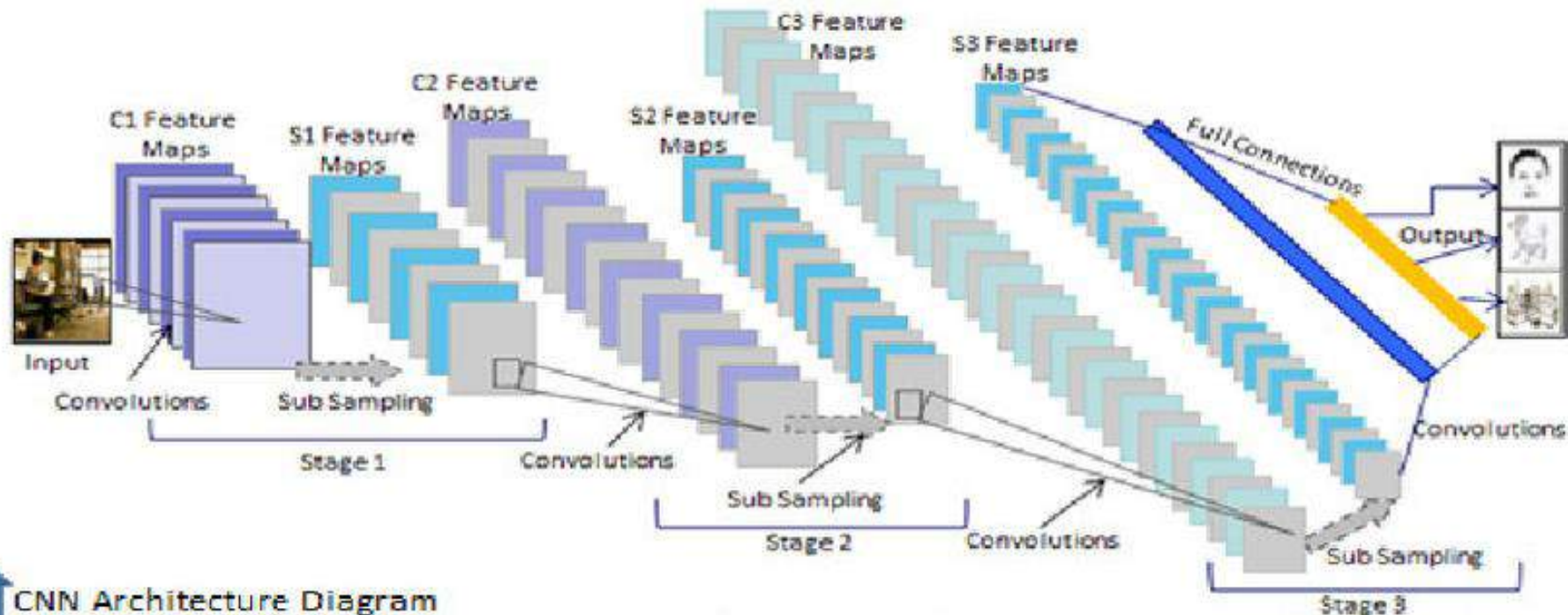
(anim)



Window
[설정]으로
합니다.







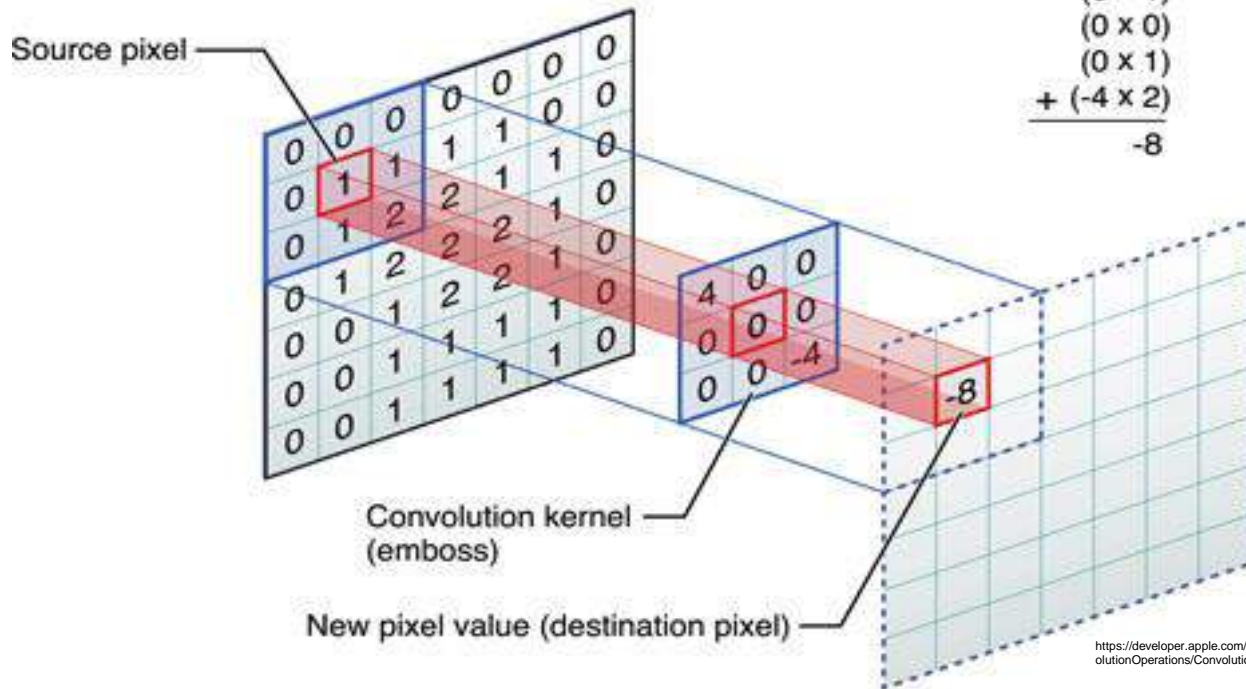
CNN Architecture Diagram

Hierarchical Feature Extraction in stages



Convolution Filter

Center element of the kernel is placed over the source pixel. The source pixel is then replaced with a weighted sum of itself and nearby pixels.



$$\begin{array}{r} (4 \times 0) \\ (0 \times 0) \\ (0 \times 0) \\ (0 \times 0) \\ (0 \times 1) \\ (0 \times 1) \\ (0 \times 0) \\ (0 \times 1) \\ + (-4 \times 2) \\ \hline -8 \end{array}$$

사람눈에는 특정 모양에 반응하는 신경세포들이 있다. 이를 구현한 가장 간단한 방법이 convolution filter이다.

1	2	3	0
0	1	2	3
3	0	1	2
2	3	0	1

입력 데이터



2	0	1
0	1	2
1	0	2

필터



15	16
6	15

1	2	1	0
0	1	2	3
3	0	1	2
2	4	0	1



2	

1	2	1	0
0	1	2	3
3	0	1	2
2	4	0	1



2	3
4	

1	2	1	0
0	1	2	3
3	0	1	2
2	4	0	1

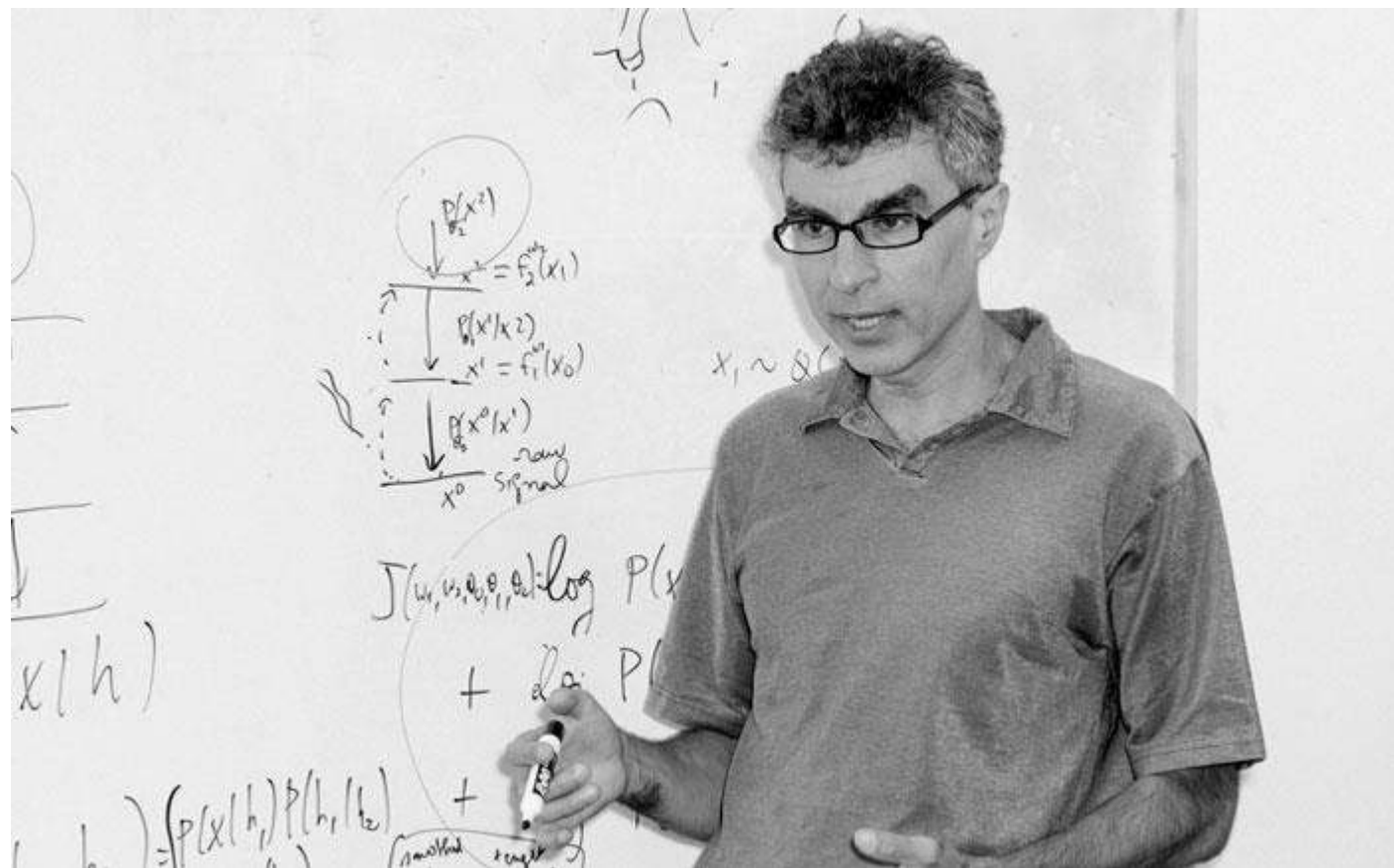


2	3

1	2	1	0
0	1	2	3
3	0	1	2
2	4	0	1



2	3
4	2





Interview with

Ian Goodfellow

Research Scientist, OpenAI

Inventor of Generative Adversarial Networks

Generative Adversarial Networks

- Created by Ian Goodfellow (OpenAI);
- Two neural networks compete (*minmax game*)
 - Discriminative network tries to distinguish between real and fake data
 - Generative network tries to generate samples to fool the discriminative one
- Use latent code (z).

Source: Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.

*Random
Input*



*Manipulator or
'Generator Network'*



Real Currency

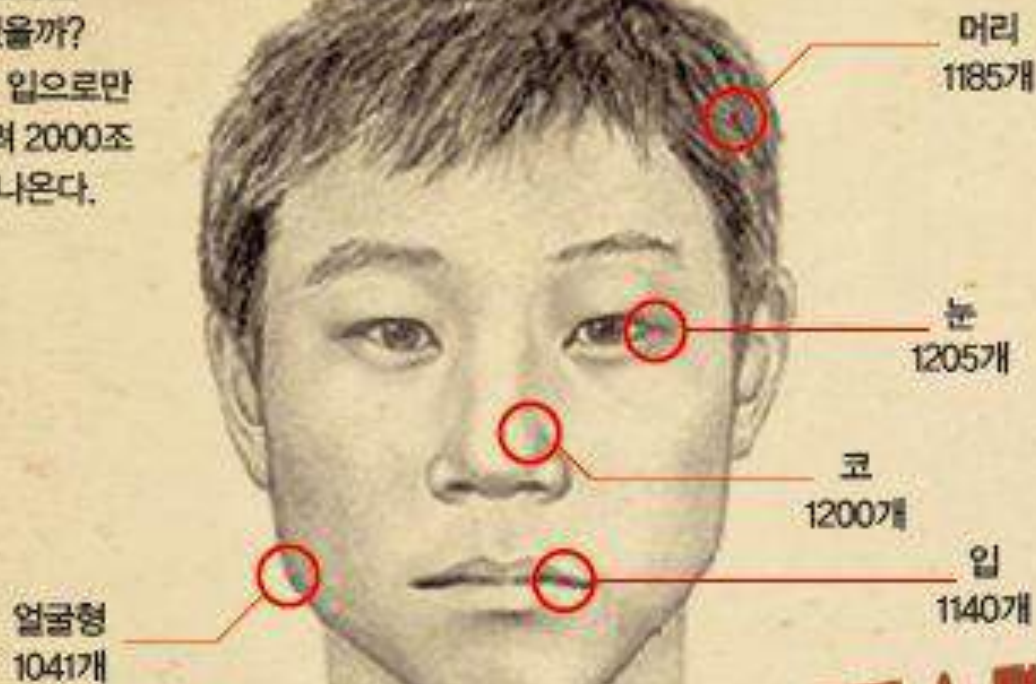


*Enforcer or
'Discriminator Network'*

Fake

Real

몽타주로 범죄자의 얼굴을
몇 명이나 그릴 수 있을까?
얼굴형, 머리, 눈, 코, 입으로만
몽타주를 그려도 무려 2000조
개 이상의 몽타주가 나온다.



WANTED

$$1041 \times 1185 \times 1205 \times 1200 \times 1140 = 2033490857400000 \text{ 개}$$

Deep learning: A brief history

In 1979, Kunihiko Fukushima invented an artificial neural network, "**Neocognitron**", which has a hierarchical multilayered architecture and acquires the ability to recognize visual patterns through learning.



In 1989, Yann LeCun et al. were able to apply the standard backpropagation algorithm, which had been around since 1974, to a deep neural network with the purpose of recognizing handwritten ZIP codes on mail.

Despite the success of applying the algorithm, the time to train the network on this dataset was **approximately 3 days**, making it impractical for general use.

/// the new rebels



Yann LeCun,
Facebook

Geoff Hinton,
Google

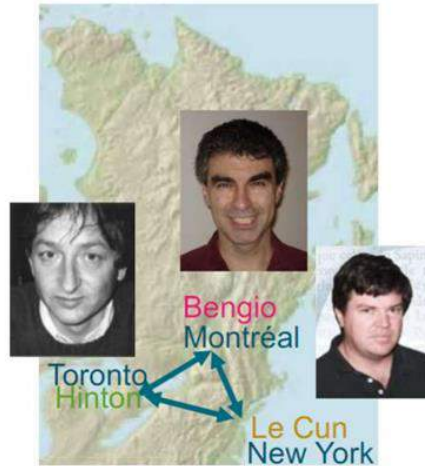
Yoshua Bengio

Andrew Ng
Baidu (Google)

Canadian Mafia (CIFAR)

@aliostad

2006: The Deep Breakthrough



- Hinton, Osindero & Teh
« A Fast Learning Algorithm for Deep Belief Nets », *Neural Computation*, 2006
- Bengio, Lamblin, Popovici, Larochelle
« Greedy Layer-Wise Training of Deep Networks », *NIPS'2006*
- Ranzato, Poultney, Chopra, LeCun
« Efficient Learning of Sparse Representations with an Energy-Based Model », *NIPS'2006*

Unsupervised greedy layer-wise

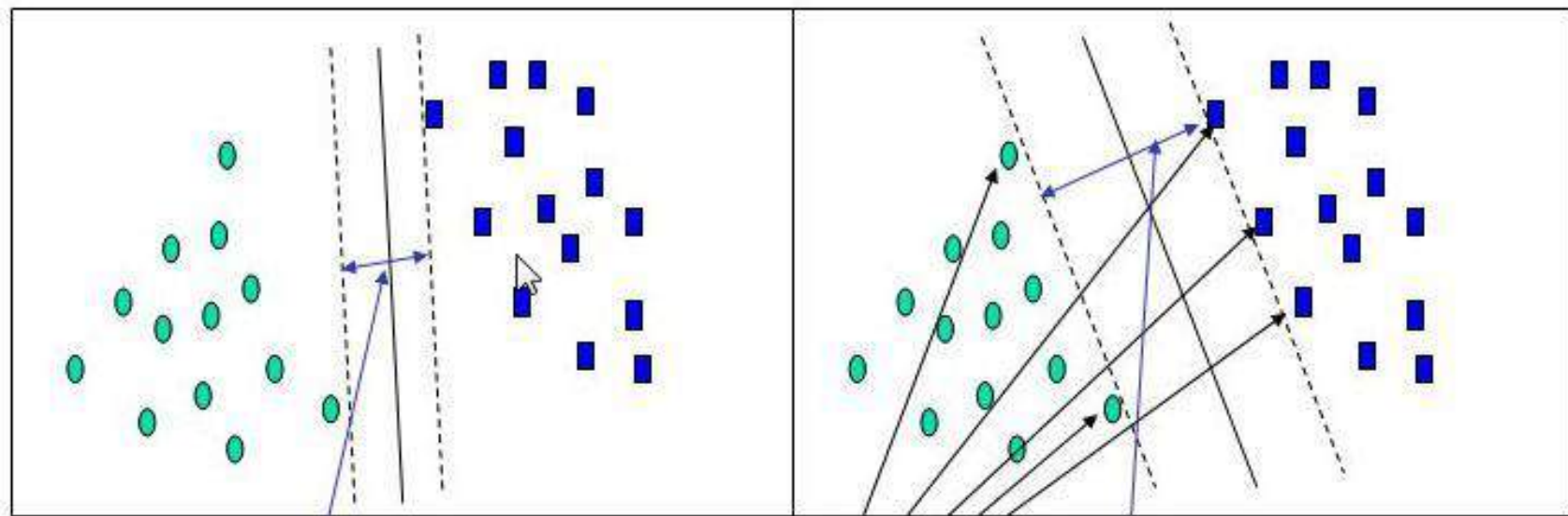
문재인과 친노 죽이기
2018년 3월 28일 오전 10시.
봉도사 카드내역이요? SBS가...
유재일 실시간 스트리밍 시작
www.youtube.com

CCN: Bitcoin, Blockchain & Cryptocurrency News
'All Hell Will Break Loose': Abra CEO Predicts Bitcoin Price Boom Will Return This Year
www.ccn.com

A close-up portrait of an elderly man with thinning grey hair and a serious expression. He is wearing a dark blue button-down shirt under a dark jacket. The background is a blurred, teal-colored architectural structure.

Prof. Vladimir Vapnik

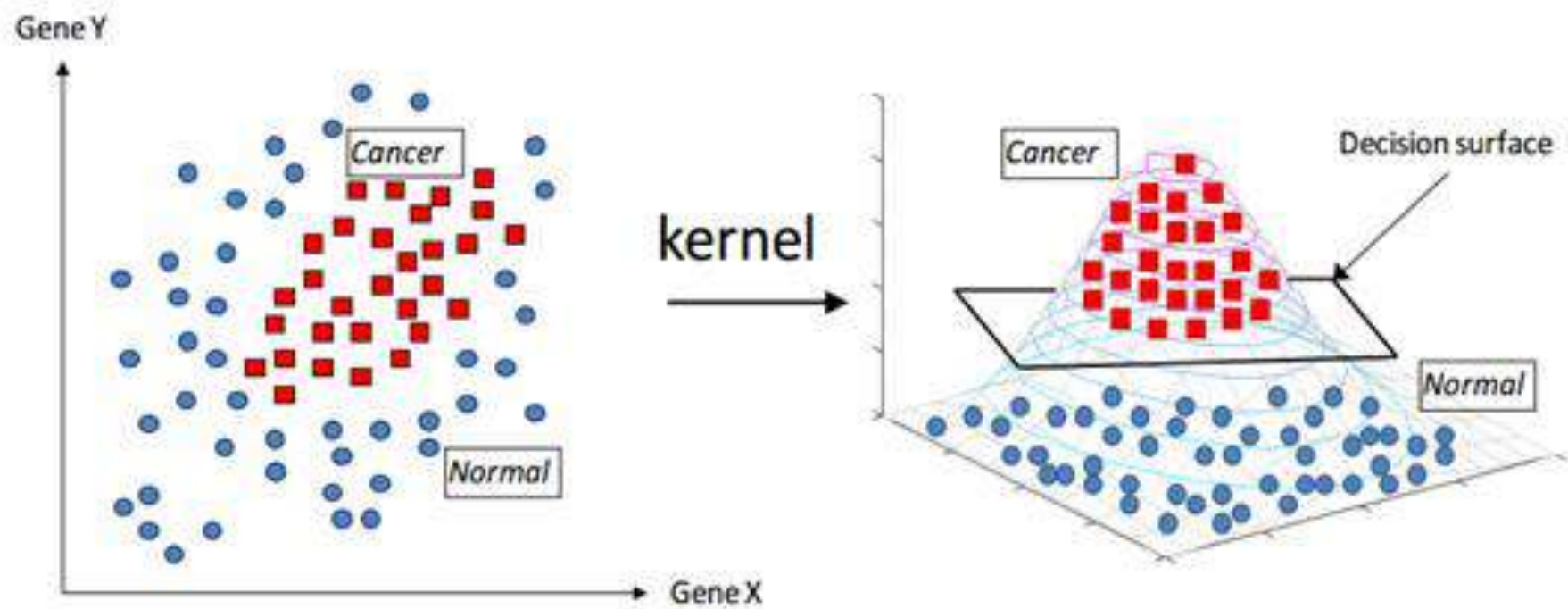
USA, Columbia University, Facebook



Small Margin

Large Margin

Support Vectors

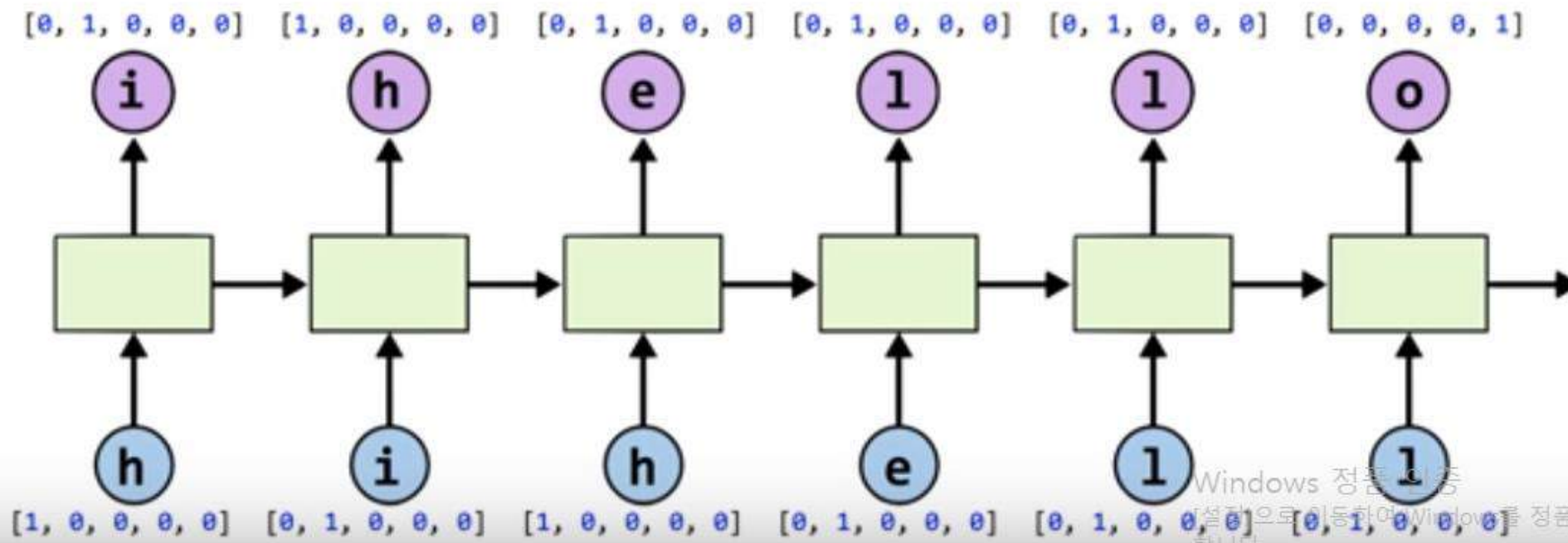


時空



Teach RNN 'hihello'

$[0, 1, 0, 0, 0]$, # i
 $[0, 0, 1, 0, 0]$, # e
 $[0, 0, 0, 1, 0]$, # l
 $[0, 0, 0, 0, 1]$, # o



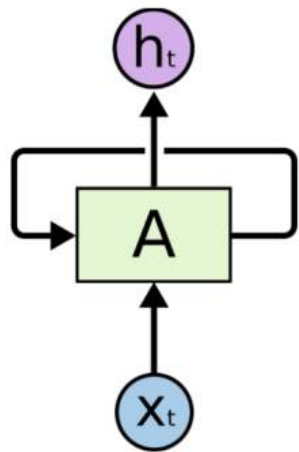
Windows 정품 인증
설치로 이동하여 Windows를 정품
합니다.



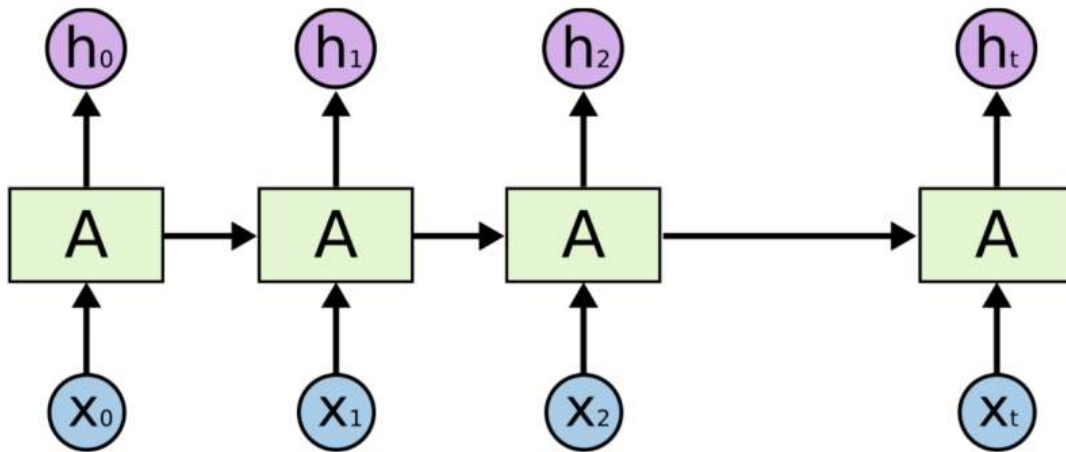


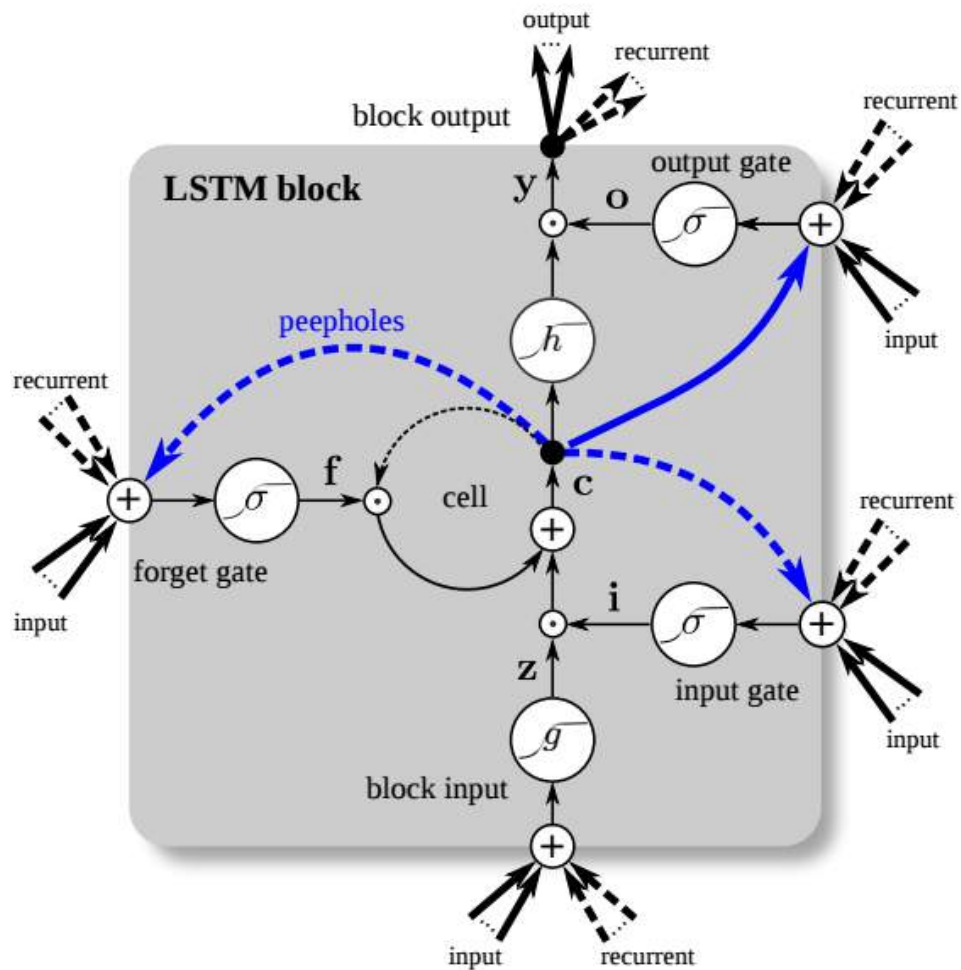
1991: SEPP HOCHREITER'S ANALYSIS OF THE
FUNDAMENTAL DEEP LEARNING PROBLEM

$$\left\| \frac{\partial e(t-q)}{\partial e(t)} \right\| = \left\| \prod_{m=1}^q W F'(Net(t-m)) \right\|$$
$$\leq (\|W\| \max_{Net} \{\|F'(Net)\|\})^q$$



=





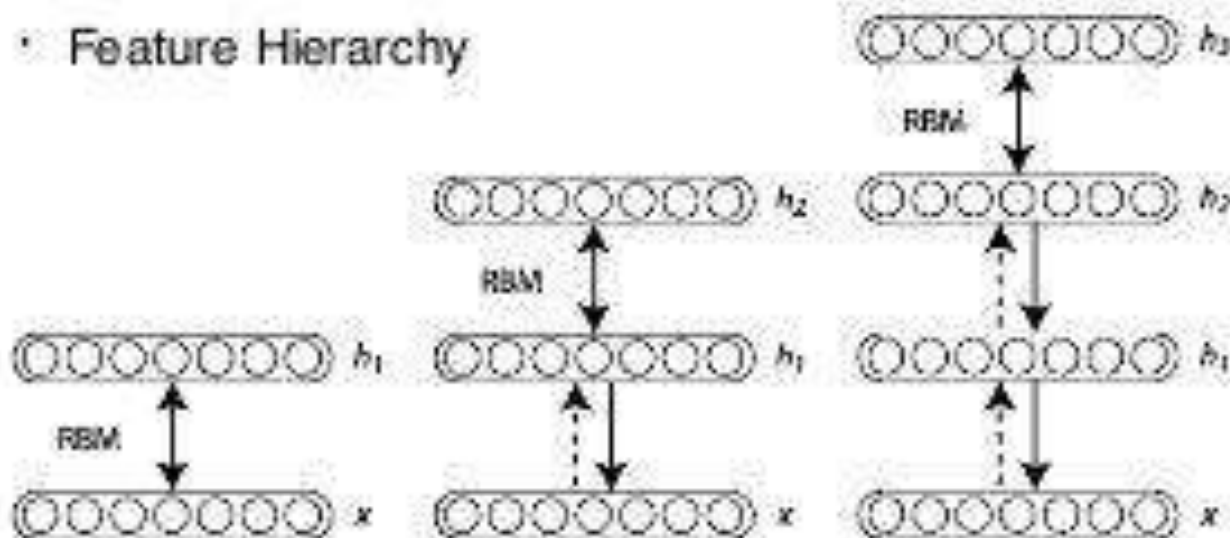
Legend

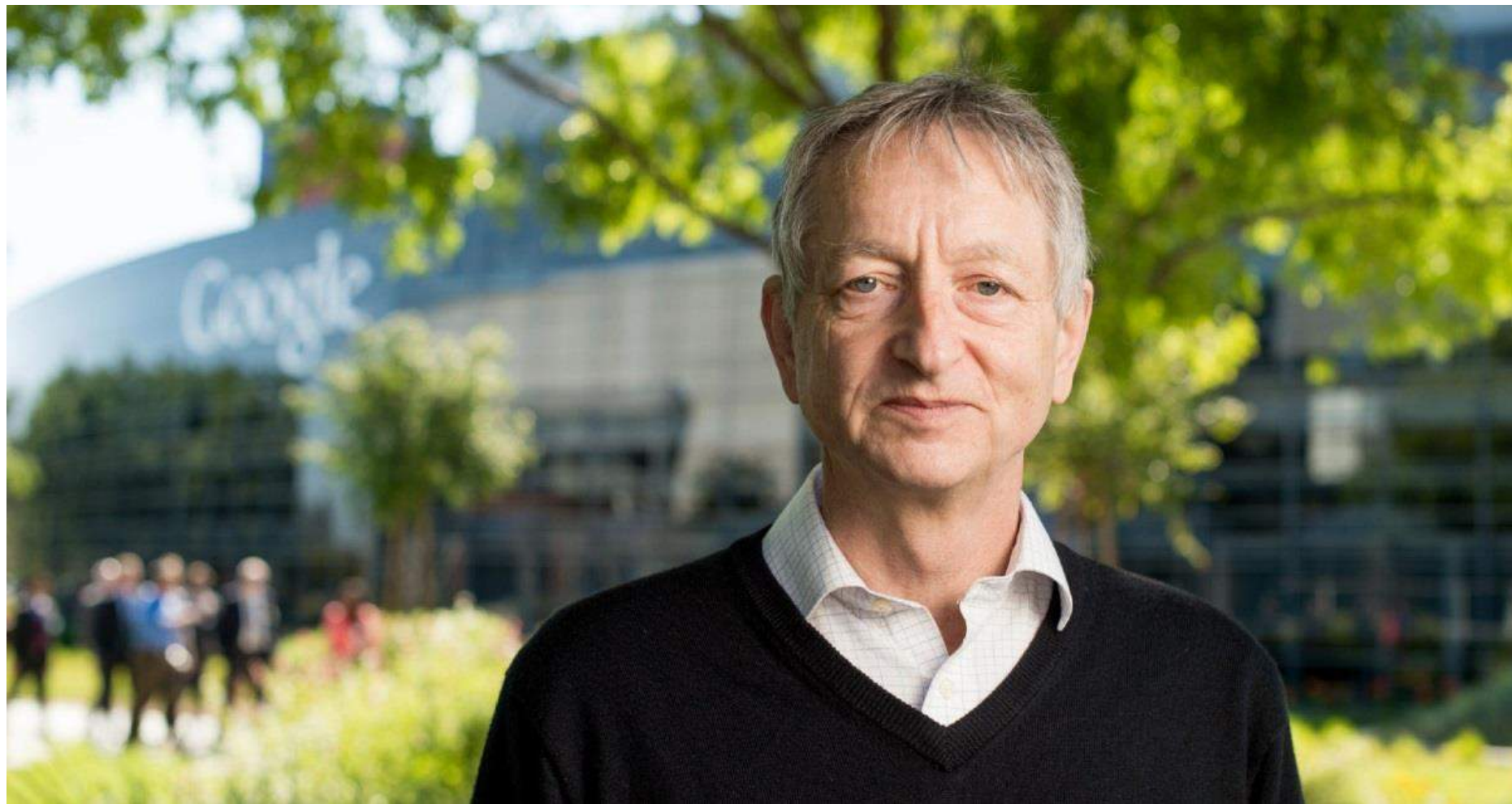
- unweighted connection
- weighted connection
- - - connection with time-lag
- branching point
- \odot multiplication
- $+$ sum over all inputs
- σ gate activation function (always sigmoid)
- g input activation function (usually tanh)
- h output activation function (usually tanh)

Deep Belief Network (DBN)



- Stacked RBM
- Introduced by Hinton et al. (2006)
- 1st RBM hidden layer == 2th RBM input layer
- Feature Hierarchy

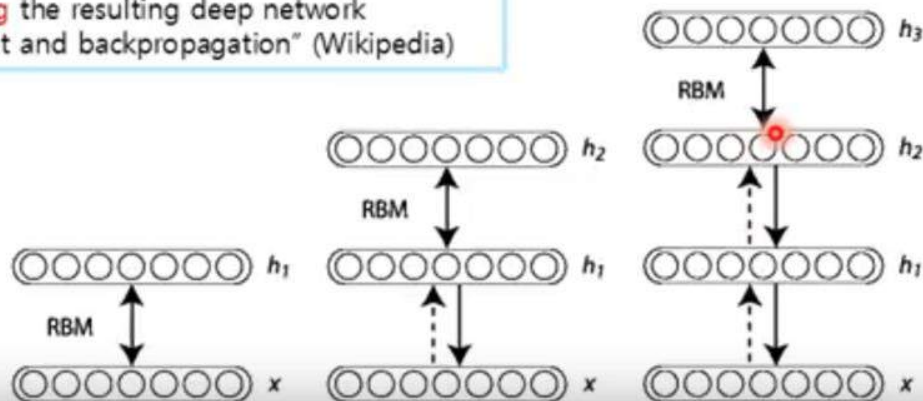




1st BreakThrough: Pre-training with RBM (Hinton, 2006)

- Deep Belief Network (DBN)
 - The network is not trained as a whole (Back-propagation)
 - Each layer is trained with unsupervised **Restricted Boltzmann machine**
 - Stochastic, Generative: Learn data distribution & can generate data-like samples

“deep belief networks can be formed by “stacking” RBMs and optionally **fine-tuning** the resulting deep network with gradient descent and backpropagation” (Wikipedia)



- No need to use complicated RBM for weight initializations
- Simple methods are OK
 - **Xavier initialization:** X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in International conference on artificial intelligence and statistics, 2010
 - **He's initialization:** K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectification: Surpassing Human-Level Performance on ImageNet Classification," 2015

Xavier/He initialization

- Makes sure the weights are 'just right', not too small, not too big
- Using number of input (fan_in) and output (fan_out)

```
# Xavier initialization
# Glorot et al. 2010
W = np.random.randn(fan_in, fan_out)/np.sqrt(fan_in)

# He et al. 2015
W = np.random.randn(fan_in, fan_out)/np.sqrt(fan_in/2)
```



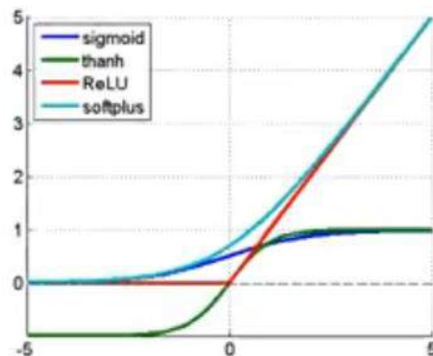
Xavier Glorot



<http://kaiminghe.com/>

2nd BreakThrough: ReLU (2010, 2011)

- Rectified Linear Unit (ReLU)
 - Gradient is defined as:
$$\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$
 - Avoids vanishing gradient
 - Eliminates the necessity of pretraining
 - Simplifies back-propagation
 - Efficient Computation: Only comparison, addition and multiplication
 - Sparse Activation → Makes learning faster
 - In a randomly initialized networks, only about 50% of hidden units are activated
 - Can also partially relieve overfitting problem



리뷰: 인공신경망에 쓰이는 활성화 함수

Neural Network Activation Functions

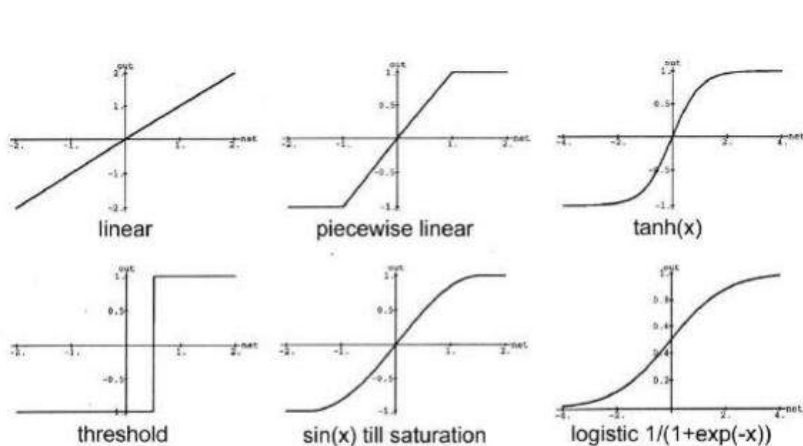
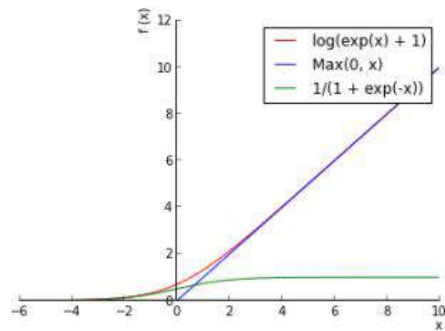


Figure 3.2: Neural network activation functions. Note the characteristic 'sigmoid' or S-shape. (A. Graves, 2012)



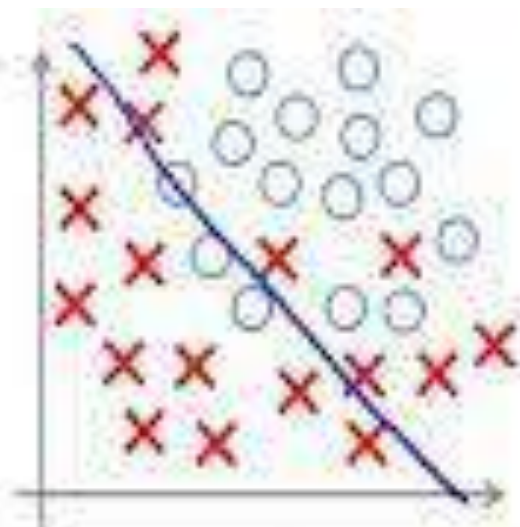
Rectified Linear Unit

$$\text{ReLU}(x) = \max(x, 0)$$

[장점]

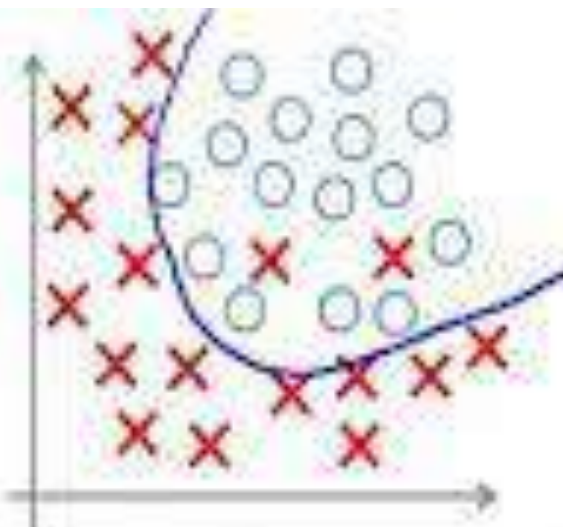
- Hidden unit에 sparsity가 나타난다
- Gradient vanishing 문제가 없다
- 계산이 단순해져 학습이 빠르면서 성능 향상에도 일조

- 입력값을 처리하여 ON 또는 OFF 출력. 보통 출력값에 상, 하한을 두려 함
- 신경망에서는 '미분 가능한' smooth한 S자 모양 함수를 적용
- 최근 딥러닝 모델에서는 ReLU를 선호
- 순환신경망 내부에서는 sigmoid와 tanh 함수를 적용함

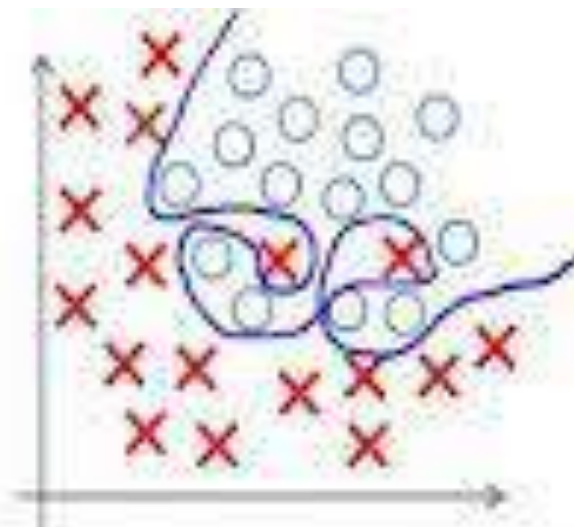


Under-fitting

(too simple to
explain the
variance)



Appropriate-fitting

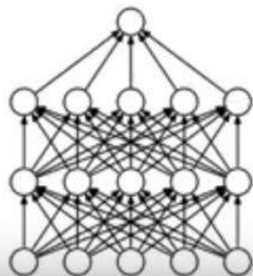


Over-fitting

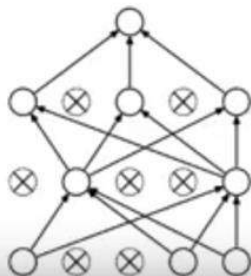
(forcefitting – too
good to be true)

3rd BreakThrough: Dropout (Hinton, 2012)

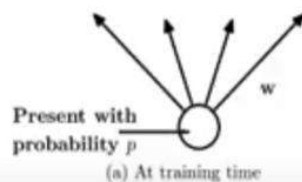
- Motivations
 - Deep neural nets with a large number of parameters are very powerful machine learning systems.
 - However, overfitting is a serious problem in such networks.
 - Large networks are also slow to use, making it difficult to deal with overfitting by combining the predictions of many different large neural nets at test time.
- The key idea is to **randomly drop units** (along with their connections) from the neural network during training.
- This prevents units from co-adapting too much.



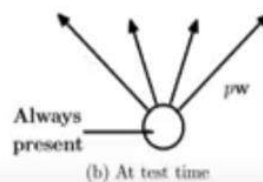
(a) Standard Neural Net



(b) After applying dropout.



(a) At training time



(b) At test time



정신없는 가을작물 속아주기

9월5일부터 10월15일까지 과정샷!

오토인코더(auto-encoder) – 특징 표현을 학습하다

● 뉴럴 네트워크의 계층

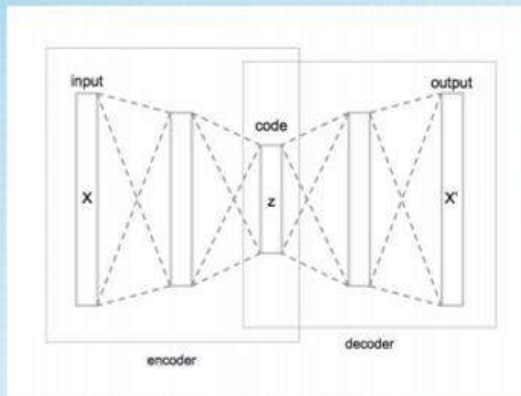
- ◆ 계층이 많으면 자유도가 높아지며 표현할 수 있는 함수도 늘어난다
- ◆ 뉴럴 네트워크를 높게 높게 쌓아 보자
 - 정밀도가 높아지지 않았다
 - 오차의 역전파가 제대로 이뤄지지 않는다

● 딥러닝은 높이 쌓아서 성능을 높였다

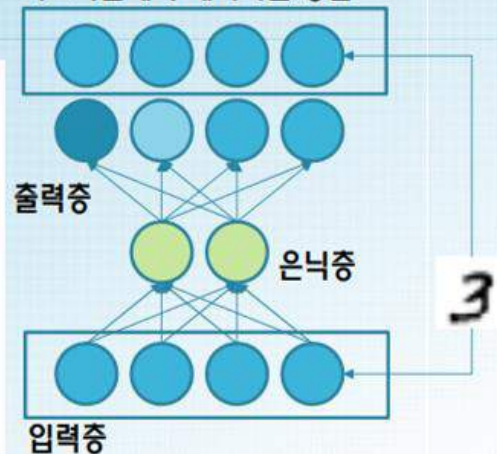
- ◆ 한 층씩 계층마다 학습이 이뤄짐
- ◆ 오토인코더 사용
 - 특징을 추출하는 역할
 - Code = 입력의 특징

● 오토 인코더

- ◆ 입력과 정답을 같이 한다
- ◆ 은닉층에 무엇이 남을까
 - 이것을 코드(code)라고 하자
 - 그것은 특징이다.



지도학습에서 제시되는 정답



손글씨 문자의 정보량

● 오토인코더(auto-encoder)

- ◆ 입력 정보가 더 작은 규모의 은닉층을 통과한 뒤 출력으로 나간다
- ◆ 가중치는 복원 에러가 최소가 되도록 조정된다
- ◆ 이것은 앞서 전국 날씨를 추정하는 것과 같다
- ◆ 은닉층은 A에게서 B로 전달된 특징 정보

● 손글씨 문자 인식

- ◆ 입력층의 노드: 28x28 픽셀 (784)
- ◆ 은닉층의 노드 100 개로 가정
 - 학습 과정에서 중복 정보들은 하나의 노드에 합쳐지는 등의 일이 이루어짐
 - 이 정보가 출력층으로 전달될 것
 - 이상적인 경우 이 정보에서 가중치가 곱해져 나오는 출력층은 입력과 동일
 - 이 경우 은닉층은 입력 데이터를 적은 규모의 데이터로 잘 표현(representation)
- ◆ 학습의 목표
 - 이 은닉층을 얻는 것 = 입력의 특징을 표현하는 방법을 배우는 것
 - “표현 학습(representation learning)” - 제프리 힌튼(Geoffrey Hinton) 교수
 - “특징 표현 학습” - 마츠오 유타카(松尾豊) 교수의 표현

Auto-encoder

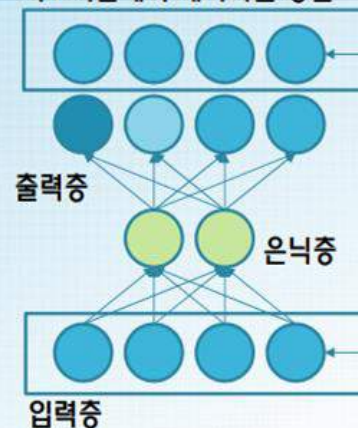
주성분 분석과 유사

- 중요한 자원 찾기
- 낮은 자원으로 투영

다양한 노이즈

- 신뢰도 높은 주성분 얻을 수 있음

지도학습에서 제시되는 정답



여러 계층을 “깊이” 탐구하다

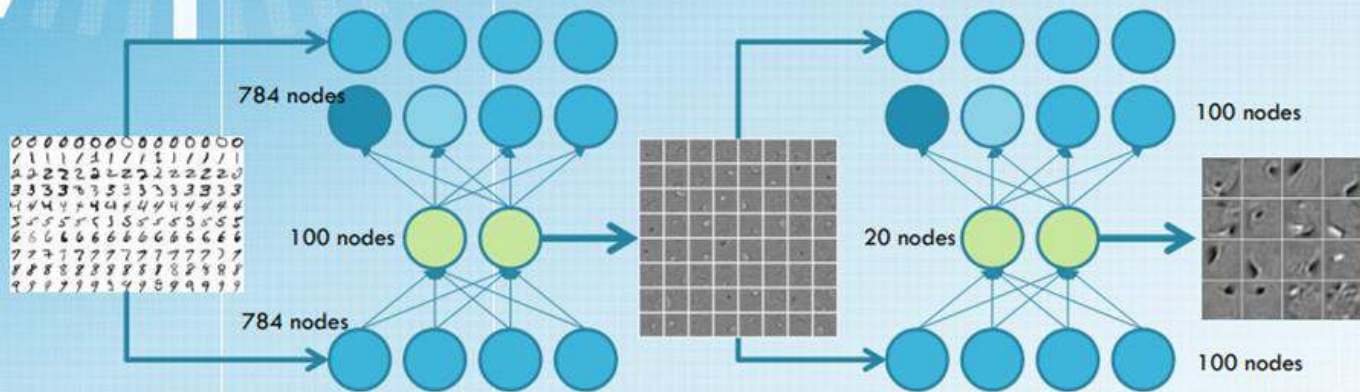
● 딥러닝

- ◆ 오토인코더의 은닉층(표현)을 다음 단계의 입력으로 제공

- End-to-end learning

● 손 글씨 인식의 예

- ◆ 1단계: 은닉층 100 개로 가정했음
- ◆ 2단계: 입력층은 1단계의 은닉층에 대응되므로 100 개의 입력
 - 2단계 은닉층이 20개인 경우
 - 이 특징을 다시 20 개의 노드로 “압축(code)” 혹은 “표현(represent)” 하는 일









Google

JEFF DEAN

제프딘과 함께하는 딥러닝(Deep Learning) 이야기

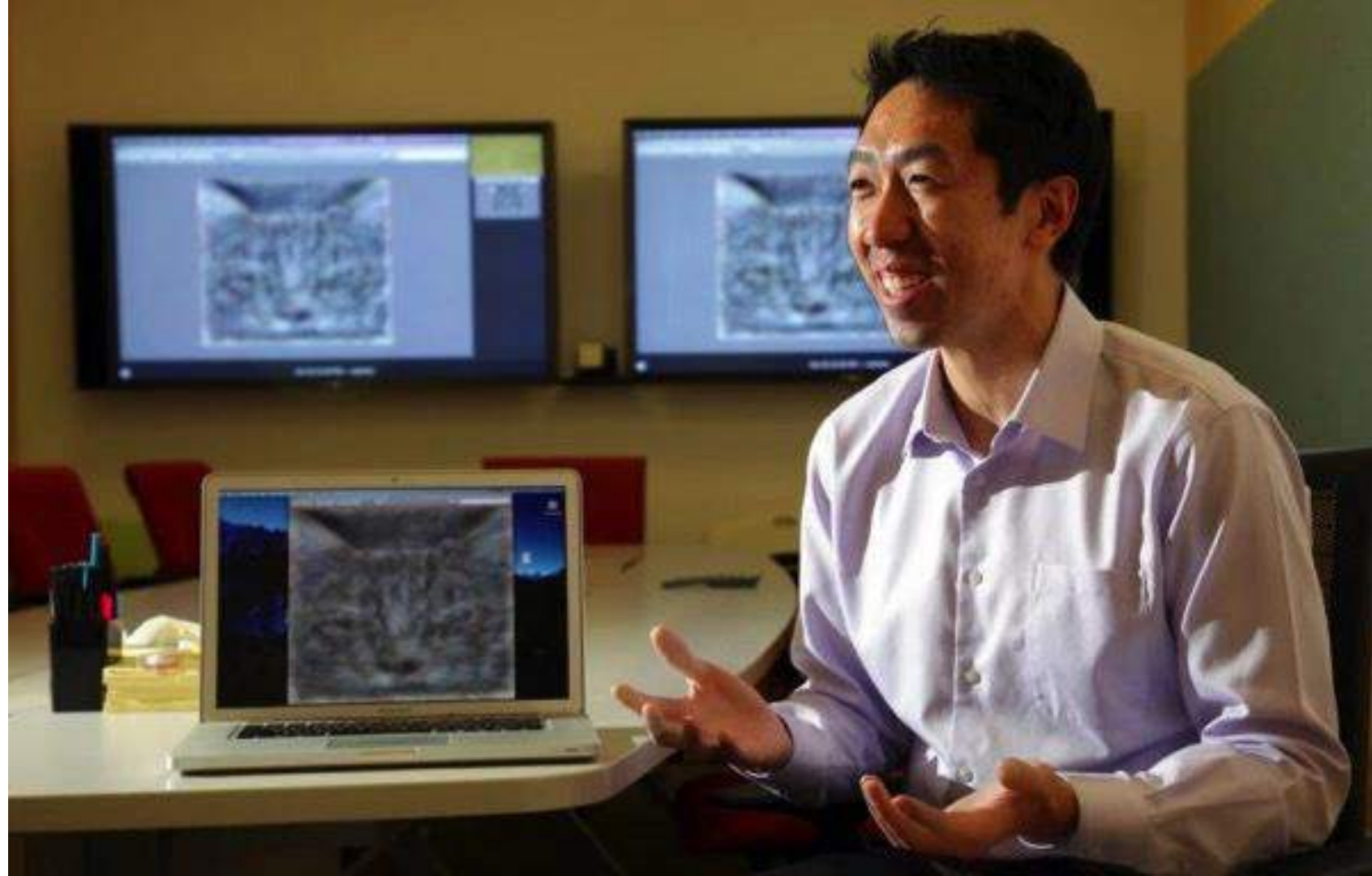
3월7일(월), 2:00-03:30p.m.

“라이브 스트리밍으로 함께 하세요”

구글의 시니어 펠로우로, 수 많은 인재들이 모이는 구글에서도 손 꼽히는 전설적인 프로그래머 중 하나로 통용된다. 현 시점에서는 구글 브레인의 수장을 맡고 있다. 2017년 3월 27일자로 구글 연구소장을 맡게 되었다.

그의 행적은 그야말로 구글의 엔지니어링 그 자체라고 할 수 있다. 구글 인프라의 핵심인 맵 리듀스, 빅테이블, 스파너 등의 가장 핵심적인 프로젝트를 전두 지휘했으며, 구글 검색, 구글 번역, 구글 광고의 초창기 버전 역시 그의 손에서 태어났다. 그리고 2012년 경 딥러닝의 존재를 파악한 뒤 전 세계 탑 레벨의 인공지능 인재들을 끌어모아 구글 브레인을 설립하는 등, 구글의 가장 핵심적인 제품들은 대부분 그의 손을 거쳐 태어났다. 또한 현 시점 임원급의 엔지니어 상당수는 제프딘이 데려온 사람들이기도 하다. 이렇게 먼치킨스러운 능력에 존경을 보내고자 구글러들은 제프딘에 대한 사실들을 정리하기에 이르르는 수준.

- 제프 딘의 PIN은 원주율의 마지막 4자리이다.
- 2002년 인텍싱 서버가 죽었을 때, 2시간 동안 제프 딘이 직접 유저의 검색에 응대했다. 그 시간 동안 검색 품질 평가 지수가 5점 올랐다고 한다.
- 구글 인사 시스템의 최대 레벨이 10일 때 제프딘은 11로 승진했다. (실제 일어난 일)
- 제프딘은 코드를 커밋하기 전에 컴파일과 실행을 해보는데, 이는 그저 CPU와 컴파일러에 버그가 있나를 확인하기 위함일 뿐이다.
- 제프딘이 스탠포드에 강의를 갔을 때 사람들이 너무나도 많이 몰려와서 도널드 커누스 교수조차 땅바닥에 앉아서 봐야했다. (실제 일어난 일)
- 제프딘의 이력서에는 그가 아직 하지 않은 일들이 적혀 있다. 그 편이 더 짧기 때문이다.
- 컴파일러는 제프딘에게 경고를 하지 않는다. 제프딘이 컴파일러에게 경고를 한다.
- 신이 "빛이여 있으라"라 말했다. 제프딘은 코드 리뷰를 해주기 위해 그 옆에 있었다.





143	179	90	39	42
86	49	179	39	139
138	90	142	106	152
116	250	96	42	80
182	144	62	183	137
244	117	161	241	25
142	148	244	9	18
66	158	90	68	21
206	19	186	2	11
29	43	170	64	1
170	1	92	145	14
98	34	133	62	5
226	88	170	76	7
219	46	191	194	1
71	106	27	125	4
98	145	201	88	1
25	200	37	225	1
126	182	188	101	2
167	194	184	108	2
205	127	97	52	1
163	37	101	196	1
152	105	58	41	1
23	168	105	160	1

ImageNet



Prof. Fei-Fei Li
Stanford University
Creator of ImageNet



IMAGENET

IMAGENET

14,847 categories, 9,349,136 images

- Animals
 - Fish
 - Bird
 - Mammal
 - Invertebrate
- Scenes
 - Indoors
 - Geological formations
- Sport Activities
- Fabric Materials
- Instrumentation
 - Tool
 - Appliances
 - ...
- Plants
 - ...

ImageNet Challenge 2012

Task 1: Classification



Car

- Predict a class label
- 5 predictions / image
- 1000 classes
- 1,200 images per class for training
- Bounding boxes for 50% of training.

Task 2: Detection (Classification + Localization)



classification

Car

- Predict a class label and a bounding box
- 5 predictions / image
- 1000 classes
- 1,200 images per class for training
- Bounding boxes for 40% of training.

Task 3: Fine-grained classification



classification

Walker hound

- Predict a class label given a bounding box in test
- 1 prediction / image
- 120 dog classes (subset)
- ~200 images per class for training (subset)
- Bounding boxes for 100% of training

Task 1: Classification



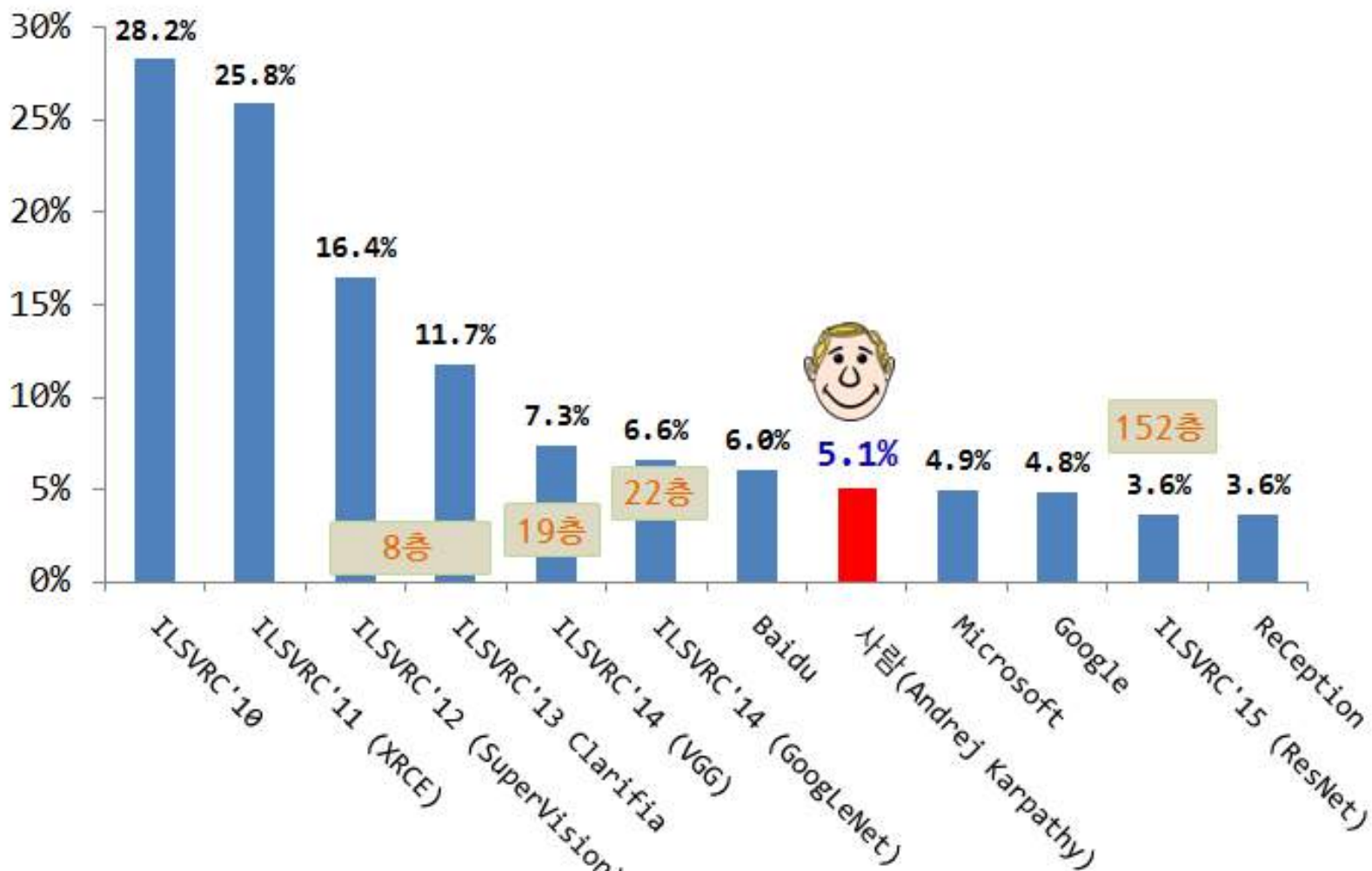
Car

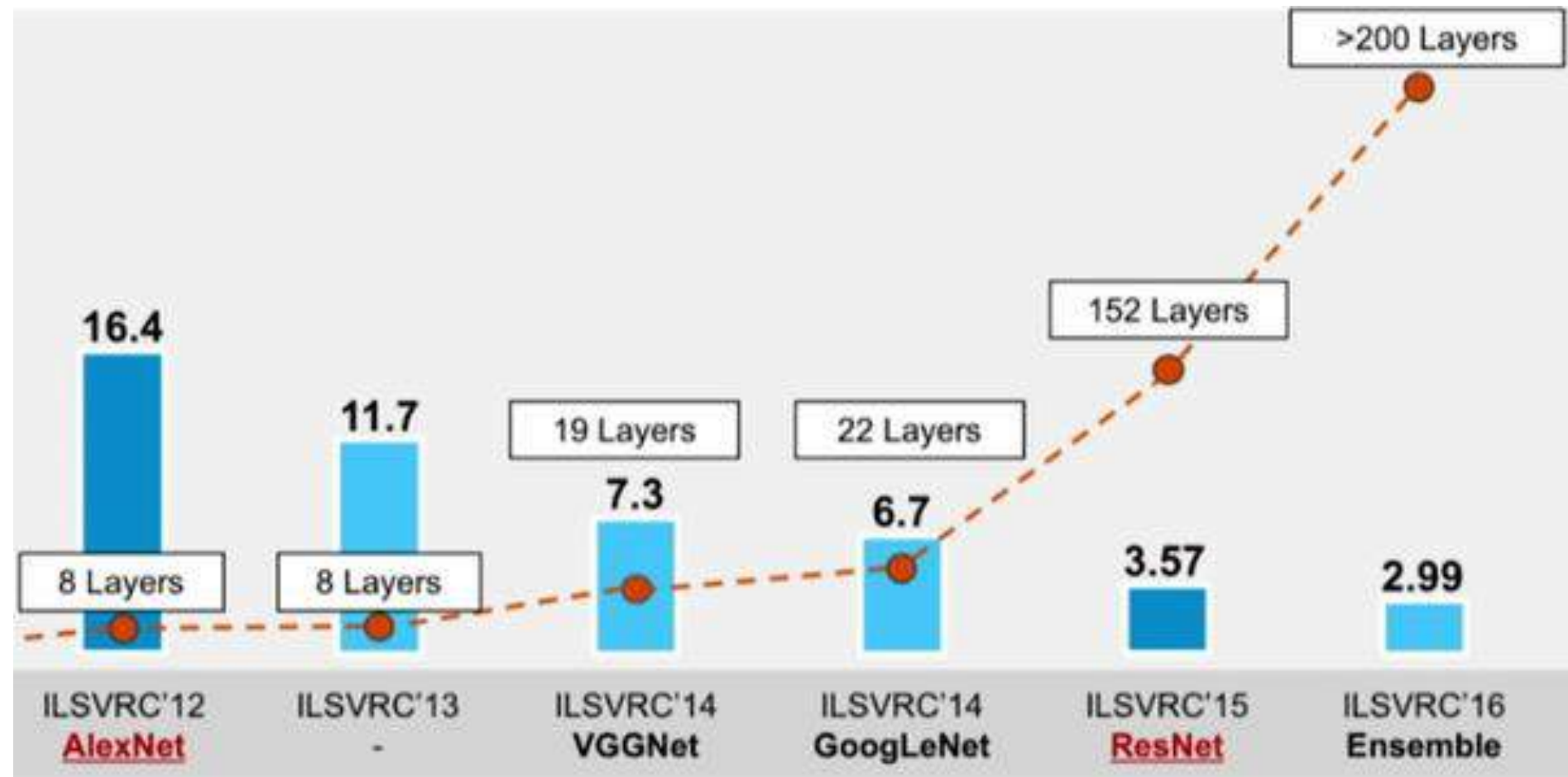
Winner

SuperVision

Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton

University of Toronto





출처 : <http://image-net.org/challenges/LSVRC/2012/ilsvrc2012.pdf>



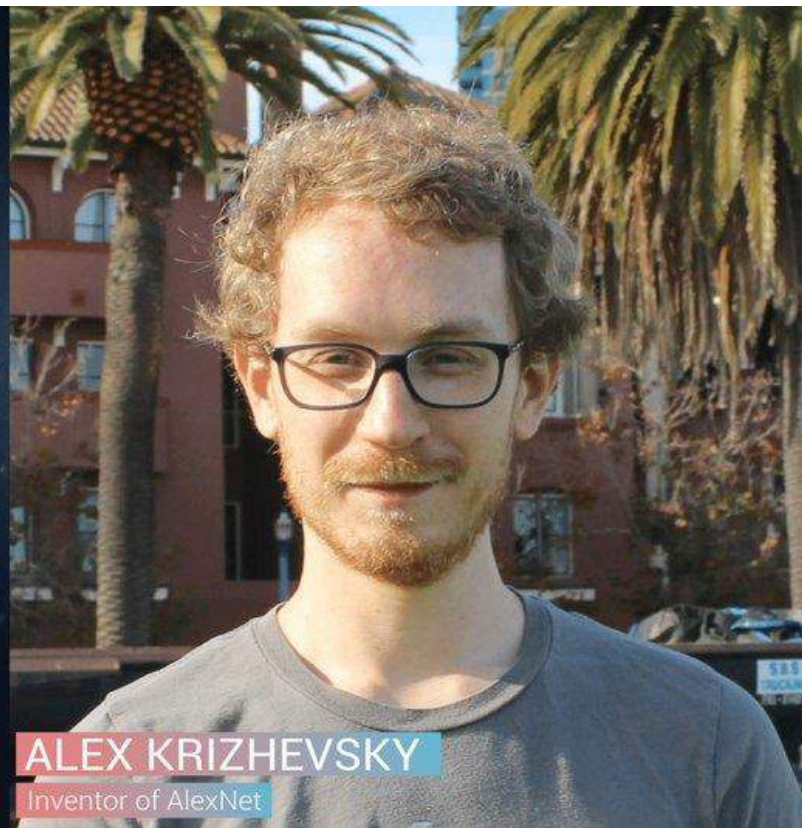
AI Meetup
Series

Age of AI

Jan 31 - Feb 1, 2018

Regency Ballroom
San Francisco CA

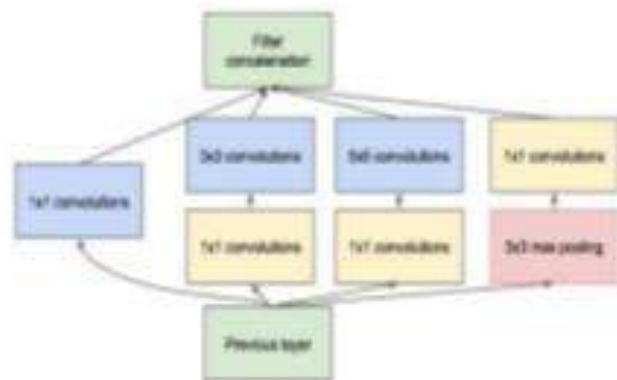
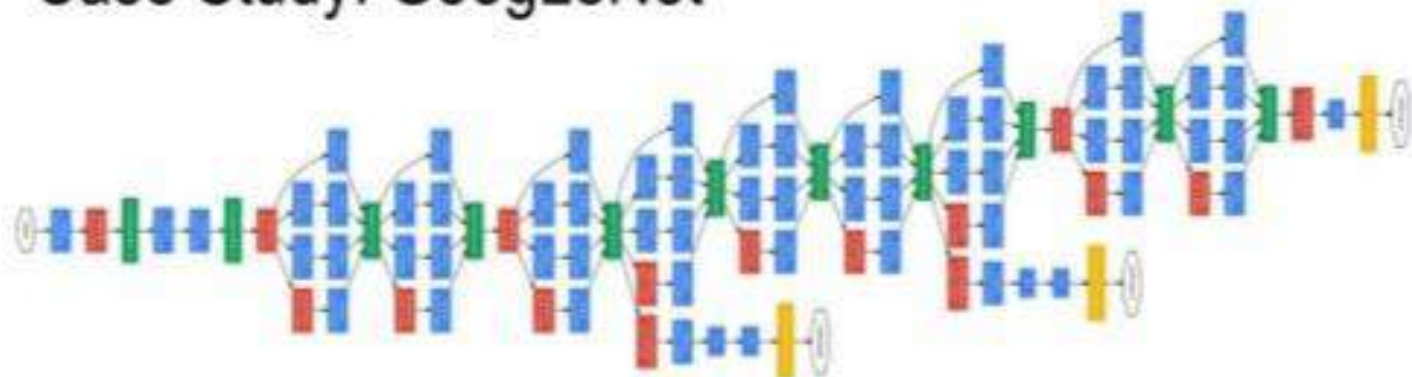
aimeetupseries.com



ALEX KRIZHEVSKY

Inventor of AlexNet

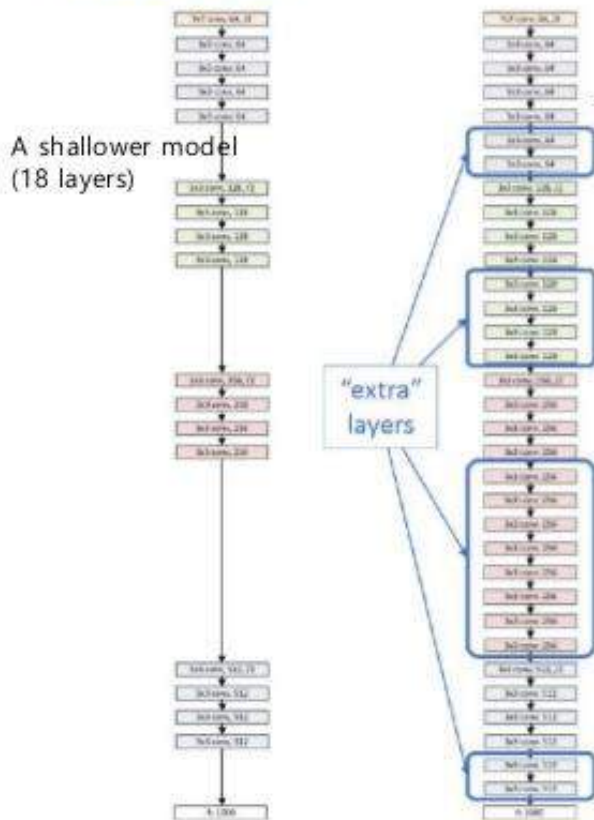
Case Study: GoogLeNet [Szegedy et al., 2014]



Inception module

ILSVRC 2014 winner (6.7% top 5 error)

ResNet



- 더 deep한 model은 training error가 더 낮아야 함
 - Shallow model의 parameter를 deeper model에 copy
 - Extra layers가 identity function의 역할만 해도 동일한 결과가 나올 수 있음
- Deep한 model은 optimization이 쉽지는 않다는 것을 발견(identity도 힘들다)

출처 :

<http://blog.naver.com/PostView.nhn?blogId=koys007&logNo=220898460208&parentCategoryNo=&categoryNo=62&viewDate=&isShowPopularPosts=false&from=postView>

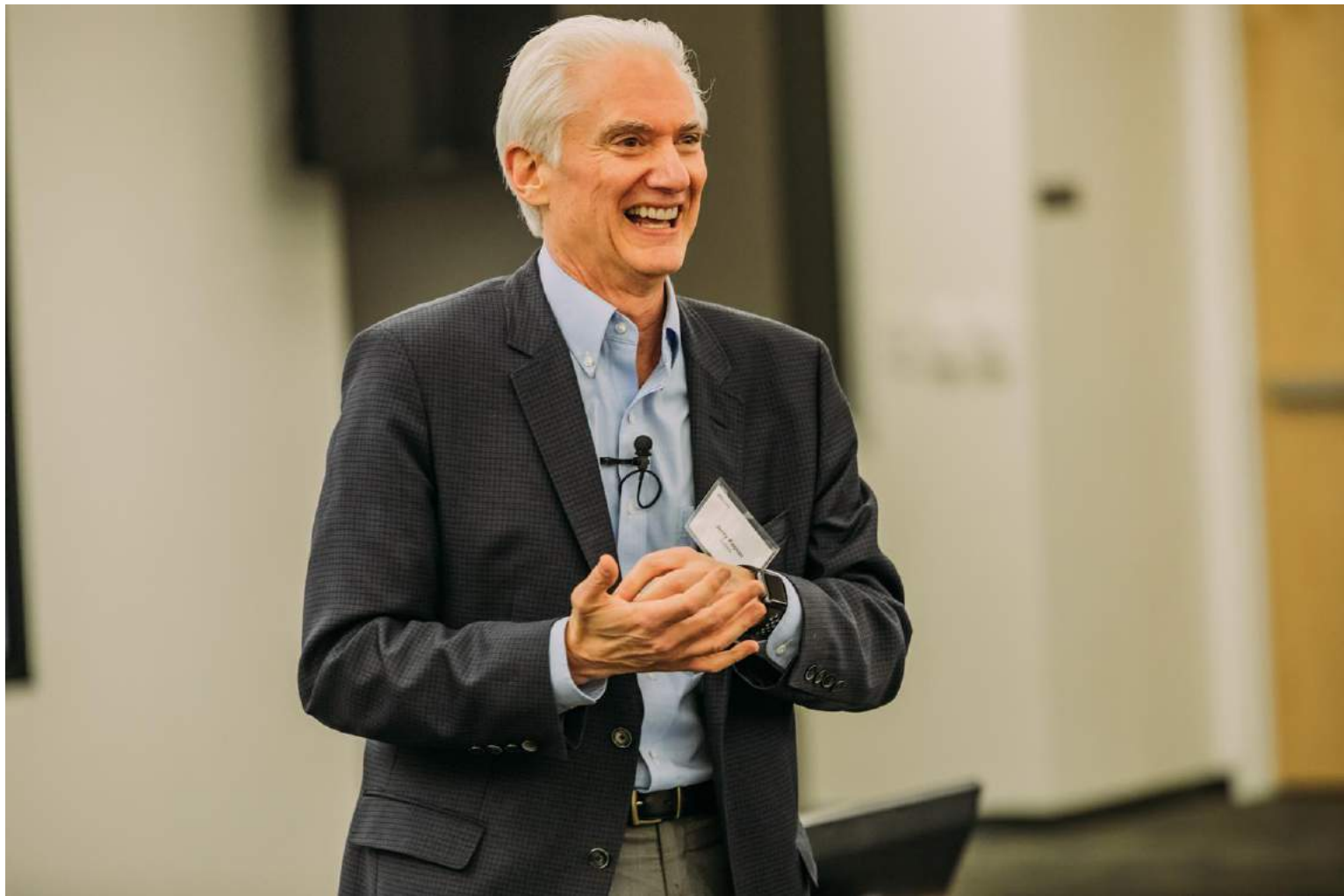
2012년 6월 스탠퍼드대 앤드루 응은 구글과 추진한 딥러닝 프로젝트인 구글 브레인(Google Brain)에서 컴퓨터가 스스로 고양이를 식별하도록 학습시키는 데 성공했다. 컴퓨터 프로세서 1만6000개와 10억개 이상의 신경망을 사용해 유튜브에 있는 1000만개 이상 동영상 중에서 고양이 사진을 골라낸 것이다.

Heroes of Deep Learning: Andrew Ng interviews Yann LeCun

<https://youtu.be/Svb1c6AkRzE>

2015년 5월 출시된 딥러닝 소프트웨어인 구글포토(Google Photos)는 스마트폰 속 수천 장의 사진을 자동으로 분류하는 응용프로그램(앱)이다. 스마트폰에 구글포토앱을 깔고, 사진을 모두 저장한 다음에 검색창에 가령 '손자'라고 쳐서 넣으면 손자 사진이 나타난다. 구글포토는 수백억 장의 사진을 학습해 사진 속 각 사물의 특징을 익히기 때문에 수억 장 사진을 몇 초 만에 판독해내는 능력을 갖게 되는 것으로 알려졌다.





Pierre Baldi

Deep Learning Applications in the Natural Sciences





기계학습의 태동

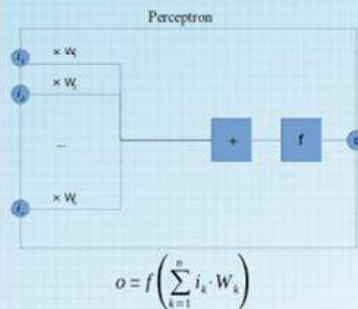
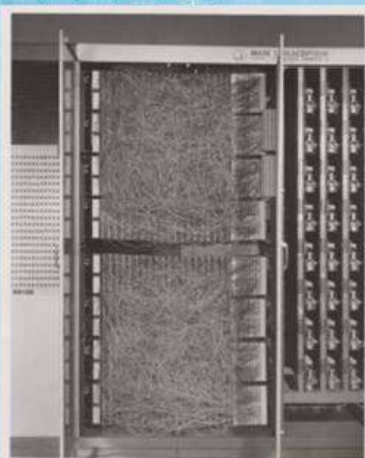
● Perceptron

◆ 프랭크 로젠블랫(Frank Rosenblatt)

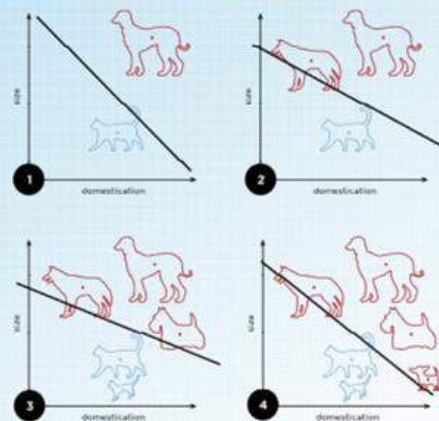
◆ 1957년 코넬대학 항공 연구실

• Mark I Perceptron

- IBM 704 컴퓨터에서 시뮬레이션 된 뒤
- 하드웨어로 구현됨



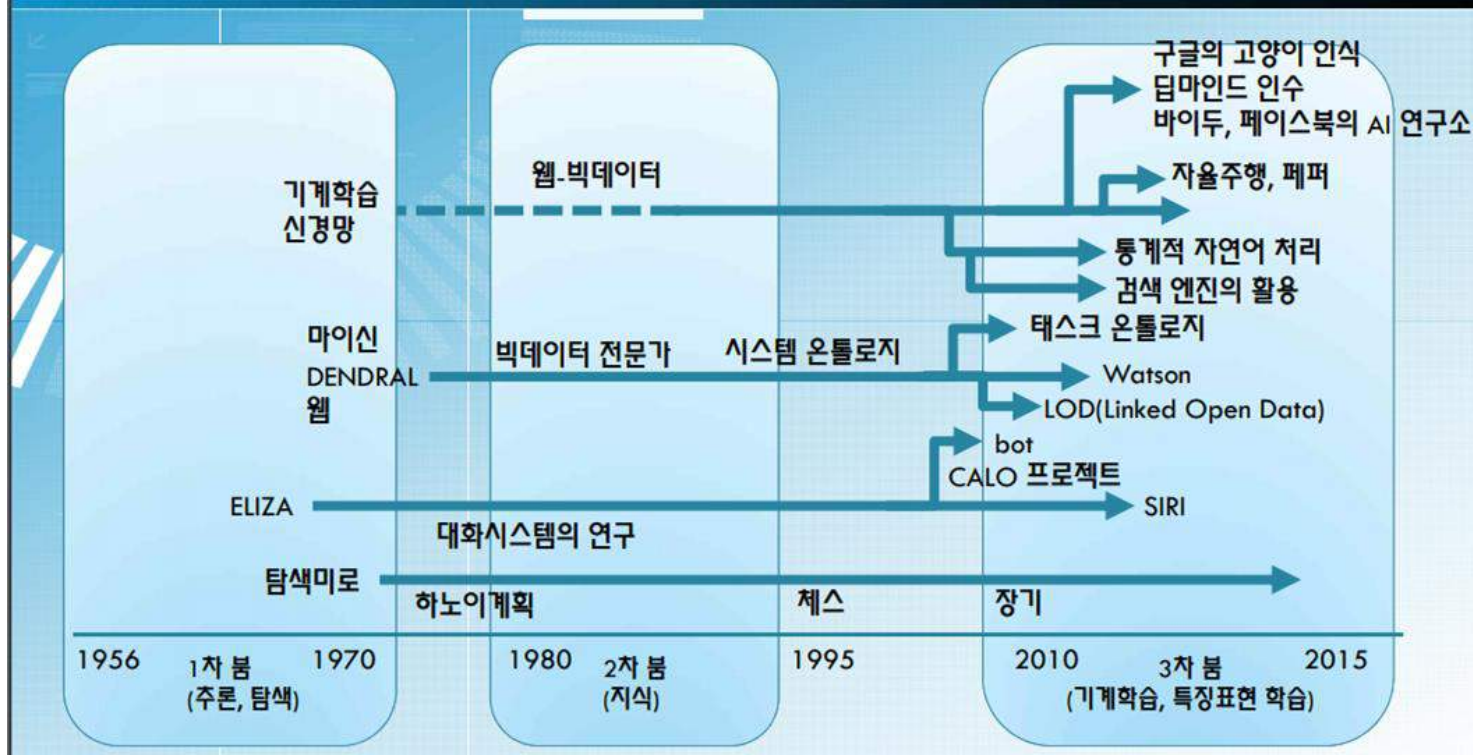
20x20개의 황화 카드용 광센서를 단 400픽셀 카메라에 연결된 Mark I Perceptron



현실의 문제를 풀지 못하는 딜레마 - 1차 봄의 끝

- 1960대에 꽃 핀 인공지능 연구
 - ◆ 낙관적 전망으로 활발한 연구
 - ◆ 컴퓨터는 지능을 얻게 될 것처럼 보였음
- 냉정한 관찰
 - ◆ 미로, 퍼즐, 장기와 바둑 = 명확히 정의된 룰 안에서 제한된 탐색 공간을 찾는 일
 - ◆ 현실의 문제 - 인공지능이 수행하기 힘든 모호한 판단
- 마빈 민스키(Marvin Minsky)의 Perceptron에 대한 비판
 - ◆ Perceptron은 본질적으로 선형 구분 → XOR 문제를 풀지 못 함
 - ◆ 세상은 쉽게 인공지능에 매료되었다가 쉽게 실망해 버렸음
 - 하나의 퍼셉트론이 XOR을 풀지 못 하는 일이 그렇게 대단한 문제라고 할 수 있을까?
 - 신경회로망은 매우 복잡한 연결로 다중의 층을 가진 다수의 네트워크
 - ◆ 1970년대 인공지능 1차 봄이 끝나고 정체기가 찾아 옴

인공지능 연구 평면도



봄과 겨울의 시대

AI 붐의 큰 파도

◆ 1차 붐

- 1950년대 후반에서 1960년대
 - 추론과 탐색을 통해 특정한 문제 해결
 - 1970년대 장남감 문제(toy problem)이 아닌 실제 문제 해결 역량 부족 드러남

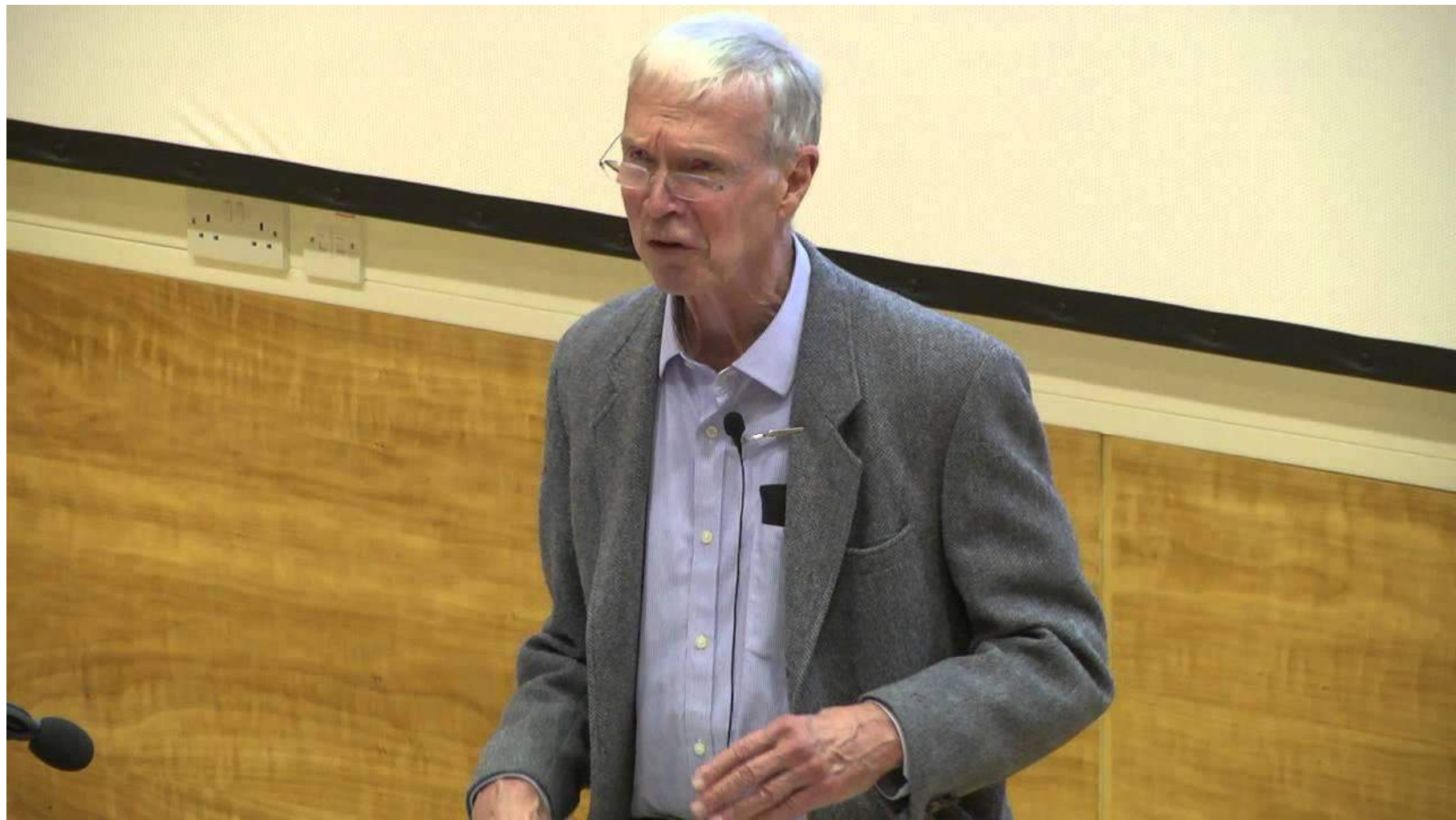
◆ 2차 붐

- 1980년대
 - 지식을 통해 지능을 구현하려는 시도. 전문가 시스템 등
 - 지능을 구현하기 위한 지식의 규모가 방대함이 드러나 1990년대 후반 다시 겨울

◆ 3차 붐

- 빅데이터 시대에 퍼진 기계학습과 딥러닝이 파도 형성
- 상징적 사건(Watson, 장기전왕전, AlphaGo)으로
- 커즈와일의 특이점 이론 등으로 대중의 관심 집중

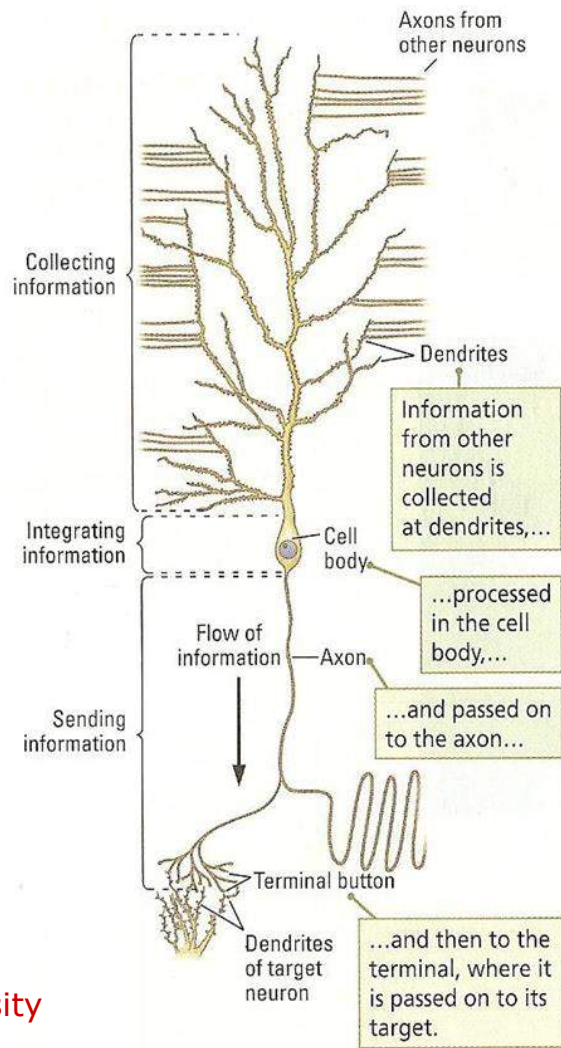




Deep learning

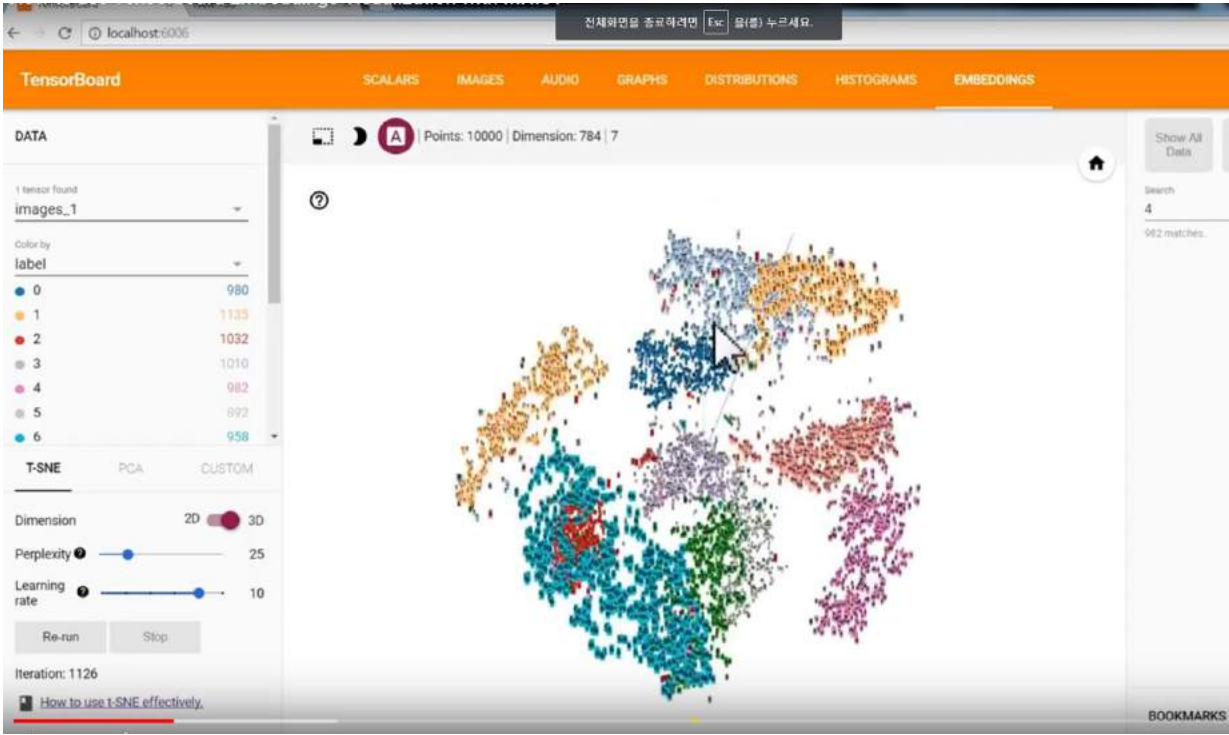
Andrew Ng





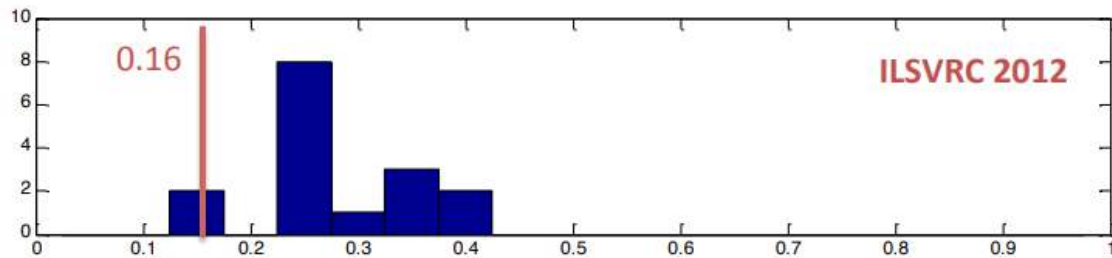
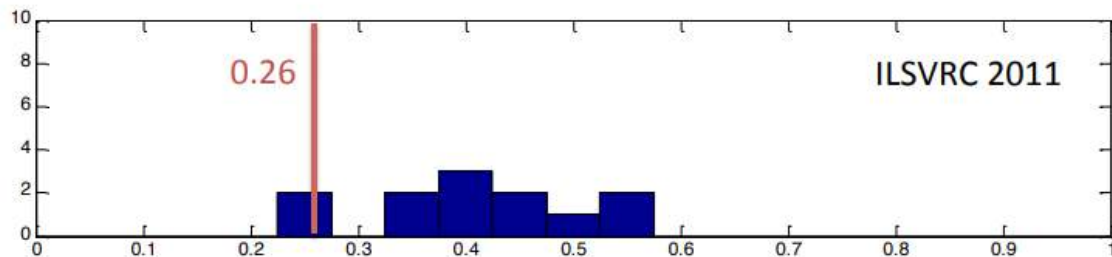
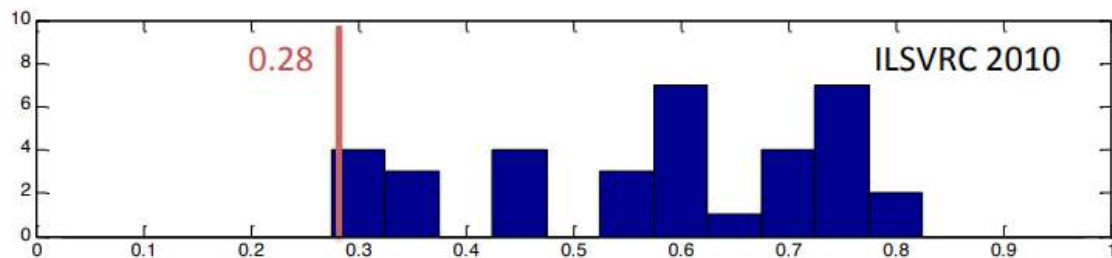
<http://www.holehouse.org/mlclass/>

How to use Tensorboard
Embeddings Visualization with
MNIST



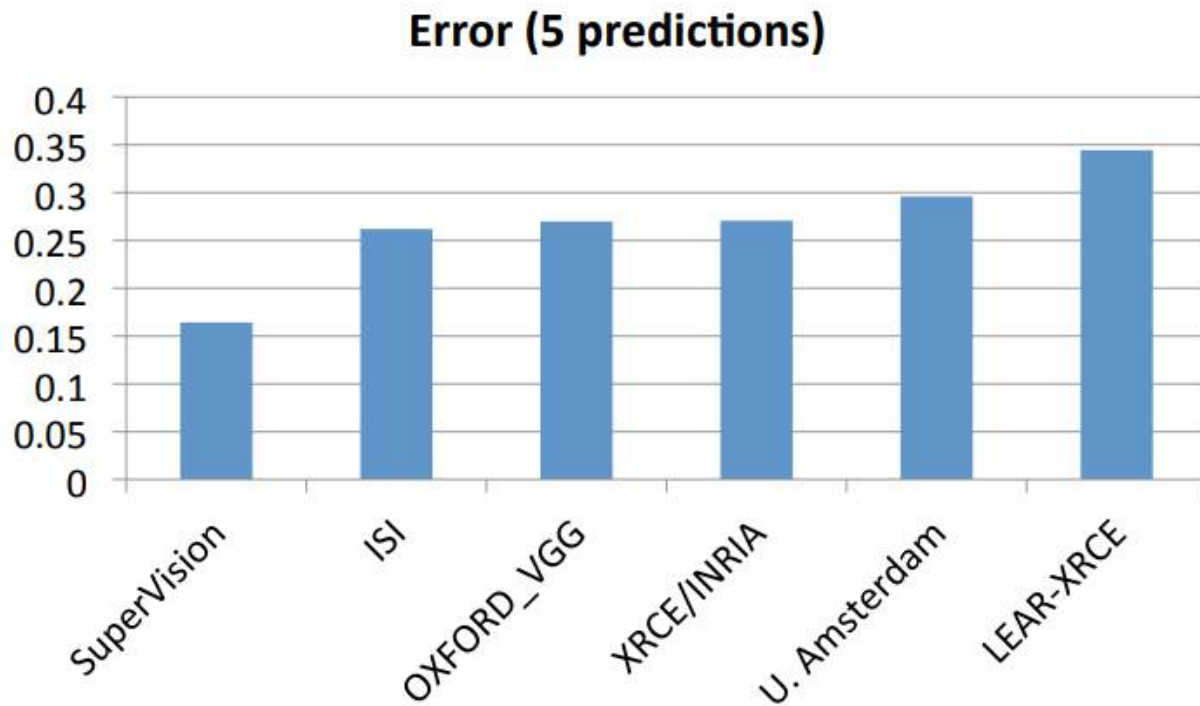
<https://distill.pub/2018/building-blocks/>

Submissions



Error (5 predictions/image)

Ranking of the best results from each team



The History of Artificial Intelligence

- The gestation of artificial intelligence (1943-1955)
 - **Artificial neurons** with on/off state (Warren McCulloch)
 - **Neural network computer** with 3000 vacuum tubes (Marvin Minsky and Dean Edmonds)
- The birth of artificial intelligence (1956)
 - Another influential figure in AI, **John McCarthy**
 - Princeton → Stanford → Dartmouth (official birthplace of AI)
 - Two-month workshop at Dartmouth with 10 attendees from Princeton, IBM, MIT and CMU.

The History of Artificial Intelligence

- Early enthusiasm, great expectation (1952-1969)
 - **General Problem Solver (GPS)** by Newell and Simon
 - The first program to embody the "thinking humanly" approach
 - Arthur Samuel (IBM) wrote a **checkers program** that eventually learned to play at a strong amateur level (1959)
 - John McCarthy at MIT
 - High-level language, **Lisp**, which was to become the dominant AI programming language
 - **Advice Taker** (1958)
 - Uses **general knowledge** of the world to search for solutions to problems
 - **Accepts new axioms** in the normal course of operation, thereby allowing it to achieve competence in new areas **without being reprogrammed**

The History of Artificial Intelligence

- Knowledge-based systems: The key to power? (1969-1979)
 - **DENDRAL** (1969): The first successful knowledge-intensive system
 - Larger reasoning with **domain-specific knowledge**
 - **Infers** molecular structure by means of molecule-related information and the mass spectrum
 - **Heuristic Programming Project (HPP)**
 - Investigates the extent to which the new methodology of **expert systems** could be applied to other areas of human expertise
 - **MYCIN** to diagnose blood infections
 - **No general theoretical model** from which the MYCIN rules could be deduced
 - The rules had to **reflect the uncertainty** associated with medical knowledge
- Understanding **natural language**
 - The importance of **domain knowledge** was apparent
 - The emphasis was more on the problems of **representing and reasoning** with the knowledge

The History of Artificial Intelligence

- The return of neural networks (1986-present)
 - Back-propagation learning algorithm (mid-1980s)
- AI adopts the scientific method (1987-present)
 - New trend started
 - Existing theories rather than brand-new one's proposal
 - Real-world applications rather than toy examples
 - Rigorous theorems or hard experimental evidence rather than intuition
 - Examples
 - HMM for speech recognition (a robust mathematical framework)
 - Data mining with neural networks
 - Bayesian network formalism for an efficient representation of, and rigorous reasoning with uncertain knowledge
 - Normative expert systems which act rationally according to the laws of decision theory

The History of Artificial Intelligence

- The emergence of intelligence agents (1995-present)
- The availability of very large data sets (2001-present)
- The state of the art
 - Robotics vehicles
 - Speech recognition
 - Autonomous planning and scheduling
 - Spam fighting
 - Logistics planning
 - Robotics
 - Machine translation

1956년 - 열린 워크숍에 참석한 존 매카시, 마빈 민스키, 앨런 뉴웰, 허버트 사이먼이라는 4명의 과학자가 컴퓨터에 관한 최신 연구 성과를 발표하며 인공지능을 거론했다.

이 중 뉴웰과 사이먼은 세계 최초로 인공지능 프로그램인 ‘로직 세오리스트’를 발표했다.

이 프로그램은 논리연산을 자동으로 증명해주는 것이다.

당시 인공지능 연구의 핵심 개념은 ‘추론’과 ‘탐색’이었다.

추론은 인간의 사고 과정을 기호로 표현하는 것이었다.

탐색은 컴퓨터가 미로에서 길을 찾을 때 경우의 수를 모두 따져 올바른 방향으로 가는 해답을 찾아내는 방법이다.

1956년 존 매카시, 마빈 민스키, 나다니엘 로체스터, 클로드 샤논 같은 당시 최고의 정보과학자들이 다트머스대학에 모여 창시했다.

1970년 - 지식을 활용한 인공지능 시스템은 1970년대 초에 미국 스탠퍼드대가 개발한 ‘마이신’을 시작으로 연구가 본격화했다. 마이신은 전염성 혈액 질환의 환자를 진단하고 항생 물질을 자동으로 처방하는 프로그램으로 감염병 전문 의사 대신 진단을 내리는 것이다. 과학자들은 컴퓨터에 학습시키는 과정에서의 한계도 깨달았다. ‘마이신’의 경우 ‘배가 아프다’거나 ‘위장이 쓰린’ 증상을 알아차리지 못했다.

인간의 경험이 집약된 상식적인 지식을 컴퓨터가 직관적으로 파악하도록 만들기 어려웠기 때문이다.

이같은 문제를 해결하기 위해 ‘시맨틱 네트워크’라는 개념이 등장했다.

이 개념은 인간이 의미를 기억할 때 구조를 나타내는 모델로, ‘개념’을 ‘노드(일종의 점)’로 표시하고 노드끼리 링크로 연결해 네트워크로 만드는 것이다.

1958년 프랭크 로젠블라트의 퍼셉트론이 등장하면서 주목을 받았다. 퍼셉트론은 학습과 의사결정을 하며, 언어를 번역할 수 있을 것이라고 기대를 모았다.

1969년 마빈 민스키와 세이무어 페퍼트가 펴낸 '퍼셉트론'이라는 책에서 퍼셉트론의 한계가 비판받으면서 이후 10년 간 이런 방식의 연결주의적 연구는 중단됐다

1980년 - 1980년 쿠니히코 후쿠시마의 네오코그니트론

1982년 -

2004년 - 2004년 캐나다 첨단 연구소(CIFAR)에서 50만 달러 정도의 적은 펀딩을 제공하자 제프리 힌튼 교수는 뉴럴 컴퓨테이션과 어댑티브 퍼셉션(NCAP) 프로그램을 만들어 컴퓨터 과학자, 생물학자, 전기공학자, 뇌과학자, 물리학자, 심리학자를 초대했다.

2006년 - 제프리 힌튼 딥러닝에 관한 기념비적인 논문인 'A fast learning algorithm for deep belief nets'를 발표한다.

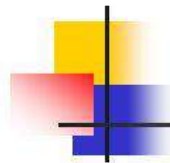
2011년 NCAP 연구자 중 하나인 스탠포드대 교수 앤드류 응은 구글 안에 딥 러닝 프로젝트를 구성했고, 음성인식과 구글 플러스의 사진 태깅에 딥러닝 기술을 활용하기 시작했다. 이 연구 그룹은 2012년 1만 6000개의 컴퓨터 프

Games as a Benchmark for AI

- Alan Turing proposed Chess as an AI benchmark.



Arthur Samuel in 1957, playing checkers against a mainframe computer.



The History of Development of Neural Networks



- The Beginning of Neural Networks (1940's)
 - McCulloch Pitts Neuron
 - Hebb Learning
- The First Golden Age of Neural Networks (1950's and 1960's)
 - Perceptrons
 - Adaline
- The Quiet Years: 1970's
 - Kohonen
 - Anderson
 - Grossberg
 - Carpenter
- Renewed Enthusiasm: 1980's
 - Backpropagation
 - Hopfield nets
 - Neocognitron
 - Boltzman machine
- Hardware Implementation



The CyC Project (Lenat 1984--)

Coding all necessary knowledge in a specialized representation

The project started in 1984 with the goal:

Assemble a comprehensive ontology and database of everyday common sense knowledge to enable AI applications to perform human-like reasoning

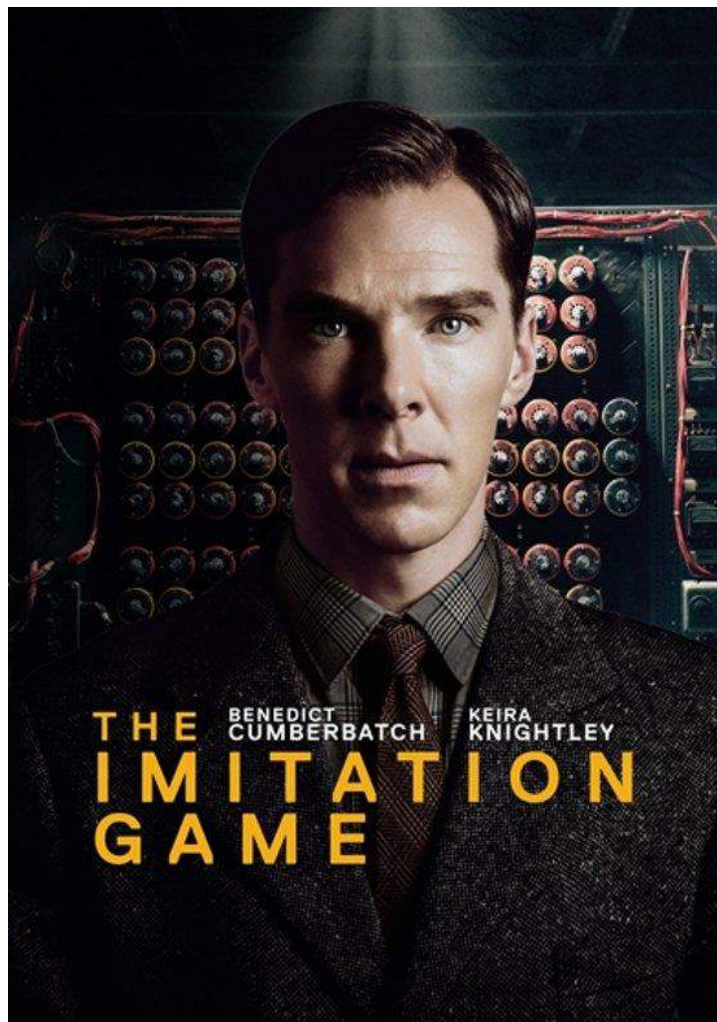
Currently the knowledge base consists of

- 3.2 million assertions (facts and rules)
- 280,000 concepts
- 12,000 concept-interrelating predicates

A smaller version of CyC was released under OpenCyc

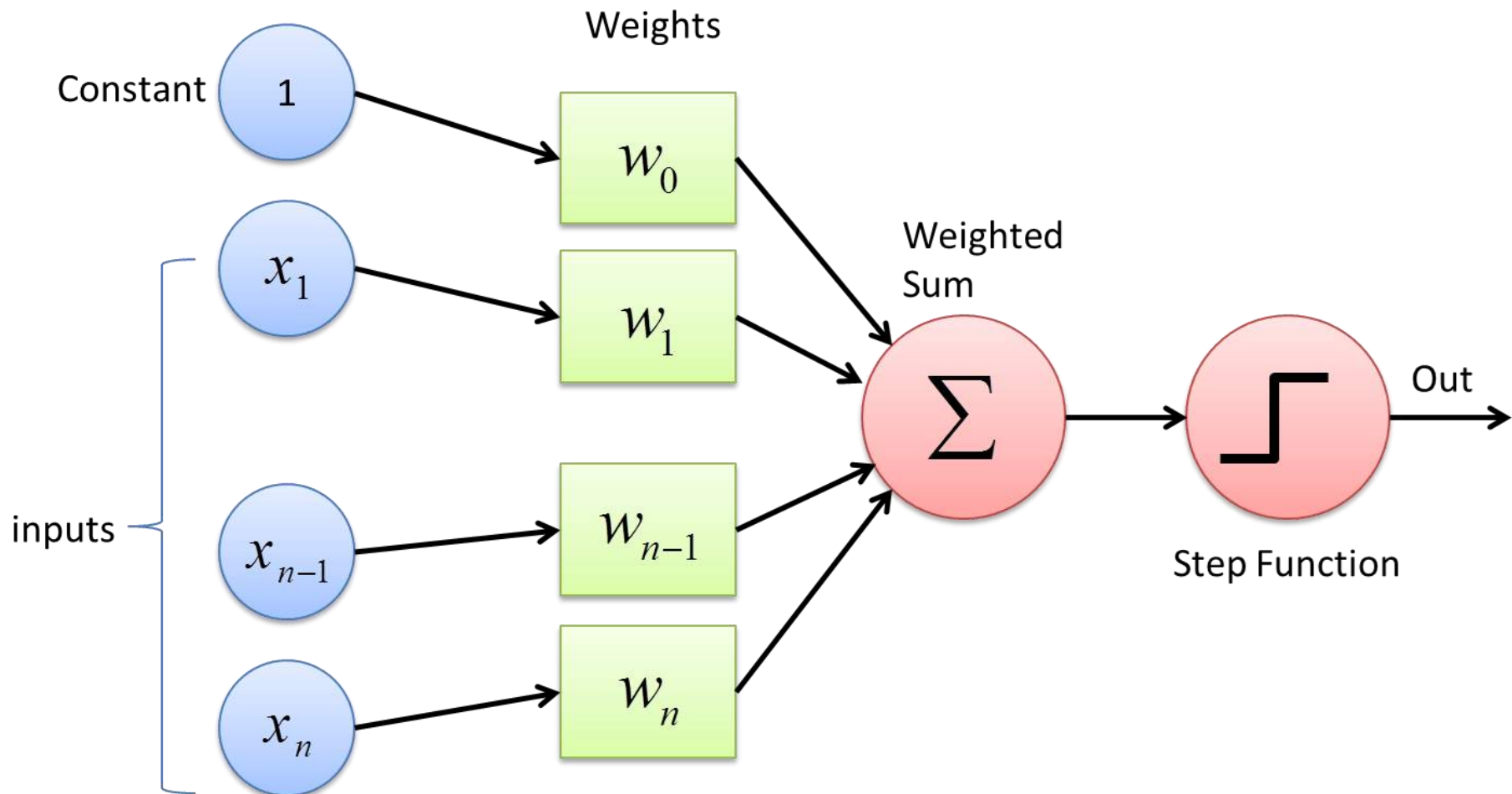
Success of CyC is still open. All knowledge is put into the machine, it does not yet have the ability to acquire knowledge by reading text.

I propose to consider the question: “Can machines think?” This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’... But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the question can be described in terms of a game which we call the **imitation game**.”



BENEDICT CUMBERBATCH KEIRA KNIGHTLEY

THE IMITATION GAME



```

=====
      EEEEEEEE L      IIIIIII ZZZZZZZ      AAA
      E        L        I        Z        A    A
      E        L        I        Z        A    A
      EEEEE    L        I        Z        A    A
      E        L        I        Z        AAAAAAA
      E        L        I        Z        A    A
      EEEEEEEE LLLLLLLL IIIIIII ZZZZZZ      A    A
=====

ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE... !
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
=====

```

사악한 초지능의 등장을 현 시점에서 걱정하는 것은
화성의 인구과잉상태를 우려하는 것과 같다

- Andrew Ng 오은달 박사 -

최윤섭의 헬스케어 이노베이션

<http://www.yoonsupchoi.com/2017/08/08/ai-medicine-4/>

<https://youtu.be/Qil4kmvm2Sw>

Deep Neural Network

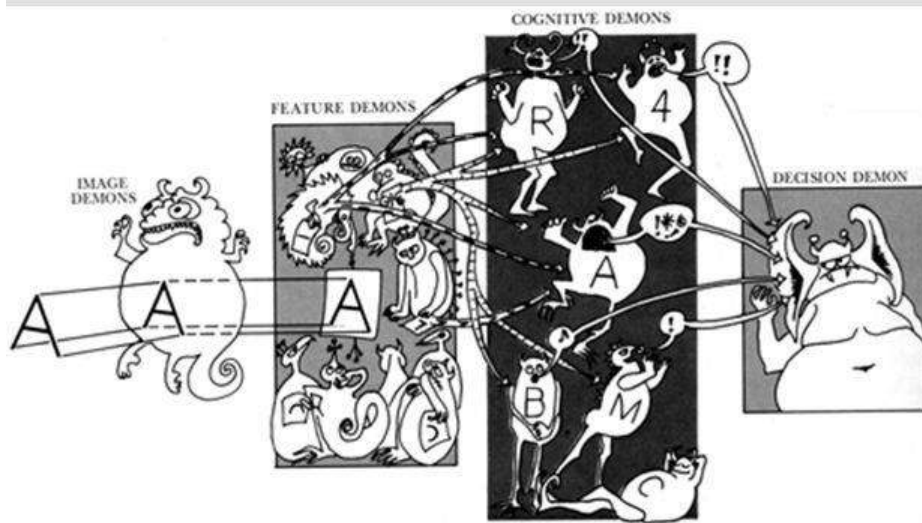
Deep learning이라는 것은 사실 deep neural network를 의미하는 것이다. Neural network에 대해서는 [이전 글](#)에서 설명하였고, 그럼 deep이란 무엇인가 하면, 다르게 아니라 feed-forward network에서 가운데 hidden layer가 1개 보다 많으면 'deep'하다고 말하는 것이다. 요즘은 layer를 무조건 1개보다는 많이 쌓기 때문에 요즘 나오는 neural network 연구는 모두 deep learning 연구라고 생각하면 된다.

그런데 사실 deep learning은 전혀 새로운 연구분야가 아니고 이미 몇 십년 전에 기본적인 연구가 끝난 분야이다.

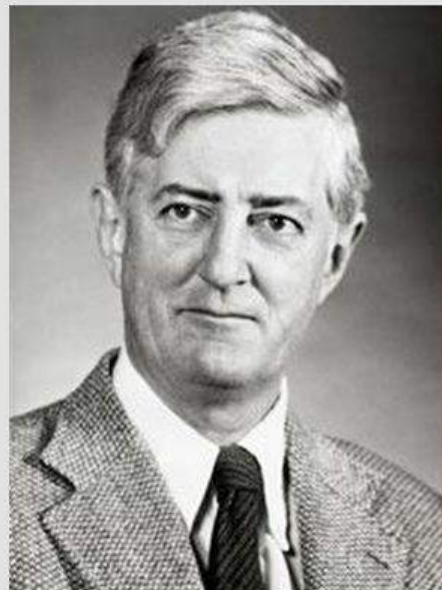
- 1958 Rosenblatt proposed perceptrons
- 1980 Neocognitron (Fukushima, 1980)
- 1982 Hopfield network, SOM (Kohonen, 1982), Neural PCA (Oja, 1982)
- 1985 Boltzmann machines (Ackley et al., 1985)
- 1986 Multilayer perceptrons and backpropagation (Rumelhart et al., 1986) 1988 RBF networks (Broomhead&Lowe, 1988)
- 1989 Autoencoders (Baldi&Hornik, 1989), Convolutional network (LeCun, 1989) 1992 Sigmoid belief network (Neal, 1992)
- 1993 Sparse coding (Field, 1993)

이렇듯 이미 가장 중요한 기초적인 연구는 예전에 다 끝났다. 지난 글에서 설명한 backpropagation 알고리즘은 이미 1986년 나온 알고리즘이고, 1989년에 나온 convolutional network가 요즘도 vision 분야에서 늘 사용하는 그 CNN이다. 그런데 정작 deep learning은 2000년도 중반이 지나고나서야 주목을 받기 시작했다. 왜 그랬을까?

Neurons: Pandemonium?



1958 idea: The mind is made up of “demons”, who shout based on other demons shouting to them. The loudest is heard...



Oliver G. Selfridge



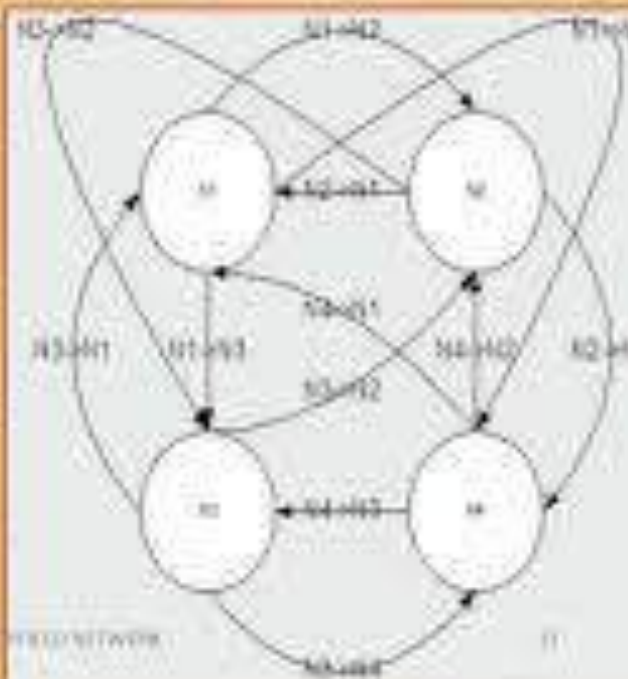
HOPFIELD NETWORK

EXAMPLE

Connection of Hopfield Neural Network

	Neuron 1 (N1)	Neuron 2 (N2)	Neuron 3 (N3)	Neuron 4 (N4)
Neuron 1 (N1)	(N/A)	N2 \rightarrow N1	N3 \rightarrow N1	N4 \rightarrow N1
Neuron 2 (N2)	N1 \rightarrow N2	(N/A)	N3 \rightarrow N2	N4 \rightarrow N2
Neuron 3 (N3)	N1 \rightarrow N3	N2 \rightarrow N3	(N/A)	N4 \rightarrow N3
Neuron 4 (N4)	N1 \rightarrow N4	N2 \rightarrow N4	N3 \rightarrow N4	(N/A)

A Hopfield Neural network:

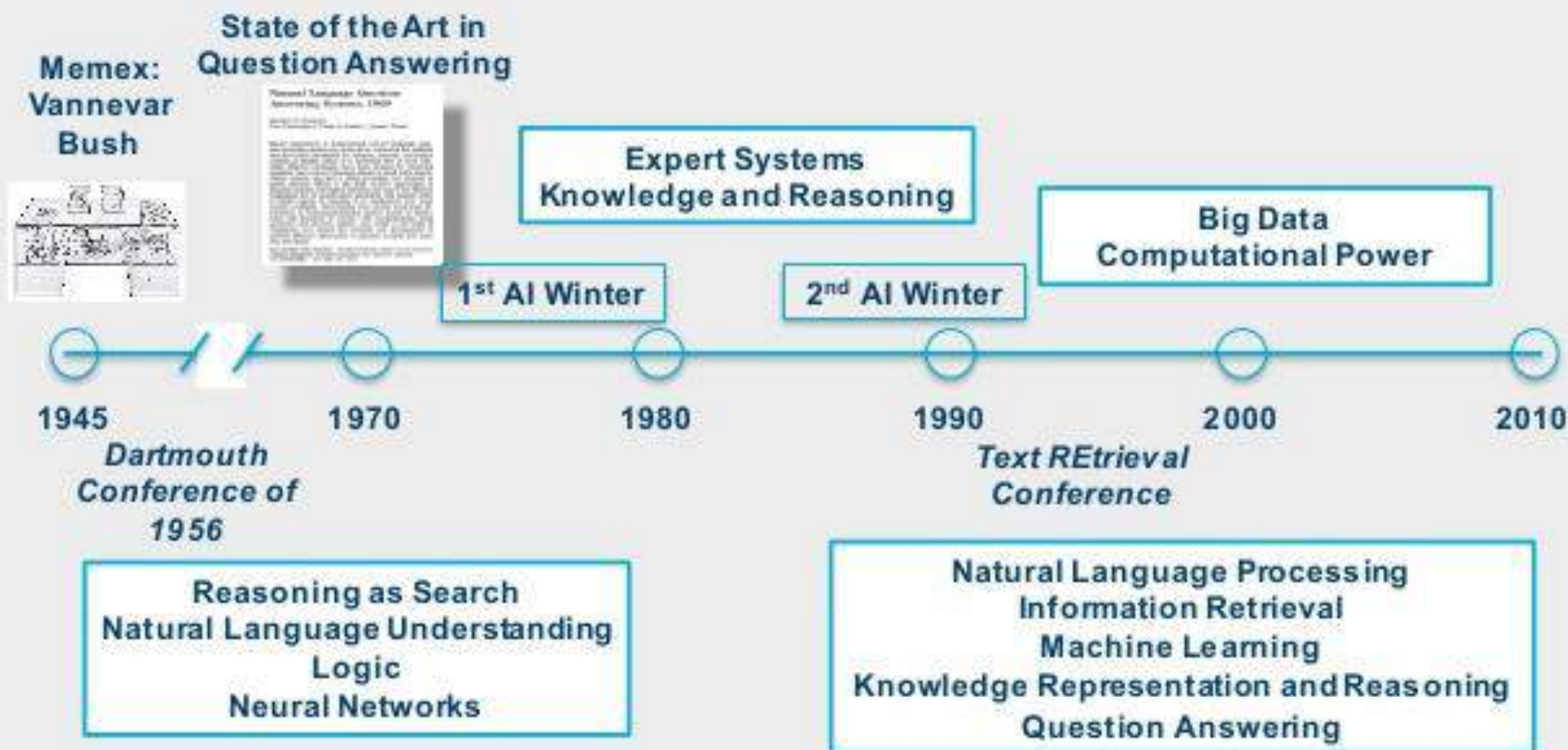


이 성공에 고무받아 스탠퍼드대학 의대생인 쇼틀리프는 이 대학 의사들의 도움을 받아서 항생제 선택에 관한 조언을 해주는 '마이신'(MYCIN)이란 의학용 전문가시스템을 개발해 낸다. 그 결과 의사의 오진율을 훨씬 밑도는 놀라운 결과를 보여주었다. 그는 더 나아가 의학용 전문가시스템에서 지식부분과 추론부분을 분리하여 추론기능만을 제공해주는 이마이신(EMYCIN)을 개발한다. 이마이신에 넣어주는 지식의 종류에 따라 어떤 경우에는 폐결핵 진단시스템이나, 암 진단시스템으로 바뀔 수가 있다는 것이다.

파이겐바움은 1936년 1월에 태어났으며, 출신 대학은 카네기 멜론 공과대학교이고, 1956년에 졸업과 동시에 경영공학대학원(GSIA : Graduate School of Industrial Administration)에 진학했다. 1960년, [Simon](#)의 지도하에 EPAM(Elementary Perceiver and Memorizer)으로 학위 논문을 취득했다. 그후, 캘리포니아 대학 버클리 분교의 비즈니스 스쿨의 조교수가 되었다. 그 당시 경영학은 보수적 성향이 매우 강해서 인간의 의사결정 과정을 시뮬레이션한다는 것은 교수들 입장에서는 상상도 할 수 없었다. 그러나 젊은 교수에 대해 카네기 사는 3년간 7만 달러라는 놀라운 연구비를 지원했다. 결국 다른 교수들의 눈총에 못이겨, 버클리 대학을 떠나야만 했고, 1965년 스탠포드 대학의 컴퓨터 사이언스 학과의 교수가 되었다.

파이겐바움이 버클리 대학시절에 올린 개가라고 하면, AI 관련 논문을 편집한 [컴퓨터와 사고(Computers and Thought)]를 1963년 McGraw-Hill사에서 출판한 일이다. 그 속에는 [Minsky](#)의 [인공지능 예의 길]등 대단히 유명한 논문들이 실려 있다. 스탠포드 대학으로 옮긴 파이겐바움은 스탠포드 대학의 컴퓨터 센터 소장이 되어 최초의 타임 셰어링 시스템을 도입했다. 파이겐바움의 이론은 인간의 경험이나 귀납을 기초로 하는 휴리스틱 어프로치(Heuristic Aproach)를 채택했다.

A Brief History of AI



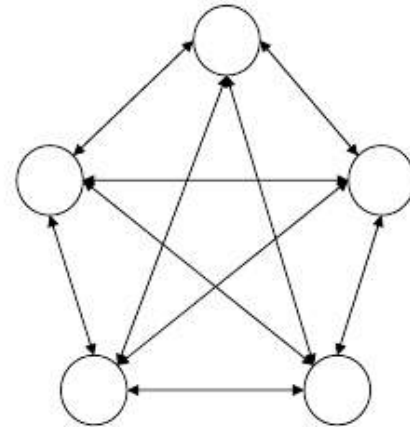
[표 1-1] AI와 지식 공학의 역사에서 주목할 만한 주요 사건

기간	주요 사건
AI의 탄생 (1943년~1956년)	McCulloch and Pitts, <i>A Logical Calculus of the Ideas Immanent in Nervous Activity</i> , 1943 Turing, <i>Computing Machinery and Intelligence</i> , 1950 The Electronic Numerical Integrator and Calculator project (von Neumann) Shannon, <i>Programming a Computer for Playing Chess</i> , 1950 The Dartmouth College summer workshop on machine intelligence, artificial neural nets and automata theory, 1956
AI의 융성 (1956년~1960년대 후반)	LISP (McCarthy) The General Problem Solver (GPR) project (Newell and Simon) Newell and Simon, <i>Human Problem Solving</i> , 1972 Minsky, <i>A Framework for Representing Knowledge</i> , 1975
AI에 관한 환멸 (1960년대 후반~1970년대 초반)	Cook, <i>The Complexity of Theorem Proving Procedures</i> , 1971 Karp, <i>Reducibility Among Combinatorial Problems</i> , 1972 The Lighthill Report, 1971
전문가 시스템의 발명 (1970년대 초반~1980년대 중반)	DENDRAL (Feigenbaum, Buchanan and Lederberg, Stanford University) MYCIN (Feigenbaum and Shortliffe, Stanford University) PROSPECTOR (Stanford Research Institute) PROLOG-a Logic Programming Language (Colmerauer, Roussel and Kowalski, France) EMYCIN (Stanford University) Waterman, <i>A Guide to Expert Systems</i> , 1986

기간	주요 사건
인공 신경망의 재탄생 (1965년~현재)	<p>Hopfield, <i>Neural Networks and Physical Systems with Emergent Collective Computational Abilities</i>, 1982</p> <p>Kohonen, <i>Self-Organized Formation of Topologically Correct Feature Maps</i>, 1982</p> <p>Rumelhart and McClelland, <i>Parallel Distributed Processing</i>, 1986</p> <p>The First IEEE International Conference on Neural Networks, 1987</p> <p>Haykin, <i>Neural Networks</i>, 1994</p> <p>Neural Network, MATLAB Application Toolbox (The MathWork, Inc.)</p>
진화 연산 (1970년대 초반~현재)	<p>Rechenberg, <i>Evolutionsstrategien-Optimierung Technischer Systeme Nach Prinzipien der Biologischen Information</i>, 1973</p> <p>Holland, <i>Adaptation in Natural and Artificial Systems</i>, 1975</p> <p>Koza, <i>Genetic Programming: On the Programming of the Computers by Means of Natural Selection</i>, 1992</p> <p>Schwefel, <i>Evolution and Optimum Seeking</i>, 1995</p> <p>Fogel, <i>Evolutionary Computation-Towards a New Philosophy of Machine Intelligence</i>, 1995</p>
단어 계산 (1980년대 후반~현재)	<p>Zadeh, <i>Fuzzy Sets</i>, 1965</p> <p>Zadeh, <i>Fuzzy Algorithms</i>, 1969</p> <p>Mamdani, <i>Application of Fuzzy Logic to Approximate Reasoning Using Linguistic Synthesis</i>, 1977</p> <p>Sugeno, <i>Fuzzy Theory</i>, 1983</p> <p>Japanese 'fuzzy' consumer products (dishwashers, washing machines, air conditioners, television sets, copiers)</p> <p>Sendai Subway System (Hitachi, Japan), 1986</p> <p>Negoita, <i>Expert Systems and Fuzzy Systems</i>, 1985</p>

Hopfield Network

- Recurrent network
 - Feedback from output to input
- Fully connected
 - Every neuron connected to every other neuron





In deep learning, the algorithms we use now are versions of the algorithms we were developing in the 1980s, the 1990s. People were very optimistic about them, but it turns out they didn't work too well.

— *Geoffrey Hinton* —

AZ QUOTES

배경 지식[편집]

인간의 [중추 신경계](#)에 대한 조사는 신경망 개념에 영감을 주었다. 생물학적 신경망을 흉내내는 네트워크를 형성하기 위해 인공지능망에서 [인공 뉴런들](#)은 서로 연결 되어있다. 인공지능망이란 무엇인가에 대한 하나의 공식적인 정의는 없다. 그러나 만약 통계학적 모델들의 집합이 다음과 같은 특징들을 가진다면 해당 집합을 신경(neural)이라고 부른다.

1. 조절이 가능한 가중치들의 집합 즉, 학습 [알고리즘](#)에 의해 조절이 가능한 숫자로 표현된 매개변수로 구성되어있다.
2. 입력의 비선형 함수를 유추할 수 있다.

조절가능한 가중치들은 뉴런 사이의 연결 강도를 의미하고 이들은 훈련 또는 예측하는 동안에 작동한다.

다양한 유닛들이 할당된 하위작업 보다 유닛들에 의한 병렬 혹은 집합적으로 함수들을 수행한다는 점에서 신경망은 생물학적 신경망과 닮았다. '신경망'이라는 단어는 보통 [통계학](#), [인지 심리학](#) 그리고 [인공지능](#)에서 사용되는 모델들을 가리킨다. 중추 신경을 모방하는 신경망 모델들은 [이론 신경과학](#)과 [계산 신경과학](#)의 한 부분이다.

인공신경망을 구현한 현대의 소프트웨어에서는 생물학적 접근법은 신호처리와 통계학에 근거한 좀 더 현실적인 접근법들로 인해 주로 사용되지 않는다. 이러한 시스템들 중 몇몇에서는 신경망 또는 신경망의 부분들(인공 신경들)은 큰 시스템을 형성하며 이러한 시스템은 조절이 가능하거나 기능하지 않은 구성 요소들로 결합되어 있다. 이러한 시스템의 일반적인 접근법은 많은 현실 문제 해결에 적합한 반면에 전통적인 인공지능 연결 모델에서는 그렇지 않다. 그러나 이들에게도 공통점이 있는데 그것은 비선형의 원리 분산, 병렬과 지역 처리 그리고 적응이다. 역사적으로 신경 모델들의 이용은 18세기 후반 if-then 규칙으로 표현된 지능을 가진 [전문가 시스템](#)을 특징으로 하는 고차원(symbolic) 인공지능에서 부터 [동적 시스템](#)의 매개변수들을 가진 지능을 특징으로 하는 저차원(sub-symbolic) [기계학습](#)으로 가는 패러다임의 변환이다.

역사[편집]

워런 맥컬록(Warren McCulloch)와 월터 피츠(Walter Pitts)는^[1] (1943) 수학과 임계 논리(threshold logic)라 불리는 알고리즘을 바탕으로 신경망을 위한 계산학 모델을 만들었다. 이 모델은 신경망 연구의 두 가지 다른 접근법에 대한 초석을 닦았다. 하나의 접근법은 뇌의 신경학적 처리에 집중하는 것이고 다른 하나는 인공 신경망의 활용에 집중하는 것이다.

1940년 후반에 심리학자 도널드 헤비안(Donald Hebb)는^[2] **헤비안 학습**(Hebbian learning)이라 불리는 신경가소성의 원리에 근거한 학습의 기본 가정을 만들었다. 헤비안 학습은 전형적인 **자율학습**으로 이것의 변형들은 **장기강화**(long term potentiation)의 초기 모델이 된다. 이러한 아이디어는 1948년 **튜링의 B-type 기계**에 계산학 모델을 적용하는데서 출발하였다.

팔리(Farley)와 웨슬리 클라크(Wesley A. Clark)는^[3](1954) MIT에서 헤비안 네트워크를 모의 실험하기 위해 처음으로 계산학 모델(후에 계산기라 불리는)을 사용하였다. 다른 신경망 계산학 기계들은 로체스터(Rochester), 홀랜드(Holland), 하빗(Habit), 두다(Duda)에 의해 만들어졌다.^[4] (1956)

프랑크 로젠블라트(Frank Rosenblatt)는^[5] (1958) **퍼셉트론** 즉, 간단한 덧셈과 뺄셈을 하는 이층구조의 학습 컴퓨터 망에 근거한 패턴 인식을 위한 알고리즘을 만들었다. 계산학 표기법과 함께 로벤블라트는 또한 기본적인 퍼셉트론에 대한 회로가 아닌예를 들면 **배타적 논리합** 회로(exclusive-or circuit)와 같은 회로를 표기하였다. 해당 회로의 수학 계산은 폴 웨어보스(Paul Werbos)에 의해 **오차역전파법**이 만들어진 후에 가능하였다.^[6] (1975)

마빈 민스키(Marvin Minsky)와 **시모어 페퍼트**(Seymour Papert)에 의해 기계학습 논문이 발표된 후에^[7] (1969) 신경망 연구는 침체되었다. 그들은 인공신경망에서 두 가지 문제점을 찾아내었다. 첫 번째로는 단층 신경망은 배타적 논리합 회로를 처리하지 못한다는 것이다. 두 번째 중요한 문제는 거대한 신경망에 의해 처리되는 긴 시간을 컴퓨터가 충분히 효과적으로 처리할 만큼 정교하지 않다는 것이다. 신경망 연구는 컴퓨터가 충분히 빨라지고, 배타적 논리합 문제를 효율적으로 처리하는 **오차역전파법**이 만들어지기까지 더디게 진행되었다.

1980년대 중반 병렬 분산 처리는 **연결주의**(connectionism)라는 이름으로 각광을 받았다. 데이비드 럼멜하트(David E. Rumelhart)와 제임스 맥클레랜드(James McClelland)가 쓴 교과서는^[8] (1986) **연결주의**를 이용해 신경 처리를 컴퓨터에서 모의 실험하기 위한 모든 것을 설명하였다.

인공신경망이 어느정도 뇌의 기능을 반영하는지 불분명하기 때문에 뇌 신경 처리의 간단한 모델과 뇌 생물학적 구조간의 상관관계에 대해 논란 중에 있으나 인공지능에서 사용되는 신경망은 전통적으로 뇌 신경 처리의 간단한 모델로 간주된다.^[9]

인공신경망은 **SVM**과 같은 다른 기계학습 방법들의 인기를 점차적으로 추월하고 있다. 2000년대 이후 **딥 러닝**의 출현이후 신경 집합의 새로운 관심은 다시 조명받고 있다.

2006년 이후 발전 동향[편집]

생물물리학 모의실험 그리고 뇌신경학 컴퓨팅을 위한 계산학 디바이스들은 [CMOS](#)를 통해 만들어졌다. 최근에는 큰 범위의 기본 요소들의 분석과 합성을 위한 [나노 디바이스](#) 제작과 같은 노력들이 있다.^[10] 만약 성공한다면 이러한 노력은 디지털 컴퓨팅을 뛰어넘는 [신경 컴퓨팅](#)의 새로운 시대를 도래하게 할 것이다.^[11] 왜냐하면 이것은 프로그래밍 보다는 학습에 기반하며 비록 첫 예시가 CMOS 디지털 디바이스와의 합작일지라도 이것은 기본적으로 [디지털](#)보다 [아날로그](#)이기 때문이다.

2009년부터 2012년동안 스위스 AI 연구실 IDSIA에서 위르겐 슈밋huber(Jürgen Schmidhuber)의 연구 그룹이 개발한 [재귀 신경망](#)과 심화 피드포워드 신경망은 여덟 번의 [패턴 인식](#)과 [기계학습](#) 국제 대회에서 우승하였다.^{[12][13]} 예를들면, 알렉스 그레이브스(Alex Graves et al)의 쌍방향 그리고 다중 차원의 [장단기 기억](#)(LSTM)은 ^{[14][15][16][17]} 2009년의 국제문서 분석 및 인식 컨퍼런스(ICDAR)의 필기 인식 부분에서 학습하게 될 세 가지 언어에 대한 아무런 사전 정보가 주어지지 않았음에도 불구하고 세 번이나 우승하였다.

IDSIA의 댄 크리슨(Dan Ciresan)과 그 동료들에 의한 빠른 [GPU](#) 기반 실행 접근 방법은 IJCNN 2011 교통 표지판 인식 대회,^{[18][18]} ISBI 2012 신경 구조의 분할의 전자 현미경 대회를 비롯하여 여러 패턴 인식 경연에서 여러 번 우승하였다.^[19] 그들의 신경망은 또한 처음으로 교통 표지판(IJCNN 2012) 또는 MINIST [필기 인식](#) 분야에서 인간과 견줄만한 또는 인간을 넘어서는 인공 [패턴 인식](#)이다.^[20]

심화 비선형 신경 아키텍처는 1980년 후쿠시마 구니히코(Kunihiko Fukushima)의 [신인식기](#)(neocognitron)와 비슷하다.^[21] 그리고 일차 [시각 피질](#)에서 데이비드 허블(David H. Hubel)과 토르스텐 비셀(Torsten Wiesel)에 의해 밝혀진 간단하고 복잡한 세포들에 영감을 받은 표준 비전 아키텍처는^[22] 토론토대학의 조프 힌턴(Geoff Hinton) 연구실의 자율 학습 방법에 의해 미리 훈련된다.^{[23][24]} 해당 연구실의 팀은 2012년 베르크(Berck)의 후원을 받는 신약 개발에 필요한 분자들을 찾는데 도움을 주는 소프트웨어 제작 대회에서 우승하였다.^[25]

모델[편집]

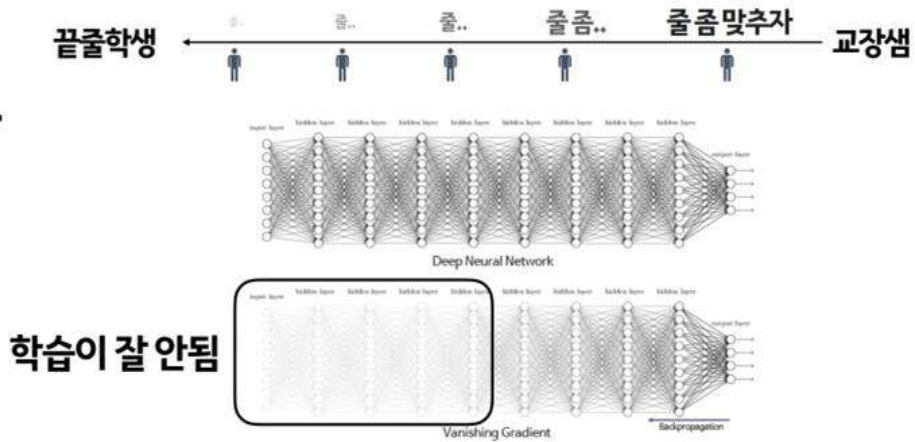


Vanishing Gradient

Back-prop에서 발생할 수 있는 문제

Layer가 많을수록 업데이트가 사라져간다
그래서 fitting이 잘 안 된다[underfitting]

Best choice: ReLU



출처: 하용호님 발표

BATCH Gradient Descent

전체 Data를 다 본 후 loss를 구해서
Weight를 업데이트 해주면서
어떤 loss 함수의 진짜 최솟값 찾기

이걸 m 번 반복

전체 데이터를 다 보고
1 걸음 간다 (update)



STOCHASTIC Gradient Descent

Data 1개 본 후 loss를 구해서
Weight를 업데이트 해주면서
어떤 loss 함수의 진짜 최솟값 찾기

반복 때마다 랜덤하게 섞는다

이걸 m번 반복



GPU !

MINI BATCH Gradient Descent

Data B개 본 후 평균 loss를 구해서
Weight를 업데이트 해주면서
어떤 loss 함수의 진짜 최솟값 찾기

반복 때마다 랜덤하게 섞는다

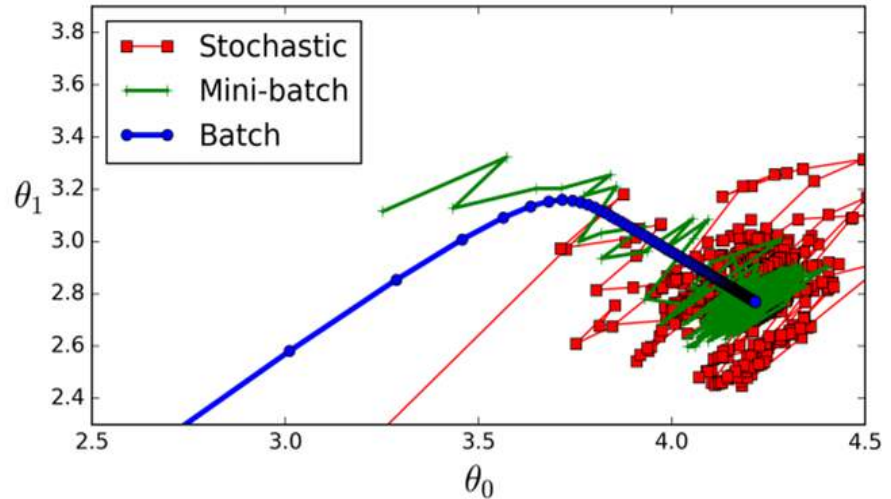
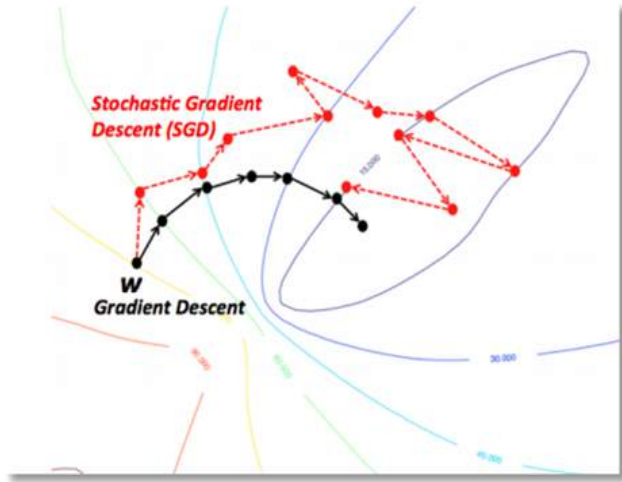
이걸 m번 반복



STOCHASTIC \approx MINI BATCH
Gradient Descent Gradient Descent

Visualizing Optimization

Different GDs



Gradient Descent

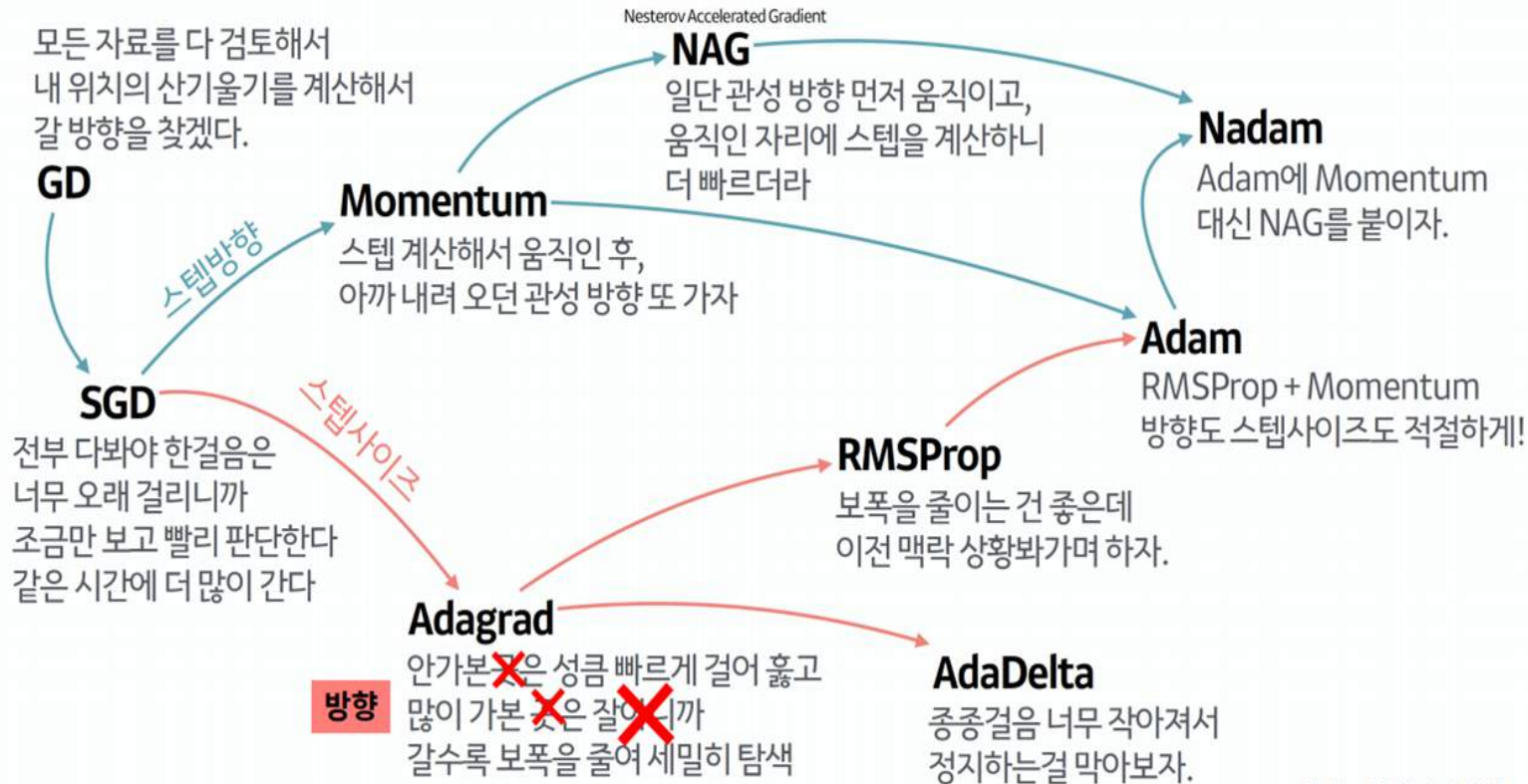
Weight를 업데이트 해주면서
어떤 loss 함수의 진짜 최솟값 찾기

1. Learning Rate 한 걸음의 크기

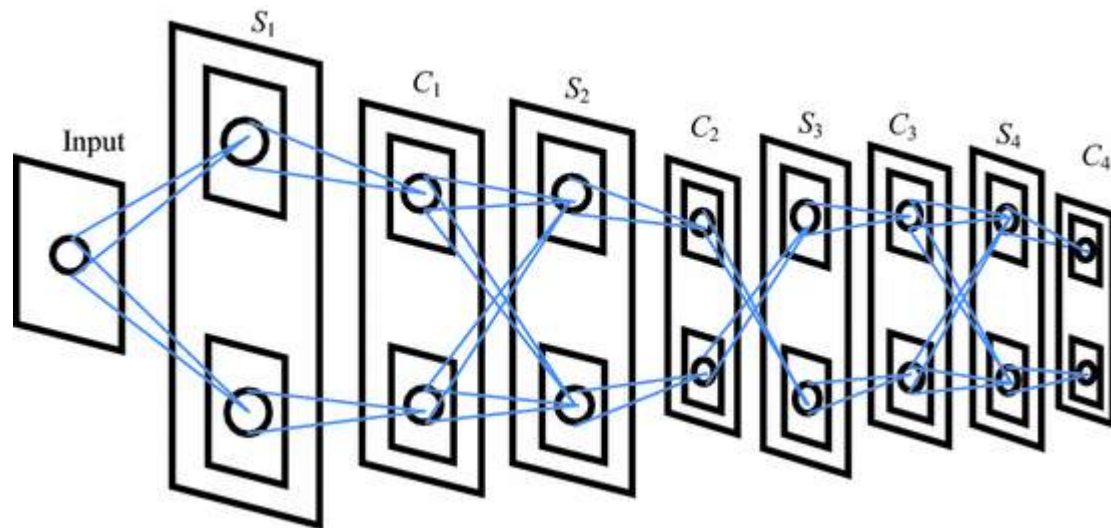
$$w_{updated} = w_{previous} - \gamma \nabla_w L(w)$$

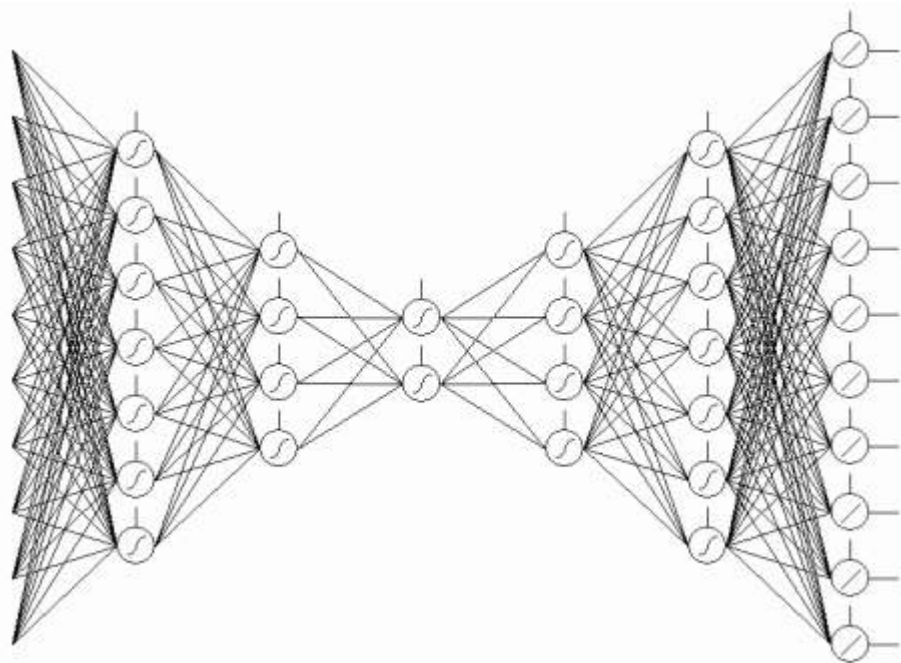
2. Gradient 어느 방향

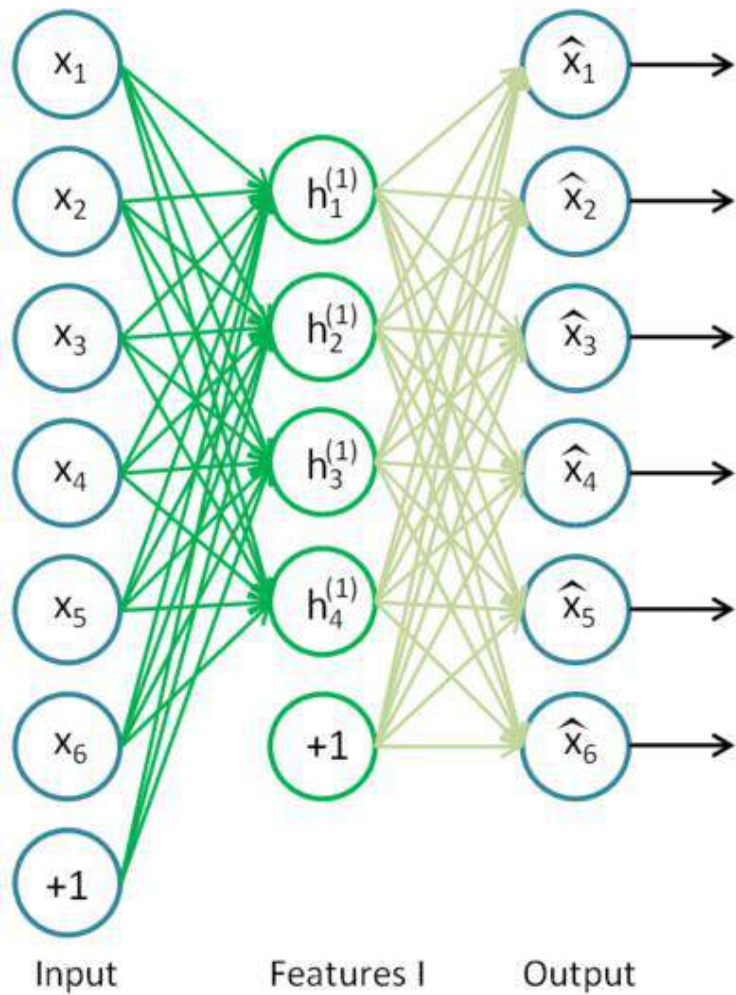
산 내려오는 작은 오솔길 찾기(Optimizer)의 발달 계보



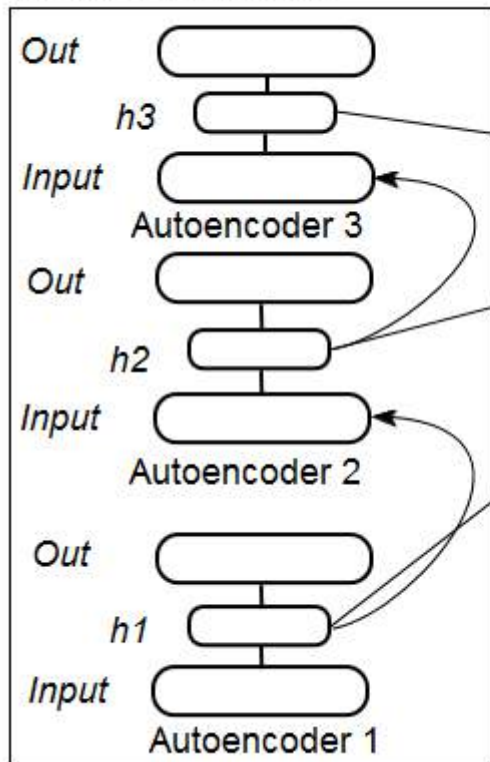
출처: 하용호님 발표



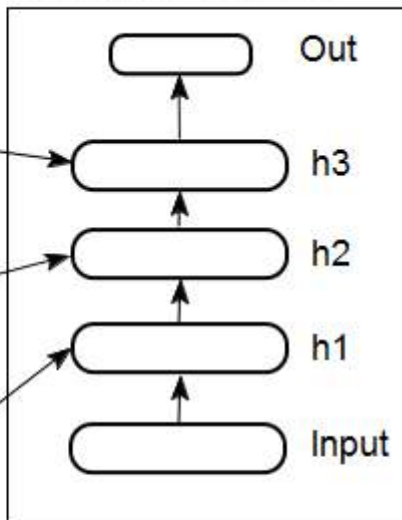




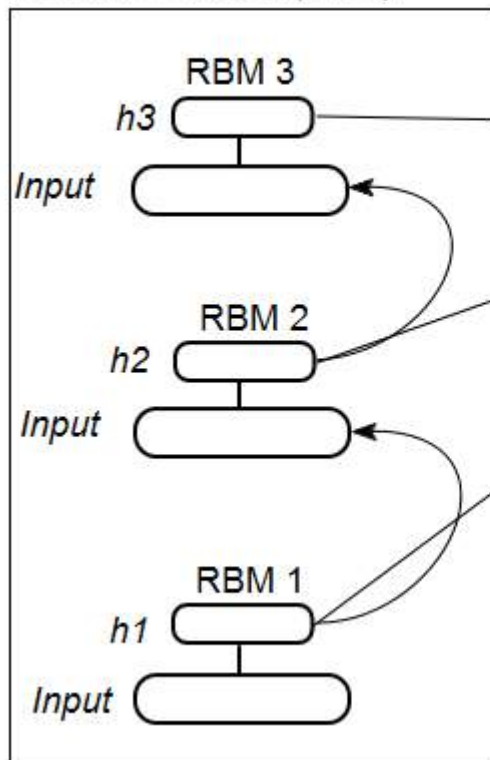
Stacked AutoEncoder



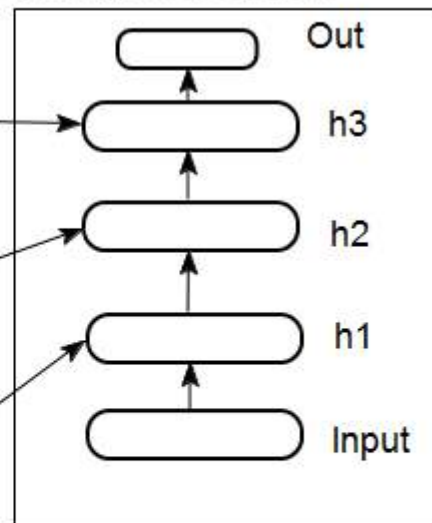
Multilayer Perceptron



Stacked Restricted
Boltzmann Machine (SRBM)



Multilayer Perceptron

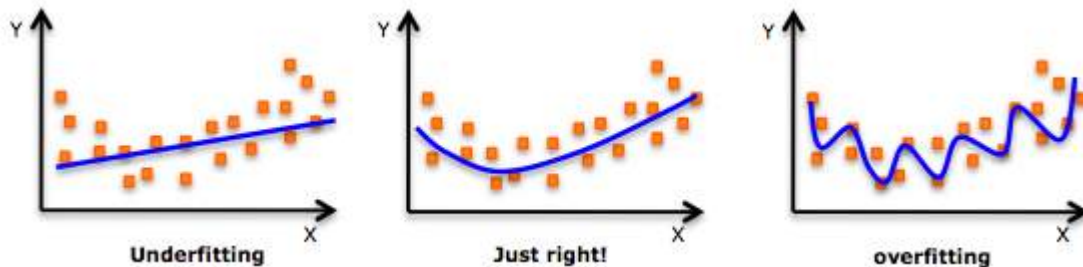


Problems to solve for deep learning

Deep learning이 흥하기까지 수 많은 연구결과들이 있었지만, 지금처럼 deep learning이 hot하게 되기까지는 앞에서 말했던 것처럼 regularization method들이나 initialization method들, 그리고 overfitting을 최대한 피할 수 있는 optimization method 등이 많이 제안되면서부터라고 할 수 있다. 이 연구들이 공통적으로 고민하고 있는 것은 [overfitting](#)이다.

Overfitting은 주어진 데이터의 양에 비해 모델의 complexity가 높으면 발생하게 된다. 안타깝게도 neural network가 깊어질수록 model의 complexity는 exponential하게 증가하게 된다. 그렇기 때문에 거의 무한한 표현형을 learning할 수 있는 deep network가 좋다는 것을 다들 알고 있음에도 불구하고, 너무나 overfitting이 심하게 발생하기 때문에 neural network 연구가 멈추게 된 것이다. 하지만 2007~8년 즈음하여 overfitting을 막기 위하여 새로운 initialization을 제안하는 work이 나오게 되는데 그 work이 바로 앞에서 설명 했던 NIPS에 발표되었던 두 work이다.

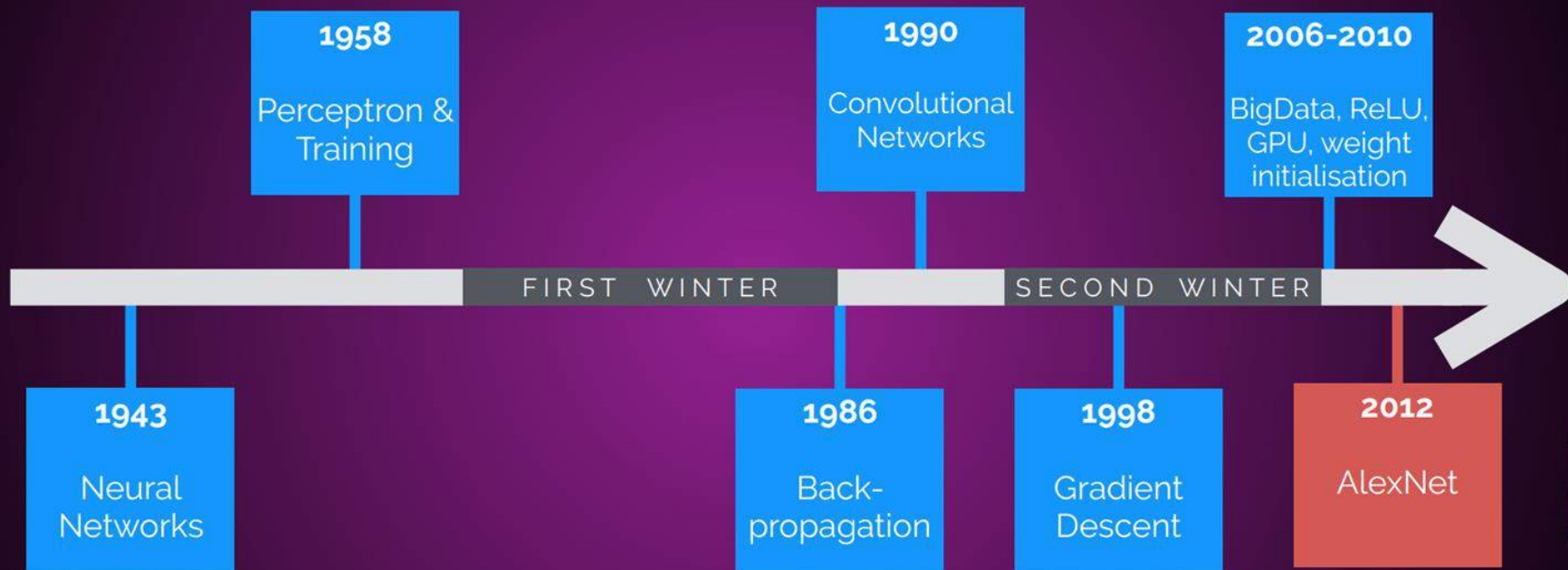
먼저 Restricted Boltzmann Machine (RBM) 에 대해 설명해보자.

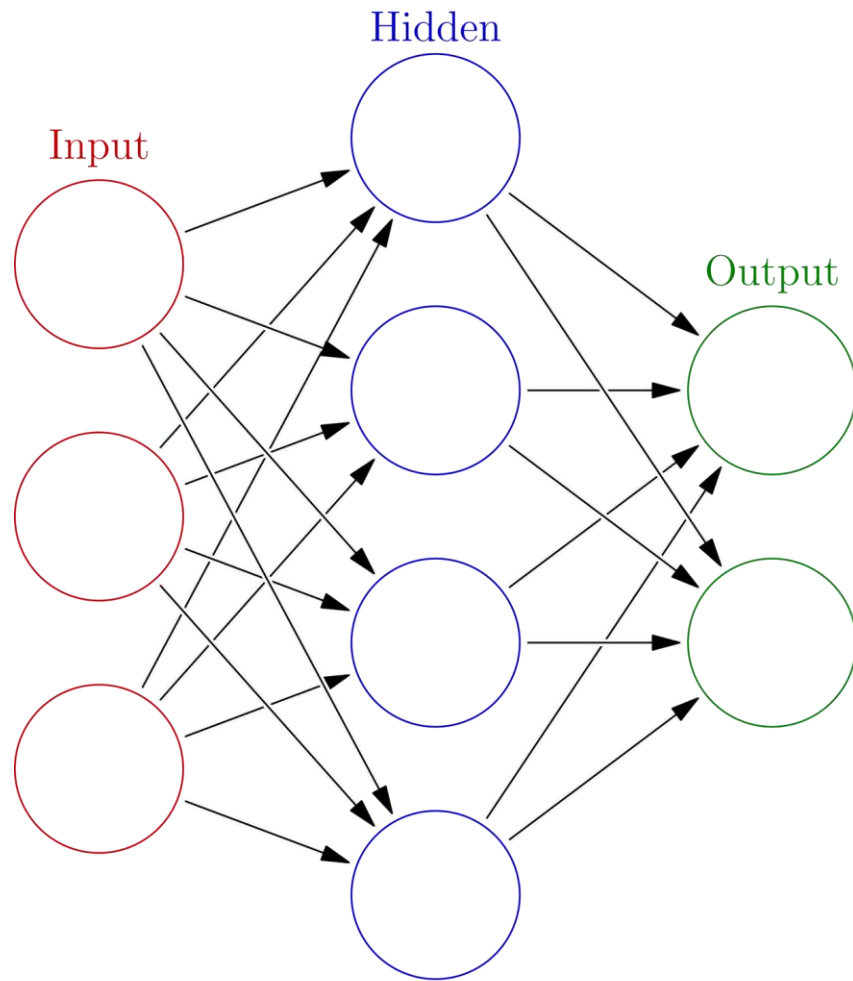






Deep Learning history







We know the past but cannot
control it. We control the future but
cannot know it.

— *Claude Shannon* —

AZ QUOTES

1959 - Bernard Widrow and Marcian Hoff of Stanford developed models called "ADALINE" and "MADALINE." - developed to recognize binary patterns and eliminate echoes on phone lines



1975 - Kohonen and Anderson developed the first multilayered network

1986 - Rumelhart, D. E., Hinton, G. E., and Williams, R. J. authored a seminal paper, which showed how backpropagation algorithm can be used to train neural networks with multiple layers quickly.





Machine Learning

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 .

~ Tom Mitchell



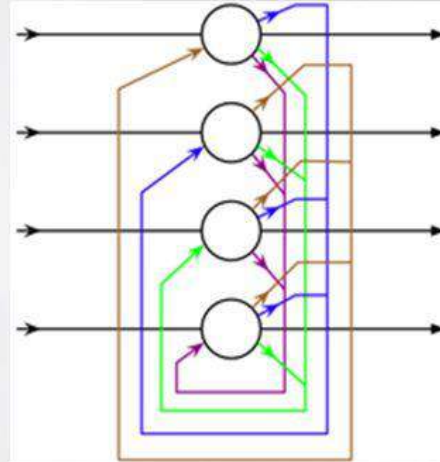


The Singularity

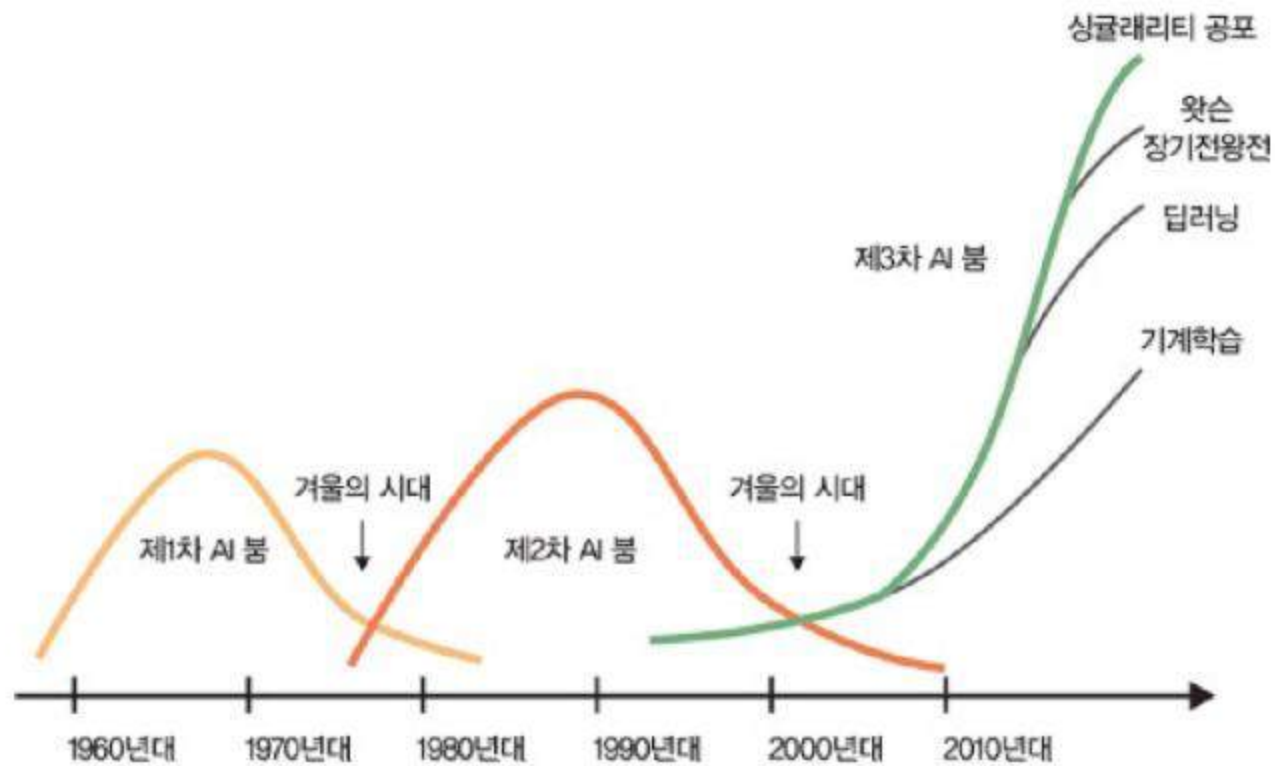
Irving John Good – 1960s • The ultraintelligent machine “A machine that can far surpass the intellectual activities of any man however clever. ... an ultraintelligent machine could design even better machines ... The intelligence of man would be left far behind.

Hopfield network (HN) model

- A Hopfield network is a form of **recurrent** artificial neural network invented by John Hopfield in 1982
- The network is entirely **interconnected**
 - All neurons are both input and output
- with binary threshold units
$$F(x) = \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{otherwise} \end{cases}$$



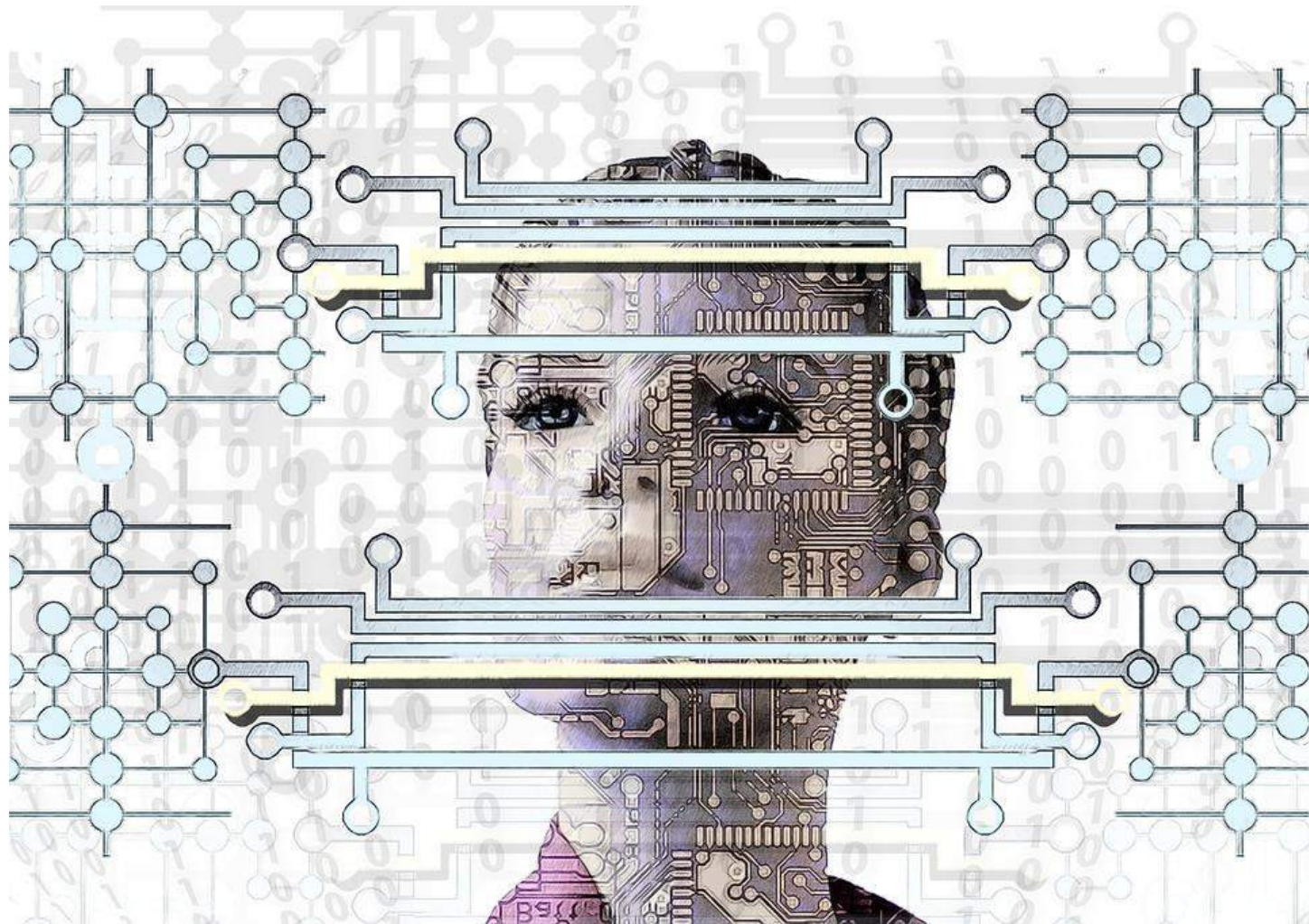
These are single layered recurrent networks



간단할 수 없기 때문에 ‘현실의 벽’은 너무나 높고 두터웠다.

이 시기는 ‘계산주의’라는 키워드로 묘사할 수 있다. 계산주의자들은 지능이란 결국 복잡한 계산에 다름아니고, 따라서 계산 능력을 이용해 논리적 추론을 하는 방법을 가르치고 세상 만물을 컴퓨터의 메모리에 올바른 기호로 저장(기억)하면 컴퓨터가 자동으로 세상에 대해 “사고”할 수 있다고 생각했다.

그런데 1971년 스티븐 쿡, 1972년 리처드 카프, 1973년 레오니드 레빈이 각각 결정적인 논문을 발표했다. 논문의 주제는 주어진 문제를 컴퓨터로 얼마나 빨리 풀 수 있는지를 다루는 ‘P-NP’ 문제로 이 추론에 따르면 어떤 문제는 입력 데이터의 크기가 증가할수록 계산에 필요한 시간이 지수적으로 증가하기 때문에 영원히 계산을 마칠 수 없다. ‘인공지능이 필요한 복잡한 문제에서는 정작 인공지능이 무용지물’이라는 뜻이다





Having just read it, it appears that the 1962 work in which Rosenblatt proposed “four layer” networks (i.e. input, output, and two hidden layers) didn’t actually do supervised learning. Instead, it uses a more Hebbian learning rule to do unsupervised learning (it relies on examples being presented in an order that means examples of the same “class” are presented in sequence; the network learns to associate examples that are presented in close temporal proximity).

I can think of three reasons that Rosenblatt's work is sometimes not cited, and even miscited. The first is that [Minsky and Papert's \(1969/1988\)](#) book is an analysis of single-layer perceptrons, and adopts the convention of referring to them as simply as perceptrons. The second is that the perceptron update rule is widely used under that name, and it applies only to single layer networks. The last is that Rosenblatt and his contemporaries were not very successful in their attempts at training multi-layer perceptrons. See Olazaran (1993, [1996](#)) for in-depth discussion of the complicated and usually oversimplified history around the loss of interest in perceptrons in the later 1960s, and the subsequent development of backpropagation for the training of multilayer nets and resurgence of interest in the 1980s.

출처 : <https://blogs.umass.edu/comphon/2017/06/15/did-frank-rosenblatt-invent-deep-learning-in-1962/>

참고자료

옥수별 - <http://sams.epaiai.com/221142104977>
<https://deeplearning4j.org/kr/restrictedboltzmannmachine>

솔라스 - <http://solarisailab.com/deep-learning>

- <https://tensorflow.blog/>

카카오 AI 레포트 - <https://brunch.co.kr/@kakao-it/162>

김태훈 - <https://github.com/carpedm20/deep-rl-tensorflow>

골빈해커 - <https://github.com/golbin/TensorFlow-Tutorials>

[Bay Area Deep Learning School](#) - <https://youtu.be/eyovmAtUx0>

<https://www.reddit.com/r/MachineLearning/>

유튜브 구루 강의

얀 레쿤 강의 <https://youtu.be/0tEhw5t6rhc>

김성훈 교수 [모두를 위한 RL강좌](#)

모두를 위한 딥러닝 강좌

https://www.youtube.com/watch?v=BS6O0zOGX4E&list=PLIMkM4tgfjnLSOjrEJN31gZATbcj_MpUm

[Bay Area Deep Learning School](#)

<http://cs231n.stanford.edu/>

바람님의 딥러닝 기초 https://www.youtube.com/playlist?list=PLsS-TVNjbU7clDOjpAZKud3uG8APHDq_M

앤디 송의 님 스탠포드 cs231n 딥러닝 영상 강좌 (페이페이 리 교수)

<https://www.youtube.com/playlist?list=PL1Kb3QTCLIVtyOuMgyVgT-OeW0PYXI3j5>

테리님의딥러닝 토크

<https://www.youtube.com/watch?v=D4zqigCb8co&list=PL0oFI08O71gKEXITQ7OG2SCCXkrtid7Fq>

<http://untitledtblog.tistory.com/35?category=667127>

David Silver <https://youtu.be/2pWv7GOvuf0>

이웅원 <https://www.gitbook.com/book/dniddnjs/rl/details>

<http://coolingoff.tistory.com/7?category=615687>

<https://deeplearning4j.org/multilayerperceptron>

<http://www.aistudy.com/neural/perceptron.htm>

Hebb, Donald O. (1949). *The Organization of Behavior*. New York: Wiley, pg. 62.

Visualizing and Understanding Deep Neural Networks by Matt Zeiler <https://youtu.be/ghEmQSxT6tw>

유투브 입문 강의

얀 레쿤 강의 <https://youtu.be/0tEhw5t6rhc>

논문

로젠블랏의 논문

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.335.3398&rep=rep1&type=pdf>

books

<http://www.deeplearningbook.org/>

Artificial Intelligence : a modern approach (3rd edition) Stuart Russel, Peter Norvig

In 1949, [Donald Hebb](#) would help to revolutionize the way that artificial neurons were perceived. In his book, *The Organization of Behavior*, he proposed what has come to be known as Hebb's rule.

2.1 First Attempts The first step toward artificial neural networks came in 1943 when Warren McCulloch, a neurophysiologist and a young mathematician, Walter Pitts, developed the first models of neural networks. They wrote a paper The Logical Calculus of the Ideas Immanent in Nervous Activity on how neurons might work [1]. Their networks were based on simple elements which were considered to be binary devices with fixed thresholds. The results of their model were simple logic functions with “all-or-none” character of nervous activity. In 1944 Joseph Erlanger together with Herbert Spencer Gasser identified several varieties of nerve fiber and established the relationship between action potential velocity and fiber diameter. In 1949, Hebb a psychologist, wrote The Organization of Behavior [2], a work which pointed out the fact that neural pathways are strengthened each time they are used, a concept fundamentally essential to the ways humans learn.

2.2 Promising and Emerging Technology In 1958, Rosenblatt a psychologist, conducted an early work on perceptrons [3]. The Perceptron was an electronic device that was constructed in accordance with biological principles and showed an ability to learn. He also wrote an early book on neurocomputing, Principles of Neurodynamics [4]. Another system was the ADALINE (ADaptive LInear Element) which was developed in 1960 by two electrical engineers Widrow and Hoff [5]. The method used for learning was different to that of the Perceptron, it employed the Least-MeanSquares learning rule. In 1962, Widrow and Hoff developed a learning procedure that examines the value before the weight adjusts it.

2.3 Period of Frustration and Disgrace

Following an initial period of enthusiasm, the field survived a period of frustration and disgrace. In 1969 Minsky and Papert wrote a book *Perceptrons: An Introduction to Computational Geometry* [6]. It was a part of a campaign to discredit neural network research showing a number of fundamental problems, and in which they generalized the limitations of single layer perceptron.

Although the authors were well aware that powerful perceptrons have multiple layers and Rosenblatt's basic feed-forward perceptrons have three layers, they defined a perceptron as a two-layer machine that can handle only linearly separable problems and, for example, cannot solve the exclusiveOR problem.

Because the public interest and available funding becoming minimal, only several researchers continued working on the problems such as pattern recognition. But, during this period several paradigms were generated which modern work continues to enhance.

Klopf in 1972, developed a basis for learning in artificial neurons based on a biological principle [7]. Paul Werbos in 1974 developed the back-propagation learning method although its importance wasn't fully appreciated until a 1986. Fukushima developed a stepwise trained multilayered neural network for the interpretation of handwritten characters. The original work Cognitron: A self organizing multilayered neural network [8] was published in 1975. In 1976 Grossberg in the paper Adaptive pattern classification and universal recoding [9] introduced the adaptive resonance as a theory of human cognitive information processing.

References

1. McCulloch, W.S., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Bioph. 5, 115–133 (1943)
 2. Hebb, D.: The Organization of Behavior. Wiley, New York (1949)
 3. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65(6), 386–408 (1958)
 4. Rosenblatt, F.: Principles of Neurodynamics. Spartan Books, Washington (1962)
 5. Widrow, B., Hoff, M.: Adaptive switching circuits. Technical report 1553-1, Stanford Electron. Labs., Stanford, June 1960
- Hubel, D. H.; Wiesel, T. N. (1959). [“Receptive fields of single neurones in the cat's striate cortex”](#). 《The Journal of Physiology》 1
- Hubel, D. H.; Wiesel, T. N. (1962). [“Receptive fields, binocular interaction and functional architecture in the cat's visual cortex”](#)

Hubel, D. H.; Wiesel, T. N. (1959). ["Receptive fields of single neurones in the cat's striate cortex"](#). 《The Journal of Physiology》 1

Hubel, D. H.; Wiesel, T. N. (1962). ["Receptive fields, binocular interaction and functional architecture in the cat's visual cortex"](#)

6. Minsky, M.L., Papert, S.: Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge (1969)

7. Klopff, A.H.: Brain function and adaptive systems - a heterostatic theory. Air Force Research Laboratories Technical Report, AFCRL-72-0164 (1972)

8. Fukushima, K.: Cognitron: a self-organizing multilayered neural network. Biol. Cyber. 20, 121–136 (1975)

9. Grossberg, S.: Adaptive pattern classification and universal recoding. Biol. Cyber. 23(3), 121–134 (1976)

10. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biol. Cyber. 43, 59–69 (1982)

11. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA 79, 2554–2558 (1982)

12. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, pp. 318–362. MIT Press, Cambridge (1986)

13. Carpenter, G.A., Grossberg, S.: A massively parallel architecture for a self-organizing neural pattern recognition machine. Comp. Vis. Graph. Image Proc. 37, 54–115 (1987)

14. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. 61, 85–117 (2014)

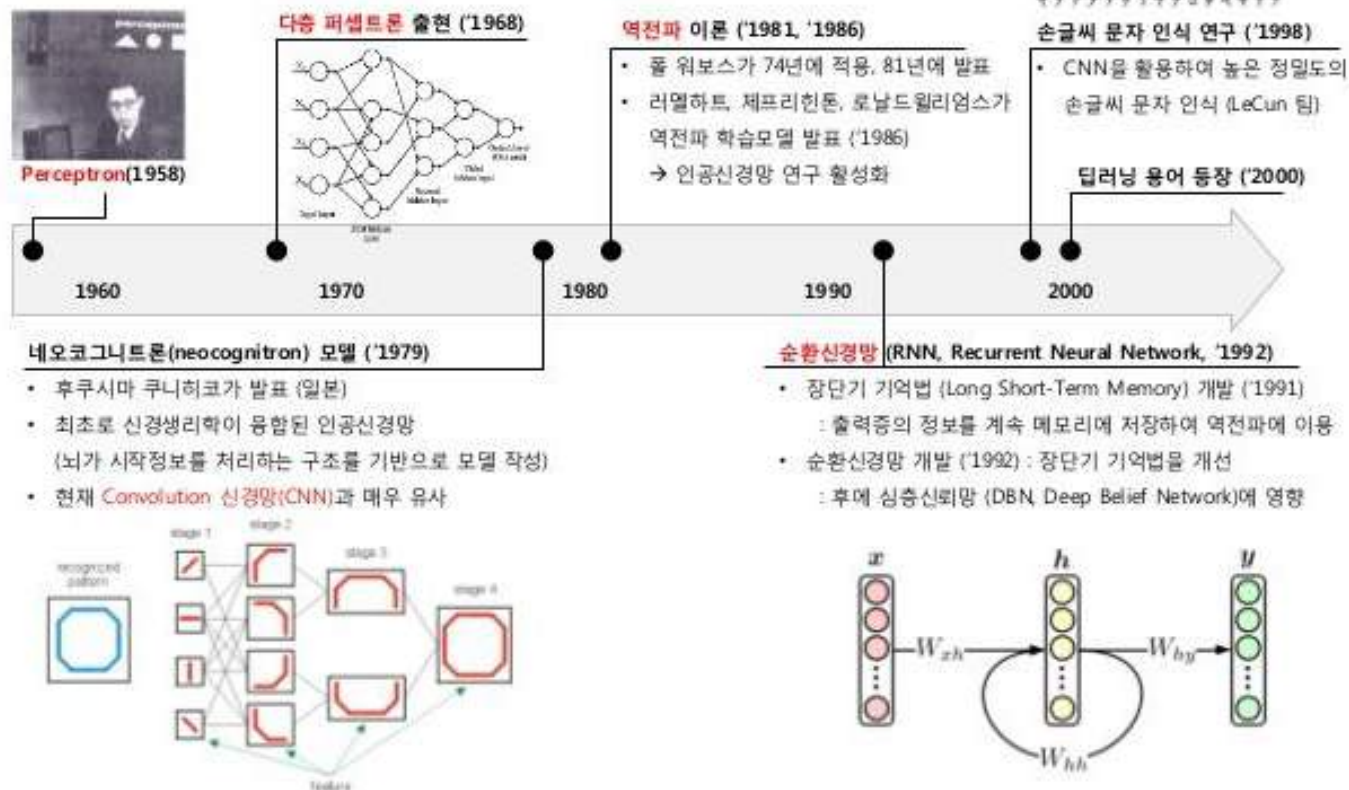
15. Gnana Sheela, K., Deepa, S.N.: Review on methods to fix number of hidden neurons in neural networks. Anna University, Regional Centre, Coimbatore 641047, India (2013)



3. 딥러닝 역사

VI. 딥러닝 개요 및 사례

딥러닝의 근간이 되는 시작된 인공지능의 연구는 1950년대부터 시작되었다.



2000년대 초 빙하기를 거치지만 2006년 부터 딥러닝 문제가 해결되면서 딥러닝이 부활되었다.



- 딥러닝 부활 ('2006)

- 심층신뢰망 (DBN)
 - ✓ 제프리 힌튼, 사이먼 오신데로, 이-화이테가 심층신뢰망 발표
 - ✓ RBM(Restricted Boltzmann Machine)이라는 사전학습(Pre-Training) 이용
- Auto-Encoder
 - ✓ 벤지오 팀이 발표, DNN을 최적화하는 사전학습(Pre-Training) 방법
 - ✓ 사전 학습을 통해 가중치 초기값을 정하면 DNN이라도 학습이 잘 됨

- ## Dropout 알고리즘 ('2012)

- Dropout 알고리즘 : 학습 시 오버피팅(overfitting)을 줄임
- Sigmoid 형 함수 대신 ReLU 형 함수 사용

- Deep Learning은 왜 주목 받는가?

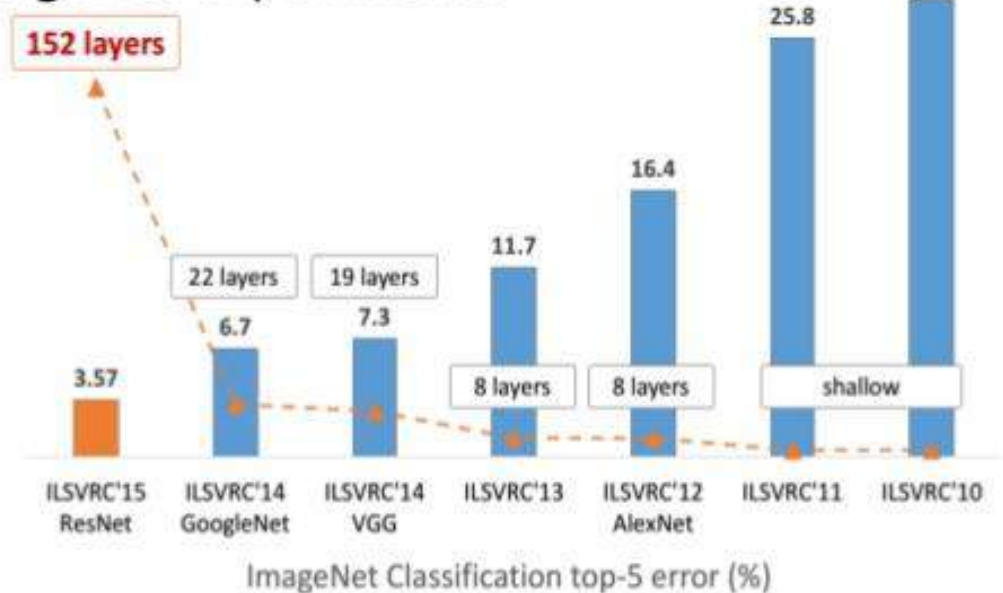
Difficulties		해결 방안
학습	DNN 학습이 잘 안 됨.	Unsupervised Pre-Training를 통한 해결
계산량	학습이 많은 계산이 필요함	H/W의 발전 및 GPU 활용
성능	다른 Machine Learning Algorithm의 높은 성능	Dropout 알고리즘 등으로 Machine Learning 대비 월등한 성능



IMAGENET WINNERS

Microsoft
Research

ImageNet experiments



Feilong Ma, Xiaoou Tang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.