

### 1.2.1 신경망(Neural Network) 변천사

- 딥러닝(deep learning)은 그 연원을 1940년대까지 거슬러 올라감. '딥러닝'이라는 용어 자체는 최근에 생김.
- 딥러닝의 발전 단계는 대략 3개 시기로 나뉨.
  - 1단계 1940~1960년대: cybernetics 시기
  - 2단계 1980~1990년대: connectionism
  - 3단계 2006~ 현재: deep learning
- 초기 학습 알고리즘들은 두뇌 학습 모델을 따른다고 해서 '인공신경망(Artificial Neural Network)'이라 불리는데, 연구자들은 아래와 같은 기대를 했음.
  - '두뇌' 자체가 '지능 행동'의 표본(proof by example)이기 때문에 뇌의 작동원리와 기능을 역공학(reverse engineering)하면 인공지능을 만들 수 있을 것이다.
  - 이런 학습 모델링이 두뇌 자체를 연구하는데도 도움이 될 것이다.
- 현대적 의미에서 '딥러닝'은 이런 뇌과학적인 관점보다, "다층 구조 학습 원리(principle of learning multiple levels of composition)"라는, 좀 더 일반적인 학습 원리에 초점을 맞춤.
- 단순 선형 모델(1단계 시기)
  - $n$ 개의 입력값:  $x_1, x_2, \dots, x_n$
  - 한 개의 출력:  $y$
  - 가중치(weight) 학습:  $w_1, w_2, \dots, w_n$   
 $\Rightarrow f(x, w) = x_1 w_1 + \dots + x_n w_n$
  - 맥클로츠-피츠 뉴런(1943)
    - $f(x, w)$  값의 음/양에 따라 카테고리 인식
    - 가중치는 사람이 결정..
  - 퍼셉트론(perceptron, 1958, 1962)
    - 입력 예시들로부터 가중치를 학습할 수 있는 최초의 모델
  - ADALINE(adaptive linear model, 1960)
    - $f(x)$  값 자체를 반환
    - 데이터들로부터  $f(x)$  값 예측하게 학습
    - 오늘날 주요 딥러닝 학습 알고리즘인 "**stochastic gradient descent**"의 특수 형태 이용.
- 단순 선형 모델의 한계
  - XOR 학습 문제(Minsky & Papert, 1969)
  - neural net 인기 하락..
- 뇌과학과 딥러닝
  - 뇌과학이 딥러닝 영감의 원천이기는 하나 단 하나의 소스는 아님(뇌 자체를 너무 모름).
  - 단 하나의 학습 알고리즘으로 많은 여러 다른 작업을 수행할 수 있다는 가설
    - 흰담비 뇌에서 청각신호 처리 영역을 통해 "볼 수"있게 학습시킬 수 있음(Von Melchner 2000).
    - NLP, vision, motion planning, speech recognition 등의 연구 커뮤니티들이 아직도 분리되어 있지만, 최근 동시에 여러 영역을 연구하는 추세임.

- 포유류의 시각처리 시스템에서 영감을 얻은 강력한 이미지 처리 모델 아키텍처 Neocognitron(Fukushima 1980)이 현대의 convolutional network(LeCun, 1986b)의 바탕이 됨.
- 최근의 신경망은 "**rectified linear unit(ReLU)**"이라는 단순한 뉴런 모델을 많이 씀.
  - Neocognitron의 최초 모델(Fukushima 1975)는 더 복잡한 뉴런 모델이었음.
  - 단순화 버전은 뇌과학에 영향을 받았다는 Nair & Hinton(2010)의 연구와 공학에서 영감을 얻었다는 Jarret(2009)의 연구 등에서 아이디어를 얻어 만들어진 것임.
- 뇌를 시뮬레이션하겠다는 시도를 딥러닝이라고 오해해서는 안됨.
  - 뇌과학만이 아니라 많은 분야, 특히 수학적으로 linear algorithm, probability, information theory, numerical optimization 등에서 도움을 많이 받음.
  - 어떤 연구자들은 뇌과학이 주요 (영감의) 원천이라고 하지만, 또 다른 연구자들은 뇌과학에 전혀 신경을 안 쓴다고 함.
- 전산뇌과학(computational neuroscience)
  - 두뇌가 알고리즘 레벨에서 어떻게 동작하는지 이해하고자 함.
  - 딥러닝과는 별개의 분야. 실제 뇌가 동작하는 방식을 좀 더 정확히 모델링하는 것이 목표임.
- Connectionism 또는 parallel distributed processing (2단계 시기)
  - 인지과학으로부터 출발
  - 1980년대 초에 상징추론(symbolic reasoning) 모델 연구
  - "connectionists"들은 1940년대 심리학 아이디어(Donald Hebb(1949))까지 차용하면서 "인식"의 신경망구현 모델링 연구 시작.
  - (단순하지만) 많은 컴퓨팅 유닛들간의 상호작용(==연결)으로 지능적인 작업이 가능하다는 개념을 형성하기 시작(뇌과학의 영향).
  - **distributed representation** (Hinton 1986)(15장에 자세히 언급)
    - {트럭,자동차,새}와 {빨강,초록,파랑}의 특성 조합은, {빨강 트럭, 빨강 자동차, 빨강 새, 초록 트럭, ...} 등 9개임.
    - 단일 뉴런으로 학습하려면 9개의 뉴런이 필요.
    - 이를 분산시켜서, 3개의 뉴런은 형태를 인식하게 하고, 3개는 색을 인식하게 하면 6개로 충분함.
  - **back-propagation 알고리즘**(Rumelhart 1986, LeCun 1987)
    - 현대 딥러닝의 주요 학습 방법
  - **Long Short Term Memory(LSTM) 모델**(Hochreiter & Schmidhuber 1997)
    - 신경망으로 연속이벤트(시퀀스)를 모델링하기 위한 방법
    - Google에서 NLP 연구 등에 사용
  - 몰락(~2007)
    - 벤처들의 (투자 유치를 위한) 허황된 약속과 그에 따른 투자자들의 실망.
    - Kernel machines(Boser 1992, Cortes & Vapnik 1995, Scholkopf 1999)이나 graphical models(Jordan 1998)의 상대적인 성공
  - 명맥(CIFAR NCAP)

- Canadian Institute For Advanced Research
  - Neural Computation & Adaptive Perception 연구 프로그램
    - ⇒ Geoffrey Hinton(토론토 대학), Yoshua Bengio(몬트리얼 대학), Yann LeCun(뉴욕대)
    - ⇒ 뇌과학자들과 컴퓨터/휴면 비전 연구자들도 참여.
- 3단계 시기(딥러닝. 2006~)
  - 1980년대 deep network 알고리즘들도 괜찮았음(계산 비용이 많이 들뿐...)
  - **greedy layer-wise pretraining**(Hinton 2006)
    - 효율적으로 deep network 학습을 할 수 있게 함.(15.1절에 소개)
    - 다른 CIFAR 관련 연구들(Bengio 2007, Ranzato 2007a)이 다른 여러 종류의 deep network에 이 전략을 적용하면서, 시스템적으로 일반화시키기 시작함.
    - "deep learning"이라는 용어가 퍼지기 시작함.
  - 초기에는 적은 데이터셋에 대한 비지도학습 관점으로 접근했는데, 요즘은 점차적으로 큰 labeled 데이터셋에 대한 지도학습으로 관심이 옮겨지고 있는 추세.