

신경회로망의 역사

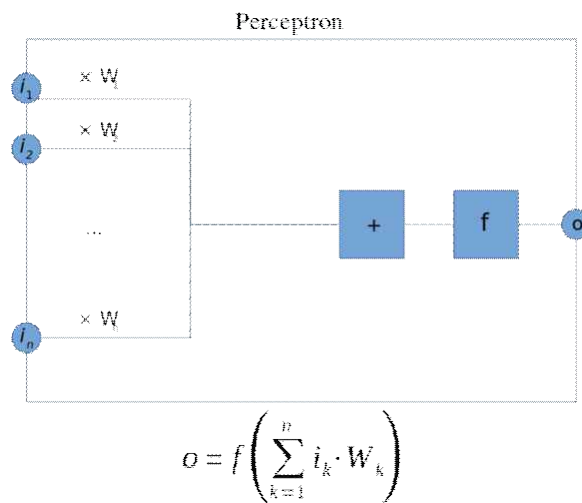
과목명	딥러닝 실제
담당교수	전명근 교수님
학과	산업인공지능학과
학번	2021254009
이름	정원용



1. 신경 회로망의 역사

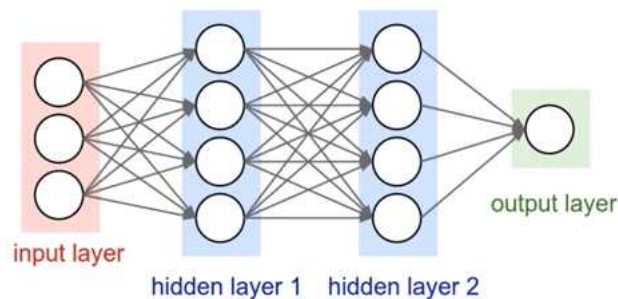
1943년 신경과학자인 Warren S. McCulloch과 논리학자인 Walter Pitts는 한 사람 뇌신경 세포를 하나의 이진 출력을 가지는 단순 논리 게이트로 설명했는데, 이를 McCulloch-Pitts 뉴런(MCP 뉴런)이라 부른다.

1957년 코넬 항공 연구소에 근무하던 Frank Rosenblatt은 MCP 뉴런 모델을 기초로 퍼셉트론(Perceptron) 학습 규칙이라는 개념을 고안하게 되는데, Rosenblatt은 하나의 MCP 뉴런이 출력신호를 발생할지 안할지 결정하기 위해 MCP 뉴런으로 들어오는 각 입력값에 곱해지는 가중치 값을 자동적으로 학습하는 알고리즘을 제안했다.



<그림1. 퍼셉트론(Perceptron) >

1958년 퍼셉트론이 발표된 후 같은 해 7월 8일자 뉴욕타임즈는 앞으로 조만간 걷고, 말하고 자아를 인식하는 단계에 이르는 컴퓨터 세상이 도래할 것이라는 다소 과격한 기사를 냈다. 하지만 1969년에 단순 퍼셉트론은 XOR 문제로 풀 수 없다는 사실을 MIT AI 랩 창시자인 Marvin Minsky 교수가 증명하였고, 아래 그림과 같은 다층 퍼셉트론(MLP)으로 신경망을 구성하면 XOR 문제를 풀 수 있다고 했다.

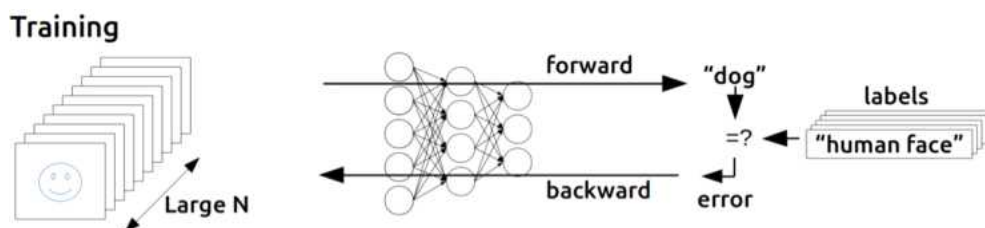


그런데, Minsky 교수는 이러한 MLP에서 hidden layer의 가중치를 계산하는, 다시 말해 MLP를 학습시키는 방법은 존재하지 않는다고 단정해 버렸다.

이로 인해 인공지능망과 관련된 학문과 기술은 더 이상 발전되지 않고 암흑기를 겪게 됩니다.

그러다가 1974년, 당시 하버드 대학교 박사과정이었던 Paul Werbos는 MLP를 학습시키는 방법을 찾게 되는데, 이 방법을 Minsky 교수에서 설명하지만 무시되어 버린다.

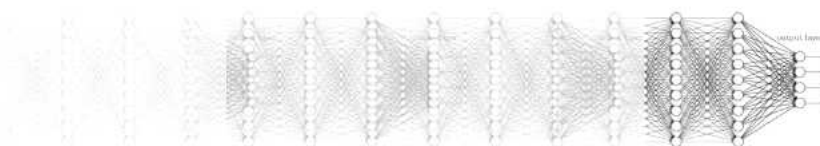
Paul Werbos가 Minsky 교수에게 설명한 MLP를 학습시킬 수 있는 획기적인 방법이 바로 오류 역전파(Backpropagation of error)라는 개념이다. 줄여서 역전파(Backpropagation)라고 부른다.



이런 획기적인 방법 역시 당시 인공지능망의 대가와 학계로부터 무시당해버린 후 Werbos는 1982년 역전파에 대한 내용을 논문으로 발표하고 마무리한다. Werbos의 역전파가 무시된 이후 10여년 넘게 AI 분야는 흑한기를 겪게 된다. 그러다가 1986년 인지심리학자이자 컴퓨터 공학자였던 Geoffrey Hinton 교수는 Werbos가 제안한 오류 역전파 알고리즘에 대한 내용을 독자적으로 제안하게 된다.

그런데 다층 구조로 되어 있는 MLP의 학습은 역전파를 통해 학습이 가능하지만 학습의 효과를 크게 하려면 인공지능망의 은닉층(hidden layer)를 많이 쌓아야 더 좋은 결과가 나올 수 있다는 것이 경험적으로 검증되었다.

하지만, 당시까지만 해도 활성화함수로써 시그모이드(sigmoid) 함수를 사용했고, 이는 은닉층의 개수가 많아질수록 역전파에 의한 가중치 계산이 불가능하게 되는 gradient vanishing 이라는 문제에 직면하게 되었다.



이 문제로 인해 인공지능망 분야의 2번째 암흑기가 시작되는 계기가 되어 버린다.

1995년에는 당시 다른 방식으로 발전되었던 보다 단순한 머신러닝 알고리즘인 SVM, RandomForest와 같은 알고리즘이 손글씨 인식 등과 같은 분야에는 더 잘 작동한다고 Lecun 교수 등이 발표하게 된다.

1998년 컴퓨터공학자인 Yann Lecun 교수는 1950년대 수행했던 고양이의 뇌파 실험에 영감을 얻어 이미지 인식을 획기적으로 개선할 수 있는 CNN(Convolutional Neural Network)라는 새로운 형태의 인공지능망을 고안하게 된다.

1950년대 수행했던 고양이의 실험은 고양이의 눈으로 보는 사물의 형태에 따라 뇌의 특정영역, 정확히 말하면 특정 뉴런만이 활성화 된다는 것을 알게 된 것이다. Lecun 교수는 이러한 실험 결과에 영감을 얻어 CNN 이라는 새로운 형태의 인공신경망을 고안하게 된 것이다.

하지만 gradient vanishing 문제 등으로 인해 한동안 인공신경망 분야가 침체기를 겪고 있던 시기에 Hinton 교수와 Bengio 교수는 2006년, 2007년 두 편의 논문을 발표하였는데, 초기 입력값을 잘 선택하면 아무리 인공신경망의 층의 개수가 많더라도 학습이 가능하며, 복잡한 문제도 층의 개수가 많게 구성된 인공신경망이라면 해결할 수 있다고 했다. 그리고 이렇게 층의 개수가 많은 신경망을 심층신경망(Deep Neural Network, DNN)이라 리브랜딩하고 심층신경망을 학습시키는 방법을 딥러닝(Deep Learning, DL)이라고 명명하게 된다.

그리고 200년 네이처(Nature)지에 시그모이드를 대신하여 사용한 Rectifier라는 활성화함수를 이용해 효과를 보았다는 내용의 생물학 분야의 논문이 발표되었다.

Rectifier는 ReLU(Rectified Linear Unit)라고 불리게 되는 활성화함수인데, 이 활성화함수를 심층신경망의 딥러닝에 사용해보니 gradient vanishing 문제가 해결 돼 버린 것이다.

물론 심층신경망의 최종 출력층에서는 여전히 시그모이드나 SoftMax를 사용하지만 말이지요.

활성함수 ReLU의 등장으로 딥러닝은 많은 발전을 이루게 되었고, 2012년 이미지 인식 분야의 유명한 대회인 ImageNet Large Scale Visual Recognition Challenge(ILVRC)에서 캐나다 토론토 대학의 AlexNet이라 불리는 CNN 인공신경망으로 우승하게 되는데 이전까지 이미지 인식 오류율이 26% 대였던 것이 15%로 줄이게 된 것입니다.

이후 CNN의 신경망구조가 지속적으로 개선되고 GPU의 발전으로 현재는 5%이내의 오류율로 학습이 가능하게 되었습니다.