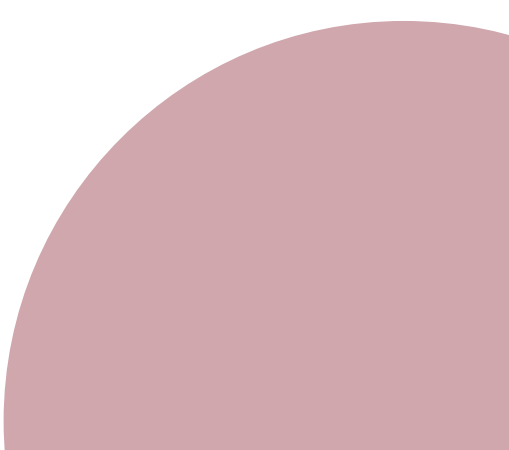


ML과 LLM의 작동 원리

인공지능 기초 심화탐구 보고서

2025-10-10

황태준
중앙고등학교
2025
2학년 7반 31번



목차

I. 인공 신경망	4
-----------------	---

Abstract

인공지능은 말그대로 인간이 컴퓨터로 만든 지능이다. 그런데, 지능이란 무엇인가? 지능이 있다는 것은 생각할 수 있다는 것인가? AI는 생각하는가? 겉으로 보기에는 윤리적 판단도 어느 정도 하고, 수학 문제도 풀어내고, 각종 사소한 건에 대해 물어보면 나름 합리적인 판단도 내려주므로 생각하는 것처럼 보이기도 한다. 하지만 중학교에서 알고리즘을 처음 배울 때, 컴퓨터는 일정한 형식의 명령으로 계산(computation)만 빠르게 할 수 있는 바보라고 하지 않았는가? 구체적이고 특정적이지 않은, 모호한 인간의 자연어를 인공지능은 어떻게 처리하고 이해하여 그로부터 양질의 출력을 제공하는가?

지능(intelligence)의 정의는 아직도 제대로 내려지지 못하고 있다. 하지만 여러 정의에서 공통적으로 언급되는 부분은, 정보나 기술을 획득하여 목표 성취에 적용하고 새로운 것을 창조한다는 부분이다. 이 과정에서는 학습이 필수적이다. 정보나 기술을 획득하여 저장하고 이해(?)하는 과정이 학습이기 때문이다. 지능을 어떻게 정의 내리냐에 따라 컴퓨터가 지능을 가질 수 있는지 없는지는 철학적인 문제가 되겠지만, 적어도 지능을 모방할 수는 있을 것이다.

인공 신경망은 이러한 발상에서 비롯한 기술이자 자료구조로, 일종의 그래프이다. 정점¹은 뉴런의 신경세포체를, 간선은 축삭을 본따 그 역할을 대신한다고 하는데, 생명과학을 배운 입장에서 이게 정말 맞는 말은 아닌 것 같고, 일반인들의 이해를 돕기 위한 비유적 설명으로 받아들일 수 있겠다.

정점들은 계층별로 모여 있는데, 직접적인 입력이 들어오는 입력 계층, 실질적인 처리가 일어나는 은닉 계층, 출력이 일어나는 출력 계층으로 나뉜다. 입력 계층과 출력 계층은 하나이며, 은닉 계층은 개수 제한은 없으나 그 개수와 해당 계층에서의 정점 개수를 어떻게 하느냐에 따라 학습 결과 성적이 좌우된다. 각 계층 간 정점은 수많은 간선들이 연결하고 있다. 최적화 알고리즘에서 모식적 모형으로 쓰이는 것들과 비슷하게, 여러 정점으로 가는 각 간선에는 서로 다른 가중치가 부여되어 있다. 인공 신경망을 통한 학습의 목적은, 궁극적으로 이 가중치들을 적절히 조절하여 주어진 입력에서 알맞은 출력을 내는 것이다. 하지만 수많은 간선들에 대해 일일이 수동으로 가중치를 설정할 수 없으니, 원하는 결과로부터 가장 가까운 출력을 내도록 역추적을 하는 방법을 사용하여 자동으

¹앞으로 노드(node)와 혼용할 것인데, 뭔가 문맥 상 자연스러운 어휘가 다르기 때문이다.

로 학습을 진행한다. 이 방법을 역전파(backpropagation)이라고 하고, 이를 실현하기 위해 다변수 미적분학과 선형대수학의 손을 빌린 수학적 아이디어를 경사하강법(gradient descent)라고 한다.

하지만 인공 신경망은 생각보다 오래된 기술이고, 단점이 명확하다. 연산의 효율성은 둘째치고, 인공 신경망은 지능이 아니라 특정 목적 달성을 위해서만 단련된 시스템이기 때문에, 예를 들어 개와 고양이 사진을 구별하는 인공 신경망에 사람의 사진을 넣으면 무조건 틀린 답을 내놓게 된다. 사람은 개도 고양이도 아무것도 아닌 것²도 아니기 때문이다. 인공 신경망은 그 태생적 한계 때문에, 범용적이 될 수 없다.

현재는 거대 언어 모델(LLM)³의 시대이다. 물론 기초적인 형태의 언어 모델은 특정 목적을 위해서만 만들어지기도 했지만, 구글의 ‘트랜스포머’ 모델 개발 이후 문장 전체의 문맥을 파악할 수 있게 된 언어 모델은 급격히 범용 인공지능을 향해 달려나가 발전하고 있다. 지금 유명한 모든 또는 거의 모든 인공지능 언어 모델은 트랜스포머에 기반한다.

I. 인공 신경망

²default 또는 empty 출력 정점

³대규모 언어 모델? 그냥 편하고 엘엘엠이라고 하겠습니다.