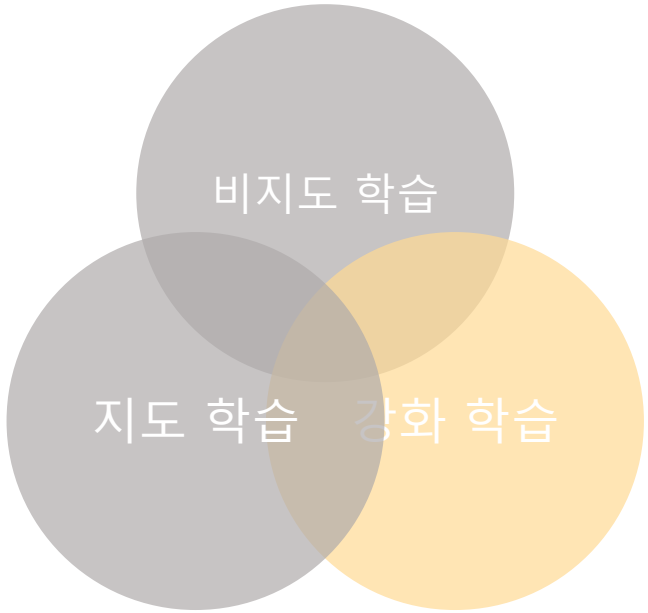
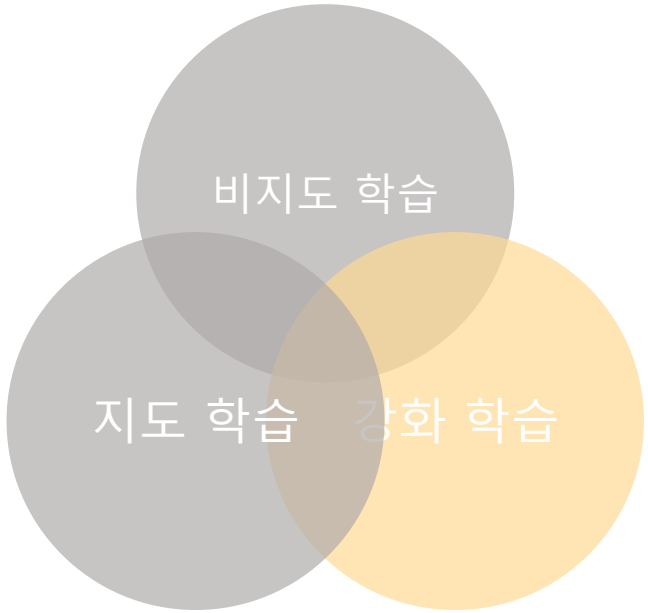


강화학습(reinforcement learning)

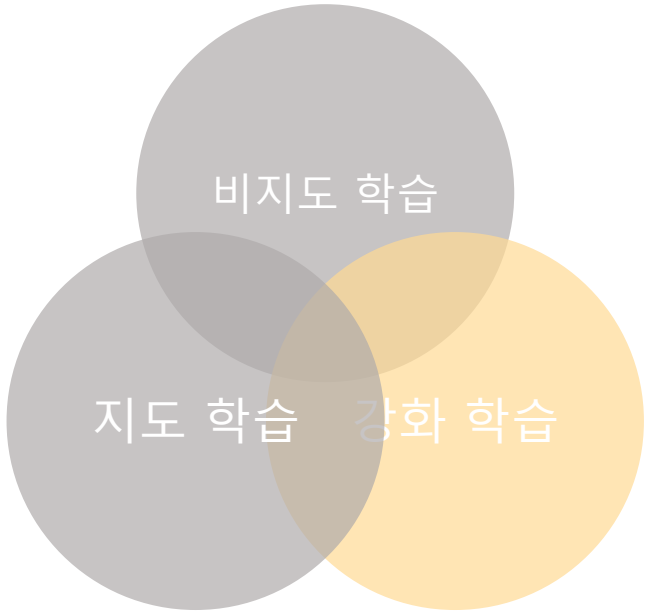
(= 결과에 대해 상과 벌을 준다.)

- 컴퓨터에게 보상을 주면서 학습을 시키는 것입니다.
- 컴퓨터에게 당근을 주면서 상을 내리는 방식입니다.
- 컴퓨터에게 알려주고 학습시키고 싶은 분야가 있을 수 있습니다.
- 강화학습은 알파고 제로를 가능하게 한 머신 러닝 기법입니다.
- 경험과 시행착오에 따른 보상체계를 시반으로 학습이 이루어집니다.
- 주어진 데이터는 없고 일단 부딪혀보면서 될 때까지 학습하는 막무가내 기법입니다.



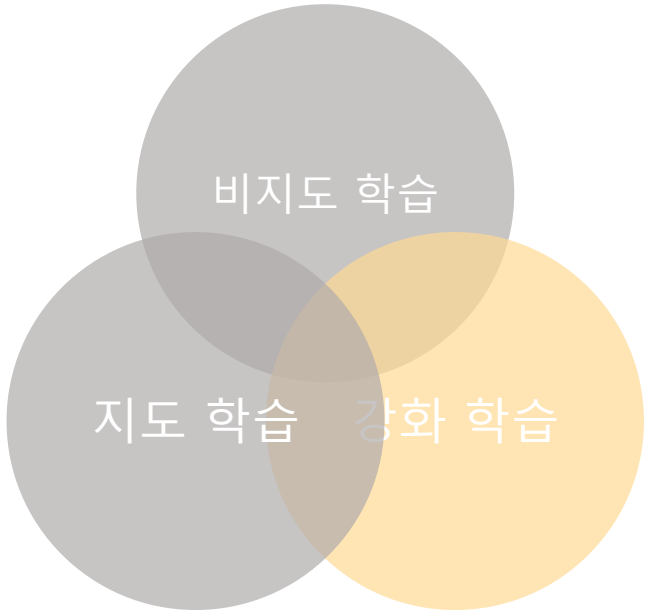
강화 학습 동작 방식

- 강화 학습은 시행착오 방법을 따라 원하는 결과를 얻습니다.
- 작업을 완료한 후 에이전트는 상을 받습니다.
- 예를 들어 개에게 공을 잡도록 훈련 시키는 것을 들 수 있습니다.
- 개가 공을 잡는 법을 배우면 비스킷과 같은 보상을 줍니다.
- 강화 학습 방법은 모델을 훈련하기 위해 외부 감독이 필요하지 않습니다.
- 강화 학습의 문제는 보상 기반입니다.
- 모든 작업 또는 완료된 모든 단계에 대해 에이전트가 받는 보상이 있습니다.
- 작업이 올바르게 수행되지 않으면 약간의 패널티가 추가됩니다.



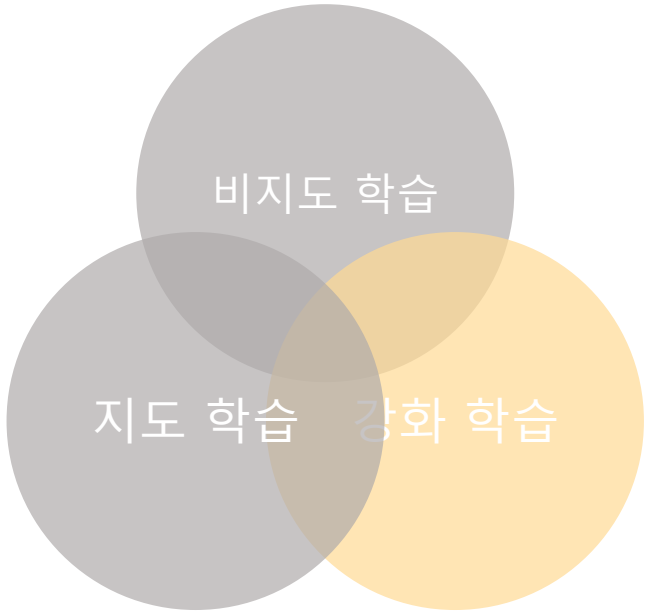
강화 학습에서 많이 사용되는 알고리즘

- Q - Learning
- Sarsa
- Monte Carlo
- Deep A network



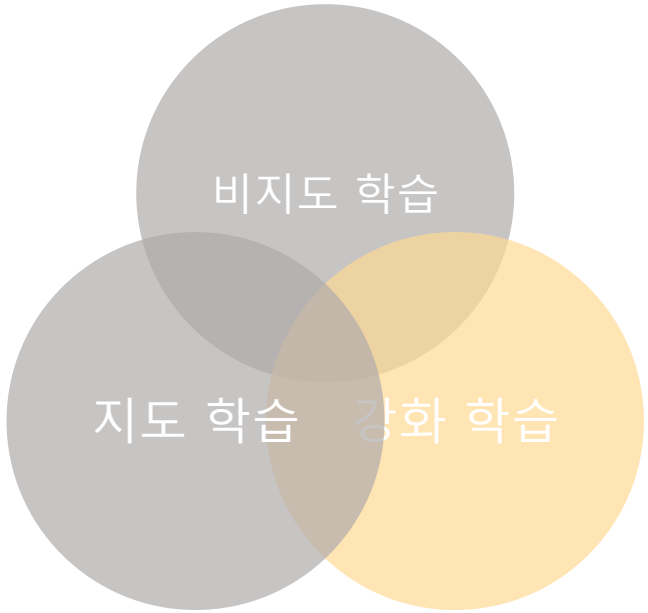
강화 학습의 두가지 형태 : Model-based

- 환경에 대한 정보가 주어진 모델이 있는 강화 학습입니다.



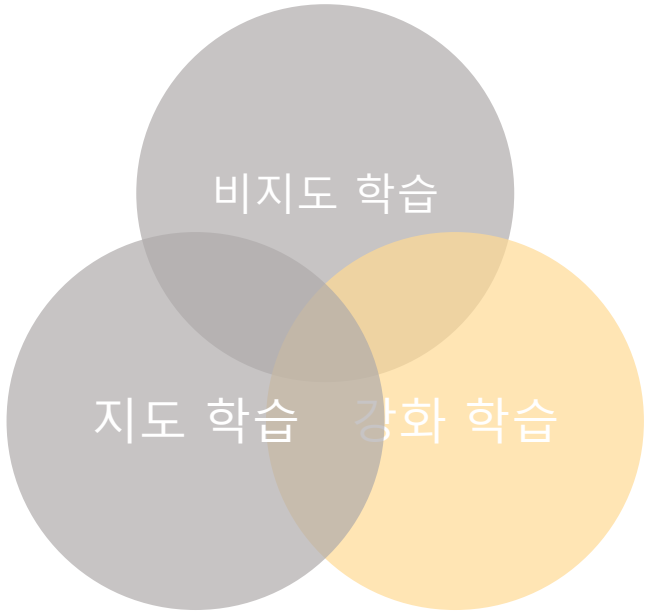
강화 학습의 두가지 형태 : Model-free

- 환경에 대한 정보가 없는 강화 학습입니다.



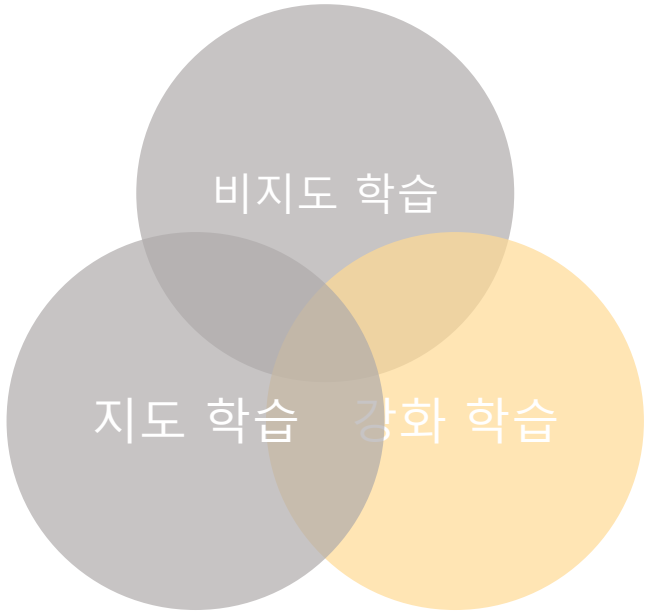
강화 학습의 두가지 형태 핵심 : 환경에 대한 정보

- 환경에 대한 정보가 중요한 이유는 어떤 Action을 했을 때 어떤 Reward를 받는지를 안다면, 쉽게 최적의 보상을 얻는 방법을 찾을 수 있겠죠.
- 하지만 대부분은 이러한 환경 정보가 주어지지 않습니다.
- 따라서, Agent는 수 많은 시행착오를 거쳐서 환경 정보를 습득하며 그 정보를 이용해서 최적의 해를 구합니다.



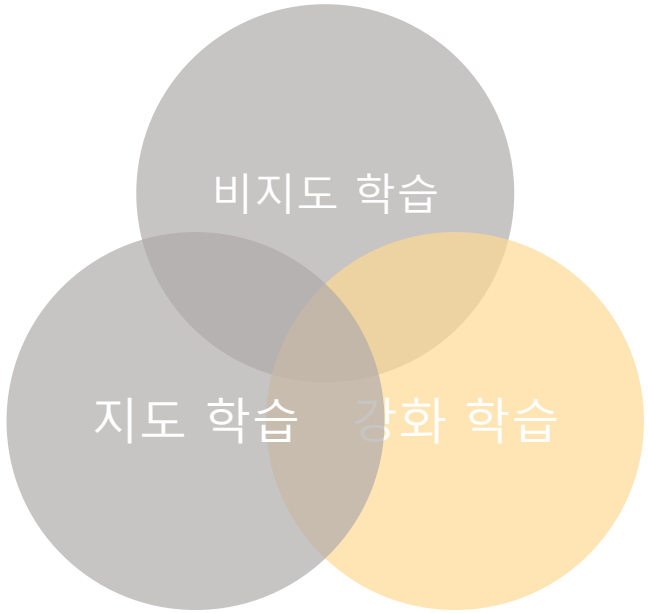
강화 학습의 두가지 형태 핵심 : 환경에 대한 정보 예시

- 미로 찾기를 생각해 볼 수 있습니다. 미로의 구조가 어떻게 되어있는지를 알고 있다면 해당 강화학습은 Model - based 학습에 해당됩니다.
- 반면 미로의 구조를 모른다면 Model - free 학습에 해당되는 것입니다.
- Model - free 강화학습의 경우에는 미로의 구조를 모르기 때문에 더 많은 시행착오를 거쳐서 미로의 구조를 파악해야 합니다.
- 따라서 Model - free 강화학습이 Model- based 보다 난해하고 어려운 경우가 많습니다.



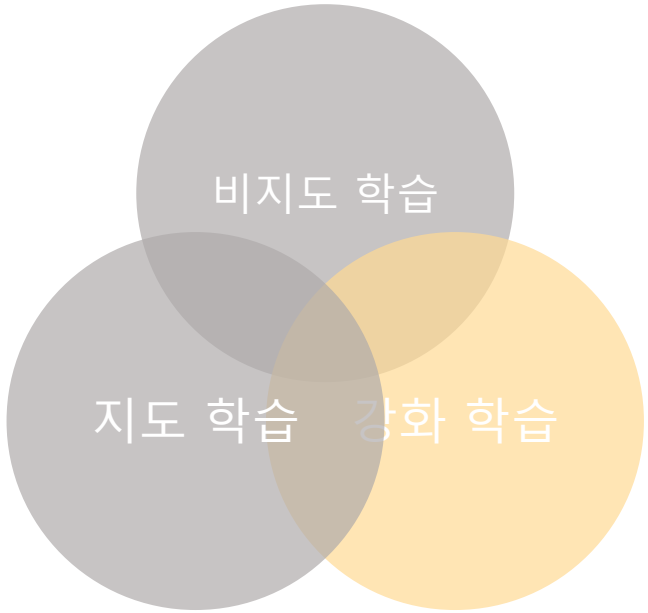
지도학습의 보완인 강화학습

- 지도학습을 보면 입력들을 이용해 규칙을 만들어 냈다.
- 하지만 한 번 만든 모델에 입력을 넣어 그 예상 결과만 받아볼 뿐 입력의 성격이 바뀌면 여기에는 적응 못하는 한계를 가지고 있는데 이걸 해결할 수 있는 방법이 강화 학습입니다.
- 강화학습은 입력을 가지고 타겟을 평가한다. 그래서 맞으면 상을 주고 틀리면 벌을 주면서 가지고 있는 규칙을 조정한다.
- 나중에 입력들의 성격이 바뀌어도 상벌에 의해서 규칙이 그 환경에 맞게 적응할 수 있다.
- 위 내용을 하나의 모델로 만든 것을 에이전트라고 부른다,
- 에이전트의 목표는 최대한 많은 보상을 받는 것이다.



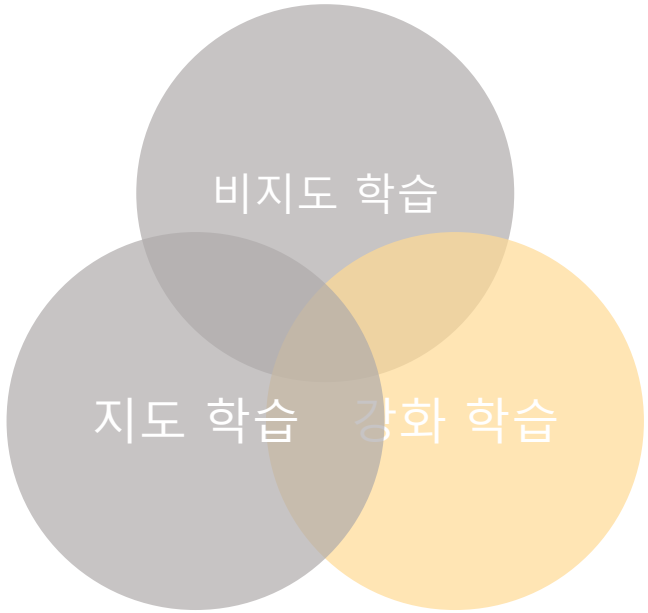
강화학습에서 상과 벌이란

- 강화학습의 상과 벌은 평가한 것에 대한 정답 오답을 알려주는 것이다.



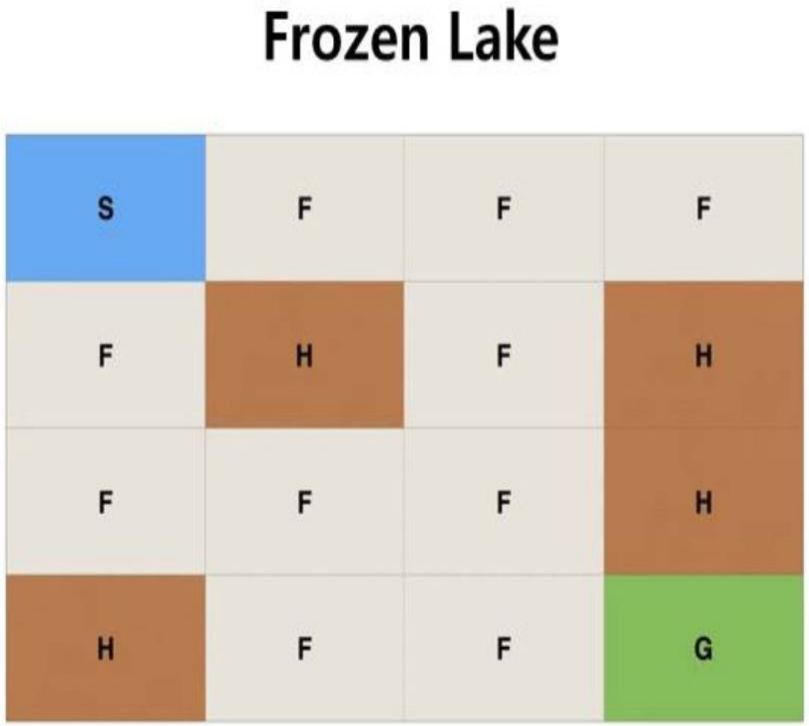
강화학습에서 어려움 : 지연되는 보상

- 행위를 선택했을 때 모든 보상이 즉시 일어나지 않아서 행동 선택의 어려움이 존재한다.
- 더구나 행동을 취했을 때 발생하는 상황과 보상에 불확실성이 있다면 의사결정은 더욱 어려워진다.
- 따라서 현재 상태만 참조하면 항상 필요한 모든 정보를 알 수 있다는 마르코프 Markov 가정한다.



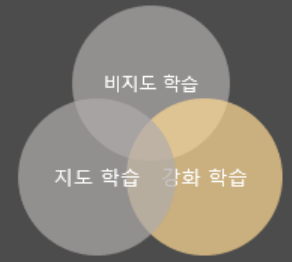
강화학습에서 어려움 : 지연되는 보상 해결

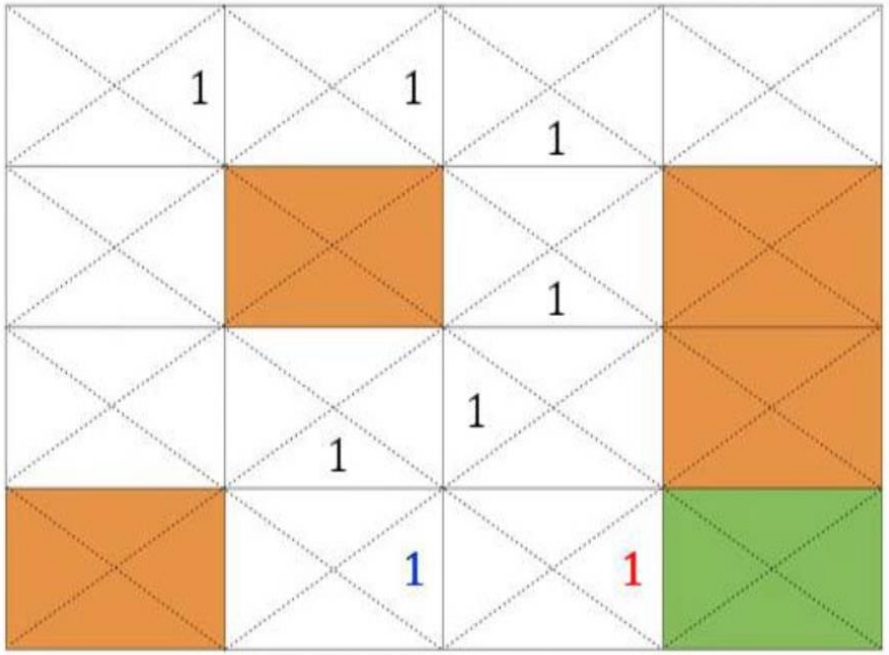
- 지능적 에이전트는 시작 상태까지 누적된 보상을 최대화하는 행동의 순서, 즉 궤적을 배워야한다. 누적 보상을 최대화하기 위하여 단기적 이익을 희생하기도 해야 한다.
- 에이전트의 바람직한 행동은 매 순간마다 가능한 여러 궤적 중에서 누적 보상의 기댓값이 가장 큰 것을 선택하는 것이다.
- Explore 와 Eploit 는 좋은 보상을 얻을 새로운 기회 탐색과 경험의 활용의 문제(기회탐색과 투자의 조화)
- 알파코도 몬테칼로 방법의 강화학습을 사용하여 매번 놓을 수의 승리 기댓 값을 계산한다.
- 자율주행차의 조정, 로봇 제어, 화학 반응 설계 등의 문제에서 사람을 능가하는 성과를 보여주고 있다.



강화학습의 예시1 : Frozen Lake

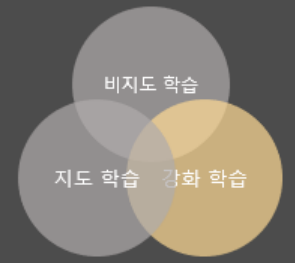
- 강화 학습의 유명한 예제인 Frozen Lake 게임이다.
- 4*4 사각형은 얼음 호수의 표면을 나타내고 파란색은 출발지, 녹색은 도착지, 갈색은 함정입니다.
- 게임의 목표는 출발점에서 시작하여 도착점까지 구멍에 빠지지 않고 도착하는 것입니다.
- 무턱대고 일단 출발하면 계속 구멍에 빠지거나 뱅뱅 돌거나 하면서 좀처럼 도착점을 찾지 못할 것입니다.
- 그렇게 계속 실패를 하다가 언젠가는 우연히 목표점에 도착하는 경우가 생길 것입니다.

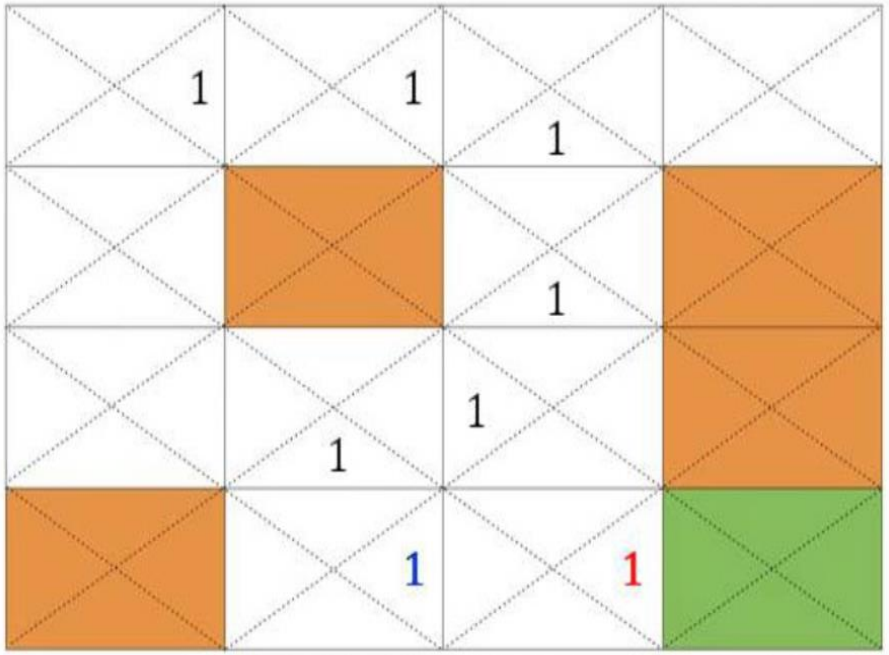




강화학습의 예시1 : Frozen Lake

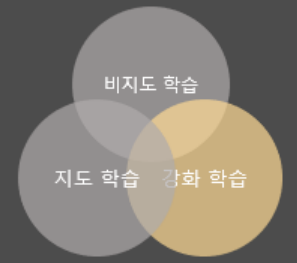
- 그렇게 계속 실패를 하다가 언젠가는 우연히 목표점에 도착하는 경우가 생길 것입니다.
- 그때 보상 점수가 부여하지만 도착점인 녹색 바로 전 까지만 점수를 부여합니다.
- 다시 학습을 시작하다 보면 또 우연히 빨간 1점이 있는 박스로 들어서는 경우가 생길 것입니다.
- 그때 보상 점수를 부여하지만 도착점인 빨간 1점 바로 전인 파란 1점까지만 점수를 부여합니다.

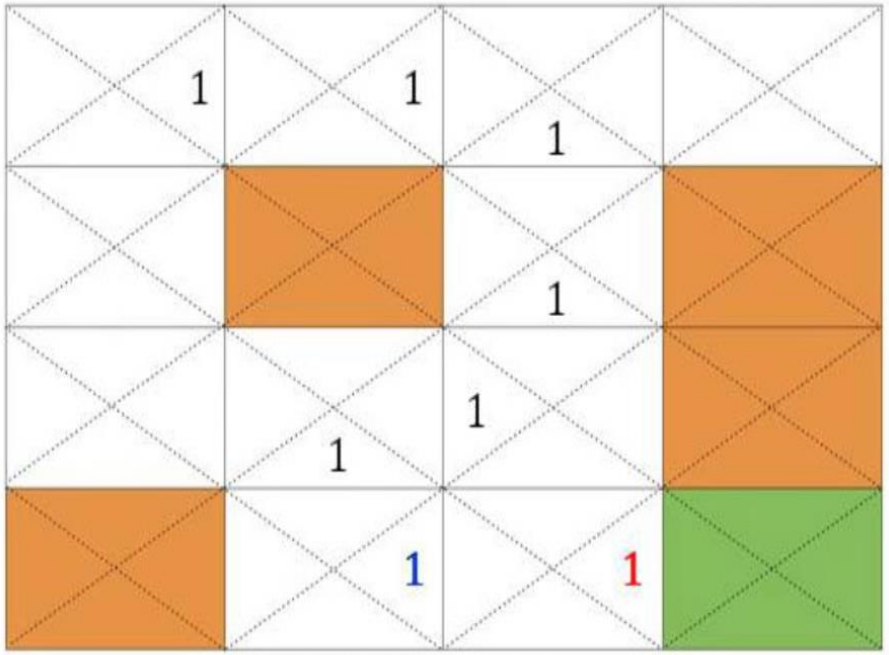




강화학습의 예시1 : Frozen Lake

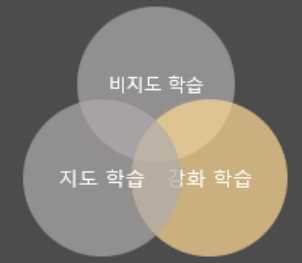
- 그때 보상 점수를 부여하지만 도착점인 빨간 1점 바로 전인 파란 1점까지만 점수를 부여합니다.
- 같은 방법을 반복하면 출발점에서 도착점으로 가는 경로가 만들어집니다.
- Frozen Lake는 간단한 게임이므로 이런 막무가내식으로 학습이 가능합니다.
- 복잡한 현실 문제는 이런 방법으로 성공하기 힘듭니다.

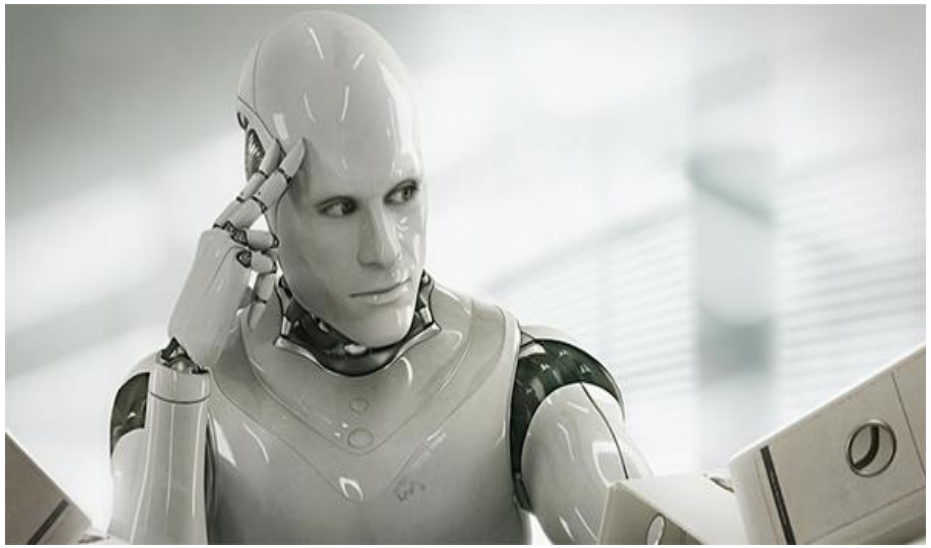




강화학습의 예시1 : Frozen Lake

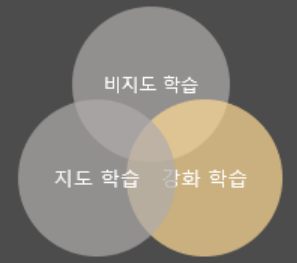
- 현실에서는 일정 기간 죽지 않고 살아만 있어도 보상을 준다 하던가 매번 확인된 길만 가는 것이 아니라 일정 비율로 안 가본 길을 가도록 하는 좀 더 복잡한 보상체계와 학습체계를 활용합니다.





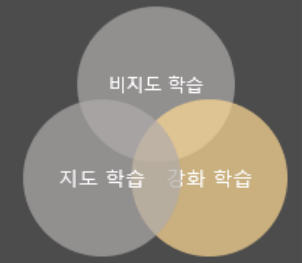
강화학습의 예시2 : 알파고

- 강화학습은 게임과 같은 제한된 조건을 가진 환경에서는 막강한 성능을 보여줍니다.
- 강화 학습을 기반으로 만들어진 인공지능인 알파고 제로가 지도 학습으로 학습한 알파고 마스터를 압도적으로 이기기도 하였습니다.
- 알파고 마스터는 3000만 기보를 보며 학습을 한 후 알파고 끼리의 셀프 대국을 통한 강화 학습을 하였습니다.



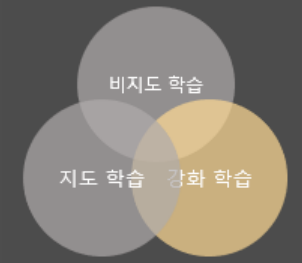
강화학습의 예시2 : 알파고

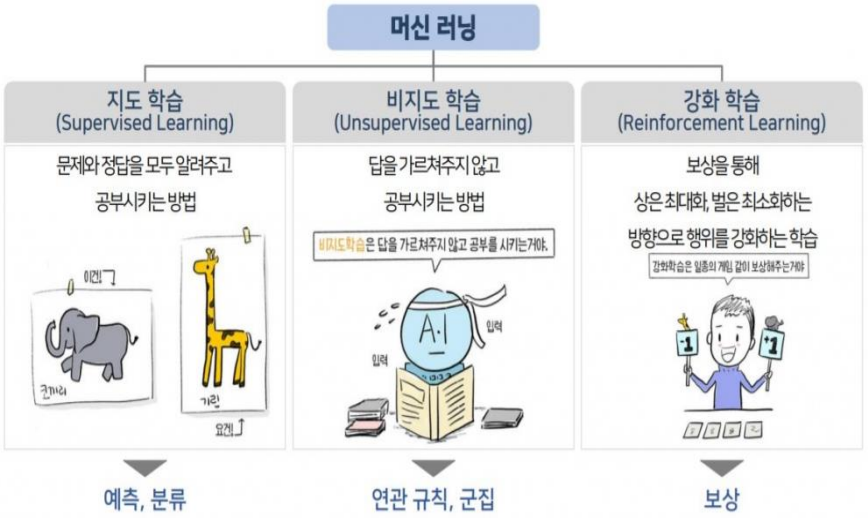
- 알파고 제로는 기보에 대한 학습과 훈련 없이 알파고 제로끼리 셀프 대국을 통한 학습을 하였습니다.
- 최종 승률이 가장 높은 수를 스스로 학습하고 바둑 이론을 업데이트 하였습니다.



강화학습의 예시2 : 자율주행 자동차

- 자율주행 기술 개발(심층 강화학습)





지도학습, 비지도학습, 강화학습 차이점

- 입출력 쌍으로 이루어진 훈련데이터 집합이 제시되지 않는다는 점에서 지도학습과 다르고 훈련데이터가 아주 없는 것이 아니라 상황 종료시에 종합적으로 주어진다는 점에서 비지도 학습과 다르다.