

기초 이론부터 실무 실습까지
머신 러닝 익히기

Part 02. 머신러닝 기초와 배경

정 정 민

Chapter 04. 머신러닝 종류

1. 레이블에 대한 이해
2. 지도 학습
3. 비지도 학습
4. 준지도 학습
5. 자기 지도 학습
6. 강화학습

레이블에 대한 이해

메타 데이터(Meta Data)와 레이블(Label)

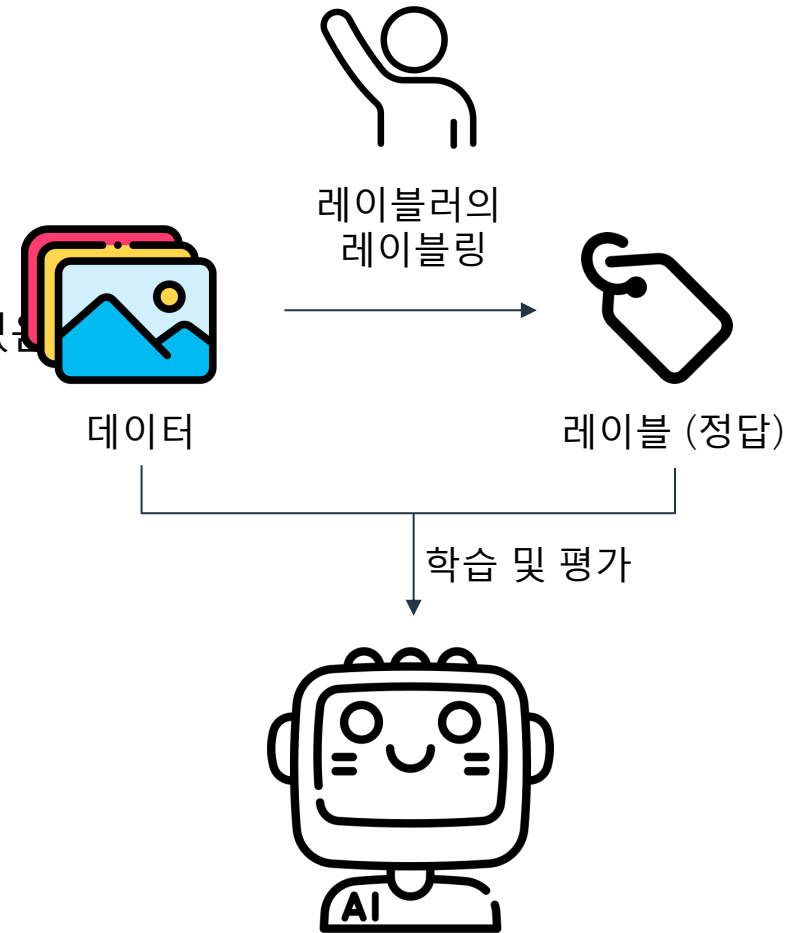
- 메타 데이터란 주어진 **기본 데이터에 추가적으로 제공하는 정보**를 의미
 - 데이터의 출처, 형식, 위치 등 데이터 간의 관계와 구조를 파악하거나
 - 데이터의 속성, 특성, 분류 등 데이터의 내용을 설명
- 레이블은 **특정 문제에 해당하는 데이터의 설명 혹은 답변**을 의미
 - 분류를 하는 문제라면 데이터가 속할 범주(클래스, class)
 - 목표 값을 찾는 회귀 문제라면 데이터가 표현할 특정 숫자
 - 등등
- 대부분 사람이 직접 생성해줘야 하는 경우가 많음
- 혹은 **타겟(target)이라고 부르기도 함**



지도 학습

지도 학습, Supervised Learning

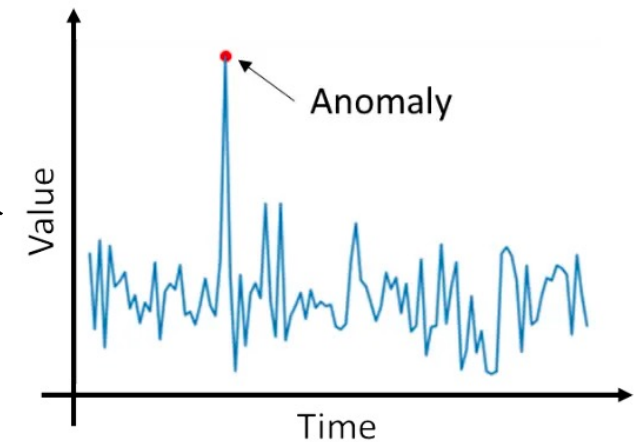
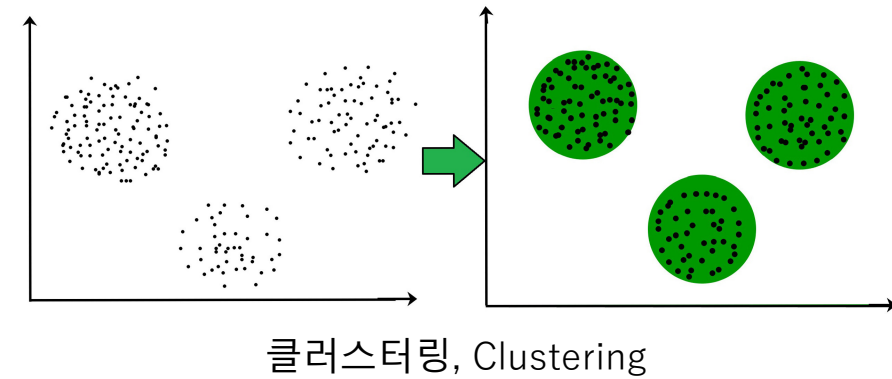
- 정답 레이블 정보를 활용해 알고리즘을 학습하는 학습 방법론
- 이 방법으로 학습되는 알고리즘은
데이터와 정답인 레이블 사이의 관계를 파악하는 목적을 갖고 있음
- 특징 및 장점
 - 정답이 존재하므로 모델이 풀어야하는 문제가 비교적 쉽고 잘 학습 됨
 - 또한, 명확한 평가 수치가 존재하며 학습된 모델의 성능을 쉽게 측정할 수 있음
- 단점
 - 정답이 필요하므로 이를 위해 추가적인 시간, 노동, 비용이 필요
 - 정답을 매기는 행위에 필요한 전문 인력과 같은 추가 비용이 발생



비지도 학습

비지도 학습, Un-Supervised Learning

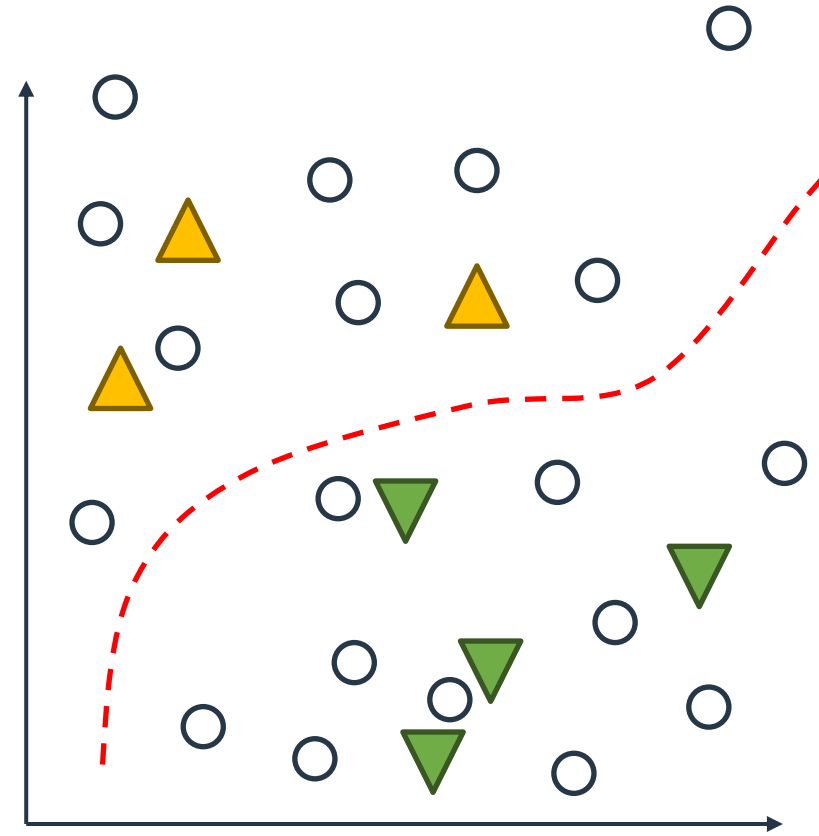
- 정답 레이블 정보가 없이 **입력 데이터만을 활용**해 알고리즘을 학습하는 학습 방법론
- 이 방법으로 학습되는 알고리즘은 **데이터 내부에 존재하는 패턴을 스스로 파악하는 목적**을 갖고 있음
- 특징 및 장점
 - 정답을 따로 준비할 필요가 없으므로 비용적인 이점이 있음
 - 사용자가 의도한 패턴 이외의 새로운 패턴을 찾을 가능성이 있으며 창작과 같은 다양한 활용 분야 사용할 수 있음
- 단점
 - 학습된 모델의 성능을 측정하기 위한 기준이 없어 결과 해석이 주관적일 수 있음
 - 신뢰할 수 있는 결과를 얻기 위해 다수의 데이터가 필요함



준지도 학습

준지도 학습, Semi-Supervised Learning

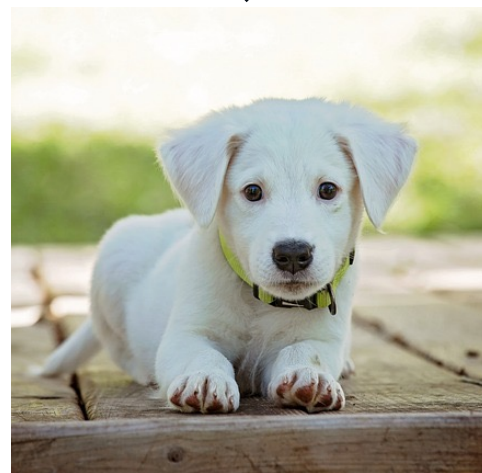
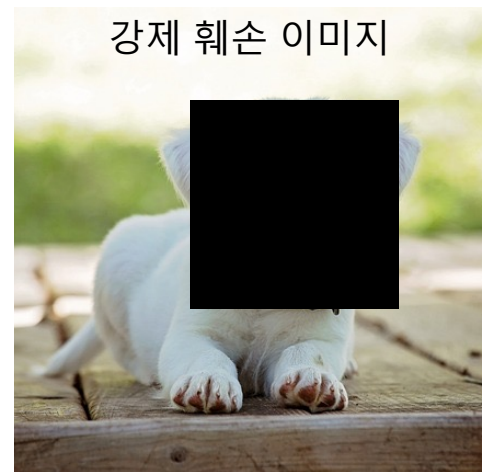
- 일부의 데이터만 정답이 존재하고, 다수의 데이터에는 레이블이 없는 상황에서 알고리즘을 학습하는 학습 방법론
- 이 방법으로 학습되는 알고리즘은 지도 학습과 비지도 학습의 특징을 일부 차용하여 일부 레이블링 된 데이터로 특성을 파악하고 레이블링 되지 않은 데이터로 전체 데이터의 패턴을 파악하는 방식으로 학습이 진행
- 특징 및 장점
 - 레이블이 부족한 데이터셋에서 유용
 - 많은 데이터를 활용할 수 있으므로 일반화 성능을 향상시킬 수 있음
- 단점
 - 품질이 낮은 레이블이나 데이터 존재에 특히 취약할 수 있음
 - 알고리즘의 복잡성이 증가하며, 구현 및 활용에 어려움이 있을 수 있음



자기 지도 학습

자기 지도 학습, Self-Supervised Learning

- 정답이 하나도 없는 데이터에서, **정답을 강제로 생성 후 학습**하는 학습 방법론
- 이 방법으로 학습되는 알고리즘은
데이터 내부를 강제로 훼손 후 복원하는 방법을 주로 사용하며
과정에서 **특정 데이터 내부의 성질을 파악하는데 사용됨**
- 이렇게 만들어진 알고리즘은 해당 데이터를 이용한 다른 문제에 적용
- 특징 및 장점
 - 레이블 없이 데이터의 특징을 파악할 수 있음
 - 다양한 데이터에 활용할 수 있음
- 단점
 - 목적하는 문제를 직접적으로 해결하는 것이 아니므로
N회 이상의 추가적인 학습 과정이 필요할 수 있음
 - 알고리즘이 잘못된 패턴을 학습할 위험이 있음



복원 이미지

강화 학습

강화 학습, Reinforcement Learning

- 어떤 환경(Environment)에서 상호작용하는 에이전트(Agent)가 보상(Reward)을 이용해 **특정 행동을 하도록 유도하는 학습 방법론**
- 우리 수업 과정에서 집중하는 학습론은 아님
- 쉬운 이해를 위해
- 강아지에게 ‘앉아’를 가르키는 강형욱 훈련사
 - 강아지 : 다양한 행동을 함
 - 강형욱 : 강아지가 앉는 행동을 하면 간식을 줌
 - 전개 상황
 - 강아지는 어쩌다가 먹은 간식을 더욱 많이 먹기위해
 - 본인이 간식을 먹을 수 있는 상황을 인지하게되고
 - 지속적으로 그런 행동을 반복
- 대표적 알고리즘 : 알파고



E.O.D