



02

기계학습의 이해

- + 학습 목표**
 - 기계학습의 개념을 설명할 수 있다.
 - 지도학습과 비지도학습의 개념을 설명하고 차이를 비교·분석할 수 있다.
- + 학습 요소**
 - 기계학습의 개념, 기계학습 유형, 지도학습, 비지도학습

생각 깨우기

다음 글과 그림을 보고 생각해 보자.

우리 주변에서 기계학습을 통해 인공지능이 어떤 값을 예측하거나 판단하는 모습을 쉽게 발견할 수 있다. 인공지능은 어떤 과정을 통해 학습하고 판단하는 것일까?



기계는 어떻게 스스로 규칙을 찾아내고 판단할 수 있을까?

1 | 기계학습의 개념

사람은 살면서 다양한 경험을 통해 지식을 축적하고, 이를 통해 결과를 판단하거나 예측한다. 직·간접적 경험을 통해 지속적으로 자각하고 인지하는 것을 학습이라고 한다. 사람이 경험을 토대로 학습하는 것과 같이 기계는 데이터를 통해 학습하는데, 이를 기계학습이라 한다.

기계학습(Machine Learning)이란 데이터로 학습한 것을 바탕으로 문제 해결을 위한 최적의 알고리즘을 찾아 특정 작업을 수행하는 인공지능의 한 분야다. 즉, 기계학습이란 컴퓨팅 시스템이 데이터로부터 규칙을 학습하고 모델을 만들고, 이를 통해 새로운 데이터에 대한 예측이나 판단을 내린다.

예를 들어, 사진을 보고 개와 고양이를 구분하고자 할 때, 기존의 프로그래밍 방식을 적용하면 강아지와 고양이의 특징이나 판단하는 방법을 직접 사람이 프로그래밍해 컴퓨터에게 알려 주어야 했다. 그러나 기계학습을 적용하면 컴퓨팅 시스템이 강아지와 고양이의 특징을 찾아내 자동으로 구분할 수 있는 규칙을 스스로 만들어내 새로운 데이터에 적용할 수 있다. 즉, 데이터와 규칙을 입력해 원하는 결과를 출력하는 전통적인 프로그래밍 방식에 비해, 기계학습은 데이터와 원하는 결과를 알려 주면 스스로 사고하기 위한 규칙을 찾아낸다.

⚙️ 기계학습에서 학습이란 해결할 문제와 관련된 충분한 양의 데이터를 모아 학습 알고리즘에 입력한 후 컴퓨팅 시스템이 이를 이용해 인공지능 모델을 만들어내는 과정을 말한다.

🔍 알고리즘

어떤 일을 처리하는 절차나 규칙, 방법을 논리적으로 표현한 것이다.



▶ 전통적인 프로그래밍과 기계학습의 문제 해결 방식의 차이



해 보기 1 | 기계학습을 이용해 손동작을 구분할 수 있을지 토론하기

🗣️ 손 모양 이미지를 입력하면 가위, 바위, 보로 구분해 주는 기계학습 프로그램을 만들려고 한다. 인공지능은 어떤 규칙을 이용해 가위, 바위, 보로 구분할 수 있을지 이야기해 보자.

2 | 기계학습의 유형

기계학습은 데이터를 이용하여 학습할 때 정답이 주어지는지 여부에 따라 지도학습과 비지도학습 등으로 구분할 수 있다. 각 데이터에 대한 정답을 함께 알려 주고 학습하는 것을 지도학습, 정답을 알려 주지 않고 데이터만을 제시해 학습하도록 하는 것을 비지도학습이라고 한다.

01 지도학습

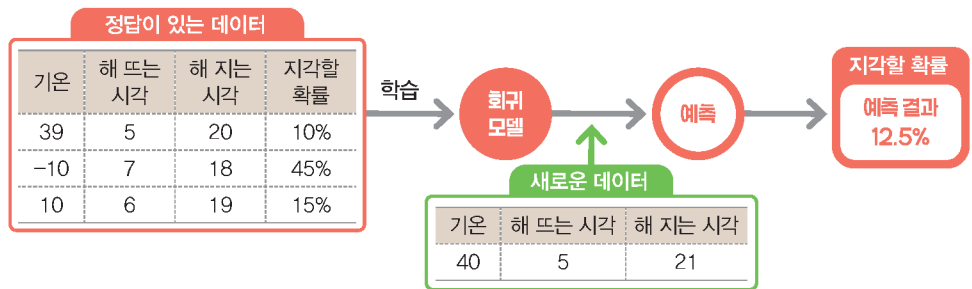
지도학습은 정답이 있는 데이터를 이용해 학습하는 것이다. 지도학습을 이용하면 데이터를 학습시켜 결과를 판단하는 모델을 만들 수 있다. 학습 모델을 만드는 데 사용하는 데이터를 '훈련 데이터'라고 한다. 만들어진 모델의 성능을 테스트할 때 사용하는 데이터를 '테스트 데이터'라고 한다. 지도학습의 경우 입력 데이터에 정답에 대한 데이터가 함께 주어지기 때문에, 학습 결과에 오류가 있는지 없는지 확실히 파악할 수 있다.

지도학습에는 회귀와 분류 두 유형이 있다.

1 회귀

회귀(Regression)는 속성들 사이의 상관 관계를 표현하는 함수식을 학습 알고리즘을 통해 구하고, 그에 따라 특정 값을 예측하는 기계학습 방법이다. 훈련 데이터로 학습한 모델을 이용해 새로운 입력 값에 대한 정답을 예측해 답을 생성하는 것으로, 예측이라고도 한다. 최근 며칠간의 기온과 해가 뜨고 지는 데이터를 바탕으로 다음 날 지각할 확률을 예상해 보는 문제는 회귀 알고리즘을 이용해 해결할 수 있다.

Q 기온과 해가 떠 있는 시간에 따라 지각할 확률은 어떻게 될까?



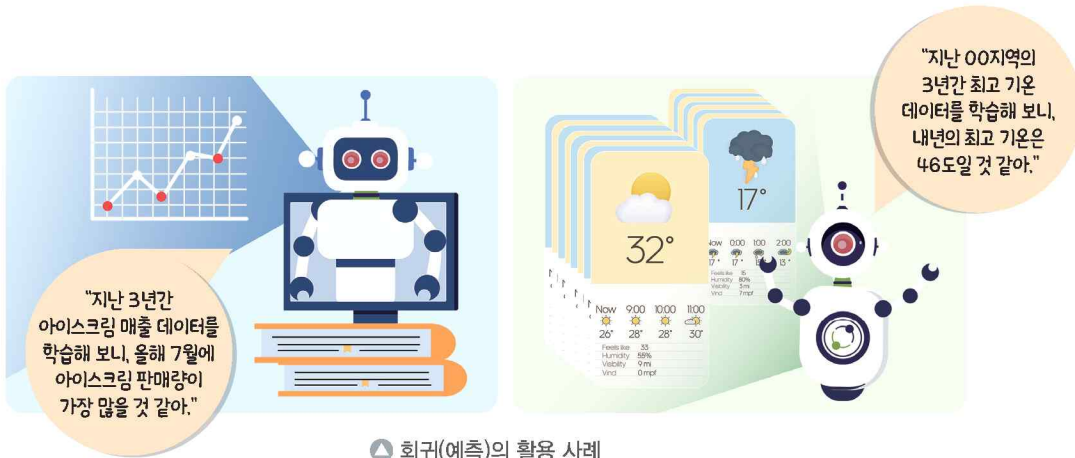
A 기온, 해 뜨는 시간, 해 지는 시간과 지각할 확률 간의 관계를 파악해 새로운 데이터에 대한 지각할 확률을 구할 수 있다.

▲ 회귀(예측)의 원리

❗ 모델이란 주어진 데이터를 기반으로 학습하고 새로운 데이터에 대한 결과를 알아내는 과정을 프로그램으로 만든 것이다.

❗ 지도학습을 위해 사용하는 데이터에는 정답에 대한 정보가 포함되어 있어야 한다.

❗ 데이터를 구별하는 속성을 레이블이라고 하고, 데이터를 구별하기 위해 정답을 부여하는 과정을 레이블링 (labeling)이라고 한다.



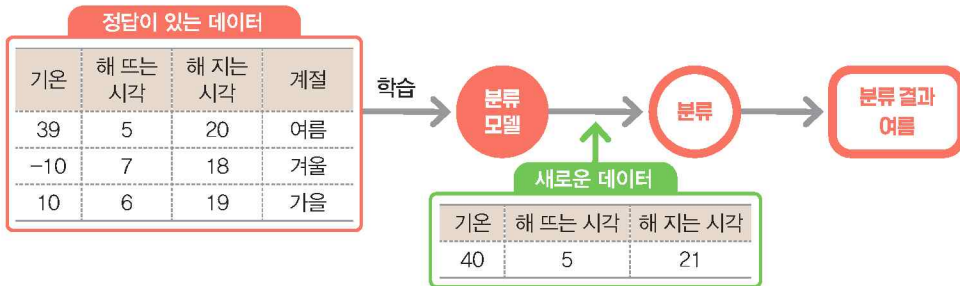
▲ 회귀(예측)의 활용 사례

2 분류

분류(Classification)는 고정된 레이블을 나누기 위한 분류 선을 학습하여 새로운 입력이 들어오면 여러 그룹 중 하나로 선택하는 과정을 통해 판단한다. 훈련 데이터를 이용해 기계학습을 진행한 후, 새로운 데이터가 어떤 그룹에 속하는지 판단하는 것이다. 즉, 데이터 속성의 관계를 파악하고 새로운 데이터가 어느 그룹에 속하는지 알아낸다.

❗ 분류에는 이진 분류와 다중 분류가 있다. 이진 분류(binary classification)는 정답 레이블이 두 클래스 중 하나로 분류되는 경우를 말하며, 다중 분류는 클래스가 3개 이상일 경우를 말한다.

❓ 기온과 해가 떠 있는 시간에 따라 계절을 구분할 수 있을까?



A 기온, 해 뜨는 시간, 해 지는 시간과 계절 간의 관계를 파악해 새로운 입력 값에 대해 계절을 분류할 수 있다.

▲ 분류의 원리



▲ 분류의 활용 사례

❗ 특정 단어들의 포함 여부, 내용 등에 따라 메일이 정상 메일인지, 스팸 메일인지 분류한다. 이는 이진 분류에 해당한다.

02 비지도학습

비지도학습은 정답이 없는 데이터를 입력하면 컴퓨팅 시스템이 자신만의 방식으로 유사한 속성값을 갖는 데이터끼리 모아 그룹을 만든다. 정답 데이터가 주어지지 않기 때문에 스스로 입력 데이터의 특징을 분석하여 패턴이나 규칙을 찾아내야 한다. 따라서 비지도학습은 훈련 데이터 속에 숨겨진 새로운 정보를 찾고자 할 때 유용하다. 비지도학습을 통해 군집 등의 작업이 가능하다.

☞ 군집

입력받은 데이터 간의 관계를 파악해서 비슷한 것끼리 묶어 주는 것을 말한다.

ⓐ 속성의 관계를 보고 비슷한 데이터를 세 그룹으로 모을 수 있을까?

ⓐ 구분되어 있지 않은 기온 관련 데이터를 여러 속성의 관점에서 비슷한 것끼리 묶어 어떻게 그룹을 이루는지 확인하고, 새로운 정보를 파악해 활용할 수 있다.

▲ 군집의 원리

데이터를 나누기 위한 명확한 기준이 없는 상태에서, 주어진 데이터의 특성을 고려해 유사한 데이터 그룹으로 나누는 것을 군집이라고 하고, 이때 나누어진 각각의 그룹 또한 군집이라고 한다. 군집은 인구 밀집도를 분석해 버스 노선을 결정하거나 SNS 조회 기록을 이용해 비슷한 성향의 게시물을 추천하는 등의 사례에 많이 활용한다.

특성이 유사한 과일끼리 나누어 볼까?
(빨강과 동그란 과일) (겉면이 보록한 과일) (노란색 과일)

▲ 군집의 활용 사례

소단원 1 요약

- 1 기계학습(Machine Learning)이란 데이터로 학습한 것을 바탕으로 문제 해결을 위한 최적의 알고리즘을 찾아 특정 작업을 수행 하는 인공지능의 한 분야다.
- 2 지도학습은 정답(레이블)이 있는 데이터를 이용해 학습하는 것으로, 회귀(예측), 분류가 이에 속한다.
- 3 비지도학습은 정답을 알려 주지 않고 데이터만을 제시하고 컴퓨터가 스스로 입력 데이터의 특징을 분석하여 패턴이나 특정 규칙을 찾아내는 것으로, 군집이 이에 속한다.

소단원 자기 평가

평가 항목	평가 기준		
	잘함	보통	노력
1. [지식·이해] 기계학습의 개념과 유형을 설명할 수 있다.			
2. [과정·기능] 기계학습으로 해결할 수 있는 문제의 유형을 비교할 수 있다.			

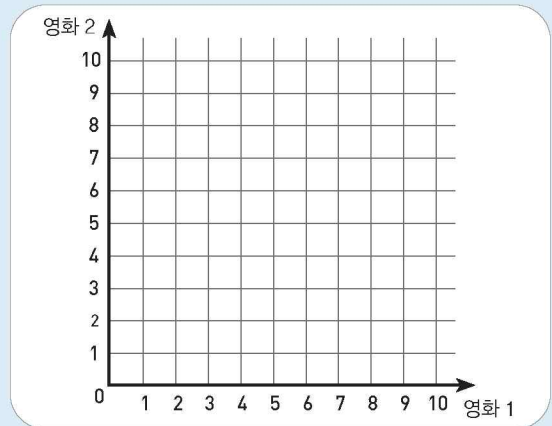


영화 평점 데이터를 통해 기계학습 이해하기

두 장르의 영화에 대해 10명의 관객이 평점을 매긴 데이터가 있다. 영화 평점 데이터를 이용하여 관객의 선호도를 2개의 군집으로 나누고, 11번째로 입력되는 새로운 데이터가 2개의 그룹 중 어떤 선호도 그룹에 속하는지 분류해 보자(평점은 0점~10점이다).

구분	관객 1	관객 2	관객 3	관객 4	관객 5	관객 6	관객 7	관객 8	관객 9	관객 10
영화 1	1	2	8	8	8	7	9	2	1	3
영화 2	10	8	3	2	4	2	6	9	9	7

1 오른쪽 그래프에서 영화 1, 영화 2의 관객 평점을 표시해 보자.



2 1에서 표시한 비슷한 위치의 점들끼리 묶어 2개의 그룹(A, B)으로 나눈 후 각 그룹의 특징을 써 보자.

A그룹에 속한 관객 번호	예 관객 1, 2, 8, 9, 10	B그룹에 속한 관객 번호	예 관객 3, 4, 5, 6, 7
A그룹의 특징	예 영화 1에 대한 평점이 낮고, 영화 2에 대한 평점이 높다.	B그룹의 특징	예 영화 1에 대한 평점이 높고, 영화 2에 대한 평점이 낮다.

3 2에서 작성한 결과 영화 1, 영화 2에 대한 평점이 모두 10인 11번째 관객이 있을 때, 11번째 관객이 어떤 그룹에 속하는지 분류하고 이유를 설명해 보자.

탐구 활동 자기 평가	평가 항목	평가 기준		
		잘함	보통	노력
	1. 군집의 원리를 이해하고 실생활의 데이터를 이용해 군집을 형성할 수 있는가?			
	2. 분류의 원리를 이해하고 실생활의 데이터를 분류한 결과를 판단할 수 있는가?			

• 잘함 (내용을 이해하고 설명함) • 보통 (내용을 이해함) • 노력 (내용을 부분적으로 이해함)