



부산대학교

PUSAN NATIONAL UNIVERSITY

Types of Learning

Labels, Structure, Rewards

What the Data Gives You

BX3600087-140 / DS3600087-140 / AB3600087-140

기계학습개론

2026학년도 2학기

오용경

데이터사이언스전공

yongkyung.oh@pusan.ac.kr

Learning Problems

같은 응용에서도 문제 설정에 따라 학습 유형이 달라진다

같은 응용, 다른 학습 문제

같은 추천 서비스에서 서로 다른 학습 문제(learning problem) 세 가지가 나온다

예측한다

과거 클릭 결과로 다음 클릭 여부를 예측

찾아낸다

사용자 이용 기록에서 비슷한 집단을 찾음

행동한다

콘텐츠를 선택하고 이후 보상(reward)을 관측

무엇을 관측하고 무엇을 알아내려는지에 따라 학습 문제가 달라진다

학습 문제의 구성

어떤 자료를 가졌고 무엇을 알아내려는가

자료

입력(input)과 함께 클릭 결과가 기록되어 있는가



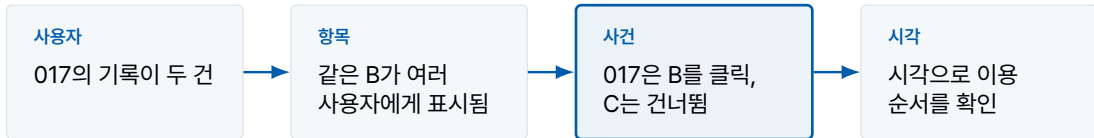
목표

클릭을 예측할 것인가, 비슷한 사용자를 찾을 것인가

클릭 예측에는 클릭 결과가 맞힐 출력으로 필요하지만, 사용자 묶기에서는 클릭 기록도 입력으로만 쓴다

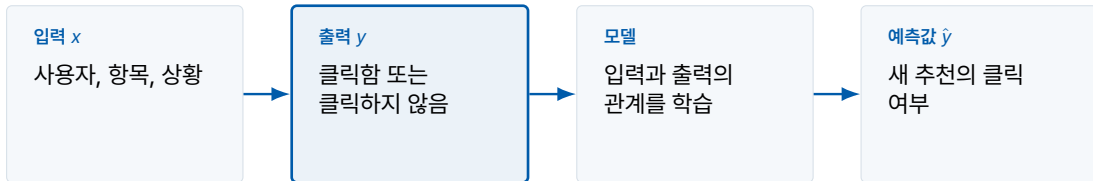
추천 기록에 남는 사건

사용자	항목	상황	사건	시각
사용자 017	항목 B	모바일, 저녁	클릭	20:14
사용자 017	항목 C	모바일, 저녁	건너뛸	20:15
사용자 204	항목 B	웹, 아침	건너뛸	08:31



클릭 예측

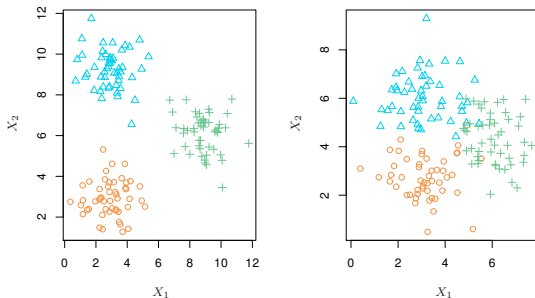
과거 사례의 클릭 여부로 모델을 학습하고, 새 추천의 클릭 여부를 예측



Source: Zhang et al. (2023), Chapter 1.

비슷한 사용자 찾기

정답 출력 없이 이용 기록이 비슷한 사용자를 찾는 경우

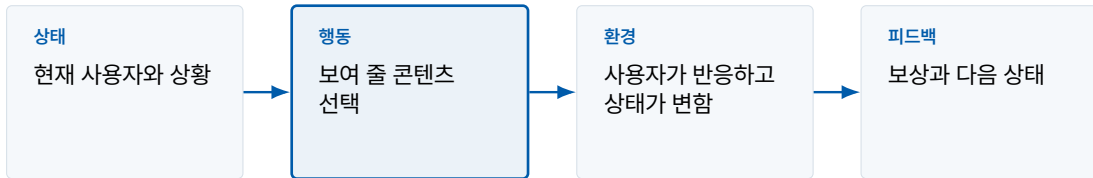


두 변수로 나타낸 예시 자료. 색은 실제 집단 표시이며 군집화는 이 표시 없이 위치만으로 집단을 찾음. 오른쪽은 집단이 겹쳐 찾기 어려움

Source: James et al. (2023), Figure 2.8.

보여 줄 콘텐츠 선택

에이전트의 선택이 이후에 관측할 자료까지 바꾸는 경우



Source: Zhang et al. (2023), Chapter 1.

응용과 학습 유형

학습 문제	관측 자료	함께 관측되는 값	유형
다음 클릭 예측	사용자, 항목, 상황	클릭 결과 y	지도학습
사용자 집단 탐색	사용자 이용 기록 벡터	정답 출력 없음	비지도학습
콘텐츠 순차 선택	상태, 행동	행동 이후 보상	강화학습

클릭 예측에는 과거 결과를, 사용자 묶기에는 이용 기록의 유사성을, 추천 선택에는 보상을 사용

Checkpoint

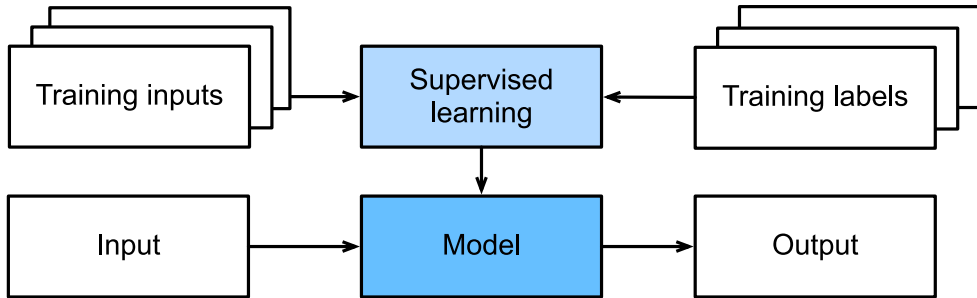
Learning Problems

- 예측하거나 선택할 학습 문제는 무엇인가
- 모델이 받는 자료에는 어떤 값이 있는가
- 출력 또는 보상은 언제 관측되는가

Section 2 • Types of Learning

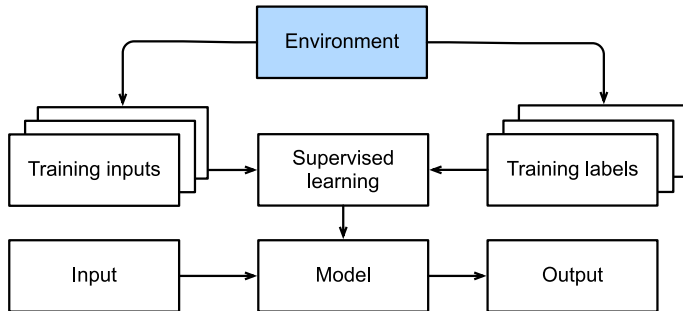
Supervised Data

학습 사례의 출력과 모델의 예측값을 연결



오른쪽 위 학습 자료의 정답 기록은 관측된 출력 y , 오른쪽 아래에서 모델이 내놓는 결과는 예측값 $\hat{y}$. 예측에는 새 입력만 넣음

학습 자료가 만들어지는 과정



실제 환경에서 입력과 그에 대응하는 출력을 모아 학습 사례를 구성. 모은 뒤에는 환경과 다시 상호작용하지 않음

Source: Zhang et al. (2023), Figure 1.3.6.

출력 y 와 예측값 $\hat{y}$

학습 자료에 기록된 출력 y (label 또는 target이라고도 부름)는 모델이 계산한 예측값 $\hat{y}$ 와 구별

관측된 출력

$$y^{(i)}$$

실제 클릭 여부. 클릭했으면 $y^{(i)} = 1$

예측값

$$\hat{y}^{(i)}$$

클릭하지 않는다고 예측하면 $\hat{y}^{(i)} = 0$

실제로 클릭했는데 모델이 클릭하지 않는다고 예측했다면 $y^{(i)}$ 와 $\hat{y}^{(i)}$ 가 다르다

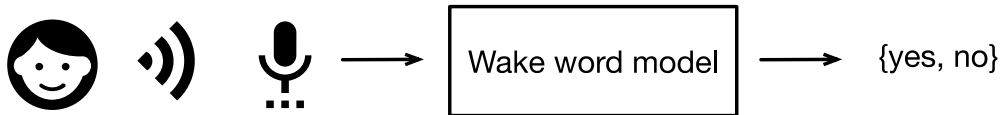
하나의 학습 사례

특성(feature)인 나이, 콘텐츠 종류, 상황이 $x^{(3)}$ 을 이루고, 클릭 여부가 $y^{(3)}$

번호	사용자 나이	콘텐츠 종류	상황	클릭 여부
$i = 3$	24	과학	모바일	$y^{(3)} = 1$

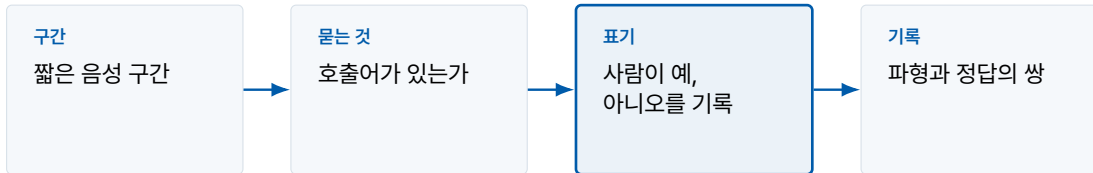
$$\mathcal{D} = \left\{ \left(x^{(1)}, y^{(1)} \right), \dots, \left(x^{(n)}, y^{(n)} \right) \right\}$$

호출어 인식 문제



음성 구간의 파형은 입력, 호출어(wake word) 포함 여부는 두 값 중 하나인 출력

음성 구간에 정답을 붙이는 과정



Adapted from Zhang et al. (2023), Section 1.1.

출력 y 는 어떻게 얻는가

예측할 값 y 는 측정하거나, 사람이 판정하거나, 이후에 발생한 결과를 기록해 얻는다

측정

센서나 장비가 수치를 기록

표기

이미지, 글, 음성을 사람이
보고 판정

결과

구매, 클릭, 회복 등 이후의
결과를 기록

관측한 출력을 입력과 짝지으면 하나의 학습 사례(example)가 된다

예측값과 관측값의 차이

$$\underbrace{\hat{y}^{(i)} = \hat{f}(x^{(i)})}_{\text{예측값}} \quad \underbrace{y^{(i)}}_{\text{관측값}} \quad \underbrace{\ell(y^{(i)}, \hat{y}^{(i)})}_{\text{손실(loss)}}$$

예측값

모델이 계산한 값

관측값

자료에 기록된 값

손실

예측이 관측값과 얼마나
어긋나는지 수치로 평가

Checkpoint

Supervised Data

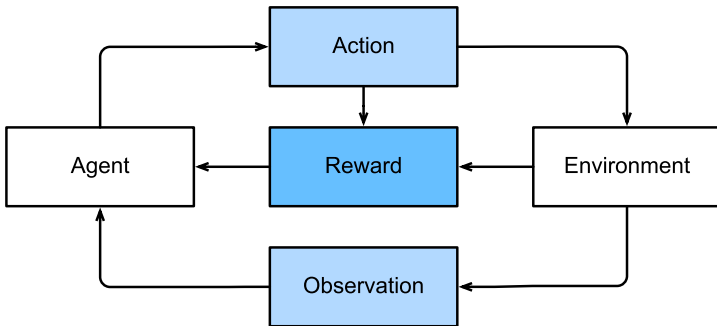
- 호출어 문제에서 예측할 출력은 무엇인가
- 학습 기록에서 관측된 $y^{(i)}$ 는 어느 값인가
- 새 음성 구간에서 모델이 내놓는 값은 무엇인가

Section 3 • Types of Learning

Actions and Rewards

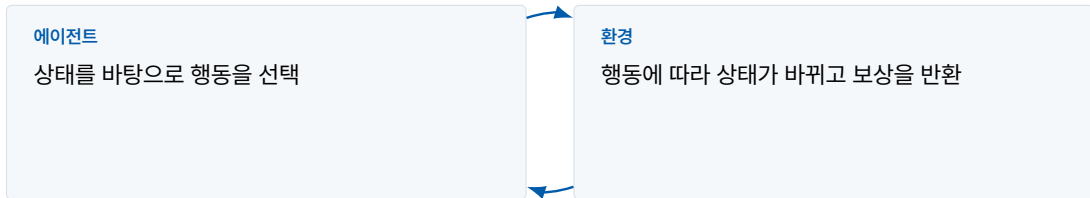
보상을 바탕으로 더 나은 행동을 고르는 학습

강화학습의 구조



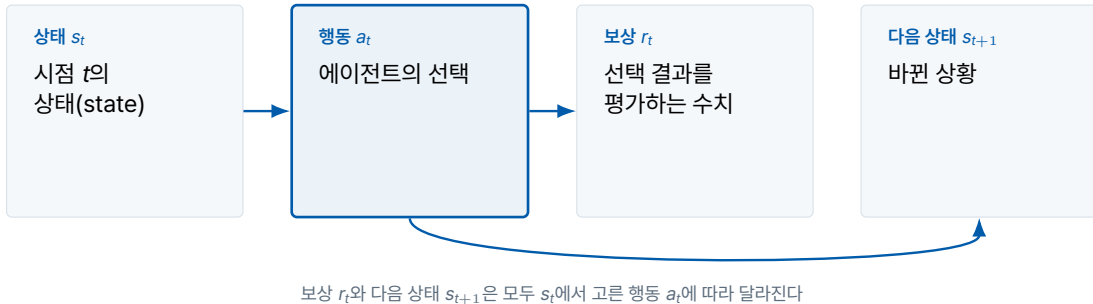
강화학습(reinforcement learning)은 에이전트가 환경에서 받은 보상으로 행동 선택을 학습. 그림의 observation은 에이전트가 받는 상태 정보

에이전트와 환경



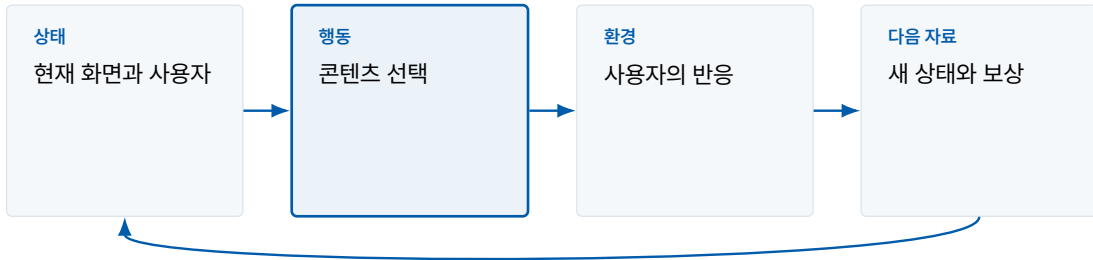
Source: MIT 6.390 (2026), Reinforcement Learning.

상호작용 한 단계



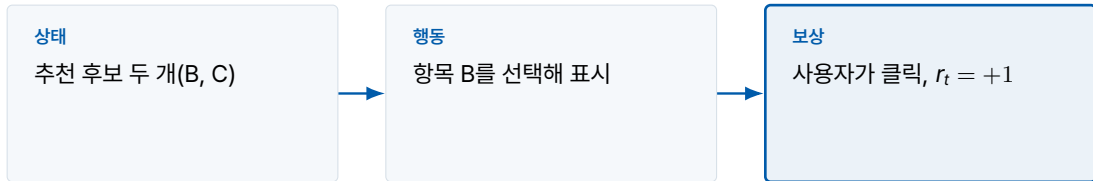
Source: MIT 6.390 (2026), Introduction.

행동이 다음 상태를 바꾼다



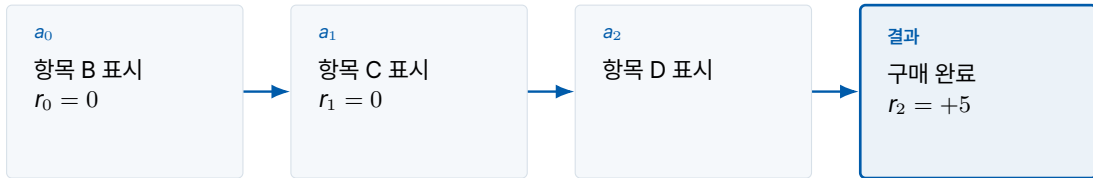
Source: Zhang et al. (2023), Figure 1.3.7.

즉시 관측되는 보상



추천한 항목을 곧바로 클릭하면, 그 선택에 대한 보상 +1을 기록

지연된 보상



B, C를 보여 준 결과는 곧바로 드러나지 않고, 여러 번 추천한 뒤 구매가 일어나야 보상을 관측

Checkpoint

Actions and Rewards

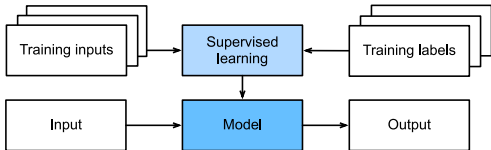
- 보상은 행동 전과 후 중 언제 관측되는가
- 보상이 여러 행동 뒤에야 관측되는 경우는 언제인가
- 행동이 다음 상태를 어떻게 바꾸는가

Section 4 • Types of Learning

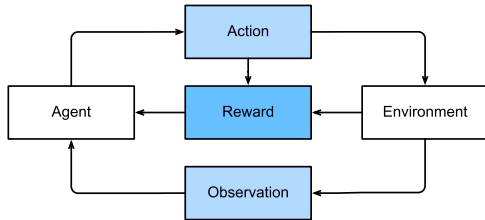
Classifying Problems

사례의 자료와 목표로 학습 유형을 판정

지도학습 자료와 보상의 구조 비교



출력 y 를 보고 입력에서 무엇을 예측해야 하는지 학습



보상을 보고 어떤 행동 선택이 좋은지 학습

지도학습의 출력과 보상의 차이

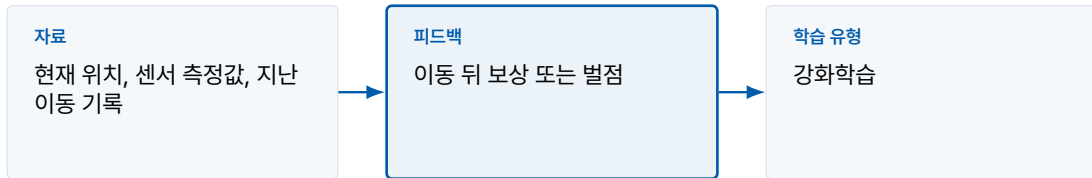
기준	출력 y	보상
역할 클릭 예시 학습 목표	입력에 대해 예측해야 할 값 이 사용자가 클릭했는가 새 입력의 출력 예측	행동의 결과를 평가하는 값 이 추천 뒤 클릭하면 보상 +1 앞으로 받을 보상이 커지도록 행동을 선택
주어진 정보	각 입력에 대응하는 출력	선택의 결과. 최선의 행동을 직접 알려 주지는 않음

늦게 관측된 값도 정답이 될 수 있으므로, 관측 시점만으로 두 학습 유형을 가르치지 않는다

Source: Zhang et al. (2023), Chapter 1.

로봇 이동

목표에 도착하면 보상을, 충돌하면 벌점을 주어 이동 선택을 학습하는 경우



다음 위치를 예측하는 것이 아니라, 충돌을 피하며 목표에 도착할 이동을 선택하는 문제

학습 유형 판정 기준

어떤 자료가 주어졌는지와 무엇을 학습하려는지를 함께 확인

입력과 출력

x, y 의 쌍으로 새 y 를
예측하면 지도학습

행동과 보상

보상을 높일 선택을 학습하면
강화학습

입력만

입력의 구조를 찾으면
비지도학습

같은 클릭 기록도 클릭 예측에 쓰는지, 추천 선택을 평가하는 데 쓰는지에 따라 역할이 달라진다

Checkpoint

Classifying Problems

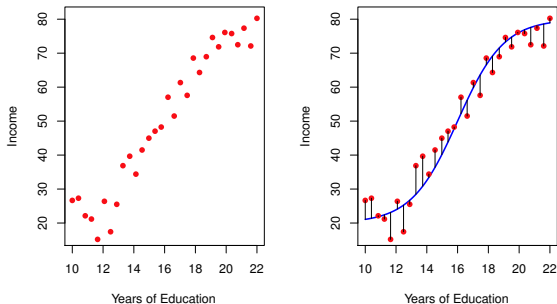
- 광고 예산으로 판매량 예측
- 암 종류 정보 없이 유전자 발현으로 집단 탐색
- 게임 상태에서 점수를 높일 행동 선택을 학습
- 거래 특성으로 다음 구매 금액 예측

사례 판정 해설

사례	학습 유형	판정 근거
광고 예산과 판매량	지도학습, 회귀	판매량 y
유전자 발현의 집단 탐색	비지도학습	입력의 구조를 탐색
게임 상태와 점수	강화학습	점수를 높일 행동 선택을 학습
다음 구매 금액 예측	지도학습, 회귀	구매 금액 y

연속값을 예측하는 문제

교육 연수(Years of Education)로 소득(Income)을 예측하면 지도학습의 회귀

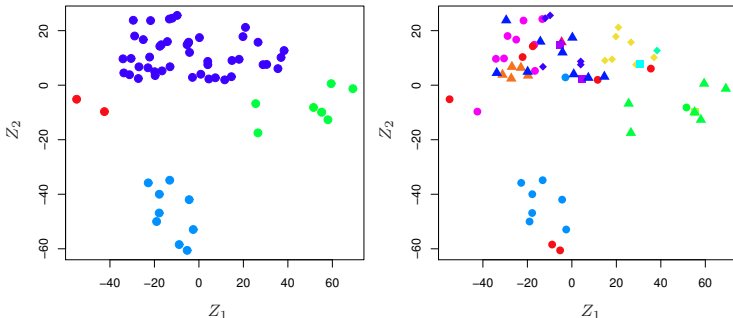


왼쪽은 관측값, 오른쪽 파란 곡선은 모의 자료를 만들 때 쓴 참 관계, 검은 선은 각 관측값과 참 관계의 차이

Source: James et al. (2023), Figure 2.2.

출력 없이 구조를 찾는 문제

유전자 발현만으로 암 세포주의 집단을 찾으면 비지도학습

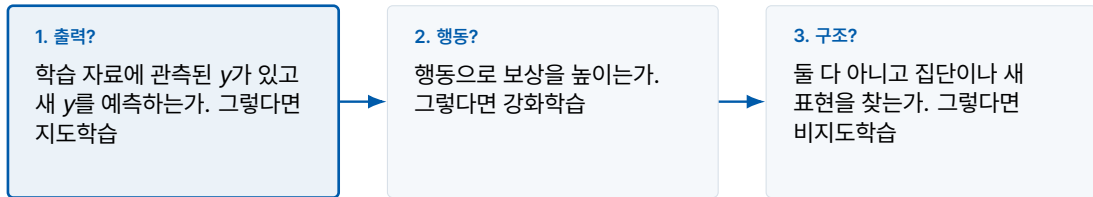


왼쪽 배치는 암 종류를 사용하지 않고 얻었고 색은 그 배치에서 보이는 네 집단, 오른쪽은 분석 뒤 암 종류와 비교한 결과

Source: James et al. (2023), Figure 1.4.

학습 유형 판정 순서

출력 예측, 행동 선택, 구조 탐색의 순서로 학습 목표를 확인



관측 시점보다 그 값을 무엇을 학습하는 데 쓰는지가 판정 기준

세 학습 유형의 비교

기준	지도학습	비지도학습	강화학습
자료 학습에 쓰는 기준 목표	입력과 출력 y 예측할 값 y 새 입력의 출력 예측	입력만 관측 입력 사이의 유사성 등 입력 속 구조 발견	상태, 행동, 보상 선택 결과의 보상 앞으로 받을 보상이 커지도록 선택
예	회귀, 분류	군집화	콘텐츠 선택, 로봇 이동

출력을 예측하면 지도학습, 입력의 구조를 찾으면 비지도학습, 보상을 높이는 선택을 배우면 강화학습

Source: Zhang et al. (2023), Chapter 1.

How Do Vectors and Matrices Represent Data and Models?

Next Lecture: Vectors and Matrices

- 벡터와 행렬의 기본 연산
- 벡터와 행렬의 모양과 인덱싱
- 내적과 행렬 곱