

강원지역혁신플랫폼

기계학습

Machine Learning

기계학습의 기본 개념



▶ 학습목표

📁 기계학습의 기본 개념을 설명할 수 있습니다.





01 | 기계학습의 기본 개념 소개



기계학습의 기본 개념 소개

△ ‘기계학습(Machine Learning)’ 이란 말을 들으면 대부분의 사람들은 로봇을 떠올림

◆ 어떤 사람은 믿음직한 집사 로봇을 생각할 것임

➢ 누구는 살인 병기 터미네이터를 생각할 것임



[집사 로봇]



[터미네이터]



01 | 기계학습의 기본 개념 소개



기계학습의 기본 개념 소개

- △ 수억 명의 생활을 편리하게 만들어 주류가 된 첫 번째 기계학습 애플리케이션은 1990년대의 스팸 필터(Spam Filter)임
 - ◆ 이 애플리케이션은 스스로 생각하는 스카이넷(Skynet) 정도는 아니지만 기술적으로 기계학습이라 할 수 있음
 - 이후 추천과 음성 검색으로 발전했음
 - ─ 매일 사용하는 많은 제품(전자 제품 등)과 기능을 소리 없이 향상시킨 수만 개의 기계학습 애플리케이션이 등장함
 - ➔ 스카이넷은 영화 '터미네이터' 시리즈에 나오는 인공지능 슈퍼컴퓨터임



01 | 기계학습의 기본 개념 소개



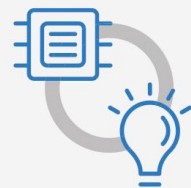
기계학습의 기본 개념 소개

△ 기계학습은 데이터를 이해하는 알고리즘의 과학이자 애플리케이션임

◆ 컴퓨터 과학 전체에서 가장 흥미진진한 분야 중의 하나임

▶ 우리는 데이터가 풍부한 시대에 살고 있음

─ 스스로 학습할 수 있는 기계학습 알고리즘을 사용하면 이 데이터를 지식으로 바꿀 수 있음



기계학습은
데이터를 지식으로
바꿀 수 있음



01 | 기계학습의 기본 개념 소개



기계학습의 기본 개념 소개

△ 기계학습의 의미에 대해서 생각해보자.

- ◆ 기계학습은 어디서 시작하고 어디서 끝나는 걸까?
- ◆ 기계가 배운다는 것이 정확히 무엇을 의미하는 걸까?
- ◆ 내가 위키백과 문서를 내려 받으면 내 컴퓨터가 실제로 무언가를 배울 수 있을까?
- ◆ 컴퓨터가 갑자기 똑똑해질 수 있을까?



이러한 질문에 답하기 위해 전체 기계학습의 그림을 조망하고
주요 영역과 핵심 개념들을 살펴보자.



02 | 기계학습이란?



기계학습이란?

△ 기계학습은 데이터에서부터 학습하도록 컴퓨터를 프로그래밍하는 과학임

◆ 조금 더 일반적인 정의는 다음과 같음

“기계학습은 명시적인 프로그래밍 없이 컴퓨터가 학습하는 능력을 갖추게 하는 연구 분야다.”

아서 새뮤얼 (Arthur Samuel, 1959)

◆ 조금 더 공학적인 정의는 다음과 같음

“어떤 작업 T 에 대한 컴퓨터 프로그램의 성능을 P 로 측정했을 때 경험 E 로 인해 성능이 향상됐다면, 이 컴퓨터 프로그램은 작업 T 와 성능 측정 P 에 대해 경험 E 로 학습한 것이다.”

톰 미첼 (Tom Mitchell, 1997)



02 | 기계학습이란?



기계학습이란?

- △ 스팸 필터는 스팸 메일과 일반 메일의 샘플을 이용해 스팸 메일 구분법을 배울 수 있는 기계학습 프로그램임
 - ◆ 시스템이 학습하는 데 사용하는 샘플을 **훈련 세트**(training set)라고 부름
 - 여기서 각 훈련 데이터를 **훈련 사례**(training instance) (혹은 샘플)라고 함
 - ◆ 스팸 필터에서 **작업 T**는 새로운 메일이 스팸인지 구분하는 것임
 - **경험 E**는 **훈련 데이터**(training data)임
 - **성능 측정 P**는 직접 정의해야 함
 - ➔ 예를 들면 **정확히 분류된 메일의 비율**을 **P**로 사용할 수 있음(정확도)



03 | 왜 기계학습을 사용하는가?

왜 기계학습을 사용하는가?

⚙️ 전통적인 프로그래밍 기법을 사용해 어떻게 스팸 필터를 만들 수 있을지 생각해보자.

1 스팸에 어떤 단어가 주로 나타나는지 살펴봄

➤ 예를 들면 ‘4U’, ‘신용카드’, ‘무료’, ‘굉장한’ 같은 단어나 구절이 제목에 많이 나타나는 경향이 있다는 것을 알 수 있음

➤ 또한, 보낸이의 이름, 메일 주소, 본문이나 이메일의 다른 요소에서 다른 패턴을 감지할 수 있음

2 발견한 각 패턴을 감지하는 알고리즘을 작성해 프로그램이 이런 패턴을 발견했을 때 그 메일을 스팸으로 분류함

3 프로그램을 테스트하고 론칭할 만큼 충분한 성능이 나올 때까지 1단계와 2단계를 반복함



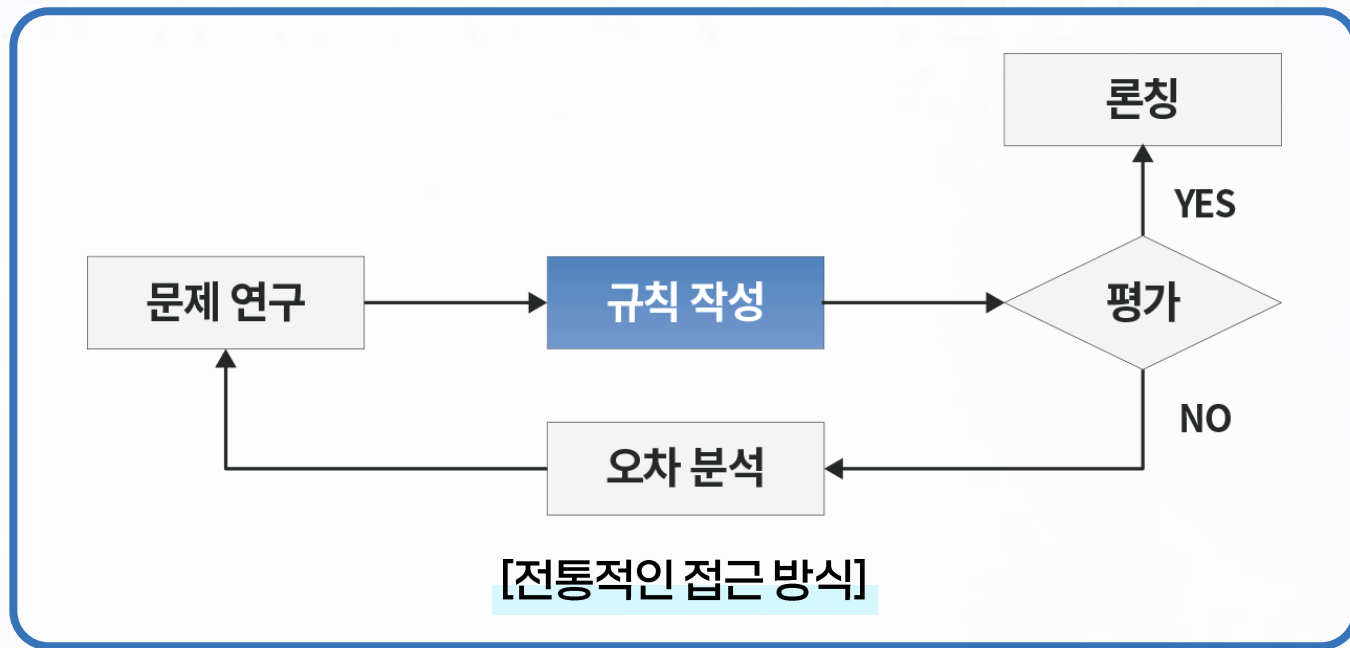
03 | 왜 기계학습을 사용하는가?



왜 기계학습을 사용하는가?

△ 아래의 그림은 전통적인 접근 방식으로 시스템을 만드는 과정을 나타냄

◆ 이러한 방식은 문제가 어렵기 때문에 규칙이 점점 길고 복잡해지므로 유지 보수하기 매우 힘들어짐



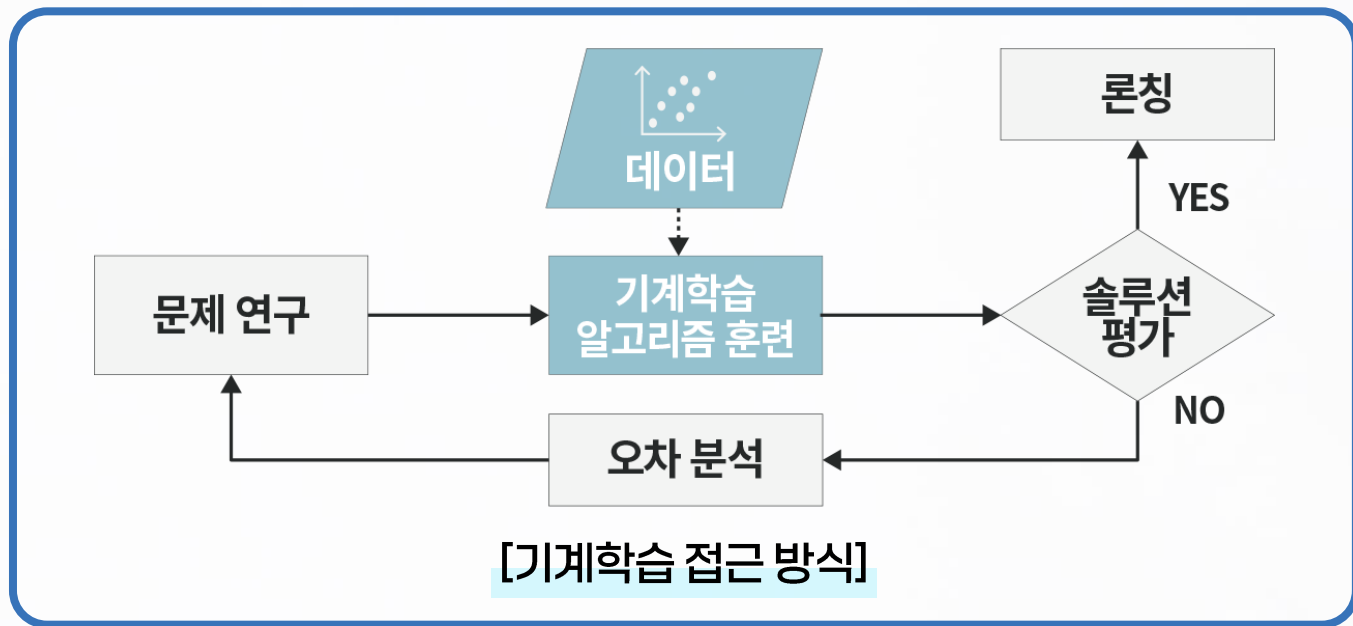


03 | 왜 기계학습을 사용하는가?

왜 기계학습을 사용하는가?

아래 그림은 기계학습 접근 방식으로 시스템을 만드는 과정을 나타냄

- 기계학습 기반의 스팸 필터는 일반 메일에 비해 스팸이 자주 나타나는 패턴을 감지하여 어떤 단어와 구절이 스팸 메일을 판단하는 데 좋은 기준인지 자동으로 학습함
- 그러므로 프로그램이 훨씬 짧아지고 유지 보수가 쉬우며 대부분 정확도가 더 높음





03 | 왜 기계학습을 사용하는가?

⚙️ 왜 기계학습을 사용하는가?

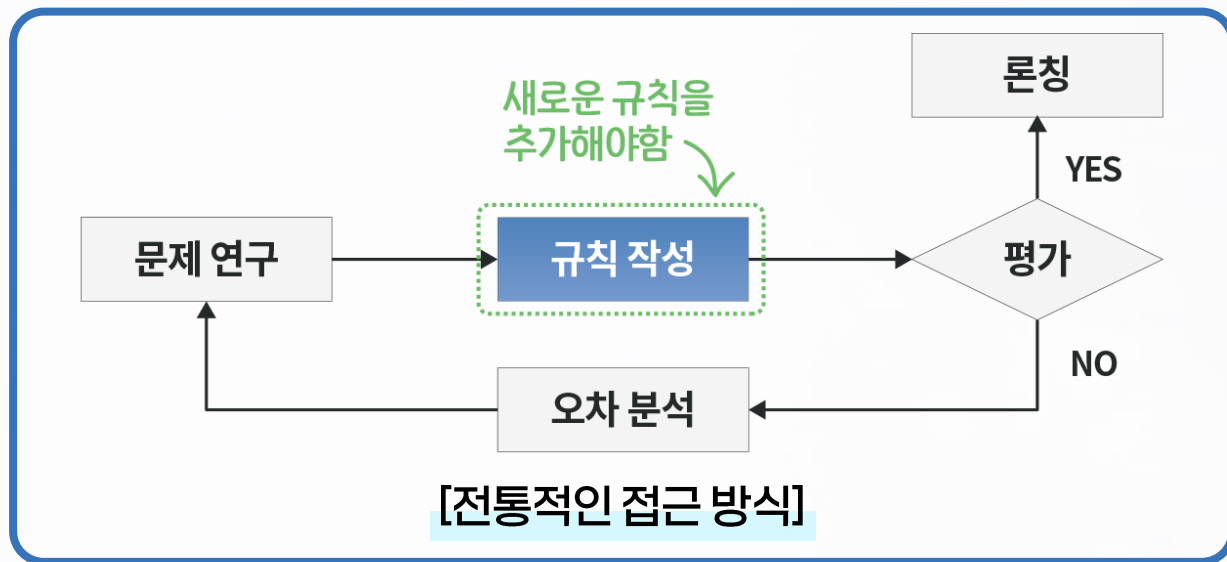
⚠️ 스팸 메일 발송자가 '4U'를 포함한 모든 메일이 차단된다는 것을 안다면 어떻게 될까?

◆ 아마도 '4U' 대신 'For U'를 쓸 수도 있을 것임

➢ 전통적인 프로그램 방식의 경우

─ 스팸 필터는 'For U' 메일을 구분하기 위해 수정이 필요함

➔ 스팸 메일 발송자가 스팸 필터에 대해 계속 단어를 바꾸면 영원히 새로운 규칙을 추가해야함





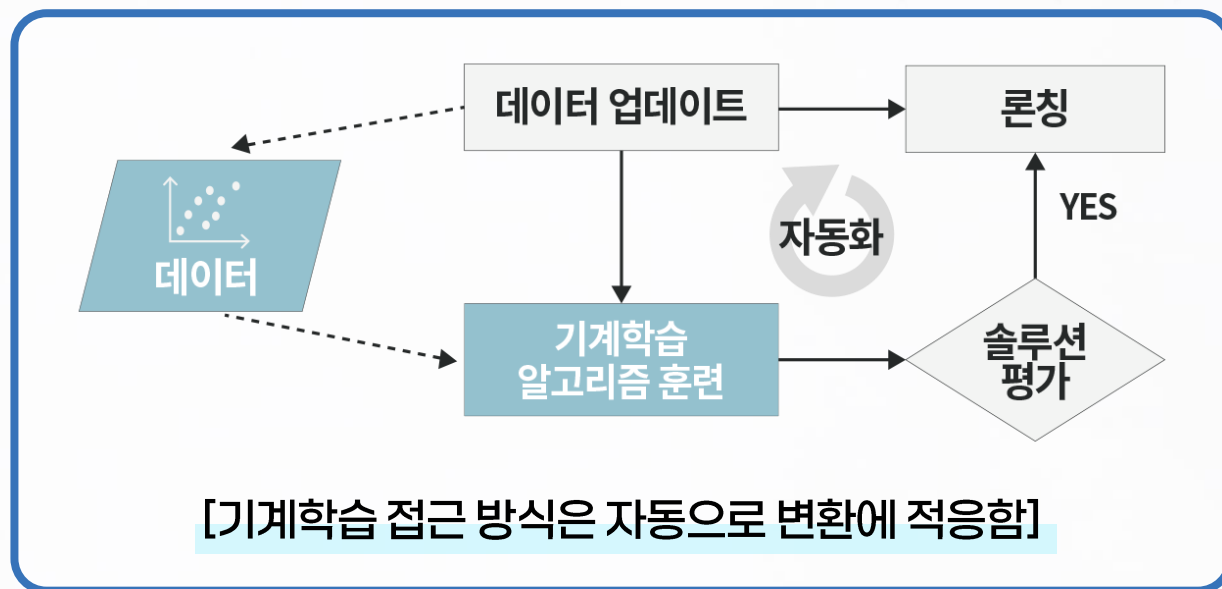
03 | 왜 기계학습을 사용하는가?

🔧 왜 기계학습을 사용하는가?

> 기계학습 기반 프로그램 방식의 경우

– 스팸 필터는 사용자가 스팸으로 지정한 메일에 유독 ‘For U’가 자주 나타나는 것을 자동으로 인식함

➔ 그러므로 별도의 작업을 하지 않아도 자동으로 이 단어를 스팸으로 분류함





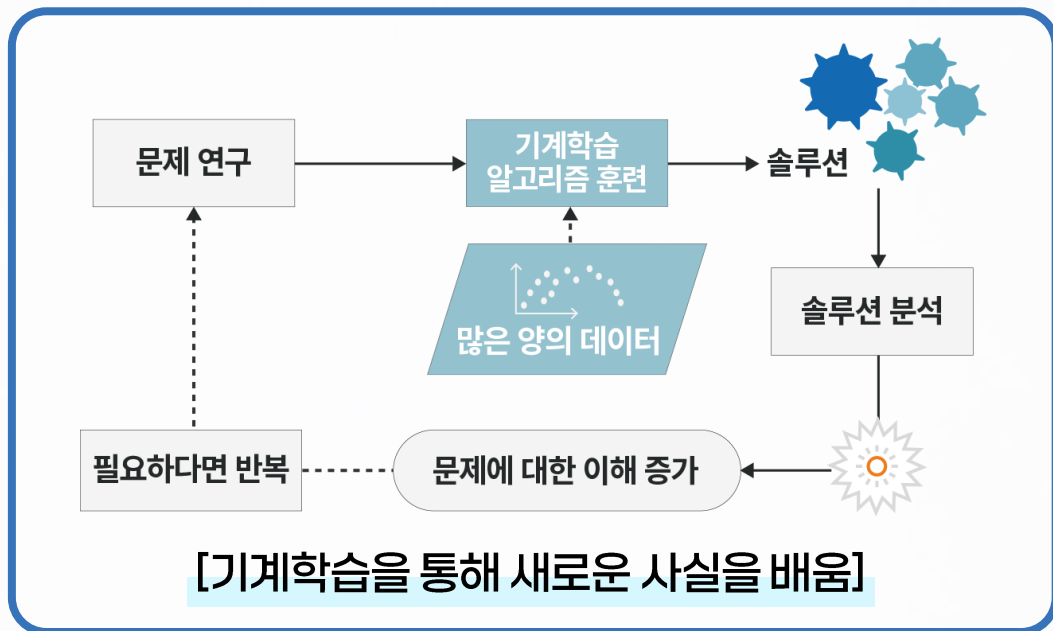
03 | 왜 기계학습을 사용하는가?

왜 기계학습을 사용하는가?

우리는 기계학습을 통해 배울 수도 있음

◆ 아래의 그림은 대량의 데이터 분석을 통해 보이지 않던 패턴을 발견함(기계학습을 통해 새로운 사실을 배움)

▶ 예를 들어 스팸 필터가 충분한 스팸 메일로 훈련되었다면 스팸을 예측하는 데 가장 좋은 단어 및 단어의 조합이 무엇인지 확인할 수 있음

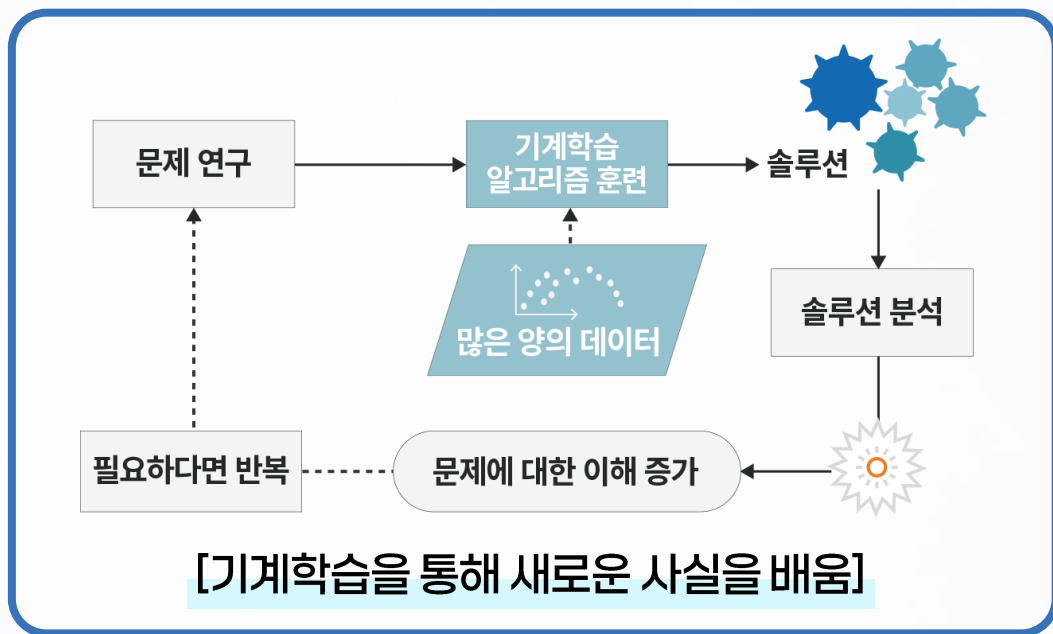




03 | 왜 기계학습을 사용하는가?

왜 기계학습을 사용하는가?

- △ 기계학습은 가끔 **예상치 못한 연관 관계**나 **새로운 추세**가 발견되기도 해서 **해당 문제를 더 잘 이해**하도록 도와줌
- ◆ 기계학습 기술을 적용해서 **대용량의 데이터**를 **분석**하면 **겉으로는 보이지 않던 패턴**을 발견할 수 있음
 - 이를 **데이터 마이닝(Data Mining)**이라고 부름





03 | 왜 기계학습을 사용하는가?



왜 기계학습을 사용하는가?

△ 기계학습은 다음 분야에 뛰어남

- 1 기존 솔루션으로는 많은 조정과 규칙이 필요한 문제
 - ▶ 하나의 기계학습 모델이 코드를 간단하게 만들고 전통적인 방법보다 더 잘 수행되도록 할 수 있음
- 2 전통적인 방식으로는 해결 방법이 없는 복잡한 문제
 - ▶ 가장 뛰어난 기계학습 기법으로 해결 방법을 찾을 수 있음
- 3 유동적인 환경
 - ▶ 기계학습 시스템은 새로운 데이터에 적응할 수 있음
- 4 복잡한 문제와 대량의 데이터에서 통찰을 얻을 수 있음