

## CHAPTER 01

# 머신러닝의 기초

**01** 머신러닝이란?

**02** 머신러닝의 학습 프로세스와 종류

**03** 머신러닝 연대기

**04** 머신러닝 환경 구축하기

# 학습목표

- 본격적인 머신러닝 시대에 머신러닝의 실제 사용 사례와 키워드에 대해 알아본다.
- 머신러닝의 학습 프로세스와 종류에 대해 학습한다.
- 데이터 시대의 시작부터 머신러닝 시대의 도래까지 머신러닝 연대기를 살펴본다.
- 머신러닝의 환경을 구축하는 방법에 대해 실습한다.

**01**

**머신러닝이란?**

### 1. 머신러닝의 시대



그림 1-1 유튜브 알고리즘과 관련된 댓글

- 사람들이 기계가 만든 어떤 '로직(logic)'에 이끌려 영상을 보게 됨

# 01 머신러닝이란?

## CHAPTER 01 머신러닝의 기초

- 유튜브는 2016년에 딥러닝 기반 추천 알고리즘을 적용
- 가장 대중적인 스마트폰 앱 1위



(a) 유튜브 앱 사용자 수 현황



(b) 주요 앱 1인당 월평균 사용 시간

그림 1-2 대한민국 유튜브 사용 현황 ©아이지에이웍스

- 알고리즘(algorithms) : 어떠한 문제를 해결하기 위한 일련의 절차나 방법
- 머신러닝(machine learning) : 기계가 패턴을 학습하여 자동화하는 알고리즘
- 유튜브는 개인이 유튜브 영상을 보는 패턴에 대해 학습하는 프로그램(머신러닝)을 만든 다음 그 패턴(알고리즘)에 맞게 다음 영상을 계속 추천

## 2. 머신러닝의 실제 사용 사례

### 2.1 구매 추천

- 인터넷 쇼핑몰에서 장바구니에 추가한 제품과 비슷한 제품 구매를 추천

The screenshot displays a recommendation interface on the YES24 website. It is divided into two main sections: '이 책을 구입하신 분들이 함께 산 책' (Books bought by those who bought this book) and '이 책을 구입하신 분들이 많이 산 책' (Books bought by many of those who bought this book). Each section lists six books with their covers, titles, authors, and prices. The books are: 1. '강원국의 어른답게 말합니다' by 강원국, 14,400원 (10% discount); 2. '2000원으로 밥상 차리기 (월간) : 7월호 [202' by 편집부, 2,850원 (5% discount); 3. '손미나의 나의 첫 외국어 수업' by 손미나, 15,120원 (10% discount); 4. '타인의 집' by 손원평, 12,600원 (10% discount); 5. '달려구르 꿈 벽화점' by 이매애, 12,420원 (10% discount); 6. '어떻게 말해줘야 할까' by 오운영, 15,750원 (10% discount).

이 책을 구입하신 분들이 함께 산 책	이 책을 구입하신 분들이 많이 산 책
 강원국의 어른답게 말합니다 강원국 지   동진지식... ★★★★★ 39건 14,400원 (10% 할인)	 2000원으로 밥상 차리기 (월간) : 7월호 [202 편집부   그리고작 ★★★★★ 21건 2,850원 (5% 할인)
 손미나의 나의 첫 외국어 수업 손미나 지   후네이도 ★★★★★ 23건 15,120원 (10% 할인)	 타인의 집 손원평 지   창비 ★★★★★ 5건 12,600원 (10% 할인)
	 달려구르 꿈 벽화점 이매애 지   책로라나연 ★★★★★ 418건 12,420원 (10% 할인)
	 어떻게 말해줘야 할까 오운영 지/차상미 그림... ★★★★★ 199건 15,750원 (10% 할인)

그림 1-3 구입하려는 책과 관련된 또 다른 책을 추천하는 YES24의 사례

### 2.2 번역

- 머신러닝에 전문 번역가의 번역을 학습시켜 새로운 문장을 번역하게 함

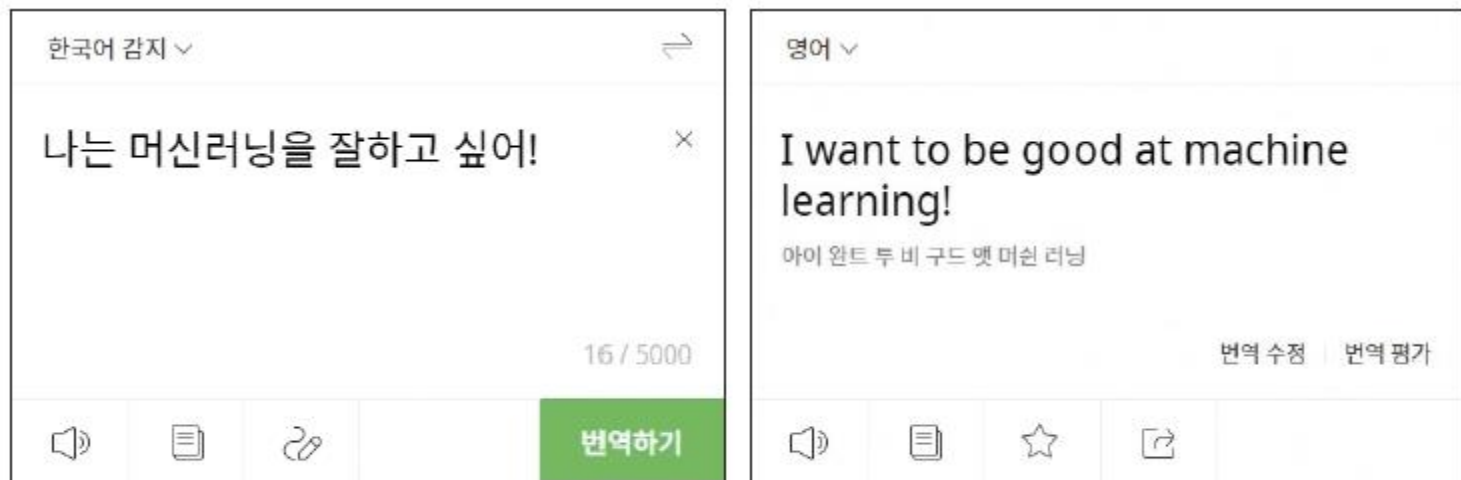


그림 1-4 네이버 번역기 파파고(Papago)



### 2.3 자율주행차

- 머신러닝 기술 중 하나인 이미지 처리 기술을 활용하여 도로상의 여러 이미지를 학습, 차량에 있는 컴퓨터가 스스로 판단하여 운행



그림 1-5 머신러닝의 이미지 처리 기술을 활용한 자율주행차

### 2.4 챗봇

- 머신러닝을 기반으로 사용자와 컴퓨터 간 대화를 지원
- 코로나19 사태로 인해 사람의 목소리를 흉내내어 응대하는 인공지능 콜센터 출현
  - 머신러닝이 인간의 대화 패턴과 목소리를 학습



그림 1-6 챗봇 사용 사례

### 3. 머신러닝의 키워드

#### 3.1 인공지능, 머신러닝, 딥러닝

- 인공지능 ⊃ 머신러닝 ⊃ 딥러닝

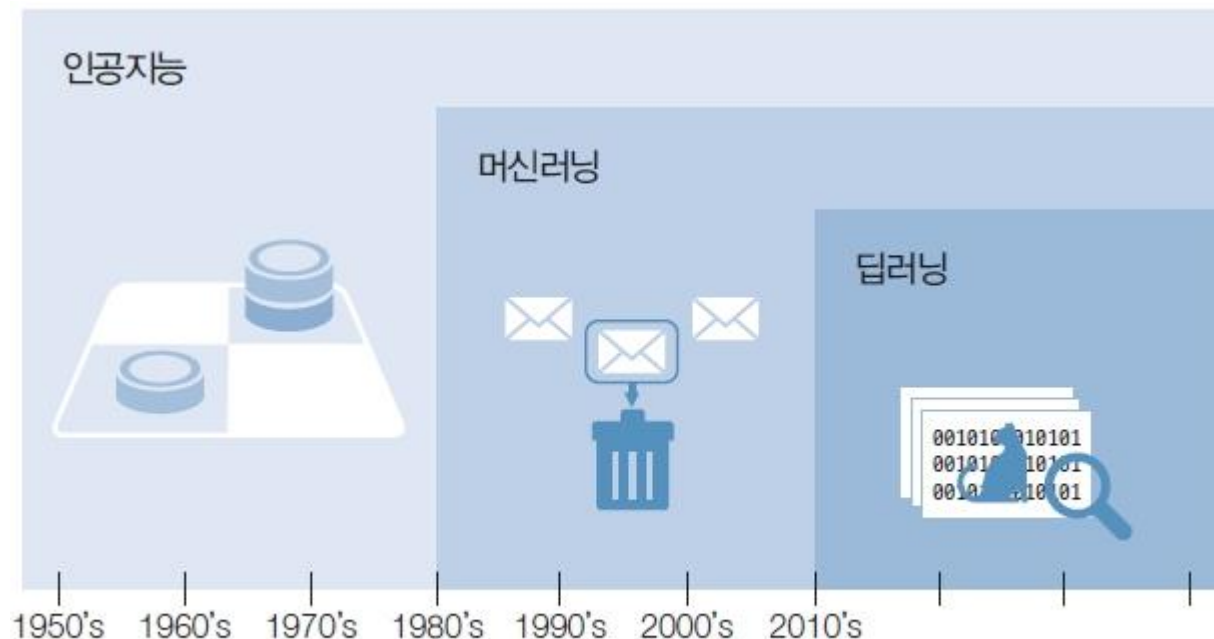


그림 1-7 인공지능, 머신러닝, 딥러닝의 관계

- 인공지능(Artificial Intelligence, AI) : 컴퓨터가 학습하고 생각하여 스스로 판단할 수 있도록 만드는 기술
  - 1956년 다트머스 컨퍼런스에서 '인공지능'이라는 표현을 처음 사용한 이후 본격적인 학문으로서 발전하기 시작



존 맥카시



클로드 섀넌



마빈 민스키



나다니엘 로체스터



레이 솔로모노프



올리버 셀프리지



트렌처드 모어



아서 사무엘



앨런 뉴얼



하버트 사이먼

그림 1-8 다트머스 컨퍼런스에 참석한 10명의 과학자들

- 머신러닝(machine learning) : 데이터를 컴퓨터에 학습시켜 그 패턴과 규칙을 컴퓨터가 스스로 학습하도록 만드는 기술
  - 이전에는 사람이 지식을 직접 데이터베이스화한 후 컴퓨터가 처리하도록 프로그램으로 만듦
  - 머신러닝은 데이터를 분류하는 수학적 모델을 프로그래밍하여, 데이터만 입력하면 이미 만들어진 수학 모델이 규칙으로 적용되어 여러 문제를 풀 수 있음
- 딥러닝(deep learning) : 머신러닝 기법 중 신경망(neural network)을 기반으로 사물이나 데이터를 군집화하거나 분류하는 데 사용하는 기술

## 3.2 통계학적 머신러닝과 딥러닝

- 딥러닝이 급속도로 발전함에 따라 전통적 머신러닝과 최근 많이 사용되는 딥러닝을 병렬 관계로 보기도 함
- 딥러닝 알고리즘은 주로 컴퓨터 과학자들이 만듦
  - 신경망 알고리즘은 제프리 힌튼 교수가 제안, 전원 컴퓨터 과학 분야 출신인 '인공지능 4대 천왕(Big4 of AI)'가 발전시킴
- 기존 머신러닝은 통계학계 중심으로 발전해 옴
  - 회귀분석이나 의사결정트리 등의 알고리즘

### [하나 더 알기] 이 책은 머신러닝 책인가? 딥러닝 책인가?

- 본 책은 통계학적 머신러닝의 방법을 알고리즘 설명과 파이썬 코드를 중심으로 풀어나가는 것을 목표로 한다
- 대부분의 딥러닝 개념은 전통적인 머신러닝에서 사용하는 통계학적 접근에 그 뿌리를 두고 있다
- 현실에서 만나게 될 데이터는 매우 다양하기 때문에 이런 데이터의 특성을 잘 이해해서 적절한 알고리즘을 사용해야 한다
- 이 책의 목표는 기존 전통적인 머신러닝의 여러 가지 기법을 학습자가 잘 이해함으로써 인공지능의 세계로 들어가는 통로가 되는 것이라고 할 수 있다

## 3.3 빅데이터와 머신러닝

- 빅데이터(big data) : 기존의 데이터베이스로는 수집·저장·분석을 수행하기 어려울 만큼 방대한 양의 데이터
  - 데이터베이스에서 기원
- 빅데이터 시스템(big data system) : 빅데이터를 다루기 위한 시스템
- 빅데이터 엔지니어링(big data engineering) : 빅데이터를 다루는 방법
- 머신러닝과 별개로 발전해왔으나 대용량 데이터가 학습 성능에 크게 영향을 미치는 오늘날 머신러닝 분야에서 의미 있게 사용됨



- GFS(Google File System) : 구글의 거대한 검색 시스템을 여러 대의 컴퓨터에 데이터를 분산해서 저장 및 관리하기 위해 설계된 시스템
- 맵리듀스(Map Reduce) : 분산되어 저장된 데이터를 여러 대의 컴퓨터가 동시에 협력해서 처리하는 방법을 알려준 시스템
- 하둡(Hadoop) : 구글의 분산 처리 시스템 개념이 확장되어 개발된 시스템. 오늘날 대용량 데이터를 처리

# 01 머신러닝이란?

## CHAPTER 01 머신러닝의 기초

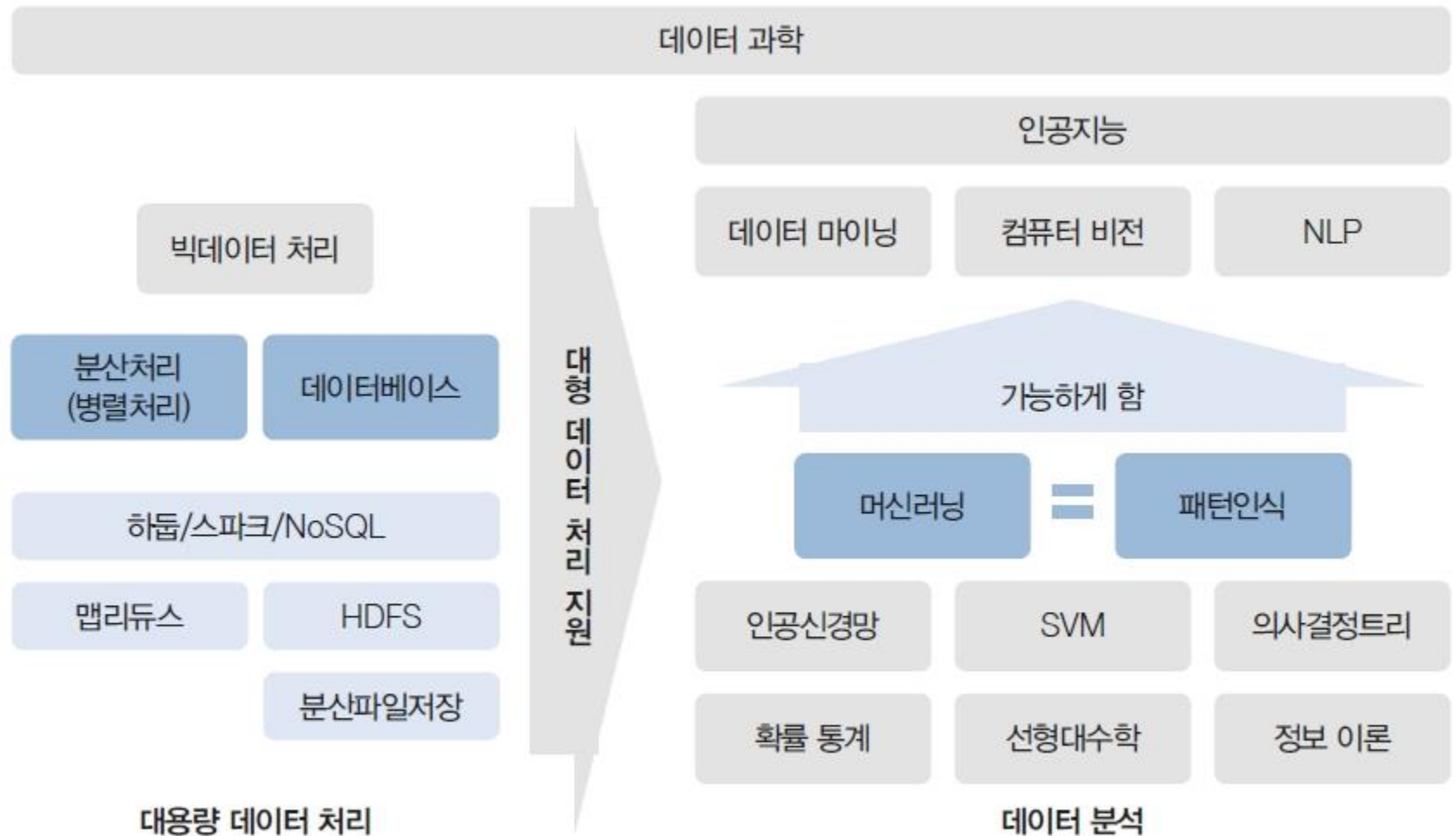


그림 1-9 인공지능 생태계

### [하나 더 알기] 데이터 과학의 의미

- 데이터 과학(data science)은 모호한 표현이다. 오늘날 세상의 거의 모든 학문은 데이터를 다루는 학문이다
- 한정해서 본다면, 머신러닝과 데이터 엔지니어링을 포함한 오늘날의 데이터 처리 및 분석과 관련된 모든 학문
- 인공지능은 데이터를 처리한 결과만을 사용하기 때문에 데이터 과학이 인공지능보다는 조금 더 큰 의미로 사용된다
- 조금 더 좁은 의미에서 기존 통계학적 분석에서 사용하는 데이터 분석의 의미로 한정하거나, 머신러닝까지 포함한 데이터 분석 관점으로 한정하기도 한다

02

# 머신러닝의 학습 프로세스와 종류

## 1. 머신러닝의 학습 프로세스

- 머신러닝의 기본 모형

$$\hat{y} = f(x)$$

- 가지고 있는 데이터  $x$ 를 우리가 학습하여 생성한 함수  $f(\cdot)$ 에 넣으면 그 결과로 어떤 문제에 대한 예측치  $\hat{y}$ 을 생성
- 머신러닝은 우리가 찾고자 하는  $f(\cdot)$ 를 제공

## 02 머신러닝의 학습 프로세스와 종류

키	158	168.7	162.8	166.5	164.9
몸무게	57.1	74.1	74.1	77.9	75.5

그림 1-10 키와 몸무게 데이터

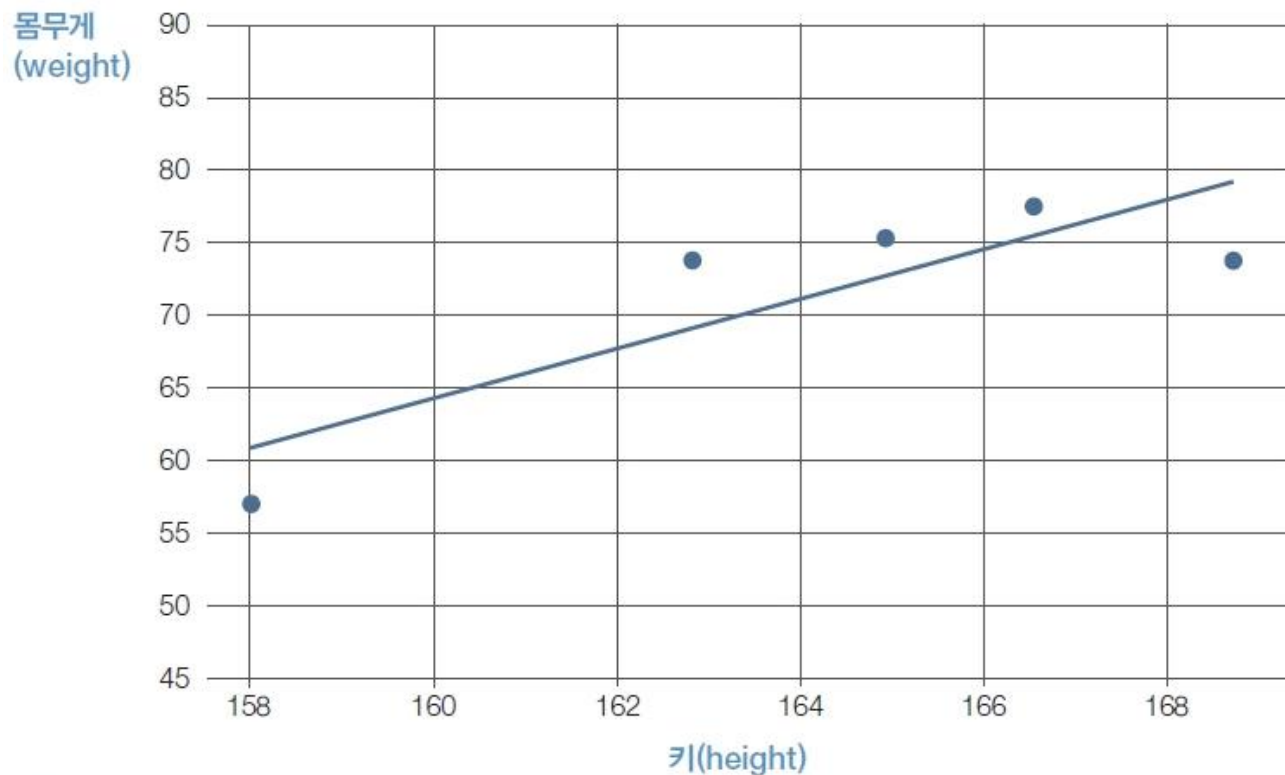


그림 1-11 키와 몸무게 간의 상관관계 그래프

- 모델(model) : 상관관계를 식으로 표현

$$y = \alpha x + \beta$$

- 어떤 모델이 키와 몸무게의 관계로 가장 적합(fitting)할지 모르는 상황에서, 이와 같은 1차방정식을 완성한다면 찾고자 하는 모델을 발견할 수 있다
  - 알고리즘을 통해 적합한  $\alpha$ 와  $\beta$ 를 찾아 새로운 키에 대한 몸무게 예측 가능

## 02 머신러닝의 학습 프로세스와 종류

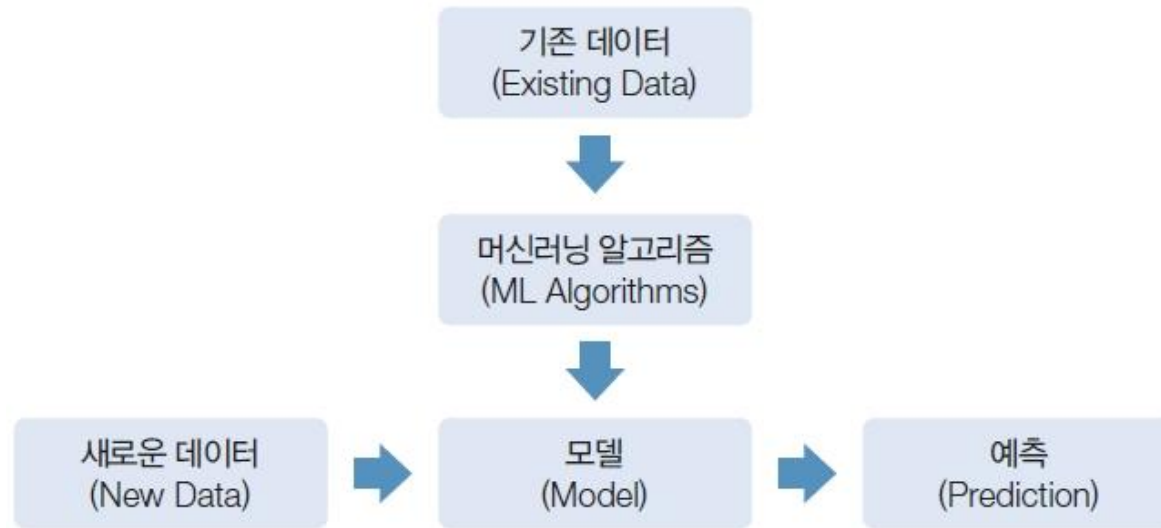


그림 1-12 머신러닝 프로세스

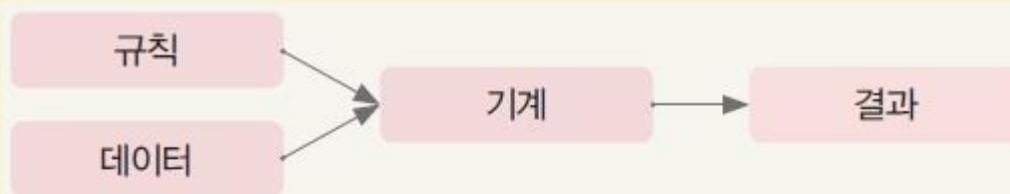
### [하나 더 알기] 모델과 알고리즘

- 본 책에서 '모델'은 '수식'이나 '통계 분포', '알고리즘'은 모델을 산출하기 위해 규정화된 과정
- 보통 하나의 모델은 다양한 알고리즘으로 표현할 수 있다
- 때때로 '알고리즘'은 하나의 '수식'으로 표현 가능하다

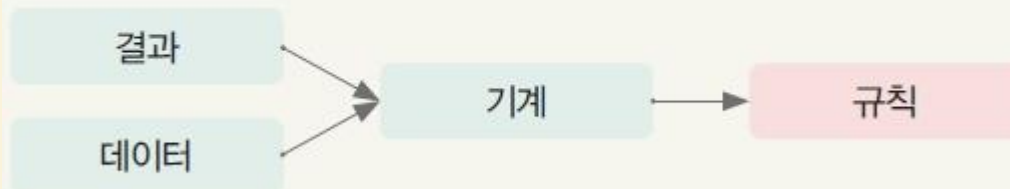


### [하나 더 알기] 머신러닝 이전의 프로그래밍

- 기존의 프로그래밍 기법들은 사람이 프로그래밍 규칙을 지정한 후 데이터를 입력하여 결과를 얻는 구조
- 머신러닝 기반 프로그래밍에서는 결과와 데이터를 입력하면 데이터와 결과로부터 규칙을 뽑아내는 구조



(a) 전통적인 프로그래밍 접근



(b) 머신러닝 기반 프로그래밍 접근

그림 1-13 전통적인 프로그래밍 접근과 머신러닝 기반 프로그래밍 접근

### 2. 머신러닝의 종류

- 지도학습(supervised learning) : 문제와 답을 함께 학습
- 비지도학습(unsupervised learning) : 조력자의 도움 없이 컴퓨터 스스로 학습. 컴퓨터가 훈련 데이터를 이용하여 데이터들 간의 규칙성을 찾아냄
- 실제 답(ground truth)  $y$ 의 존재 여부에 따라 구분

표 1-1 전통적인 관점에서 머신러닝의 종류

머신러닝 대분류	머신러닝 종류	설명
지도학습	회귀	연속형 값인 $y$ 의 특징을 찾아 데이터 $x$ 를 사용하여 $y$ 값을 예측하는 기법
	분류	이산형 값인 $y$ 의 특징을 찾아 데이터 $x$ 를 사용하여 $y$ 값을 예측하는 기법
비지도학습	군집	$y$ 값이 주어지지 않고, 데이터의 특징이 유사한 값들의 모임을 군집으로 표현하는 기법

## 02 머신러닝의 학습 프로세스와 종류

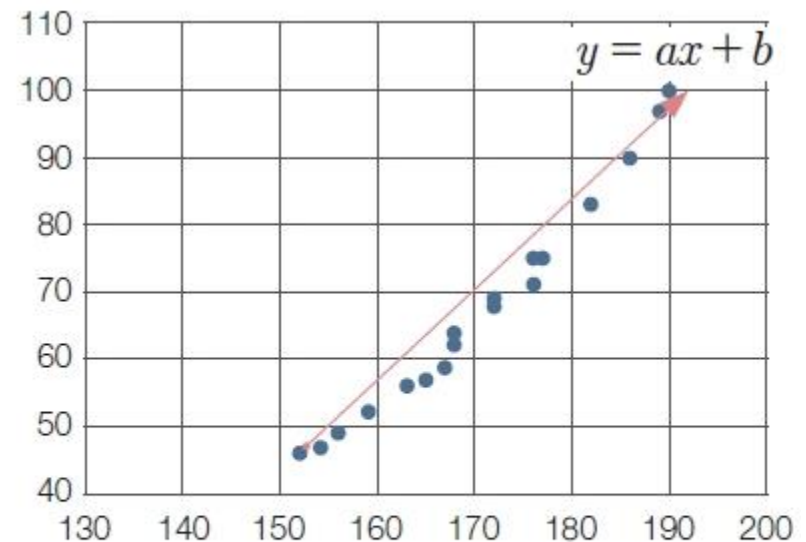
- 강화학습(reinforcement learning) : 컴퓨터가 세상에 존재하는 규칙을 스스로 시뮬레이션하면서 게임처럼 규칙을 학습
- 생성(generation) : 세상에 존재하는 다양한 규칙들을 학습한 모델이 세상에 존재하지 않는 새로운 무엇인가를 창조해냄
  - 컴퓨터로 사람의 얼굴을 생성
  - 컴퓨터가 챗봇 형태로 인간과 대화

## 2.1 회귀

- 회귀(regression) : 독립변수  $x$ 와 종속변수  $y$ 의 관계를 함수식으로 설명
- 추세선을 표현하는 수학적 모델을 만드는 기법

키	몸무게	키	몸무게
152	46	172	69
154	47	172	68
156	49	176	71
159	58	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100

(a) 키와 몸무게 데이터



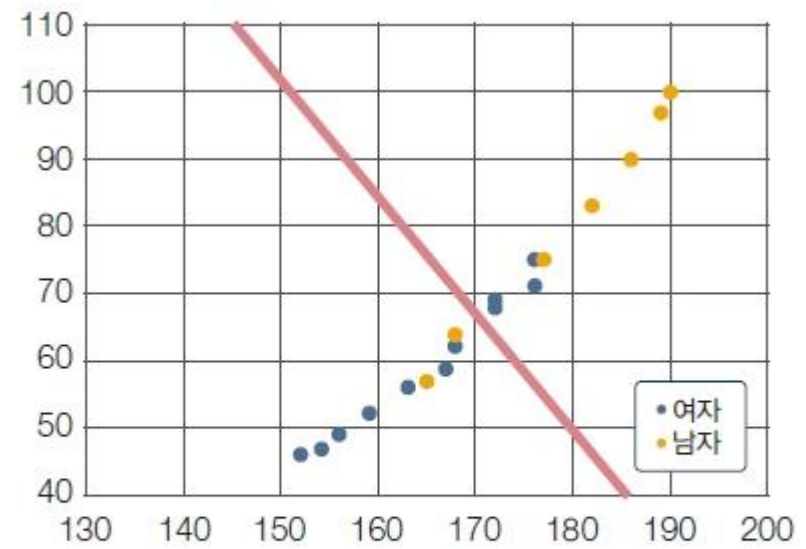
(b) 회귀 표현

### 2.2 분류

- 분류(classification) : 데이터를 어떤 기준에 따라 나눔
- 이진분류(binary classification) : 2개의 값 중 1개를 분류
- 다중분류(multi-class classification) : 3개 이상 분류 실행

성별	키	몸무게	성별	키	몸무게
여자	152	46	남자	172	69
여자	154	47	여자	172	68
여자	156	49	여자	176	71
여자	159	58	여자	176	75
여자	163	56	남자	177	75
남자	165	57	남자	182	83
남자	167	59	남자	186	90
남자	168	64	남자	189	97
여자	168	62	남자	190	100

(a) 성별, 키, 몸무게 데이터



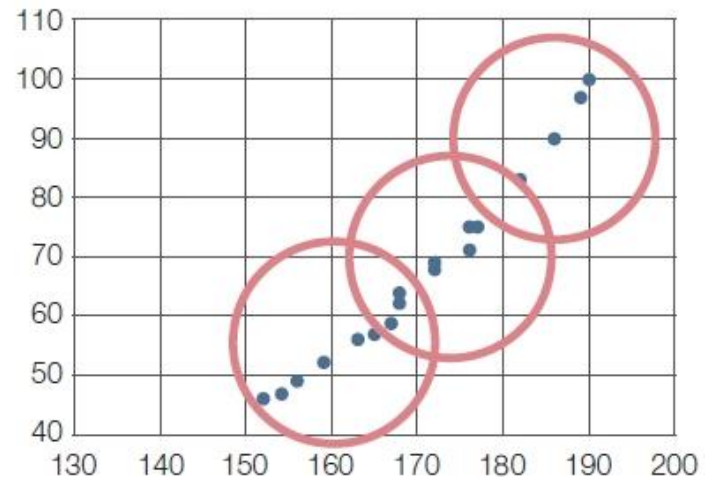
(b) 분류 표현

### 2.3 군집

- 군집(clustering) : 기존에 모여 있던 데이터에 대해 따로 분류 기준을 주지 않고 모델이 스스로 분류 기준을 찾아 집단을 모으는 기법
- 비슷한 수준의 농구팀 3개 만들기

키	몸무게	키	몸무게
152	46	172	69
154	47	172	68
156	49	176	71
159	58	176	75
163	56	177	75
165	57	182	83
167	59	186	90
168	64	189	97
168	62	190	100

(a) 키와 몸무게 데이터



(b) 군집 표현

**03**

# **머신러닝 연대기**

### 1. 데이터 시대의 시작 :

#### 제2차 세계대전과 컴퓨터의 출현

- 제2차 세계대전 중 연합군은 독일군의 암호 해독 방법을 연구하였고 영국의 수학자 앨런 튜링이 현재의 컴퓨터와 유사한 형태의 기계 장치를 만들



그림 1-17 영화 《이미테이션 게임》



### [하나 더 알기] 앨런 튜링(Alan Turing)

- 알고리즘과 수학 원리를 이용한 튜링 머신으로 컴퓨터 과학 발전에 큰 공헌을 하였으며 튜링 테스트를 고안해 낸 것으로 유명하다
- 컴퓨터 과학 분야의 노벨상에 해당하는 상 이름을 튜링상(Turing Award)이라고 부른다
- 앨런 튜링은 당시 금기시되었던 동성애자였고 이것이 발각되어 화학적 거세를 당했다
- 그는 독이 든 사과를 먹고 스스로 목숨을 끊었다. 일설에 의하면 스티브 잡스는 앨런 튜링의 업적을 기념하기 위해 회사의 명칭을 애플로 짓고, 로고를 일부 베어 먹은 사과의 모습으로 정했다고 한다

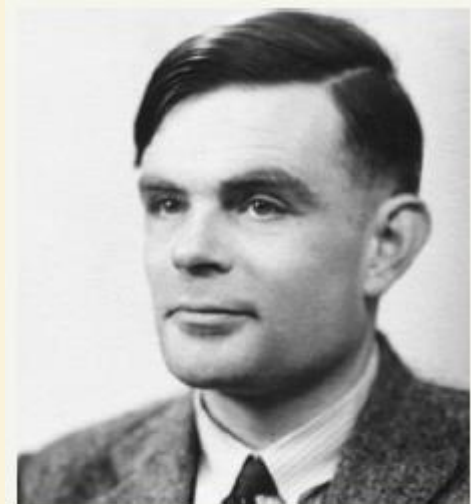


그림 1-18 앨런 튜링

### 2. 기술의 전이 : 비즈니스를 위한 기술

- 제2차 세계대전이 끝나고 전쟁에서 사용되었던 다양한 기술들이 민간으로 이전됨
- 서버 컴퓨터 메인프레임에 터미널을 사용하여 접속
  - 전자정보처리시스템 또는 거래처리시스템으로 은행에서 거래 처리 정보를 데이터로 저장



그림 1-19 1950년대 사용한 메인프레임 서버 IBM 700 시리즈와 당시 사용하던 터미널

- 1960년대 정보시스템들이 기업을 중심으로 더욱 확산
  - 경영정보시스템 도입하여 거래정보 외에도 직원들의 월급부터 고객의 개인정보까지 처리
  - 스프레드시트 프로그램이 보급되어 복잡한 수식을 처리
- 1980년대 개인용 컴퓨터(Personal Computer, PC)가 서서히 보급되며 기업 업무에 컴퓨터 적극 활용
- 데이터를 활용한 간단한 형태의 인공지능시스템이나 의사결정지원시스템 도입
  - 기업이 데이터 기반 의사결정을 시도

### 3. 정보화 혁명의 시작 : 데이터의 확장

- 1989년 월드와이드웹 WWW 개발 이후 인터넷상에서 많은 시스템 개발
- 이전에 많은 시스템이 사내에서 직원들을 위해 개발되었다면 이제는 전 세계 고객들이 시스템을 통해 서비스 이용

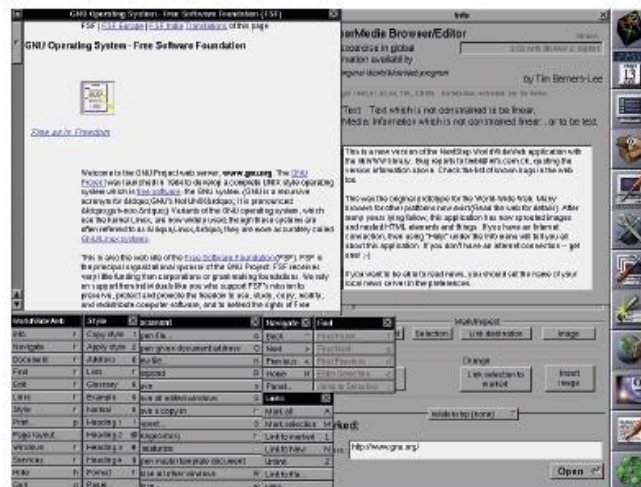
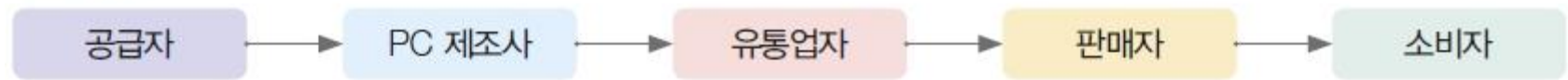


그림 1-20 WWW의 등장

전통 모델



델(Dell)의 모델



그림 1-21 델의 공급사슬망 관리(SCM) 전략의 변화

- 정보화 혁명은 단순히 데이터를 저장하는 것을 넘어 기업이 일하는 방식을 바꾸는 데 일조를 함
- 사업 프로세스를 정보화 시스템에 결합하여 업무효율을 높이는 시도를 지속
  - 전사적자원관리나 지식정보시스템 도입하여 PC의 정보들을 사내 전산망에서 관리

- 고객의 정보를 관리하는 고객관계관리 시스템 도입

### [하나 더 알기] 기저귀와 맥주

- 1993년 IBM의 라케시 아그라왈이 영국의 대형마켓인 막스앤스펜서에서 고객관계관리(CRM)를 바탕으로 소비자 분석을 진행하였다
- '기저귀'와 '맥주'의 구매 여부에 대한 상관관계를 분석했다
- '기저귀를 산 사람 중 40%(50명 중 20명)'가 맥주를 구매했다는 사실에 주목했고, 기저귀를 산 사람에게 맥주를 추천하면 매출이 오를 것이라고 판단했다
- 이때부터 맥주와 기저귀를 나란히 배치하여 큰 매출을 올렸다

### 4. 모바일 시대의 시작 : 데이터의 폭발

- 2007년 스마트폰 출시로 본격적인 모바일 시대가 열림
- 개인의 만족을 극대화시키는 스마트폰 기업, 다국적 IT 기업들이 각광받음
- 기업들이 개인에게서 생성되는 다양한 데이터를 수집하며 빠르게 성장

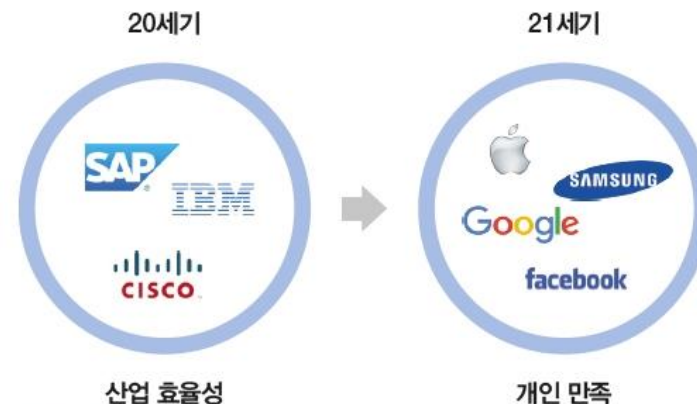


그림 1-22 20세기와 21세기의 IT 기업 패러다임의 변화

### 5. 머신러닝 시대의 도래 : 데이터의 관리 및 학습

- 빅데이터를 다루기 위한 도구들이 출현
  - 구글의 하둡은 병렬처리 방식으로 서버 비용을 획기적으로 절감
- 방대한 빅데이터를 클라우드(cloud)에 저장



그림 1-23 아마존의 AWS 데이터 센터



- 사물인터넷(Internet of Things, IoT) 출현
- 인간 지식 노동의 자동화

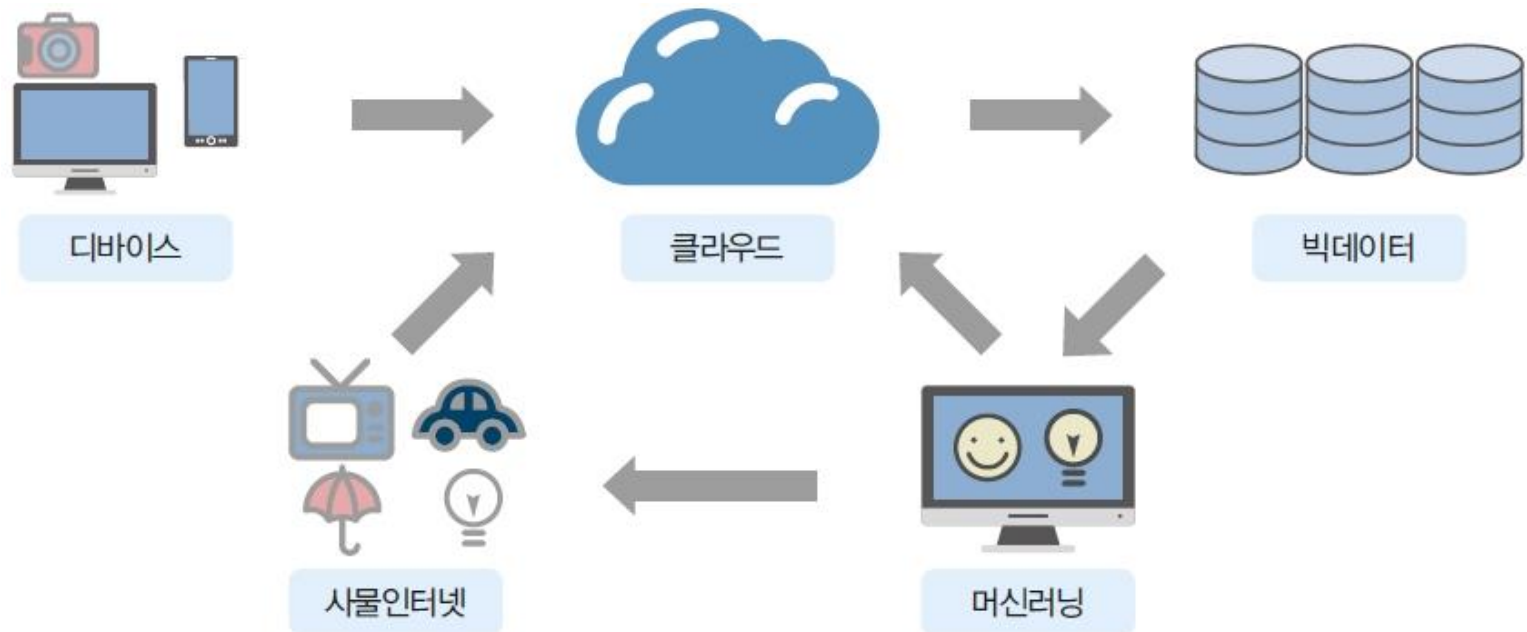


그림 1-24 데이터가 흐르는 길에 개념이 발생 © 〈What will come next〉, 하용호, 2015년

**04**

# **머신러닝 환경 구축하기**

## 1. 파이썬으로 머신러닝 환경을 구축하는 이유

- ① 스크립트 언어 : 파이썬은 컴파일러를 쓰지 않는 스크립트 언어로, 수정이 쉽고 간단하며 사용자와 지속적인 인터랙션을 할 수 있음
  - 그래프를 수시로 확인하기, 특정 데이터를 화면에 찍어보기
- ② 쉬운 언어 : 문법 자체가 간단하고 배우기 쉬움
- ③ 머신러닝과 딥러닝의 표준 언어 : 머신러닝과 딥러닝을 지원하는 프레임워크들은 대부분 파이썬 언어를 지원함
  - 머신러닝 사이킷런, 딥러닝 텐서플로우, 파이토치

### 2. 파이썬 머신러닝 환경 구축하기

표 1-2 파이썬 머신러닝의 도구들

도구 종류	설명	사용 도구
파이썬 인터프리터	파이썬 코드를 해석하고 실행시키기 위한 실행 프로그램	아나콘다(Anaconda)
코드 편집기	파이썬 코드를 수정할 때 사용하는 프로그램	주피터(Jupyter), VS코드(VSCode)
통계 분석 및 전처리 도구	데이터를 로드하고 전처리하기 위한 도구들	넘파이(NumPy), 판다스(Pandas), 사이파이(SciPy)
시각화 도구	데이터의 상태를 파악하기 위해 시각화를 지원하는 도구	맷플롯립(Matplotlib), 시본(Seaborn), 플롯리(Plotly)
머신러닝 프레임워크	실제 머신러닝 모델을 생성하고 데이터에 적용할 수 있도록 도와주는 도구	사이킷런(Scikit-learn)

## 2.1 파이썬 인터프리터

- 파이썬을 실행하는 파이썬 해석기 프로그램
- 파이썬 인터프리터(Python interpreter) : 파이썬 재단에서 배포하는 가장 기본적인 인터프리터
- 최근에는 과학 계산용 파이썬 인터프리터를 표방한 아나콘다(Anaconda) 버전이 많이 사용됨



그림 1-25 파이썬 인터프리터, 아나콘다(Anaconda)

## 2.2 코드 편집기

- 단순한 코드 편집기는 노트패드처럼 코드를 문서처럼 취급하여 수정
- 통합개발환경(Integrated Development Environment, IDE)은 코드 저장, 코드 전송, 코드 배포 등 개발의 전 과정을 지원하는 도구
- 일반적으로 인터랙티브 셸(interactive shell)을 사용하여 코드를 계속 수정, 데이터 변화를 관찰하면서 코딩

- 주피터(Jupyter) : 데이터 분석에 있어 가장 많이 쓰이고 사실상의 표준으로 사용되는 코드 편집기
  - 웹 기반의 코드편집기
  - 코드를 입력하면 화면에 바로 결과를 출력하고 그래프나 수식들도 웹상에 바로 보여줌
- VS코드(VSCoDe) : 마이크로소프트에서 개발
  - 처음에는 단순 코드 편집기로 시작되었지만 최근 거의 완벽한 수준의 IDE로 받아들여짐

```
In [1]: a = 10  
        b = 50  
        a * b
```

```
Out[1]: 500
```

그림 1-26 VS코드의 주피터 노트북 기능

## 2.3 통계 분석 및 전처리 도구

- 넘파이(NumPy) : 선형대수의 계산식을 파이썬으로 구현할 수 있도록 도와주는 라이브러리
- 판다스(Pandas) : 파이썬 데이터 처리를 위한 사실상의 표준 라이브러리
  - 데이터를 전처리하거나 통계정보, 또는 엑셀의 피벗 테이블과 같은 기능이 필요할 때 사용
  - 넘파이 기반으로 개발됨
- 사이파이(SciPy) : 파이썬에서 통계나 미분 연산 등 공업 수학 레벨에서 사용하는 거의 모든 수학 연산을 지원
  - 넘파이 기반으로 개발됨



## 2.4 시각화 도구

- 맷플롯립(Matplotlib) : 대중적으로 사용되어 왔으며 많은 파이썬 라이브러리의 근간이 됨
- 시본(Seaborn) : 그래프를 생성하는 라이브러리
  - 모든 기능은 맷플롯립을 기반으로 제공되며 맷플롯립과 상호 호환됨
- 플롯리(Plotly) : 사용자의 마우스에 반응하는 인터랙션 플로팅 기능을 제공해 필요에 따라 상세히 데이터를 볼 수 있음

## 2.5 머신러닝 프레임워크

- 실제 머신러닝 모델을 생성하고 데이터에 적용할 수 있도록 도와주는 도구
- 사이키런이 대표적인 머신러닝 프레임워크로 사실상 표준으로 사용됨
- 딥러닝에서 사용되는 신경망을 제외하고 기본적인 통계기반 라이브러리를 매우 다양하게 제공

## 3. 파이썬 머신러닝 환경 설치하기

- 구글에서 제공하는 코랩(Colab)을 사용하거나 다음 과정을 따라 머신러닝 환경을 직접 설치

### 3.1 미니콘다 설치하기

**01** 미니콘다 공식 페이지에 접속하여 윈도우용 미니콘다 (64bit) 설치 파일을 다운로드(Python 3.9 버전)

### Miniconda

Miniconda is a free minimal installer for conda. It is a small, bootstrap version of Anaconda that includes only conda, Python, the packages they depend on, and a small number of other useful packages, including pip, zlib and a few others. Use the

`conda install command` to install 720+ additional conda packages from the Anaconda repository.

See if Miniconda is right for you.

### Windows Installers

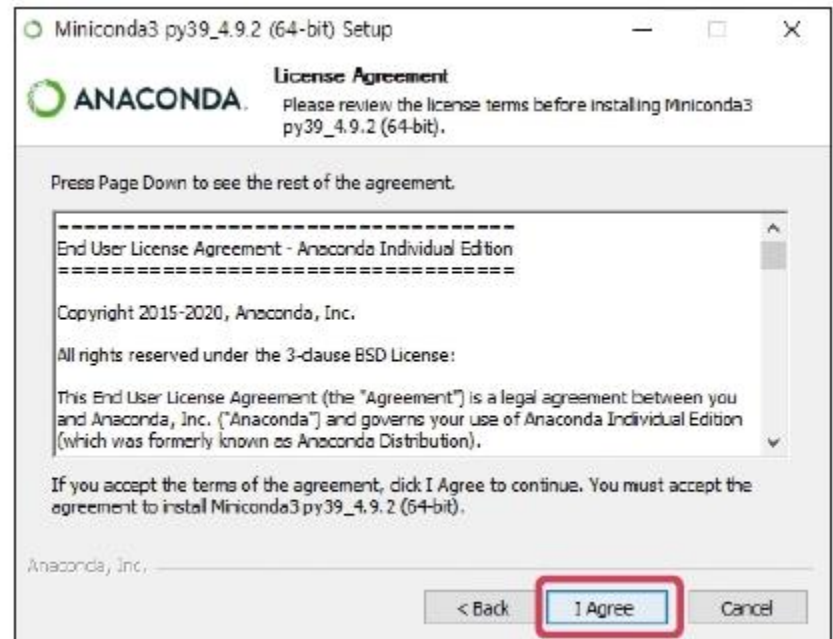
#### Windows

Python version	Name	Size	SHA256 hash
Python 3.9	Miniconda3 Windows 64-bit	57.7 MiB	c3a43d5bc4c4fa92454dbfa62c6cc8599045d875d7f602b31a271b9e8c3fec2b8
	Miniconda3 Windows 32-bit	54.9 MiB	5045fb90c4405d0ba21054262b7d1040a61a0739c1a56038ccb0250f232ad545
Python 3.8	Miniconda3 Windows 64-bit	57.0 MiB	4fa22b0a0497ba0b506608cb8043545372a99f5321c0120099a0e1d003f627c61
	Miniconda3 Windows 32-bit	54.2 MiB	9c2ef75bae97249c85c105793ca90fd1fe08a403f90a2a511feae81ce7ebc661
Python 2.7	Miniconda2 Windows 64-bit	54.1 MiB	6973015404832944e074bfe02bdac4594902eed4707bb51baa8f0dba4bf326c
	Miniconda2 Windows 32-bit	47.7 MiB	c8049d26f8b6b954b57bcd4e998d72d1ffa13f4a6b218e64e641504427b2817d

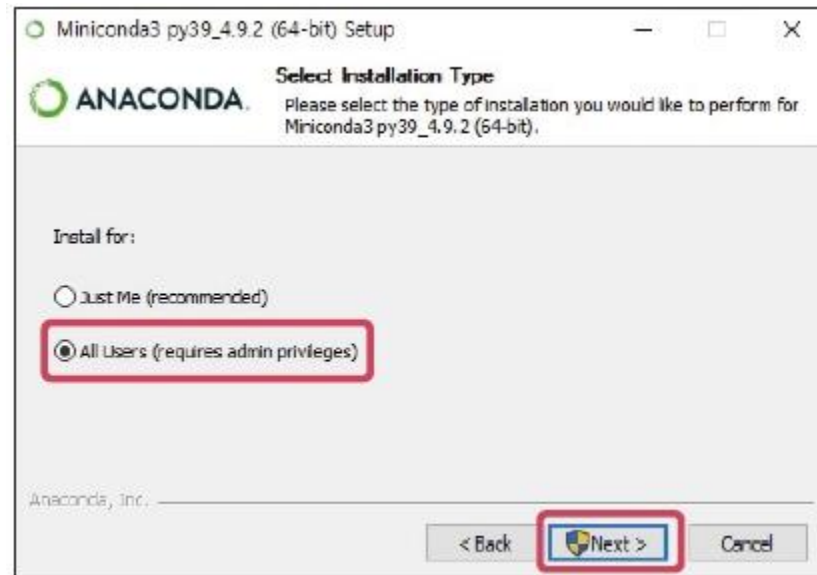
# 04 머신러닝 환경 구축하기

## CHAPTER 01 머신러닝의 기초

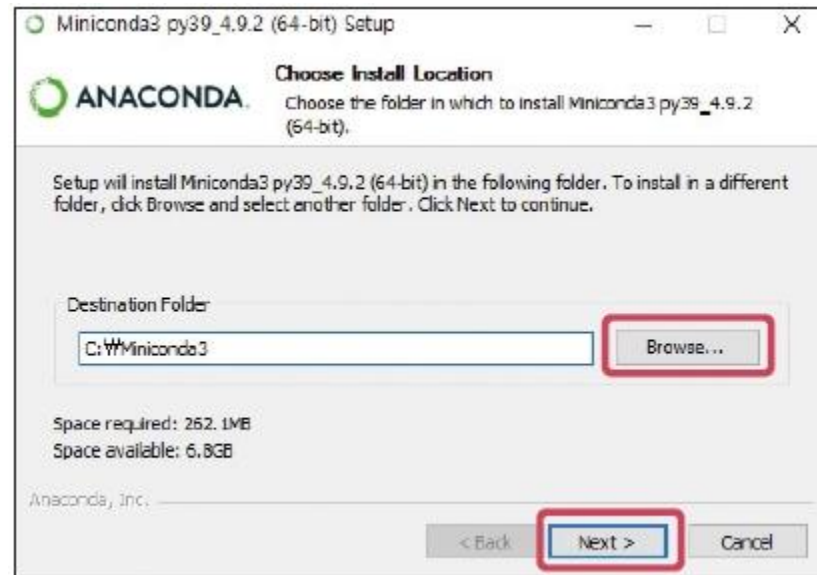
02 다운로드한 인스톨러를 실행하고 [Next]를 클릭한 후,  
다음 화면에서 [I Agree]를 클릭



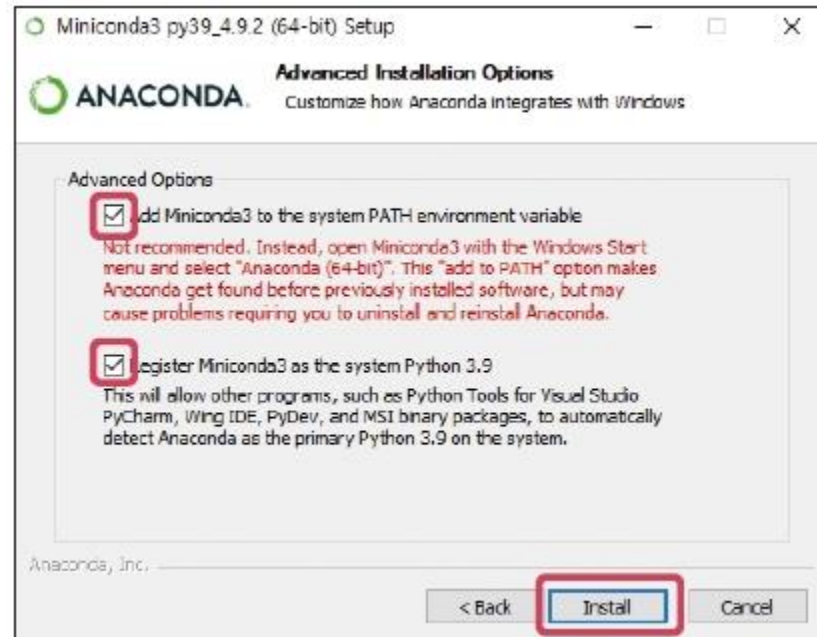
**03** 인스톨 타입 설정 창에서 'All Users'를 선택하고, [Next]를 클릭



**04** 인스톨 경로 설정 창에서 [Browse]를 클릭하여 적절한 경로를 지정하고, [Next]를 클릭



**05** Advanced Installation Options 항목을 모두 선택하고  
[Install]을 클릭





## 3.2 conda 가상환경 구성하기

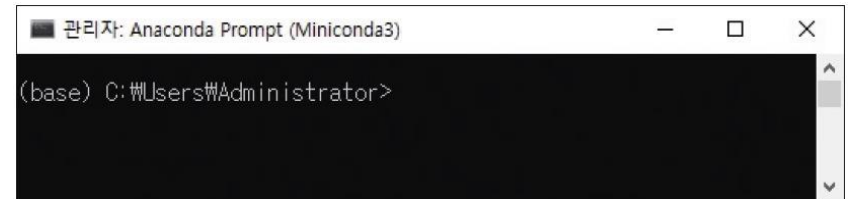
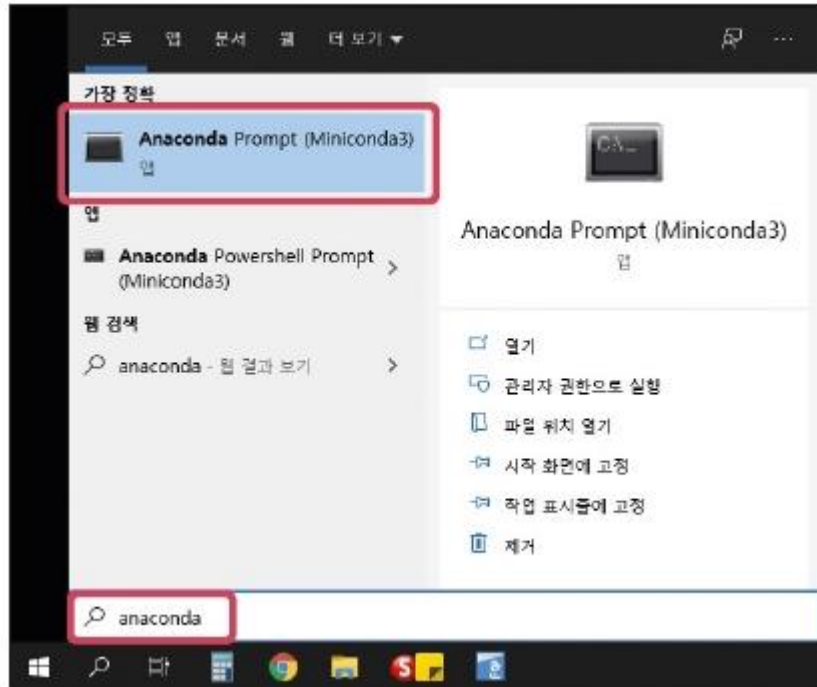
- conda 가상환경 : conda를 사용하기 위해 필요한 여러 라이브러리들을 하나의 환경에 설치하여 손쉽게 사용할 수 있도록 만들어줌

### 01 conda 가상환경이 실행될 수 있는 터미널 환경을 실행

- 윈도우 작업표시줄의 검색바에 anaconda를 입력하고 Anaconda Prompt (Miniconda3)를 실행하면 conda 전용 프롬프트가 나타남

# 04 머신러닝 환경 구축하기

## CHAPTER 01 머신러닝의 기초

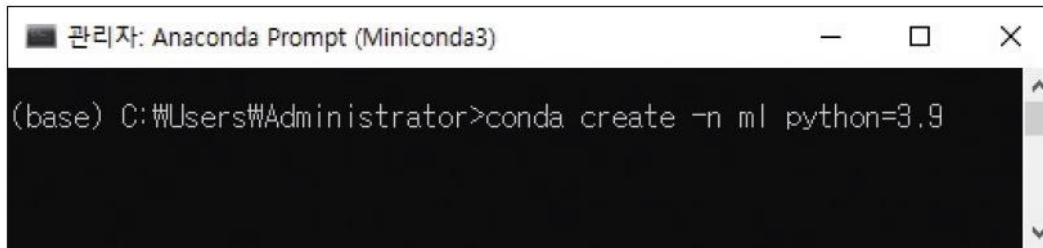


[TIP] 리눅스나 맥에서는 미니콘다를 설치하면 기본 가상환경이 바로 실행된다. 기본 가상환경은 터미널이 처음 시작될 때 '(base)'라고 표시된다.

## 04 머신러닝 환경 구축하기

**02** 실행된 터미널 환경에서 다음과 같은 명령어를 입력

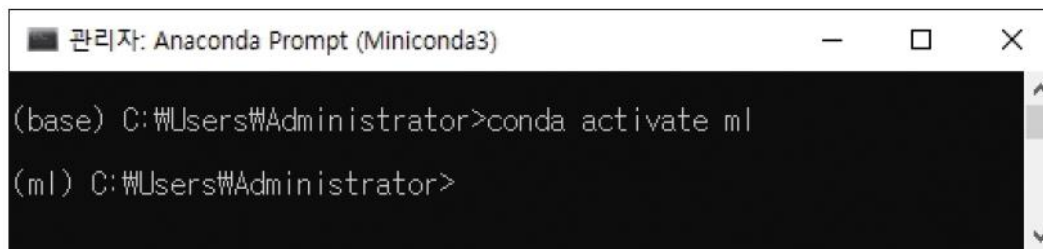
```
(base) C:\...>conda create -n ml python=3.9
```



A screenshot of the Anaconda Prompt window. The title bar reads '관리자: Anaconda Prompt (Miniconda3)'. The command prompt shows the command `(base) C:\Users\Administrator>conda create -n ml python=3.9` being entered. The prompt is currently at the end of the command line.

**03** 해당 가상환경을 수행하는 명령어를 실행하면, 가상환경의 명칭이 (base)에서 ml로 변경됨

```
(base) C:\...>conda activate ml  
(ml) C:\...>
```



A screenshot of the Anaconda Prompt window. The title bar reads '관리자: Anaconda Prompt (Miniconda3)'. The command prompt shows the command `(base) C:\Users\Administrator>conda activate ml` being entered. The prompt has moved to the next line, showing `(ml) C:\Users\Administrator>`, indicating the environment has been successfully activated.

## 3.3 각 라이브러리 설치하기

- 머신러닝에 필요한 모듈을 설치
- 각 라이브러리들은 서로 의존성을 가지고 있기 때문에 필요에 따라 서로 맞는 버전을 설치해야 한다

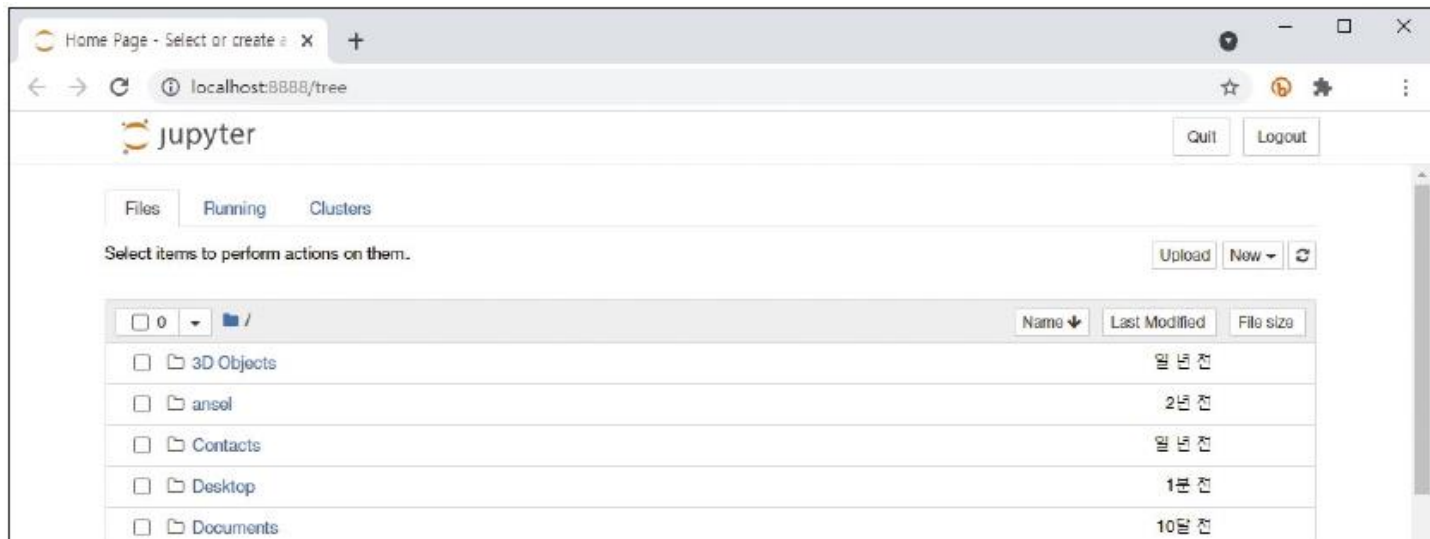
```
(ml) C:\...>conda install numpy pandas scikit-learn  
matplotlib seaborn jupyter  
(ml) C:\...>conda install -c plotly plotly  
(ml) C:\...>conda install -c plotly chart-studio
```

### 4. 주피터 노트북 사용하기

#### 4.1 주피터 노트북 실행하기

**01** 터미널 환경에서 주피터를 실행하면 현재 실행된 디렉토리를 기준으로 파일 리스트가 나타남

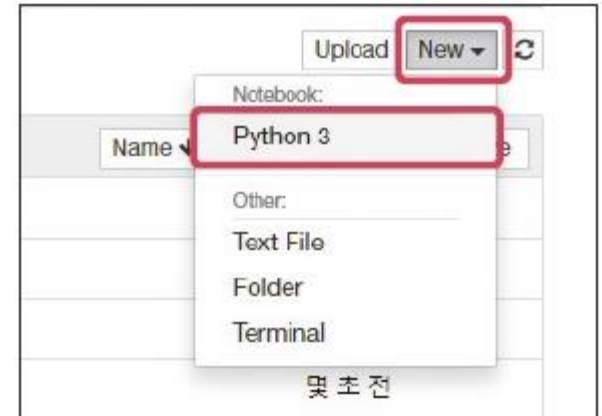
```
(ml) C:\...>jupyter notebook
```



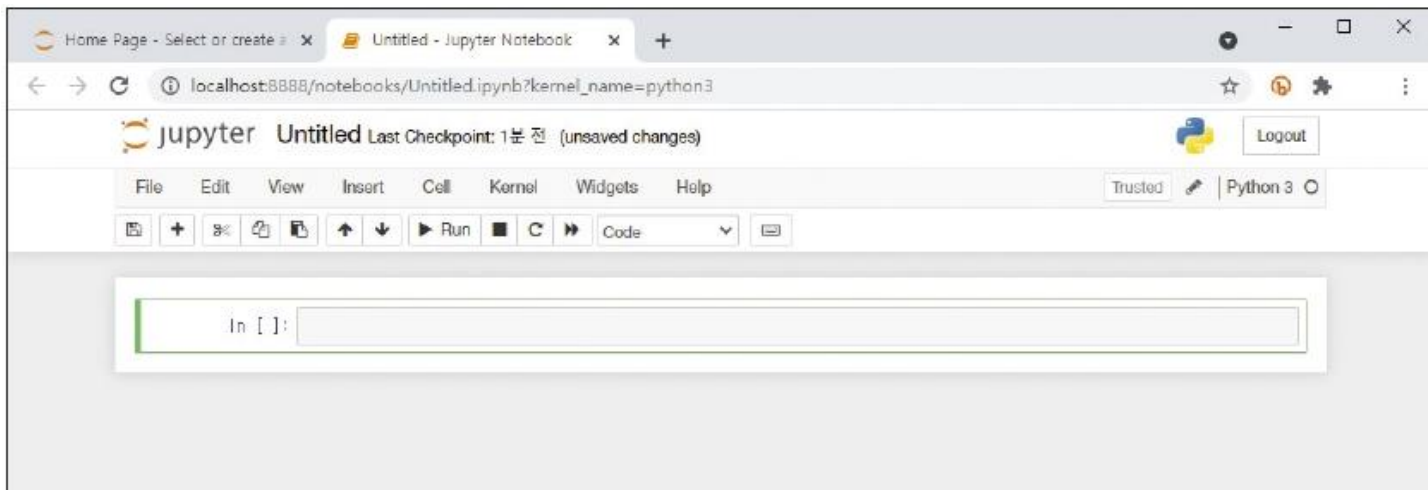
## 04 머신러닝 환경 구축하기

### CHAPTER 01 머신러닝의 기초

**02** 오른쪽 상단의 [New] 버튼을 누르고 'Python 3' 커널을 실행할 수 있는 가상환경 버튼인 [Python 3]을 클릭

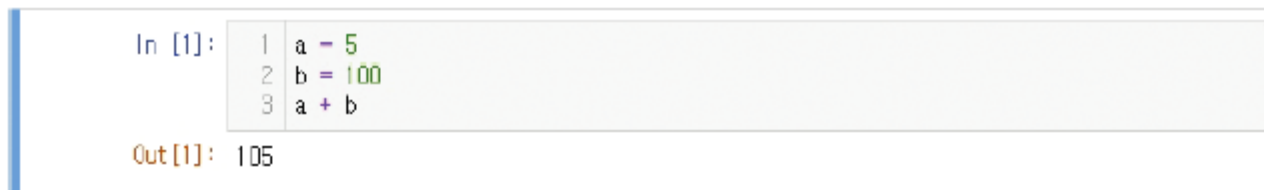


**03** 코딩을 할 수 있는 주피터 노트북이 실행됨



## 4.2 주피터 노트북에서 코드 실행하기

- 노트북(notebook) : 대화형 파이썬 셸인 주피터에서 대화를 하는 공간
- 셀(cell) : 노트북에서 한 칸
- 셀에는 필요한 코드를 붙여서 입력할 수 있고, 코드 작성 후에 [Ctrl] + [Enter]를 눌러 코드 실행

A screenshot of a Jupyter Notebook cell. The input area shows three lines of code: '1 a = 5', '2 b = 100', and '3 a + b'. The output area shows 'Out[1]: 105'.

```
In [1]: 1 a = 5
        2 b = 100
        3 a + b

Out[1]: 105
```

그림 1-27 주피터 노트북 실행 화면

### 4.3 주피터 노트북에서 다양한 단축키 활용

- 에디터 모드(editor mode) : 셀 안에서 코드를 작성할 수 있는 상태로, 메모장과 거의 같은 기능을 함
  - 커맨드 모드에서 [Enter]를 눌러 에디터 모드로 진입

표 1-3 에디터 모드와 관련된 단축키

단축키	내용
<b>Ctrl</b> + <b>Enter</b>	선택한 셀의 코드를 실행(다음 셀 추가 없음)
<b>Shift</b> + <b>Enter</b>	선택한 셀의 코드를 실행한 후 다음 셀을 추가하거나 다음 셀로 이동
<b>Enter</b>	줄 바꿈
<b>Shift</b> + <b>Tab</b>	함수가 포함된 줄에서 해당 함수의 기능을 상세히 설명
<b>Ctrl</b> + <b>Shift</b> + <b>-</b>	커서가 있는 지점을 기준으로 셀 분할



- 커맨드 모드(command mode) : 선택한 셀을 기준으로 추가·자르기·복사하기·붙여넣기·합치기 등 셀과 관련된 작업을 할 수 있음
  - 에디터 모드에서 [ESC]를 누르면 커맨드 모드로 진입

표 1-4 커맨드 모드와 관련된 단축키

단축키	내용
[A]	위로 셀 추가
[B]	아래로 셀 추가
[X]	셀 자르기
[C]	셀 복사하기
[V]	선택한 셀 아래에 붙여넣기
[D]+[D]	셀 삭제하기
[Shift]+[M]	선택한 셀과 아래 셀 합치기

단축키	내용
<b>M</b>	선택한 셀을 마크다운 셀로 변경
<b>Y</b>	마크다운 셀을 코드 셀로 변경(코드 입력 상태로 복구)

### [하나 더 알기] 마크다운과 레이텍

- 주피터는 단순히 코드뿐만 아니라 다양한 이미지, 텍스트, 수식 등을 하나의 문서로 나타낼 수 있다
- 하나의 노트북은 단순히 코드의 집합이라기보다 일종의 보고서라고 말할 수 있다
- 마크다운(Markdown) : 간단히 문서를 작성하기 위해 사용되는 마크업 언어로, 복잡하지 않은 문서를 HTML 형태로 표현할 때 사용
- 레이텍(LaTeX) : 데이터 분석을 하다가 수식을 적어야 할 때 사용