



기계학습 (1): 기계학습 개요 및 Linear Regression

임 경 태

CONTENTS

—

- ① 데이터분석과 기계학습
- ② Linear Regression
- ③ Multi-variable
Linear Regression

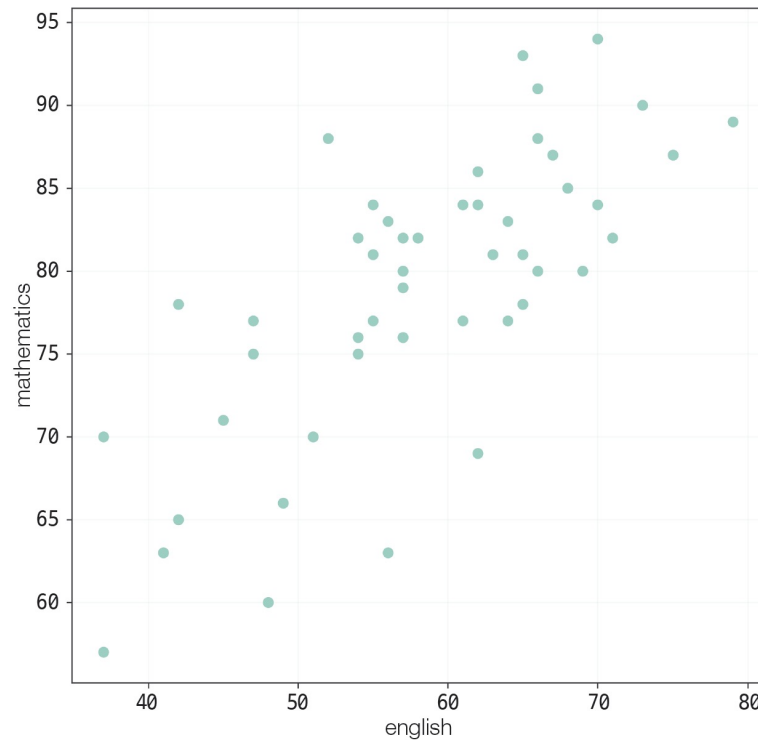
CONTENTS

- ① 데이터분석과 기계학습
- ② Linear Regression
- ③ Multi-variable
Linear Regression

01. 데이터 분석과 기계학습 (데이터 분석이란?)

□ 데이터 분석: **관측된** 데이터를 설명할 수 있는 **특징** (대푯값)을 찾아내고 **분석**하는 것

- 우리반 영어, 수학 점수는 어떤 상관관계가 있을까?
- 영어를 잘하는 친구들이 수학도 잘할까?

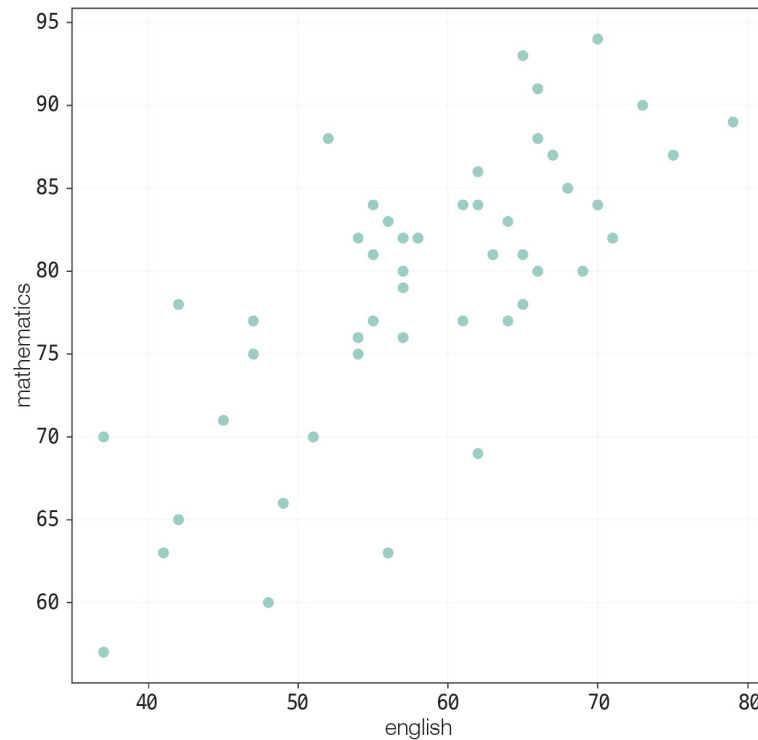


[그림 3-4] 산점도

- 대푯값
- 중위값
 - 분산
 - 공분산
 - 상관계수 등

01. 데이터 분석과 기계학습 (데이터 분석의 확장)

- 전체 데이터 (모집단)를 획득할 수 없을 때 **관측된** 데이터로 전체 데이터의 **특징**을 알 수 있을까? 얼마나 정확할까?

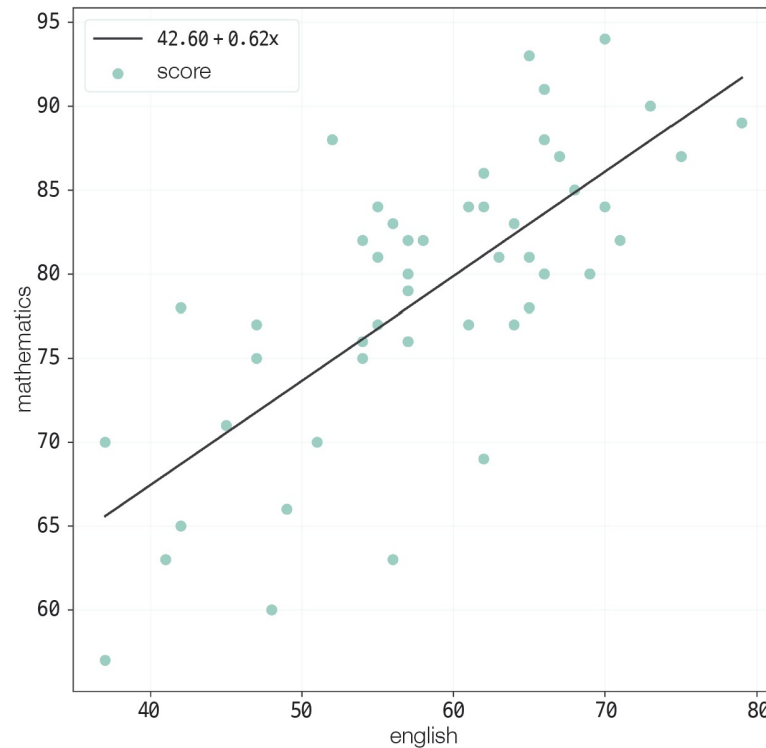


[그림 3-4] 산점도

- 대푯값
 - 중위값
 - 분산
 - 공분산
 - 상관계수 등

01. 데이터 분석과 기계학습 (좁은 의미의 기계학습)

□ 기계학습: 관측 데이터 (표본)를 활용해 전체 데이터(모집단)를 최적으로 모사하는 프로그램



[그림 3-5] 산점도와 회귀직선

[국어사전]

모사 (模寫)

[모사] ㄴㅇ

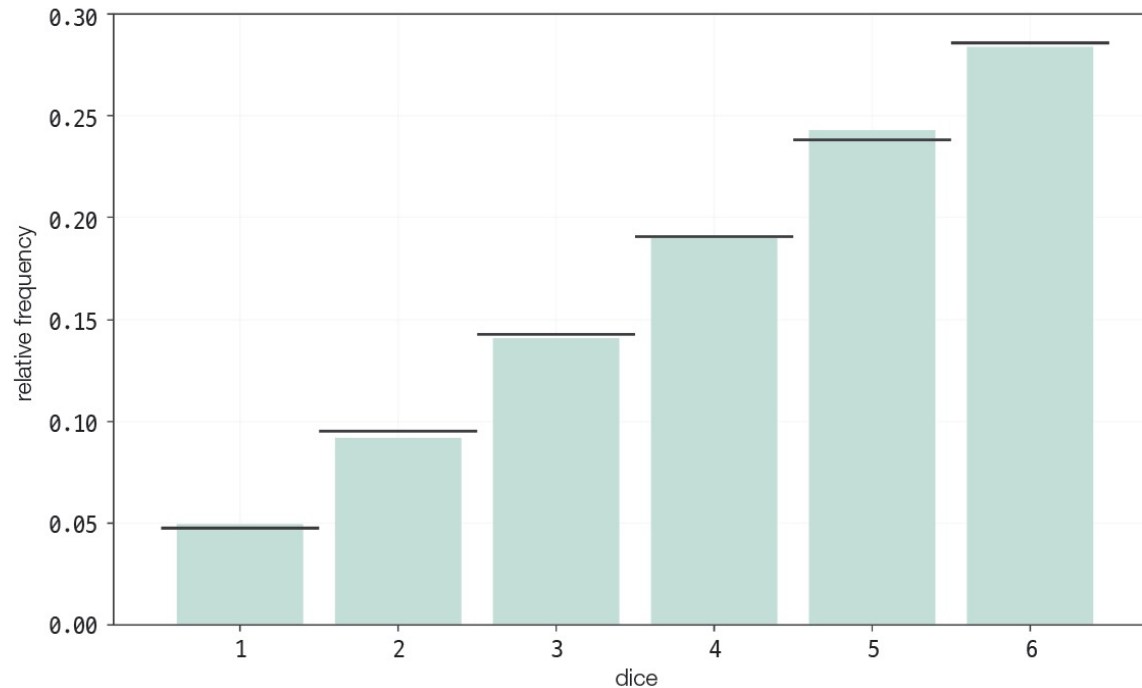
- 1 사물을 형체 그대로 그림. 또는 그런 그림.
- 2 원본을 베끼어 씬.
- 3 어떤 그림의 본을 떠서 똑같이 그림.

□ 모사의 방법
- 회귀직선 등

01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

❑ 그런데 모사(==모델링)가 되기는 함?

- 불공정한 주사위를 10000번 돌리고 그 눈금을 relative frequency로 확인해보면 확률 모형이 나옴



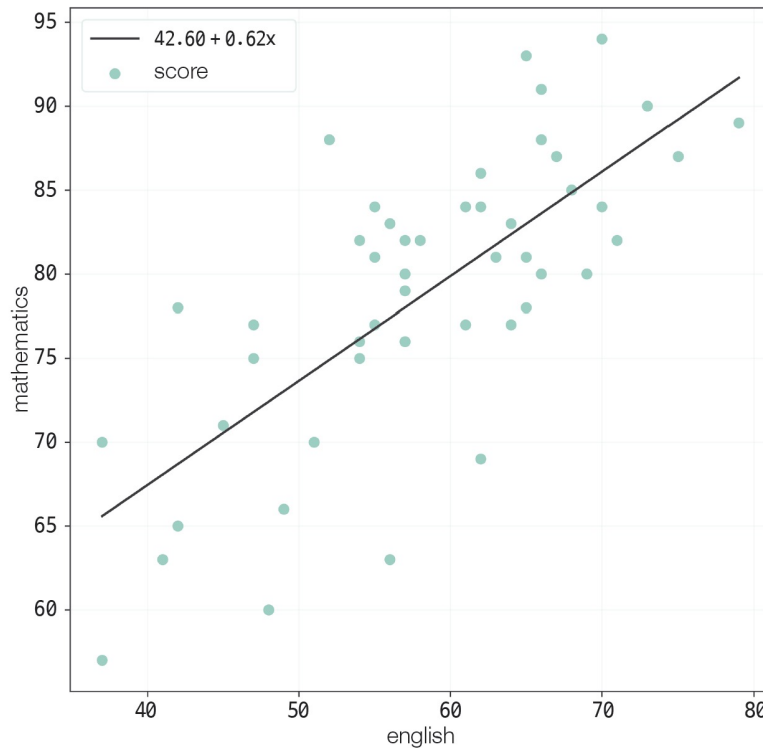
[그림 4-3] 10000번 시행했을 때 주사위 눈의 히스토그램 [SAMPLE CODE](#)

- ❑ 모사의 방법
 - 확률모형

01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

□ 그런데 모사(==모델링)가 되기는 함?

- 우리반 영어, 수학 점수 (관측) 데이터를 이용해 **영어** 점수와 **수학** 점수 간의 상관관계를 가장 잘 나타내는 **회귀모형**을 그릴 수 있음



[그림 3-5] 산점도와 회귀직선

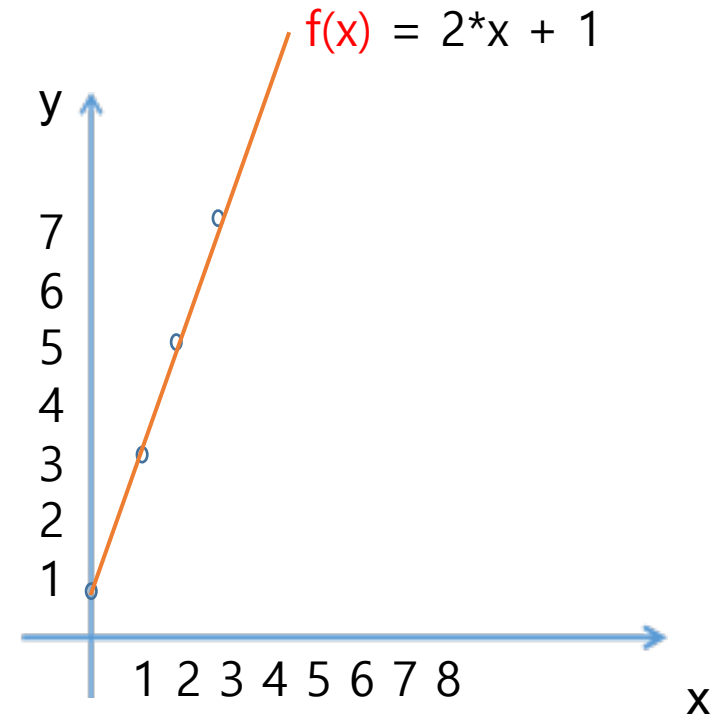
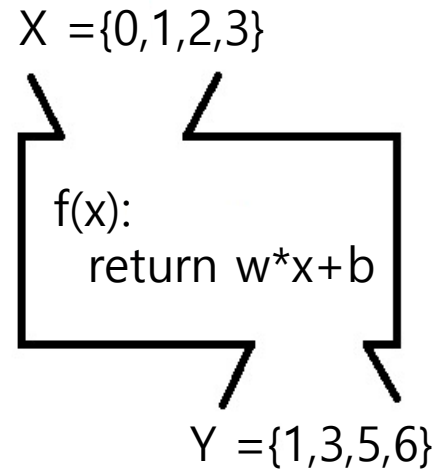
- 모사의 방법
- 회귀직선 등

01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

□ 그러면 모델링은 어떻게 함?

- 관측 데이터를 가장 잘 설명하는 함수 $f(x)$ 를 만들면 됨. 여기서 x = input data

$$y = f(x)$$
$$f(x) = w * x + b$$



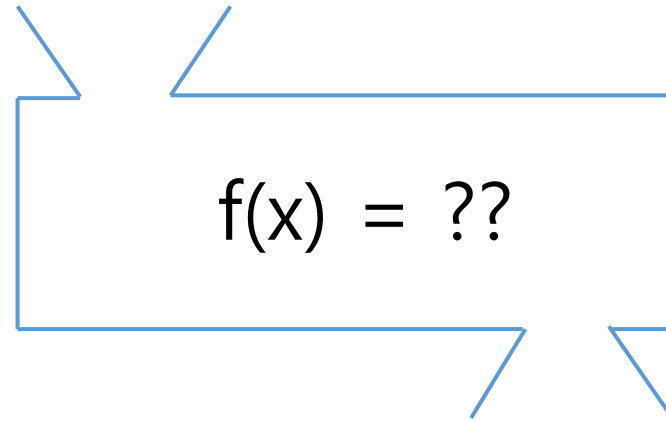
01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

□ 기계학습 = Data(D)로 Model(M)을 만드는 과정

01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

□ 기계학습 = Data(D)로 Model(M)을 만드는 과정

X : 입력 데이터

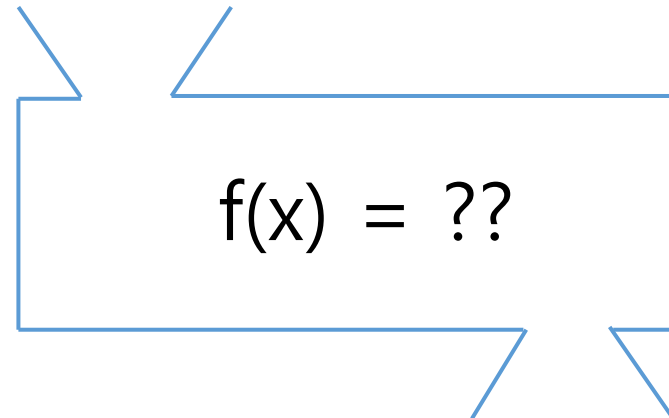


Y : 예측, 분류 하고싶은 것

01. 데이터 분석과 기계학습 (기계학습의 분류)

□ 기계학습 = Data(D)로 Model(M)을 만드는 과정

X : 입력 데이터



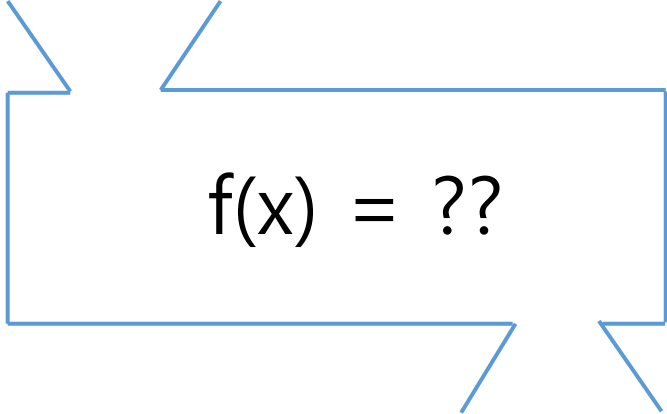
Y : 예측, 분류 하고싶은 것

- 데이터에 Y값이 있을 때: 지도학습
- Y값이 없을 때: 비지도학습
- Y값이 없는데 잘잘못을 판단할 수 있을 때: 강화학습

01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

- 기계학습 = Data(D)로 Model(M)을 만드는 과정
 - 데이터가 숫자라면? (선형회귀 예제)

$X = \text{수학 점수 } \{45, 29, 83, 66\}$



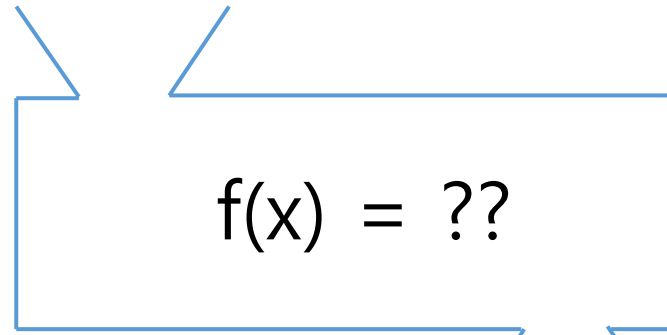
$f(x) = ??$

$Y = \text{영어 점수 } \{65, 49, 93, 56\}$

01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

- 기계학습 = Data(D)로 Model(M)을 만드는 과정
 - 데이터가 글자라면? (자연어 처리)

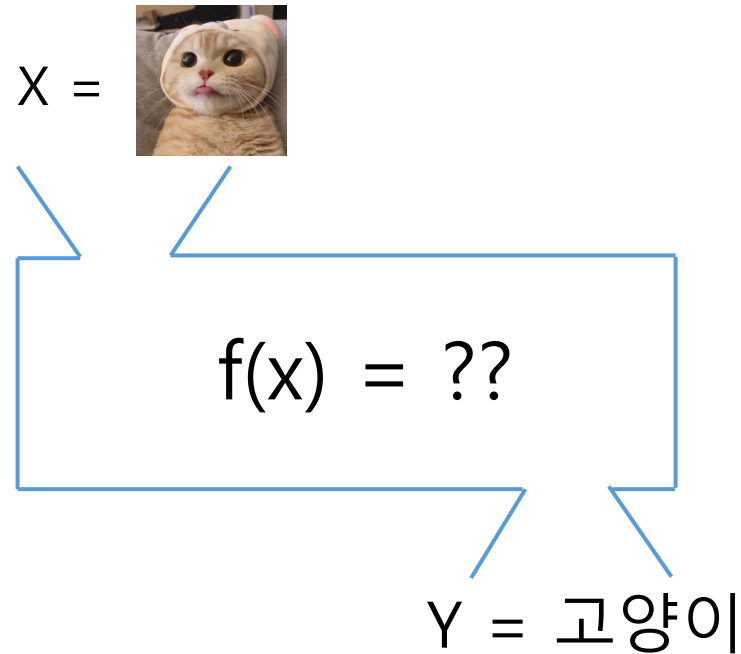
X = 이 면도기 최악임 비추



Y = 부정적

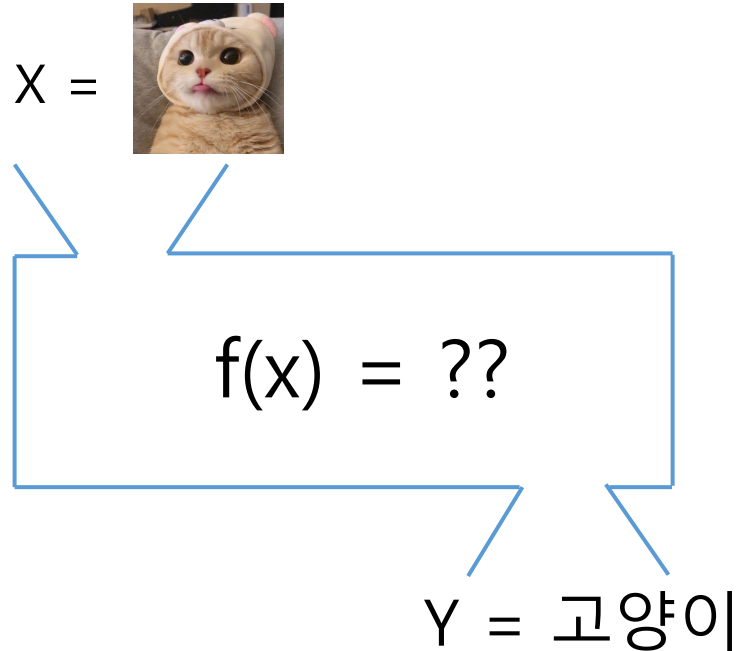
01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

- 기계학습 = Data(D)로 Model(M)을 만드는 과정
 - 데이터가 이미지라면? (이미지 처리)



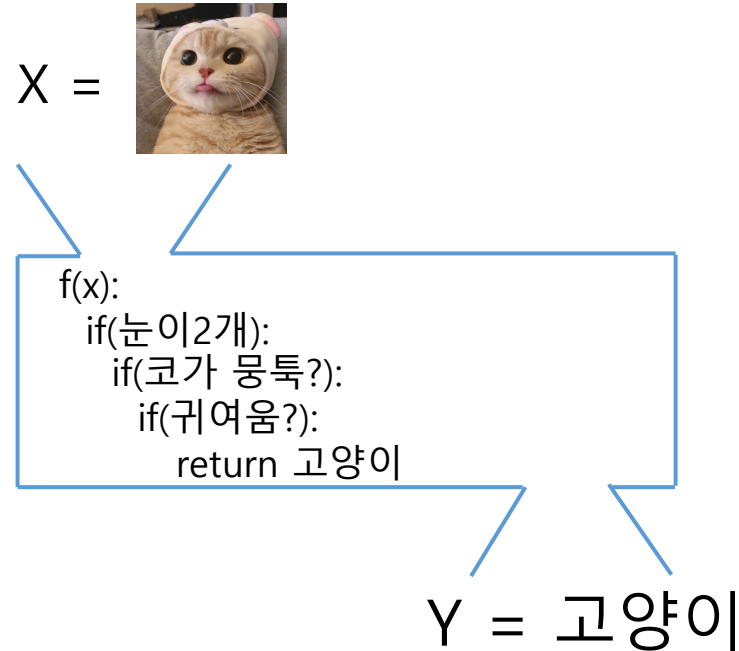
01. 데이터 분석과 기계학습 (기계학습의 기본 개념)

- 기계학습 = Data(D)로 Model(M)을 만드는 과정
 - 결국 **데이터 유형**에 따라 **$f(x)=??$ 부분을 구현**하면 되는 것! **어떻게?**



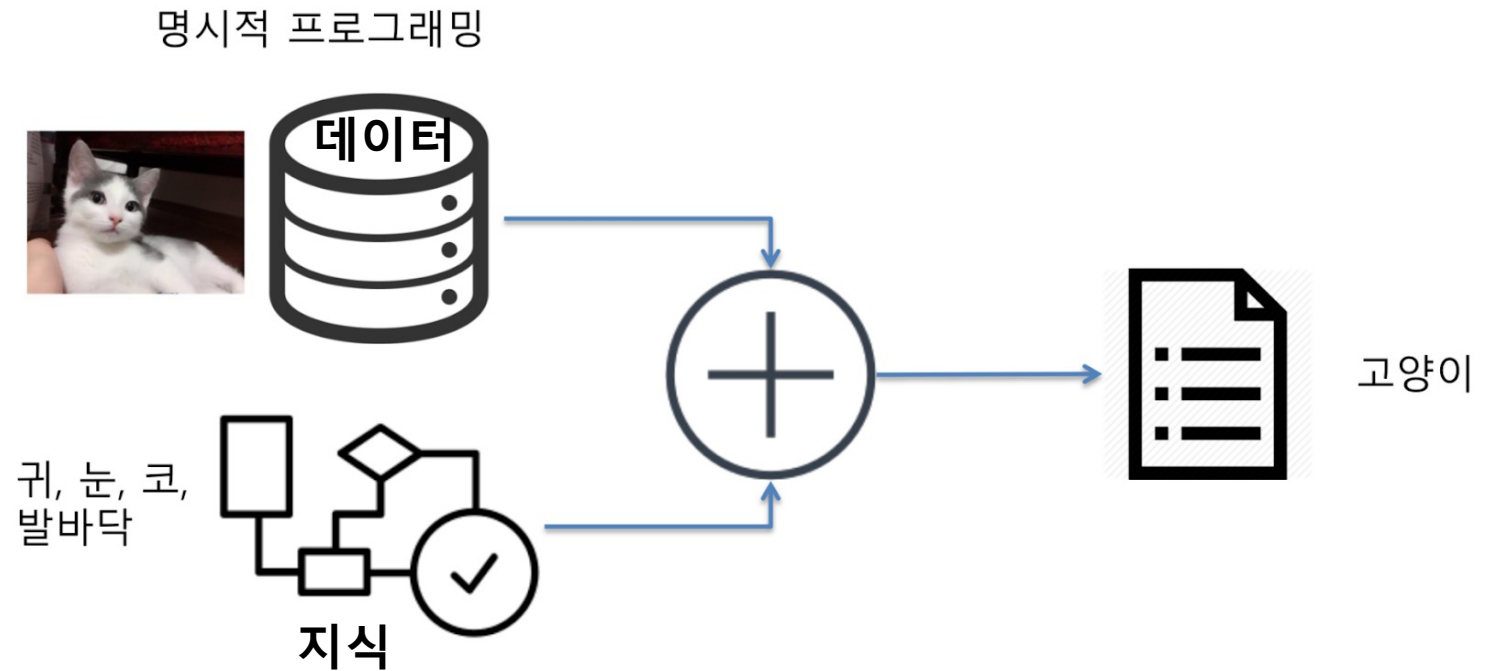
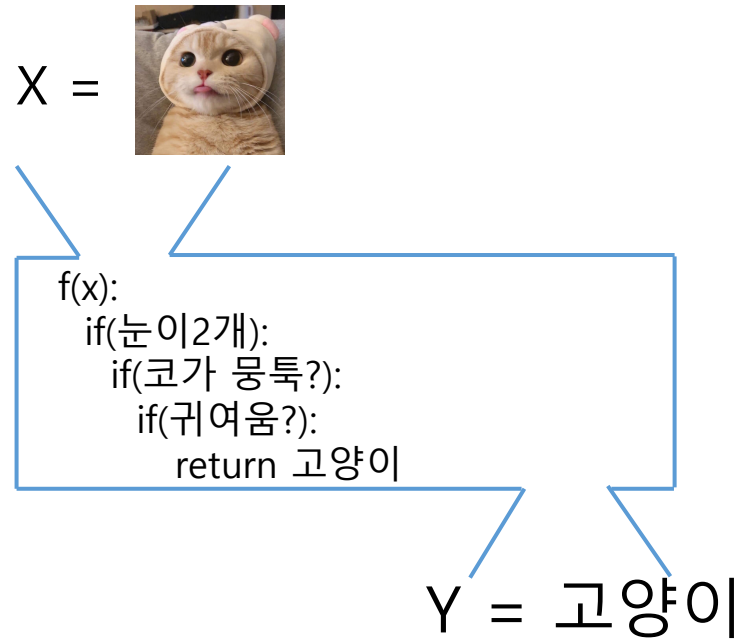
01. 데이터 분석과 기계학습 (데이터 모델링 방법)

- ❑ 어떻게 데이터로부터 모델링= $f(x)$ 을 할까?
 - 우리는 이미 너무 익숙하게 $f(x)$ 를 구현해 왔다. **명시적 프로그래밍으로!**



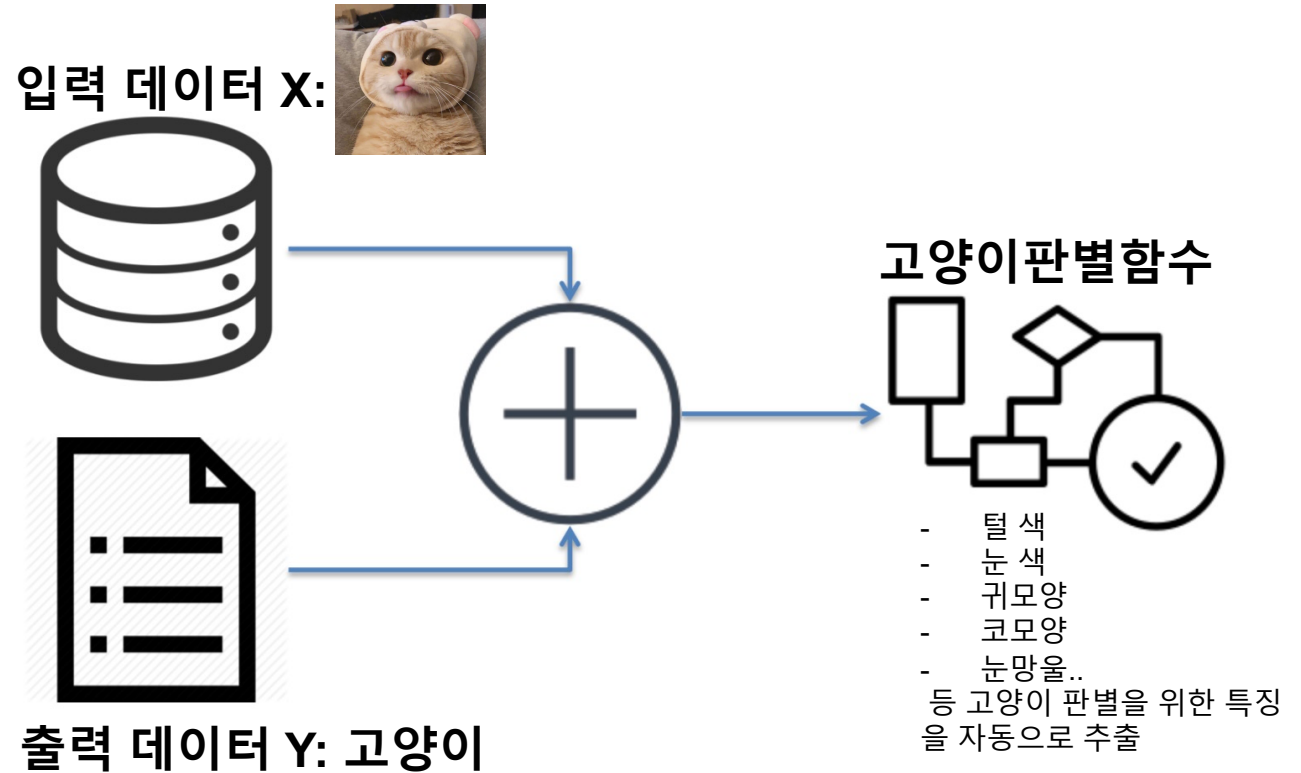
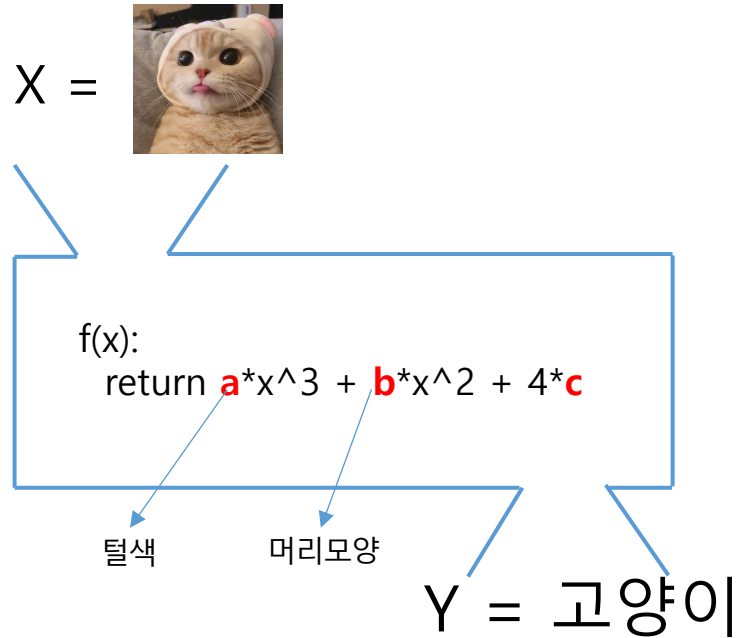
01. 데이터 분석과 기계학습 (명시적 프로그래밍의 모델링 방법)

- 명시적 프로그래밍: 사람이 데이터를 관측하고 지식을 추출해서 규칙을 정함
 - (의문점) 고양이의 모든 특징을 넣을 수 있을까? 개와 고양이는 어떻게 구분하지?



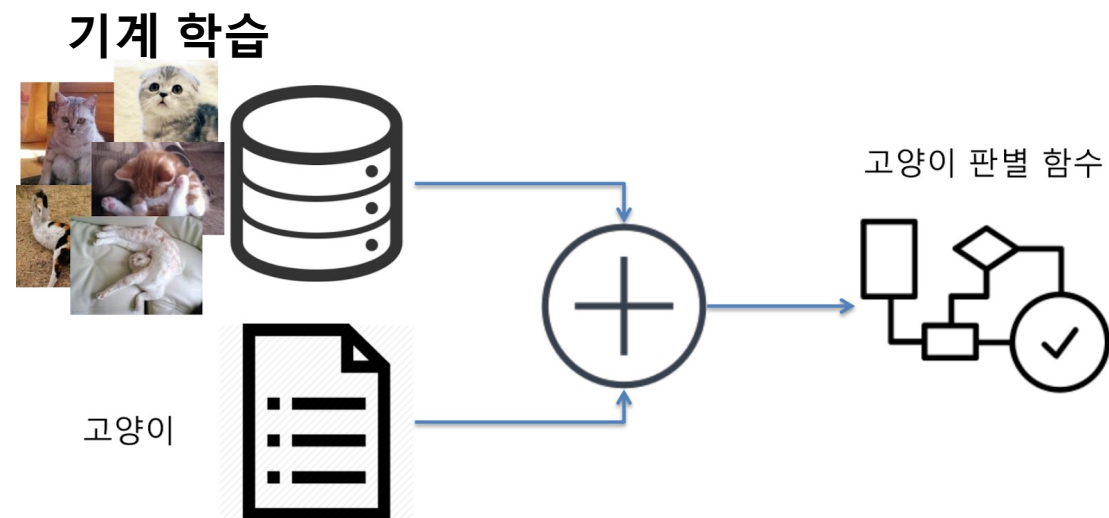
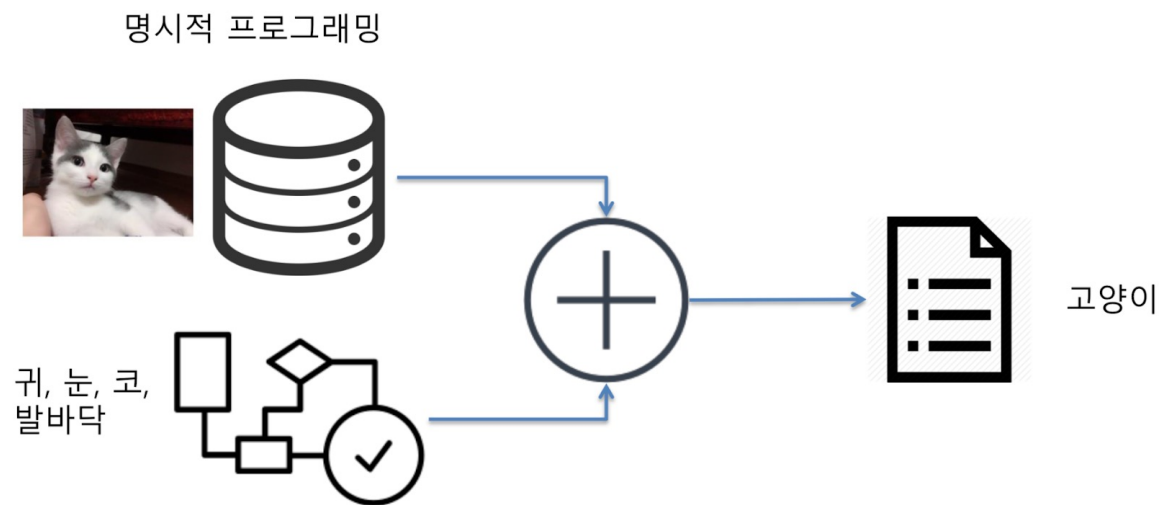
01. 데이터 분석과 기계학습 (기계학습의 모델링 방법)

- 기계학습의 모델링: 데이터로부터 문제 해결을 위한 지식을 생성
 - 고양이 사진으로부터 고양이를 분별할 수 있는 규칙을 자동으로 생성하도록 해보자



01. 데이터 분석과 기계학습 (지식기반 프로그래밍 vs 기계학습)

□ 명시적 프로그래밍과 기계학습의 비교



다 좋은데 고양이 판별 지식은 어떻게 자동으로 생성하나?

01. 데이터 분석과 기계학습 (기계학습의 모델링 방법)

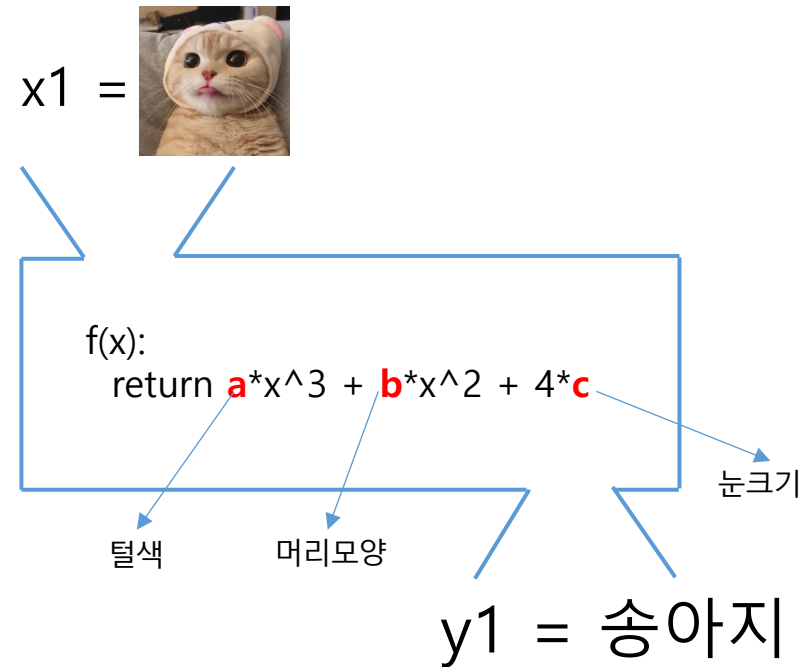
□ 기계학습에서 자동으로 생성될 지식은 어떻게 찾을 수 있을까?

- 강하게 키우자. 문제를 주고 예측해보게 한 뒤에 틀리면 혼내고 정답 알려주고 복습 시키자

- 0) 입력으로 x1을 받았다
- 1) 최초 f(x)를 a=2, b=3, c=9로 그냥 찍었다.
- 2) 송아지와 가까운 return이 나왔다
- 3) 틀렸다.. 혼났다. 정답이 고양이라고 한다.
- 4) 정답과 비교해보니 a=털색이 좀더 gray일때 고양이 라고 한다.
- 5) a,b,c값을 x1과 비슷한 사진이 들어오면 고양이 값이 나올 수 있도록 **조정**해본다



앞으로 우리가 배울 전부





감사합니다.