

Part.01

Machine Learning의 개념과 종류

I Machine Learning의 개념

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

I Machine Learning

커리큘럼

- Part1 : Machine Learning의 개념과 종류
- Part2 : 회귀분석 – 통계학기초, 회귀분석, PCA
- Part3 : Machine Learning알고리즘 – KNN, LDA, SVM, DecisionTree, ANN
- Part4 : Ensemble Learning – Bagging, Boosting, RandomForest, Stacking
- Part5 : Clustering – Kmeans, Hierarchical, DBSCAN
- Part6 : Class imbalanced problem – Oversampling, Undersampling
- Part7 : 빅콘테스트 후기

I Machine Learning의 개념

- “무엇(X)으로 무엇(Y)을 예측하고 싶다”

데이터(행렬)

X						Y
성별	키	몸무게	체지방	BMI지수	폐활량	흡연여부
남	182	78	18	15	86	Y
여	156	52	25	17	95	N
여	165	58	21	19	98	N
...

I Machine Learning의 개념

- 기계 학습 또는 머신 러닝(영어: machine learning)은 인공지능의 한 분야로, 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야를 말한다 (위키피디아)

모형 (머신 러닝 알고리즘)

$$Y = f(X)$$

출력 변수 (종속변수, 반응변수) 입력 변수 (독립변수, feature)

- 주어진 데이터를 통해서 입력변수와 출력변수간의 관계를 만드는 함수 f 를 만드는 것
- 주어진 데이터 속에서 데이터의 특징을 찾아내는 함수 f 를 만드는 것

I Machine Learning의 개념

- Machine Learning으로 할 수 있는 것들

X : 고객들의 개인 정보 및 금융 관련 정보, Y : 대출 연체 여부 -> 대출 연체자 예측 탐지 모델, 대출 연체 관련 주요 feature 추출

X : 게임 유저들의 게임 내 활동 정보, Y : 게임 이탈 여부 / 어뷰징 여부 -> 이상 탐지 모델

X : 숫자 손 글씨 데이터, Y : 숫자 라벨(0~9) -> 숫자 이미지 분류 모델

X : 상품 구매 고객 특성 정보 -> 군집화를 통한 고객 특성에 따른 segmentation

X : 고객들의 상품 구매 내역 -> 매장내 상품 진열 위치 리뉴얼을 통한 매출 증대

X : 쇼핑몰 페이지 검색 및 클릭 로그 기록 -> 맞춤 상품 추천 시스템

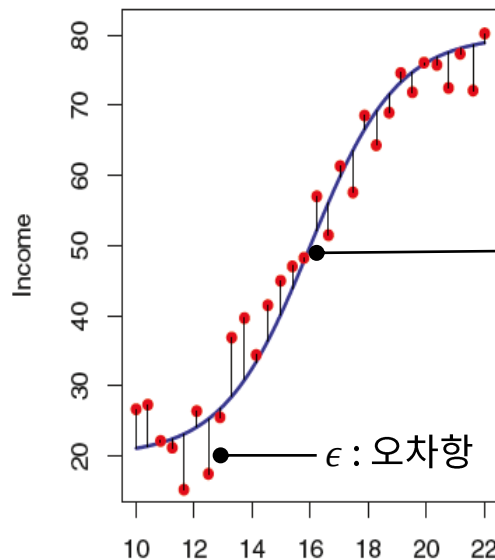
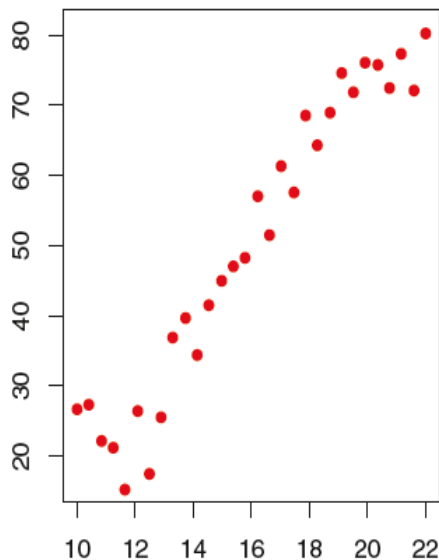
X : SNS데이터 및 뉴스 데이터 -> 소셜 및 사회 이슈 파악

I Machine Learning의 개념

■ f 란 무엇인가 (회귀 분석인 경우)

- f 를 구하기 위해서 입력 변수와 출력 변수가 필요함
- p 개의 입력 변수 X_1, X_2, \dots, X_p 가 있고, 출력 변수 Y 가 있을 때, $X = (X_1, X_2, \dots, X_p)$ 라 하면 입력 변수와 출력 변수의 관계를 나타내는 식은 다음과 같음

$$Y = f(X) + \epsilon$$



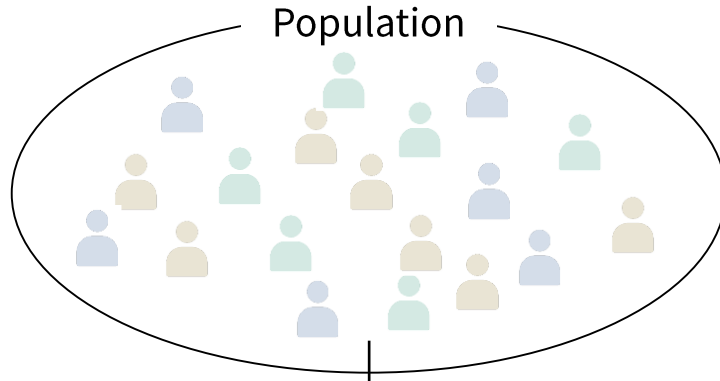
f : 정해졌지만 알 수 없는 함수

ϵ : 오차항

I Machine Learning의 개념

- f 란 무엇인가 (회귀 분석인 경우)

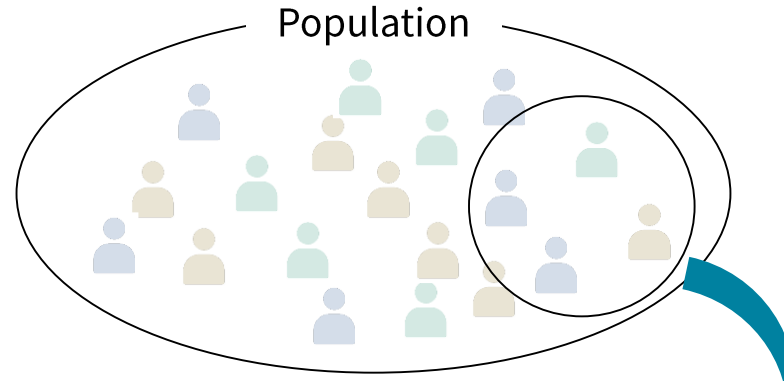
Population을 알 때



$$Y = f(X)$$

'신'만이 알고 있음!

Population을 모를 때



학습 데이터



$$\hat{Y} = \hat{f}(X)$$

using machine learning
(statistics)