

Statistical Machine Learning

1주차

담당: 15기 박지우

1. How does a Machine learn?

2. End to end ML Project

1. How does a machine Learn?

Machine Learning

[머신러닝은] 명시적인 프로그래밍 없이 컴퓨터가 학습하는 능력을 갖추게 하는 연구 분야다.

_아서 새뮤얼 Arthur Samuel, 1959

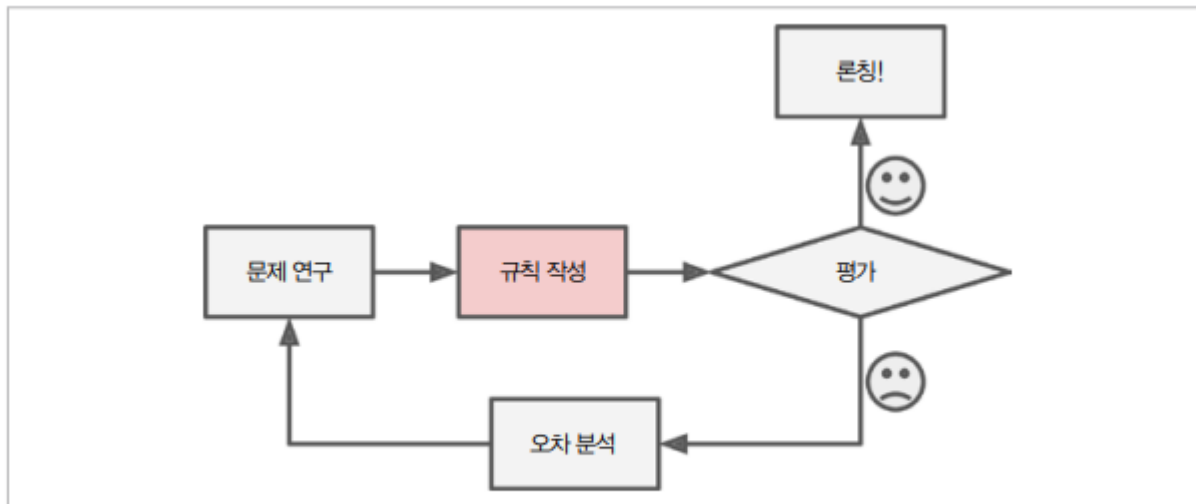


그림 1-1 전통적인 접근 방법

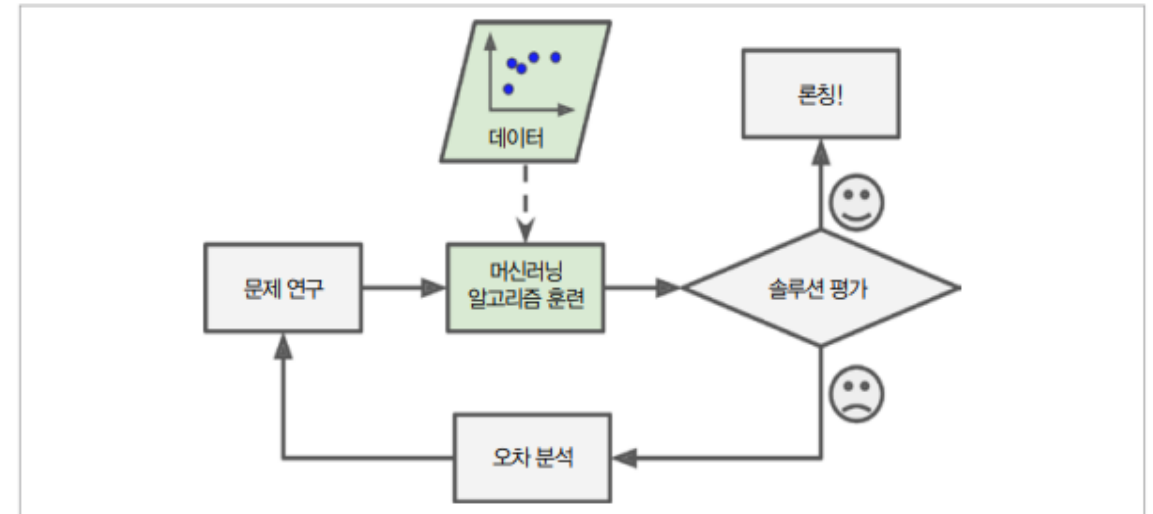
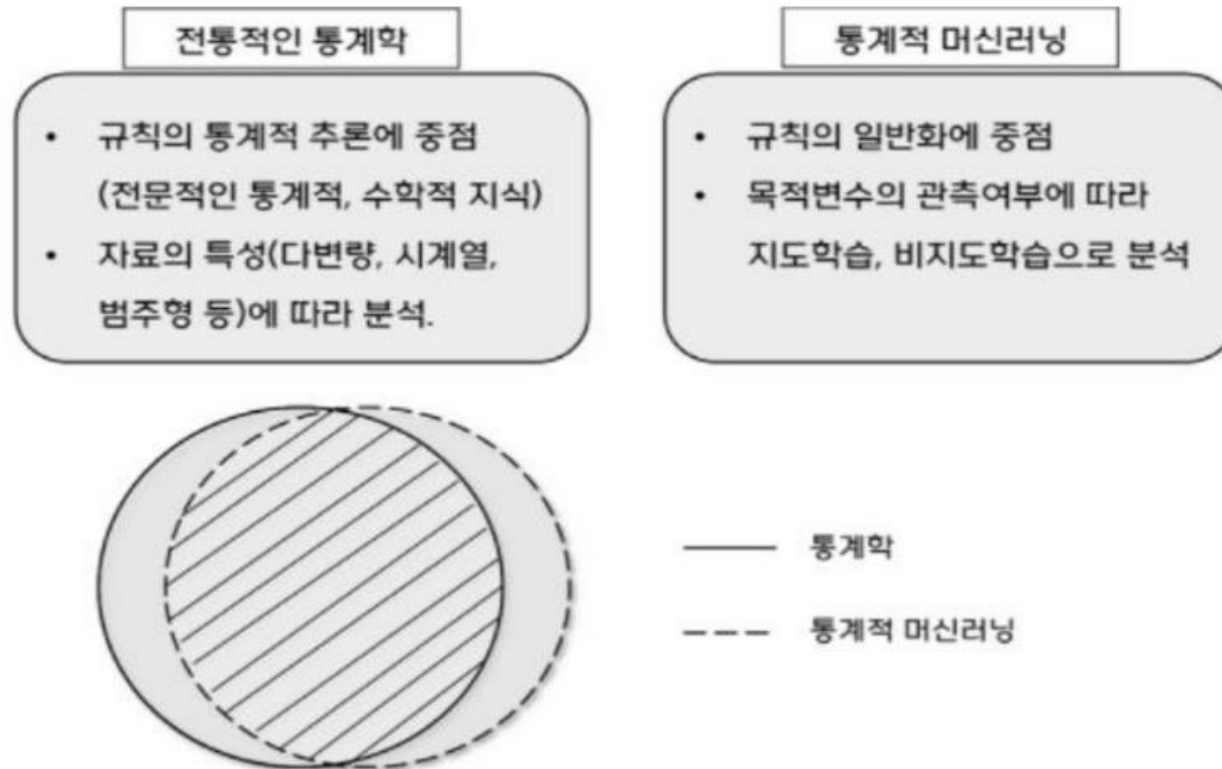


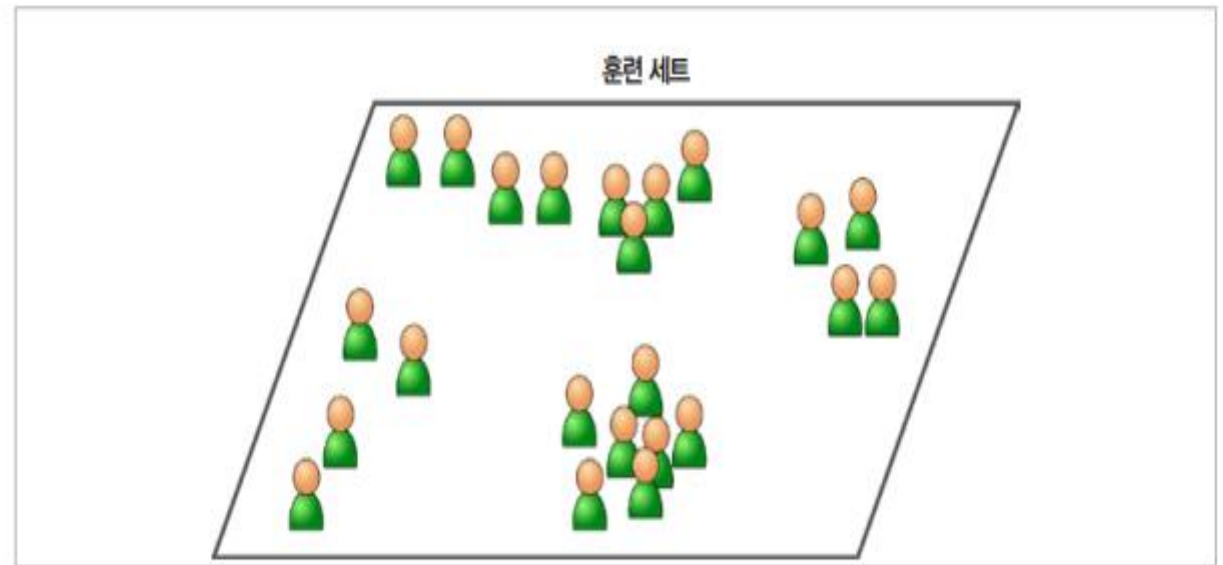
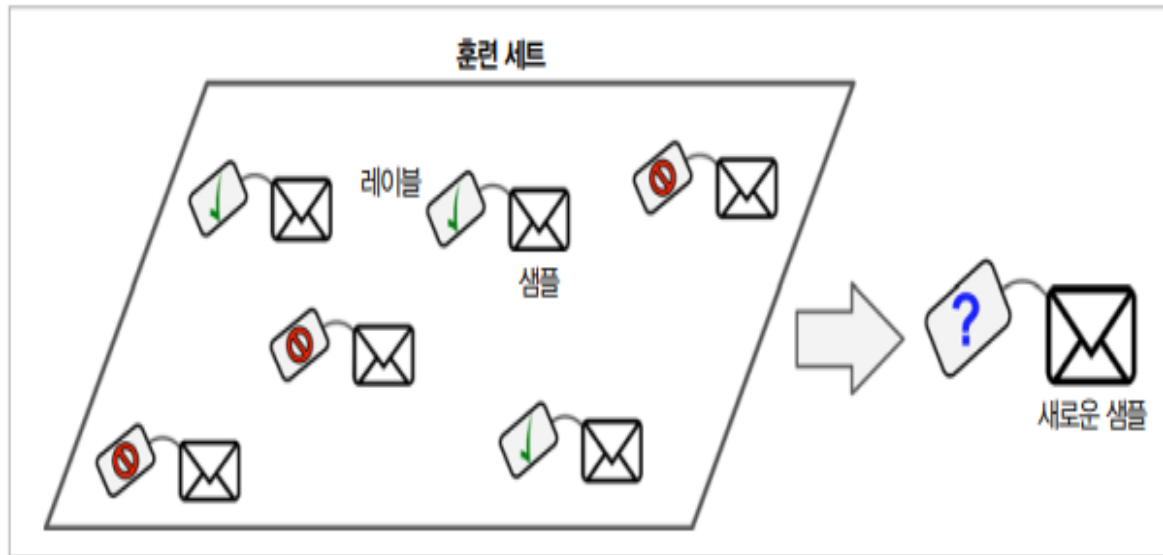
그림 1-2 머신러닝 접근 방법

Statistical Machine Learning



Machine Learning

- 지도 학습 VS 비지도 학습



Machine Learning

- 배치 학습 VS 온라인 학습
- 사례 기반 VS 모형 기반

Machine Learning



2. End to end ML Project

EDA(Exploratory Data Analysis)

- EDA란?

수집한 데이터를 다양한 시각에서 이해하는 과정

-> 데이터 분석 전에 그래프, 통계적 방법으로 자료를 바라보는 것

- 필요성

EDA(Exploratory Data Analysis)

- 시각화
- 상관계수
- 결측치

Data Preprocess

보다 높은 정확성을 갖는 분석을 위해 원자료에 대해 전환 및 가공을 거치는 단계
(AutoML의 등장으로 그 중요도 및 비중이 높아지고 있음)

정규화와 표준화

특성변수의 단위 등에서 나타나는
차이를 조정해주는 역할

One-Hot Encoding

범주형 변수를 수치형 변수로 변환

Data Preprocess

차원축소

Feature이 많을 때 발생하는 overfitting을
방지하기 위하여 진행

이상치/결측치 처리

머신러닝 모형은 직접 결측치를 처리할
수 없음

불균형 자료처리

불균형 자료 문제 해소를 위한 과대표집
방법

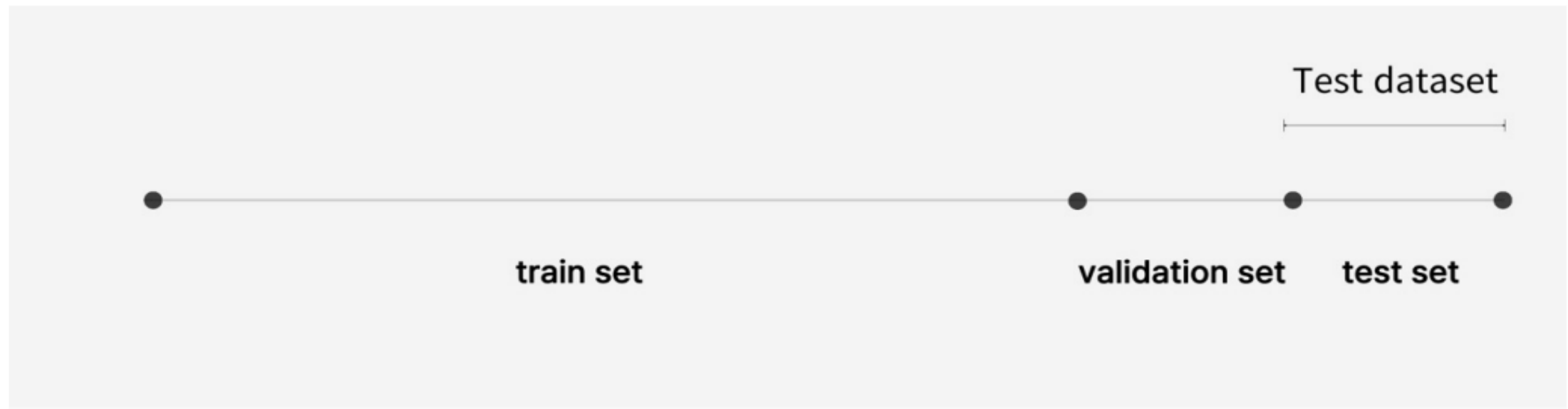
Learning Process

Use train set to fit a model + adjust hyperparameters of model



Model Evaluation

To overcome overfitting: Increase data size, regularization, ensemble method etc.



Model Evaluation



Regression

지도학습: y 가 연속형



Classification

지도학습: y 가 범주형

Prediction

