




# Statistical Machine Learning

1주차  
담당: 13기 박주영





How does a machine learn?



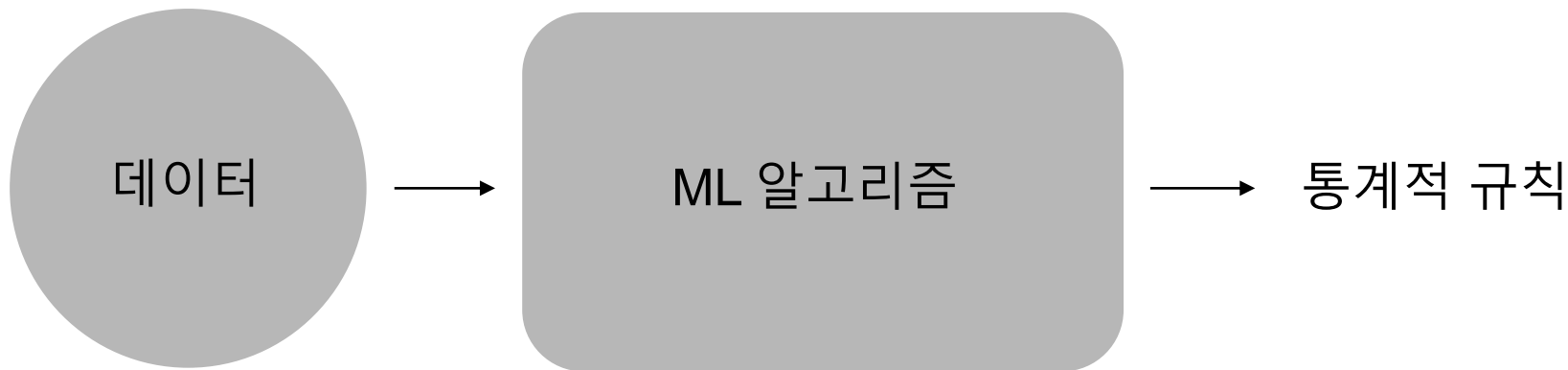
# Big Data

---

- Volume
- Velocity
- Variety

# What is ML?

---



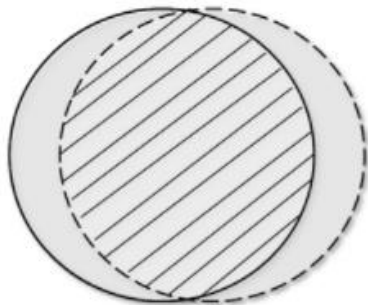
# What is Statistical ML?

## 전통적인 통계학

- 규칙의 통계적 추론에 중점  
(전문적인 통계적, 수학적 지식)
- 자료의 특성(다변량, 시계열, 범주형 등)에 따라 분석.

## 통계적 머신러닝

- 규칙의 일반화에 중점
- 목적변수의 관측여부에 따라 지도학습, 비지도학습으로 분석

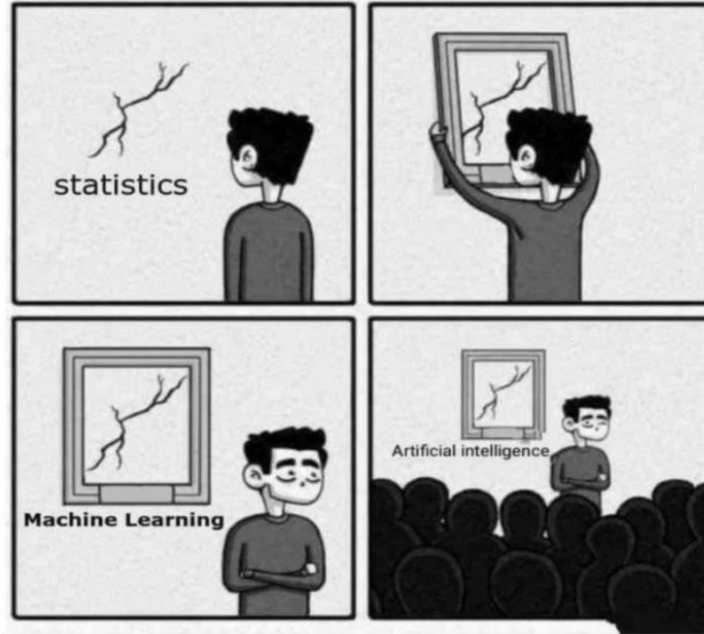


—— 통계학

--- 통계적 머신러닝

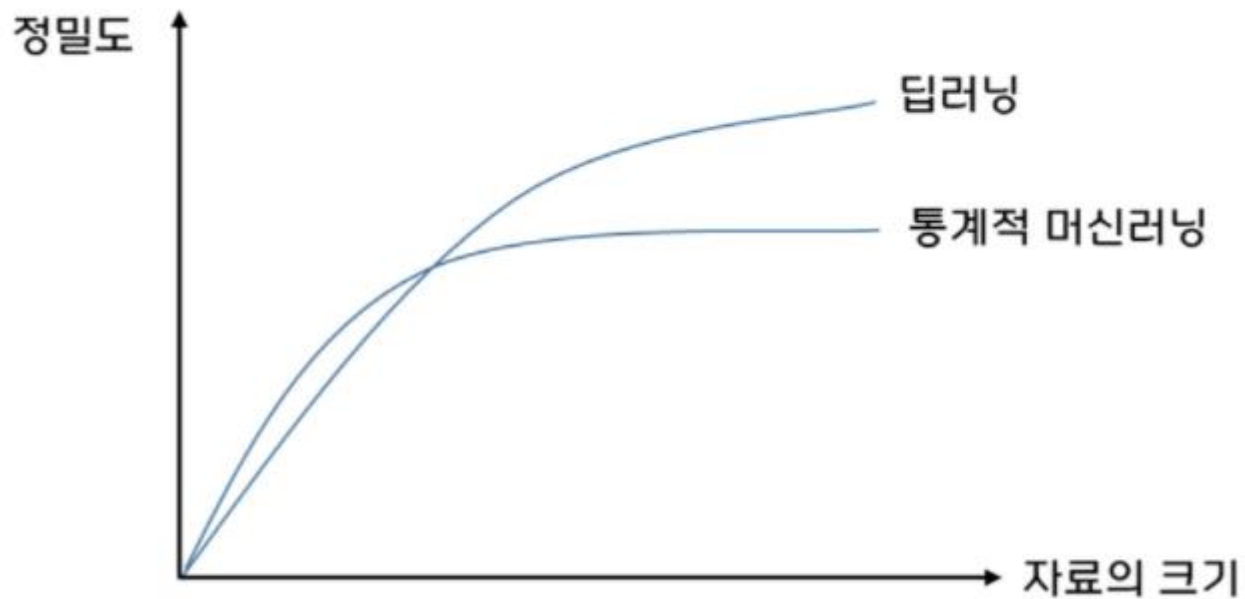
# What is Statistical ML?

---



# Machine Learning VS Deep Learning

---



# Machine Learning VS Deep Learning

구분	통계적 머신러닝	딥러닝
데이터 크기	중/소 크기	빅데이터
분석자료 형태	2차원 텐서	2차원 텐서이상
강점을 갖는 자료	정형화된 자료	비정형자료
특성변수	특성변수를 만들어야 함	특성변수가 만들어짐
특성변수의 정규화 및 표준화	선택	필요
모형	매우 많음	기본적으로 3 개의 모형
최적화	일반적으로 전체 데이터 사용	배치데이터
해석여부	해석이 쉬움 (단, SVM과 boosting 제외)	어렵거나 불가능
하드웨어	중급	고성능(GPU 요구)
실행요구시간	최대 시간 단위	최대 주단위 시간



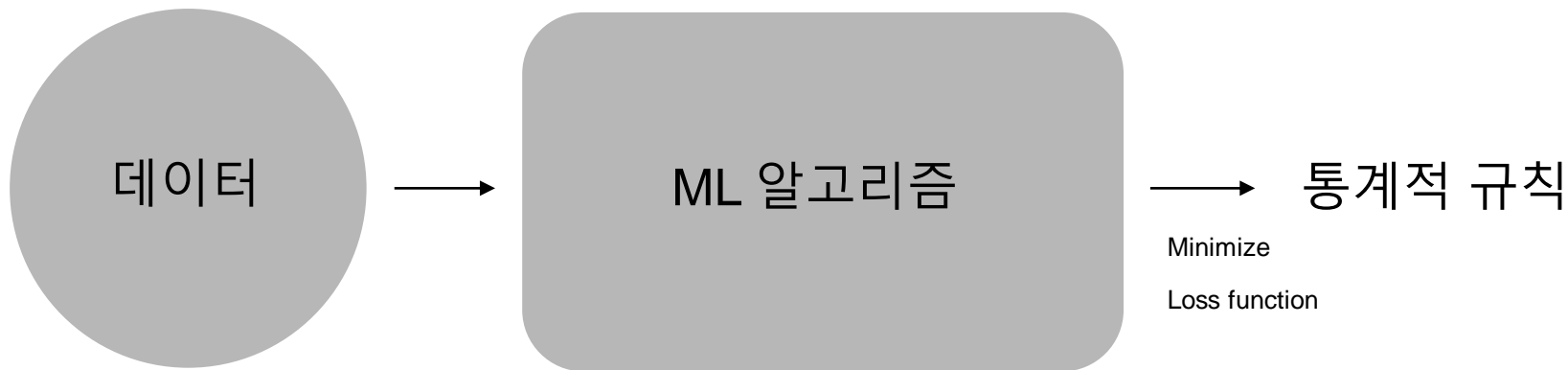
# ML type

---

- 지도학습 VS 비지도학습 VS (강화학습)
- 배치학습 VS 온라인학습
- 사례기반 VS 모형기반

# What is ML?

---



# MLE and Loss function

---

- What is MLE?

Definition (Maximum likelihood estimator, MLE)

For  $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x; \theta)$ , the MLE of  $\theta$  is

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} L(\theta; \mathbf{x}).$$

which is equivalent to maximize the logarithm of  $L(\theta; \mathbf{x})$  which we call the log-likelihood

$$\ell(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}).$$

# Likelihood function and Optimization

---

- What is Likelihood function?

## Definition (Likelihood)

For  $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x; \theta)$ , where  $\theta$  denotes a parameter of interest. The **likelihood function** is

$$L(\theta; \mathbf{X}) = L(\theta; X_1, \dots, X_n) = \prod_{i=1}^n f_X(X_i; \theta)$$

# Likelihood functions of different probability distributions

---

- Bernoulli distribution

$$\log L(p) = \sum_{i=1}^n (y_i \log p + (1 - y_i) \log (1 - p))$$

# Likelihood functions of different probability distributions

---

- Binomial distribution

$$\log \binom{n}{c} + \sum_{i=1}^n (y_i \log p + (1 - y_i) \log (1 - p))$$

# Likelihood functions of different probability distributions

---

- Multinomial distribution

$$\sum_{i=1}^n \sum_{j=1}^c y_{ij} \log p_j$$

# Likelihood functions of different probability distributions

---

- Normal distribution

$$\log L(\mu) \approx - \frac{\sum_{i=1}^n (y_i - \mu)}{\sigma^2}$$



# Likelihood function and Loss function

---

- Binary Cross Entropy
- Categorical Cross Entropy
- MSE