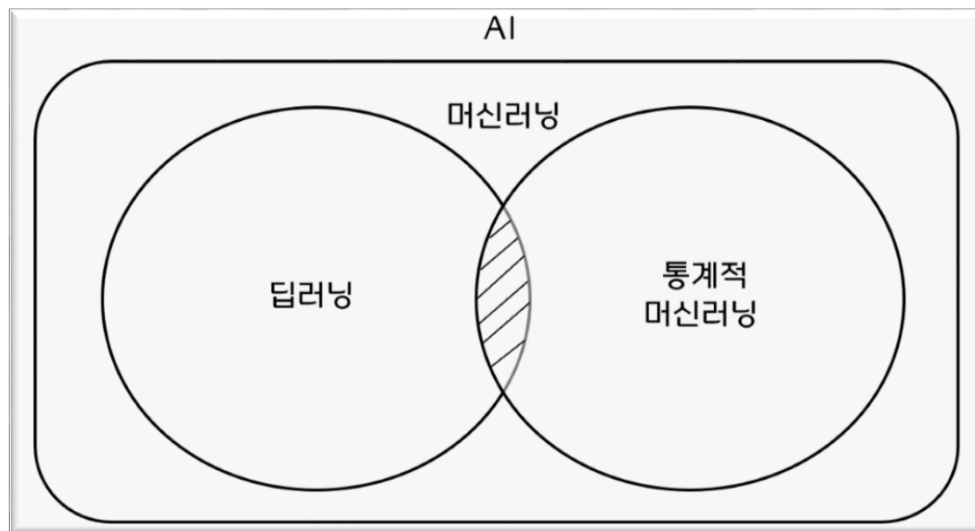


Statistical Machine Learning

1주차

담당: 11기 명재성

Statistical Machine Learning



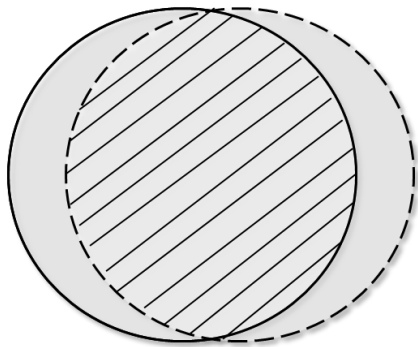
Statistical Machine Learning

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

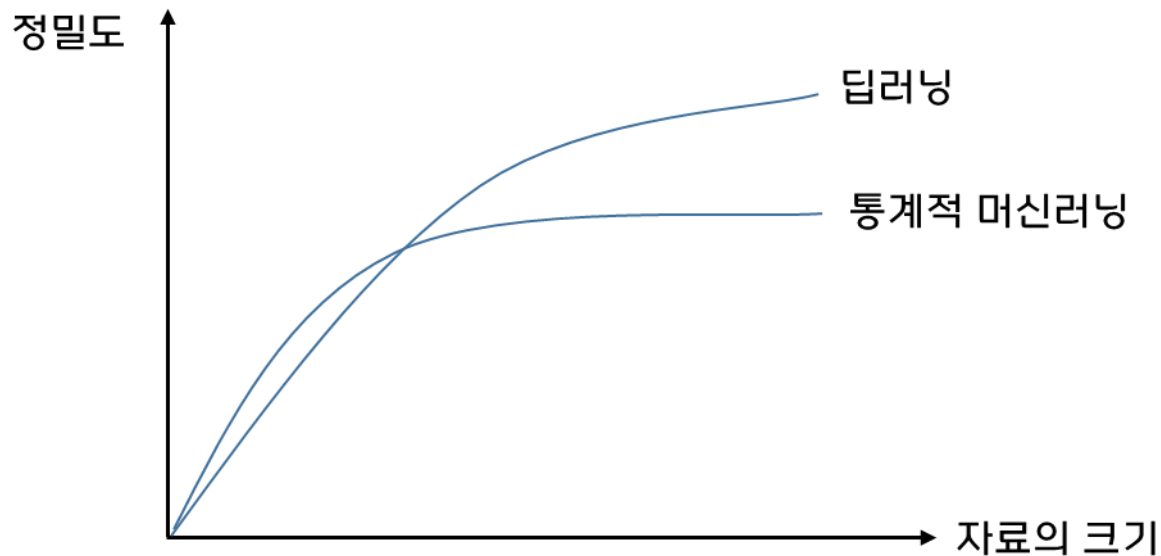
--- 통계적 머신러닝

BigData

- Volume
- Velocity
- Variety

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

Statistical Machine Learning and Deep Learning



Statistical Machine Learning Type

- 지도학습(supervised learning)
비지도학습(unsupervised learning)
강화학습(Reinforcement learning)
- 배치학습(Batch learning)
온라인학습(Online learning)
- 사례기반(Instance-based learning)
모형기반(Model-based learning)

Linear Regression

- Linearity?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \epsilon_i$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \epsilon_i$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_p X_i^p + \epsilon_i$$

Linear Model

- Linearity? \longrightarrow Linear Model

$$Y_i \stackrel{iid}{\sim} (\mu(\mathbf{X}), \sigma) \quad \text{where} \quad E[Y] = \mu(\mathbf{X})$$

$$\begin{aligned} \mu(\mathbf{X}) &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \\ &= \mathbf{X} \boldsymbol{\beta} \end{aligned}$$

Regression

- Least Square Estimator

$$\sum \epsilon_i^2 = \sum (Y_i - \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})^2$$

$$\frac{\partial}{\partial \beta_0} \sum (Y_i - \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})^2 \stackrel{set}{=} 0$$

$$\frac{\partial}{\partial \beta_1} \sum (Y_i - \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})^2 \stackrel{set}{=} 0$$

⋮

$$\frac{\partial}{\partial \beta_p} \sum (Y_i - \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi})^2 \stackrel{set}{=} 0$$

Regression

- Least Square Estimator

```
> summary(model.a<-lm(exp~income+ factor(Region)))

Call:
lm(formula = exp ~ income + factor(Region))

Residuals:
    Min       1Q   Median       3Q      Max
-77.624 -26.431  -8.821  19.391 174.548

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   21.94531    60.05982   0.365   0.7165
income         0.05337     0.01169   4.566 3.84e-05 ***
factor(Region)2  1.21498    20.02606   0.061   0.9519
factor(Region)3 -0.44452    20.91222  -0.021   0.9831
factor(Region)4 49.92487    19.78310   2.524   0.0152 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
```

Regression

- Risk Function

$$R(\theta, T(X)) = E[L(\tau(\theta), T(X))]$$

- Loss Function

$$\begin{aligned} L[\tau(\theta), T(X)] &= \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum |Y_i - \hat{Y}_i| \end{aligned}$$

Regression

- Error term?
 - Mean 0
 - Identical, Independent
 - Normal?

Regression

- Likelihood function

Definition (Likelihood)

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

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

Regression

- Maximum Likelihood Estimator

Definition (Maximum likelihood estimator, MLE)

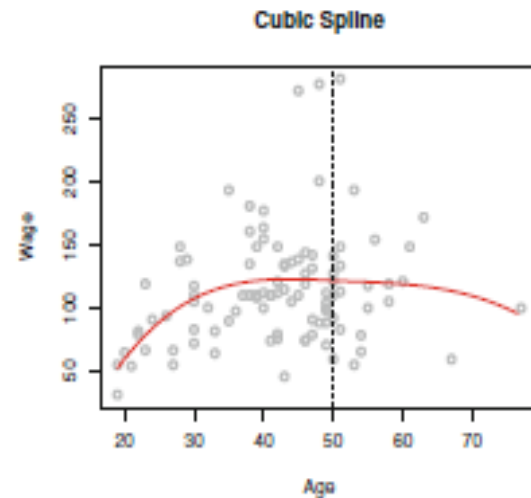
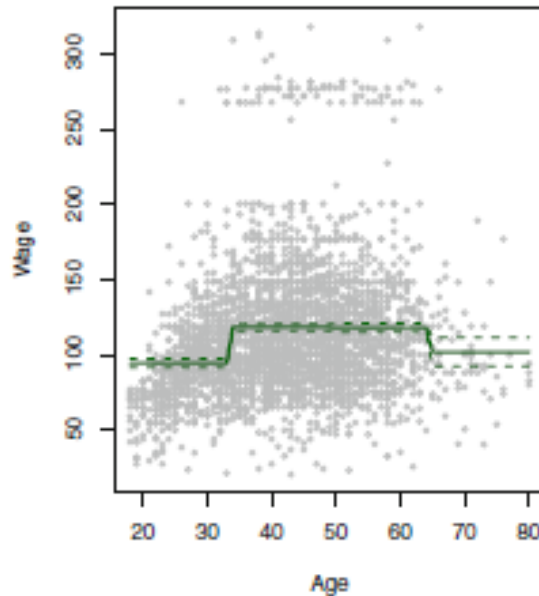
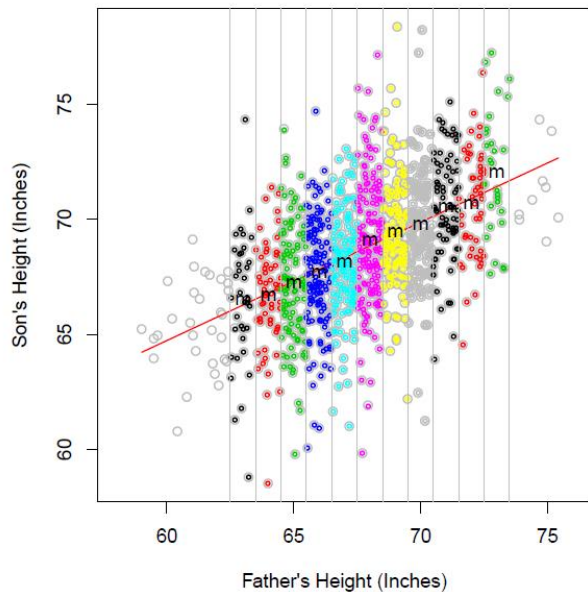
For $X_1, \dots, X_n \stackrel{iid}{\sim} f_X(x; \theta)$, the MLE of θ is

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

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

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

Other Regression...?



Logistic Regression

$$\log\left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}$$

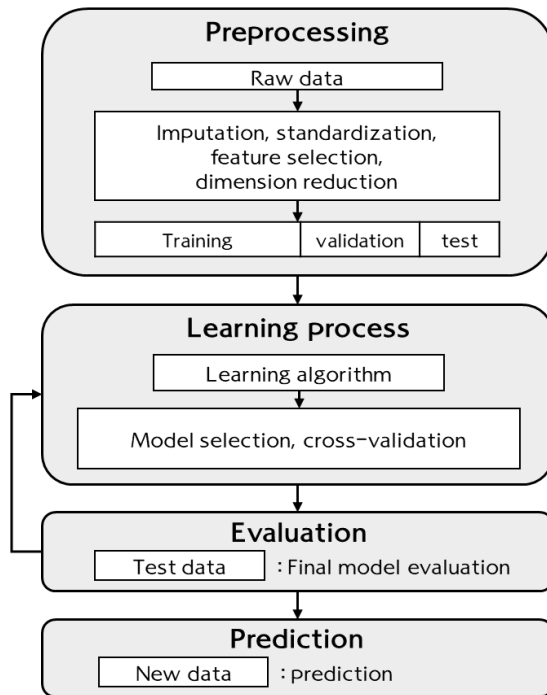
$$\begin{aligned} P(Y = 1|\mathbf{X}) = \pi(\mathbf{X}) &= \frac{e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}}} \\ &= \frac{e^{\boldsymbol{\beta}^T \mathbf{x}}}{1 + e^{\boldsymbol{\beta}^T \mathbf{x}}} \end{aligned}$$

Generalized Linear Model

	Normal	Poisson	Binomial	Gamma	Inv Gaussian
Notation	$N(\mu, \sigma^2)$	$P(\mu)$	$B(n, \pi)/n$	$G(\mu, v)$	$IG(\mu, \sigma^2)$
Support	$(-\infty, \infty)$	$\{0, 1, \dots\}$	$\{0, \dots, n\}/n$	$(0, \infty)$	$(0, \infty)$
$a(\phi)$	$\phi = \sigma^2$	1	$1/m$	v^{-1}	σ^2
$b(\theta)$	$\theta^2/2$	e^θ	$\log(1 + e^\theta)$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$
$b'(\theta) = E(Y)$	θ	e^θ	$\frac{e^\theta}{1+e^\theta}$	$-1/\theta$	$(-2\theta)^{-1/2}$
$(b')^{-1}(\mu) = g(\mu)$	μ	$\log(\mu)$	$\log \frac{\mu}{1-\mu}$	μ^{-1}	μ^{-2}
$b''(\theta)$	1	μ	$\mu(1 - \mu)$	μ^2	μ^3

Table: Summary of some popular GLM models.

Summary



reference

자료

19-2 STAT424 통계적 머신러닝 - 박유성 교수님

교재

파이썬을 이용한 통계적 머신러닝 (2020) - 박유성

ISLR (2013) - G. James, D. Witten, T. Hastie, R. Tibshirani

The elements of Statistical Learning (2001) - J. Friedman, T. Hastie, R. Tibshirani

Hands on Machine Learning (2017) - Aurelien Geron