

PAC.



21.03.31.,

# • PAC learning (Probably-Approximately-Correct Learning)

## - Contents

### - Stat-Learning-Framework

$X$ : domain set

$Y$ : Label set

$S = \{(x_1, y_1), \dots, (x_n, y_n)\}$  Training Data

$(x_i, y_i) \sim D$

$x, y$

### - Learner

- Hypothesis, prediction rule  $h: X \rightarrow Y$

## Data Generation Model

$D$ : distribution over  $X$ , unknown

$f$ : "correct" labeling function  $y_i := f(x_i)$

## Prediction error:

$$L_{D,f}(h) := P_{X \sim D}(h(x) \neq f(x)) := D(\{x : h(x) \neq f(x)\})$$

## Empirical risk

$$L_S(h) = \frac{|\{i \in [n] : h(x_i) \neq f(x_i)\}|}{n}$$

$[n] = \{1, 2, \dots, n\}$

## Generalization error

comes from  $L_{D,f}(h) - L_S(h)$

## Empirical risk minimization

- $\mathcal{H}$ : hypothesis class
- each  $h \in \mathcal{H} : X \rightarrow Y$

$$\text{ERM}_{\mathcal{H}}(S) \in \arg \min_{h \in \mathcal{H}} L_S(h)$$

## • Realizability Assumption

$h^* \in \mathcal{H}$ , such that  $L_{D,f}(h^*) = 0$

$\Rightarrow L_S(h^*) = 0$  with probability 1

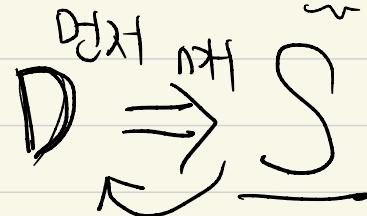
## • Sample Complexity $\epsilon, \delta$

$(\epsilon, \delta)$ -sample complexity is the smallest number  $n$  w.p  $(1-\delta)$   $L_{D,f}(h) \leq \epsilon$ .

$\begin{cases} \epsilon : \text{error} \\ \delta : \text{uncertainty} \end{cases}$

## • I.I.D

Identical, Independent,



$S \sim D^n$ ,  $x_i \in S$ ,  $x_i$  i.i.d sampled

## • PAC Learnability (Pr - A - C - Learning)

$\mathcal{H}$ : Hypothesis

—  $\mathcal{H}$  is PAC-learnable if —

$\forall \epsilon, \delta \in (0,1), \exists n \geq n_{\mathcal{H}}(\epsilon, \delta)$

$L_{D,f}(h) \leq \epsilon$  with probability  $1 - \delta$

$n_{\mathcal{H}}(\epsilon, \delta)$  exists

- Finite Hypothesis

Thm. PAC-Guarantee

$$|\mathcal{H}| = n \quad n \geq \frac{1}{\varepsilon} (\log |\mathcal{H}| + \log \frac{1}{\delta}) \quad \checkmark$$

With prob. greater than or equal to  $1 - \delta$ ,

every  $h \in \mathcal{H}$  with zero training error has ✓

test error less than  $1 - \varepsilon$ .

$$\left( \begin{array}{l} L_S(H) = 0 \text{ 인 한 경우 생각} \\ L_{Df}(H) \geq \varepsilon \end{array} \right)$$

$$D \rightarrow S \xrightarrow{n \in H} P(h(x_i) \neq f(x_i))$$

$n \in H$  sample  $x_1 \sim x_n$   
 $\vdash \varepsilon \quad (1 - \varepsilon)$

$A_i :=$  "Bad Event",  $H_i$ .

$$P(A_i) = \prod_{j=1}^n P(h(x_j) \neq f(x_j)) \leq (1 - \varepsilon)^n.$$

$$P(A) \leq \sum_{i=1}^{|\mathcal{H}|} P(A_i) \leq |\mathcal{H}| (1 - \varepsilon)^n \leq \delta$$

$$|\mathcal{H}| (1 - \varepsilon)^n \leq |\mathcal{H}| e^{-n\varepsilon} \leq \delta$$

$$e^{-n\varepsilon} \leq \frac{\delta}{|\mathcal{H}|}$$

✓

Rmk

① Zero Training Error? ✓

Assumption

②  $|H| \rightarrow \infty$  ?

PAC-Guarantee

On:  $\frac{m}{|H|}$  parameter,  $\underline{R}$ .

$\underline{R}$  hypothesis

$|H| \rightarrow \infty$

~~$|H| \rightarrow \infty$~~

$|H| < \infty$  ✓

Thm. Uniform Convergence ✓

$$n \geq \frac{1}{2\varepsilon^2} (\log |\mathcal{H}| + \log \frac{2}{\delta})$$

With Probability greater than  $1 - \delta$ , every  $h \in \mathcal{H}$  satisfies

$$|\mathcal{L}_S(h) - \mathcal{L}_D(h)| \leq \varepsilon.$$

$S = \{x_1, \dots, x_n\}$  Fix  $H_i \in \mathcal{H}$ .  $x_j = H_i$ 가  $x_j$ 를 같은 높이에

$X_j \sim \text{Bernoulli}(p = \mathcal{L}_D(H_i))$

$$\frac{1}{n} \sum_{j=1}^n X_j = \text{err}_S(H)$$

$$P(|\text{err}_S(H) - \text{err}_D(H)| > \varepsilon)$$

$$= P\left(\left|\frac{1}{n} \sum_{j=1}^n X_j - p\right| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}.$$

$$\therefore P(VA_S) \leq 2|\mathcal{H}| e^{-2n\varepsilon^2} \leq \delta.$$

$$|\mathcal{H}| < \infty$$

$$|\mathcal{H}| = \infty \times$$

VC-dimension

Radamacher-complexity ✓

