PAC.

21.03.31.

- **PAC learning** (Probably-Approximately-Correct Learning)

- Contents

- Stat-Learning-Framework

$\mathcal{X}$: domain set

$\mathcal{Y}$: Label set

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

- Learner

  - Hypothesis, prediction rule $h: \mathcal{X} \to \mathcal{Y}$

$(x_i, y_i) \sim \mathcal{D}$

$\mathcal{X}, \tilde{\mathcal{Y}}$.

Data Generation model

$\mathcal{D}$: distribution over $\mathcal{X}$, unknown

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

Prediction error

$$L_{\mathcal{D}, f}(h) := \mathbb{P}_{x \sim \mathcal{D}}(h(x) \neq f(x)) := \mathcal{D}(\{x : h(x) \neq f(x)\})$$

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 $L_{\mathcal{D}, f}(h) - L_S(h)$

Empirical risk minimization

- $\mathcal{H}$: hypothesis class
- each $h \in \mathcal{H}$ $\mathcal{X} \to \mathcal{Y}$
- $ERM_{\mathcal{H}}(S) \in \arg\min_{h \in \mathcal{H}} L_S(h)$

- Realizability Assumption

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

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

- Sample Complexity ε, δ

  (ε, δ) - sample complexity is the smallest number n

  w.p $(1-\delta)$ $L_{\mathcal{D},f}(h) \leq \varepsilon$.

  (ε : error.
    δ : uncertainty.

- I.I.D

  Identical, Independent, $D \overset{먼저\ 넣저}{\Longrightarrow} S$

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

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

  $\mathcal{H}$: Hypothesis

— $\mathcal{H}$ is PAC-learnable if —

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

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

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

– Finite Hypothesis

Thm. PAC-Guarantee

$|S| = n$. $\quad n \geq \frac{1}{\varepsilon}\left(\log|\mathcal{H}| + \log\frac{1}{\delta}\right)$ ✓✓

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

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

test error less than $1-\delta$.

$$\begin{pmatrix} L_S(H) = 0 \text{ 인 한 경우를 생각} \\ L_{D,f}(H) \geq \varepsilon \end{pmatrix}$$

$$L_{D,f}(H) = \mathbb{P}(h(x_i) \neq f(x_i))$$

n개

$D \to S$ $\quad h_i$ ✓ $\quad$ n개 sample $x_1 \sim x_n$

$1-\varepsilon \quad (1-\varepsilon)$

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

$$\mathbb{P}(A_i) = \prod_{i=1}^{n} \mathbb{P}(h(x) = f(x)) \leq (1-\varepsilon)^n.$$

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

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

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

Rmk

① Zero Training Error ? Assumption

PAC-Guarantee

② $|\mathcal{H}| \to \infty$?

예: m개 parameter, $\mathbb{R}$.

$|\mathcal{H}|$ " $\mathbb{R}$ hypothesis

"

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

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

Thm. Uniform Convergence

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

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

$$|L_S(h) - L_D(h)| \leq \varepsilon.$$

$S = \{x_1, \cdots, x_n\}$ Fix $H_i \in \mathcal{H}$. $X_j = H_i$가 x_j를 잘못 분류

$X_j \sim \text{Bernoulli}(p = L_D(H_i))$.

$$\frac{1}{n}\sum_{j=1}^{n} X_j = err_S(H)$$

$$\mathbb{P}(|err_S(H) - err_D(H)| > \varepsilon)$$

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

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

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

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

VC-dimension

Radamacher-complexity