

Validation \rightarrow test set is fixed

problem by h , $h: X \rightarrow Y$ \rightarrow $c_t(x_i)$ has shape now, $c_t(x_i)$ \in C_t

? h for m \rightarrow Y \in C_t

validation set \in X^m : m \rightarrow C_t

\Rightarrow m \rightarrow C_t \rightarrow m \rightarrow C_t \in Y \in C_t

$$\{x_1, \dots, x_m\} \in X^m \quad m = 1000$$

$X \sim D$ \rightarrow iid x_1, \dots, x_m \in $P_{X|D}$ \rightarrow C_t

$$\text{err}_v(h) = \frac{1}{m} \sum_{i=1}^m [\mathbb{I}[c_t(x_i) \neq h(x_i)]] \quad \rightarrow$$

misclassified sample \rightarrow h \in C_t

: \rightarrow m \rightarrow C_t \rightarrow m \rightarrow C_t

Hoeffding Bound

$D(\hat{f}(x)) \sim D$ \rightarrow iid \rightarrow $\{x_1, \dots, x_m\} \in X^m$ \rightarrow C_t

\rightarrow $\epsilon > 0$ \rightarrow $f: X \rightarrow [0, 1]$

$$P_r \left[\left| \frac{1}{m} \sum_{i=1}^m f(x_i) - \mathbb{E}_{x \sim D}[f(x)] \right| > \epsilon \right] \leq 2e^{-2m\epsilon^2}$$

: \rightarrow m \rightarrow C_t

$$\rightarrow f(x) = [\mathbb{I}[h(x) \neq c_t(x)]] \quad \rightarrow$$

$$P_r \left[|\text{err}_v(h) - \text{err}(h)| > \epsilon \right] \leq 2e^{-2m\epsilon^2} \quad \downarrow \quad \rightarrow$$

$$\epsilon = \sqrt{\frac{\ln(\frac{2}{\delta})}{2m}} = 0.0515 \quad \rightarrow \quad \delta = 0.01 \quad \rightarrow \quad m = 1000$$