

Discussion 7

No free lunch theory

田鹏超

tianpch@shanghaitech.edu.cn

PAC/SLT models for Supervised Learning

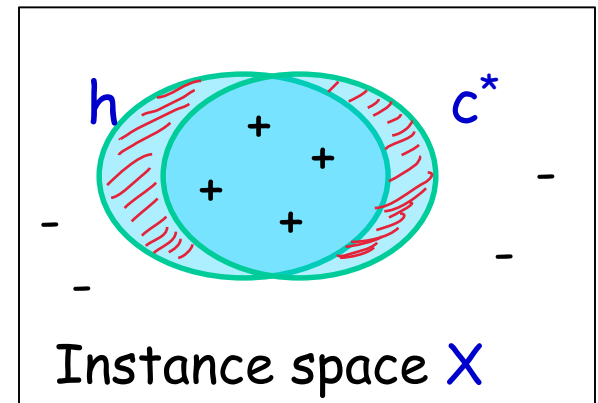
- X - feature or instance space; distribution D over X
e.g., $X = \mathbb{R}^d$ or $X = \{0,1\}^d$
- Algo sees training sample $S: (x_1, c^*(x_1)), \dots, (x_m, c^*(x_m))$, x_i i.i.d. from D
 - **labeled** examples - assumed to be drawn i.i.d. from some distr. D over X and labeled by some target concept c^*
 - labels $\in \{-1,1\}$ - **binary** classification
- Algo does **optimization over S** , find hypothesis h .
- Goal: h has small error over D .

$$c^*: X \rightarrow \{-1,1\}$$

$$err_D(h) = \Pr_{x \sim D}(h(x) \neq c^*(x))$$

$$L(h(x), c^*(x)) = \begin{cases} 1, & h(x) \neq c^*(x) \\ 0, & h(x) = c^*(x) \end{cases}$$

Need a bias: **no free lunch.**



$$c^* \in H$$



NFL (no free lunch) Theorem:

(无万金油算法定理)

- $P(h | X, \mathcal{A})$: the probability of finding hypothesis h when applying learning algorithm \mathcal{A} on training set $X \subset \mathcal{X}$
- f is the target function, $X \subset \mathcal{X}$ is noise-free
- Consider the 0-1 error
- Purpose: compare the performance of different \mathcal{A}

$err_D(h|\mathcal{A})$

The out-of-sample error of \mathcal{A} is given by:

$$E(\mathcal{A} | X, f) = \sum_h \sum_{\underline{x \in \mathcal{X} - X}} \underbrace{P(x)}_{\text{red}} \underbrace{\delta(h(x) \neq f(x))}_{\text{red}} \underbrace{P(h | X, \mathcal{A})}_{\text{red}}$$

-
1. D. H. Wolpert, "The Lack of A Priori Distinctions Between Learning Algorithms," in *Neural Computation*, vol. 8, no. 7, pp. 1341-1390, Oct. 1996.
 2. David H. Wolpert and William G. Macready, "No Free Lunch Theorems for Optimization", *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, VOL. 1, NO. 1, APRIL 1997
 3. <https://www.kdnuggets.com/2019/09/no-free-lunch-data-science.html>

无万金油算法定理

$$E(\mathcal{A} | X, f) = \sum_h \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) P(h | X, \mathcal{A})$$

If you apply \mathcal{A} to **any** problem (any f), the overall out-of-sample error of \mathcal{A} is given by:

$$\begin{aligned} \sum_f E(\mathcal{A} | X, f) &= \sum_f \sum_h \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) P(h | X, \mathcal{A}) \\ &= \sum_h \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) P(h | X, \mathcal{A}) \sum_f \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) \\ &= \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \sum_h P(h | X, \mathcal{A}) \sum_f \mathbb{I}(h(\mathbf{x}) \neq f(\mathbf{x})) \\ &= \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \sum_h P(h | X, \mathcal{A}) \frac{1}{2} 2^{|\mathcal{X}|} \\ &= \frac{1}{2} 2^{|\mathcal{X}|} \sum_{\mathbf{x} \in \mathcal{X} - X} P(\mathbf{x}) \cdot 1 = 0.5^n 2^{|\mathcal{X}|} \end{aligned}$$