No-free lunch theorem

PAC-learning the class of all functions over an infinite domain is impossible.

<u>Theorem</u> Let X be an infinite domain. Let A be any learning algorithm. Then for any $m \geq 0$ there exists a distribution D and target function

$h: X \to \{0,1\}$ such that

if $S = ((x_1, h(x_1)), \dots, (x_m, h(x_m)))$ where $x_1, x_2 \dots x_m \sim_{i.i.d} D$ then

$$\Pr\left[err_{h,D}(A(S)) > 1/8\right] \geq 1/7.$$

Proof:

Let $C \subseteq X$ be of size $2m$.

Pick a distribution D to be uniform distribution over C.

There are 2^{2m} functions from C to $\{0,1\}$.

Let's call them

$h_1, h_2, \dots, h_{2^{2m}}$.

Let $S_i = ((x_1, h_i(x_1)), \dots, (x_m, h_i(x_m)))$

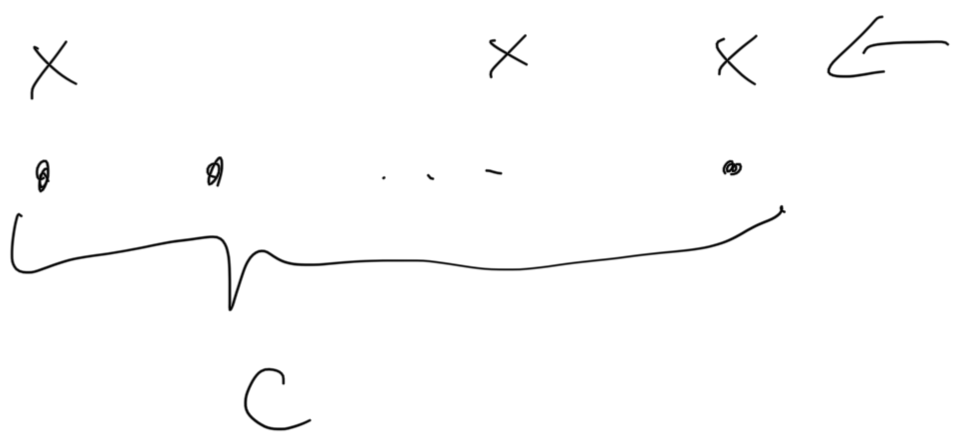
We will show that

$$\max E\left[\mathrm{err}_{h_i,D}(A(S_i))\right] \geq 1/4 \quad (*)$$

$$\max_i \mathbb{E}\left[\mathrm{err}_{h_i,D}(A(S_i))\right]$$

$$\geq \frac{1}{2^{2m}} \sum_{i=1}^{2^m} \mathbb{E}\left[\mathrm{err}_{h_i,D}(A(S_i))\right]$$

$$= \frac{1}{2^{2m}} \sum_{i=1}^{2^m} \mathbb{E}\left[\mathbb{E}\left[\mathrm{err}_{h_i,D}(A(S_i)) \,\middle|\, x_1 \dots x_m\right]\right]$$



$$= \mathbb{E}\left[\frac{1}{2^{2m}} \sum_{i=1}^{2^m} \frac{1}{2m} \sum_{x \in C} \mathbb{1}\left[A(S_i)(x) \neq h_i(x)\right] \,\middle|\, x_1 \ldots x_m \right]$$

$$= \mathbb{E}\left[\frac{1}{2m} \sum_{x \in C} \underbrace{\frac{1}{2^m} \sum_{i=1}^{2^m} \mathbb{1}\left[A(S_i)(x) \neq h_i(x)\right]} \,\middle|\, x_1 \ldots x_m \right]$$

if $x \in \{x_1 \ldots x_m\}$

for every h_i there is h_k such that

$S_i = S_k$ but $h_k(x) \neq h_i(x)$.

So $A(S_i)(x) = A(S_k)(x)$

but

$A(S_i)(x) = h_i(x)$

$A(S_k)(x) \neq h_k(x)$

$$= \mathbb{E}\left[\frac{1}{2m} \sum_{\substack{x \in C \\ x \notin \{x_1, \dots x_m\}}} \frac{1}{2} \;\middle|\; x_1, \dots x_m\right]$$

$$\geq \mathbb{E}\left[\frac{1}{2m} \cdot m \cdot \frac{1}{2} \;\middle|\; x_1, \dots x_m\right]$$

$$= \frac{1}{4}$$

So there exists $\underline{h}$ and $\underline{D}$ such that

$$\mathbb{E}\left[\mathrm{err}_{h,D}(A(S))\right] \geq 1/4 \text{ .}$$

Let $Z = 1 - \mathrm{err}_{h,D}(A(S))$.

We know

- Z is non-negative
- $\mathbb{E}[Z] \leq 3/4$

$$\Pr\left[Z \geq \frac{7}{8}\right] \leq \frac{8}{7} \cdot \mathbb{E}[Z]$$

$$= \frac{8}{7} \cdot \frac{3}{4} = \frac{24}{28}$$

$$= \frac{6}{7}$$

$$\Pr\left[Z < \frac{7}{8} \right] \geq \frac{1}{7}$$

$$\Pr\left[1 - \mathrm{err}_{D,h}(A(S)) < \frac{7}{8} \right] \geq \frac{1}{7}$$

$$\Pr\left[\mathrm{err}_{D,h}(A(S)) > \frac{1}{8} \right] \geq 1/7 .$$

$\blacksquare$