Homework #4

Due: 2024-12-8 23:59 | 7 Problems, 100 Pts

Name: XXX, ID: XXX

Problem 1 (10'). Find out and prove the VC-dimension of the hypothesis class $\mathcal{H}$ on instance space $\mathbb{R}^2$ where

$$\mathcal{H} = \{\{\boldsymbol{x} = (x_1, x_2) \mid x_1 \geq c_1, x_2 \geq c_2\} \mid \boldsymbol{c} = (c_1, c_2)\}.$$

◀

Problem 2 (10'). Find out and prove the VC-dimension of the hypothesis class $\mathcal{H}_n$ on instance space $\mathbb{R}$ where

$$\mathcal{H}_n = \left\{\left\{x | c_0 + c_1 x + c_2 x^2 + ... + c_n x^n > 0\right\} \mid c_0, c_1, ..., c_n \in \mathbb{R}\right\}.$$

Express the answer as a function of n. ◀

Problem 3 (16'). Find out and prove the VC-dimension of the hypothesis class $\mathcal{H}_n$ on instance space $\mathbb{R}^2$ where

$$\mathcal{H}_n = \{\{\boldsymbol{x} = (x_1, x_2) \mid \forall i \in [n],\ a_i x_1 + b_i x_2 + c_i \geq 0\} \mid a_1, \cdots, a_n, b_1, \cdots, b_n, c_1, \cdots, c_n \in \mathbb{R}\}.$$

Express the answer as a function of n. ◀

Problem 4 (16'). Find out and prove the VC-dimension of the hypothesis class $\mathcal{H}_n$ on instance space $\mathbb{R}^n$ $(n \geq 2)$ where

$$\mathcal{H}_n = \{\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{c}\|_2 \leq r\} \mid \boldsymbol{c} \in \mathbb{R}^n, r \geq 0\}.$$

Express the answer as a function of n. ◀

Problem 5 (16'). Find out and prove the VC-dimension of the hypothesis class $\mathcal{H}_n$ on instance space $\{0,1\}^n$ $(n \geq 1)$ where

$$\mathcal{H}_n = \{\{\boldsymbol{x} \in \{0,1\}^n \mid f_S(\boldsymbol{x}) = -1\} \mid S \subseteq \{1, 2, \cdots, n\}\}.$$

Here, $f_S(\boldsymbol{x}) : \{0,1\}^n \to \{-1, +1\}$ is defined as

$$f_S(\boldsymbol{x}) := \begin{cases} -1, & S = \varnothing; \\ (-1)^{\prod_{j \in S} x_j}, & S \neq \varnothing. \end{cases}$$

Express the answer as a function of n. ◀

Problem 6 (14'). The shatter function $\pi_{\mathcal{H}}(n)$ is the maximum number of subsets of any set A of size n that can be expressed as $A \cap h$ for $h \in \mathcal{H}$. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two hypothesis classes and $\mathcal{H} = \{h_1 \cap h_2 \mid h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2\}$. Recall that we have proved $\pi_{\mathcal{H}}(n) \leq \pi_{\mathcal{H}_1}(n)\pi_{\mathcal{H}_2}(n)$ in class.

(1) (6') Recall the Sauer's lemma we have learned in class. Sauer's lemma tells that for a hypothesis class $\mathcal{H}$ with VC-dimension d, $\pi_{\mathcal{H}}(m) \le \sum_{i=0}^{d} \binom{m}{i}$. Prove that $\sum_{i=0}^{d} \binom{m}{i} \le \left(\frac{\mathrm{e}m}{d}\right)^d$ when $m \ge d$.

(2) (8') For a hypothesis class $\mathcal{H}$ with VC-dimension d, define the hypothesis class $\mathcal{H}^k$ $(k \ge 2)$ as

$$\mathcal{H}^k = \left\{ \bigcap_{i=1}^{k} h_i \mid h_i \in \mathcal{H} \right\}.$$

Prove that, the VC dimension of $\mathcal{H}^k$ is no more than $7dk \ln k$. You may use the assertions above. $(\ln 2 \approx 0.693,\ \mathrm{e} \approx 2.718,\ \ln 7 \approx 1.946,\ \ln\ln 2 \approx -0.367)$

◀

Problem 7 (18'). Recall online learning and the Halving Algorithm we have introduced in class.

Problem setting: There are N experts. Suppose that we have access to the predictions of N experts. At each time $t = 1, 2, \cdots, T$, we observe the experts' predictions $f_{1,t}, f_{2,t}, \cdots, f_{N,t} \in \{0,1\}$ and predict $p_t \in \{0,1\}$. We then observe the outcome $y_t \in \{0,1\}$ and suffer loss $\mathbf{1}_{p_t \neq y_t}$. Suppose $\exists j$ such that $f_{j,t} = y_t$ for all t.

Halving Algorithm: Every time, we eliminate experts who make mistakes. That is, initially $C_1 = [N]$ and $C_t = C_{t-1} \cap \{i | f_{i,t-1} = y_{t-1}\}$. Let r_t be the fraction of experts in C_t predicting 1. We predict p_t as $\mathbf{1}_{r_t \ge 1/2}$.

In class we showed that the number of mistakes made by Halving algorithm is upper bounded by $\log_2 N$. Here, we consider a randomized version of Halving Algorithm.

Randomized Halving Algorithm: Define $C_1 = [N]$ and $C_t = C_{t-1} \cap \{i | f_{i,t-1} = y_{t-1}\}$. Let r_t be the fraction of experts in C_t predicting 1. We predict $p_t = 1$ with probability

$$\min\left\{1, \frac{1}{2}\log_2 \frac{1}{1-r_t}\right\},$$

and $p_t = 0$ otherwise.

Prove that, the expected number of mistakes made by Randomized Halving Algorithm is at most $\frac{1}{2}\log_2 N$.

[Hint: Consider potential function $\Phi_t = \log_2(|C_t|)$.] ◀