Homework #4

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

Name: 徐靖, ID: 2200012917

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

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

◀

Answer. The VC dimension of $\mathcal{H}$ is 2.

Shattering 2 points : Let $x_1 = (0,1), x_2 = (1,0)$, then elements of set $\{\{\boldsymbol{x} = (x_1, x_2) \mid x_1 \geq c_1, x_2 \geq c_2\} \mid \boldsymbol{c} = (c_1, c_2)$ realize these labels.

Attempting to shatter 3 points : $\forall (x_i, y_i), i \in [3]$, consider labels $(0,1,1)$, we have

$$x_2, x_3 \geq c_1, y_2, y_3 \geq c_2, c_1 > x_1 \text{ or } c_2 > y_1$$

which means (x_1, y_1) is strictly smaller than the other two points in at least one coordinate. The same is true for points (x_2, y_2) and (x_3, y_3). This contradicts because there are only 2 dimensions.

◁

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. ◀

Answer. The VC dimension of $\mathcal{H}$ is $n+1$.

Shattering $n+1$ points : $\forall \{x_i\}_{n+1} \subset \mathbb{R}$ s.t. $x_{i+1} > x_i$, we can find $\{c_i\}_n$ s.t. $\forall i, x_i < c_i < x_{i+1}$. For label $l = (l_1, l_2, \ldots, l_{n+1})$, we construct a polynomial of degree at most n: $\epsilon \sum_{i \in [n]} (x - c_i)^{\epsilon_i} \in \mathcal{H}$, where

$$\epsilon = 2l_1 - 1, \epsilon_i = \begin{cases} 1, & l_i \neq l_{i+1}, \\ 0, & l_i = l_{i+1} \end{cases}$$

Using induction, we find that the inequality at x_i is consistent with whether l_i is 1. That means all labels are implemented by $\mathcal{H}$

Attempting to shatter $n+2$ points : $\forall \{x_i\}_{n+2} \subset \mathbb{R}$ s.t. $x_{i+1} > x_i$. If the labels are staggered, then according to the Intermediate Value Theorem, there must be a zero of the polynomial between every two points. A total of $n+1$ different zeros means that the polynomial that implements the label must be at least $n+1$, which is a contradiction.

◁

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. ◀

Answer. The VC dimension of $\mathcal{H}$ is $2n + 1$.

For a finite number of points on a plane, we consider its convex hull. If there is a point A in the convex hull, we select a label l that is only 0 at this point and 1 at other points. It is easy to find that label I is not realized by any $h \in \mathcal{H}_n$, because if it is realized, there exists i such that $a_i x_1 + b_i x_2 + c_i < 0$ is only true for point A, and the intersection of the half plane and the convex hull boundary must have an endpoint, otherwise A is also on the convex hull boundary, which is a contradiction.

Shattering $2n + 1$ points : Consider the endpoints of the convex hull arranged along its boundary as $\{X_i\}_{2n+1}$, and assume that the indices are taken modulo $2n + 1$. For any label such that $l_i = l_{i+1} = \cdots = l_j = 0$, we connect the points $(\frac{x_{1,i}+x_{1,i-1}}{2}, \frac{x_{2,i}+x_{2,i-1}}{2})$ and $(\frac{x_{1,j}+x_{1,j+1}}{2}, \frac{x_{2,j}+x_{2,j+1}}{2})$ with a straight line. This line divides the plane into a half-plane that excludes the points $x_i, \ldots, x_j$.

For the sequence $\{X_i\}_{2n+1}$, the number of such continuous zero-labeled point subsequences is at most n, and hence it is possible to use n half-planes to exclude them. In other words, there exists a hyperplane $h \in \mathcal{H}_n$ that achieves this labeling.

Attempting to shatter $2n + 2$ points : From the above discussion, we know that for any $\{X_i\}_{2n+2}$, if there are points not on the convex hull boundary, they cannot be shattered. If all points are on the convex hull boundary, we assign alternating labels 0 and 1. In this case, there is no half-plane that can separate the two points labeled 0, since these two points are adjacent on the convex hull. Therefore, to shatter X_{i2n+2}, at least $n + 1$ half-planes are required, which cannot be achieved by $\mathcal{H}_n$. ◁

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. ◀

Answer. Let $B(\boldsymbol{c}, r)$ be the ball with radius r and center $\boldsymbol{c}$. Thus $\boldsymbol{x} \in B(\boldsymbol{c}, r)$ iff

$$\|\boldsymbol{x}\|_2^2 - 2\sum_{i\in[n]} c_i x_i + \sum_{i\in[n]} c_i^2 - r^2 \leq 0$$

which is equivalent to $\langle W, X\rangle + B \leq 0$, where $W = \begin{bmatrix} 1 & -2c_1 & \cdots & -2c_n \end{bmatrix}^\top, X = \begin{bmatrix} \|\boldsymbol{x}\|_2^2 & x_1 & \cdots & x_n \end{bmatrix}^\top$, $B = \sum_{i\in[n]} c_i^2 - r^2$. Thus the VC dimension of $\mathcal{H}$ is no more than VC dimension of hyperplanes in $\mathbb{R}^{n+1}$, which is $n+2$. On the other hand, any $n+2$ points on the hyperplane can be realized, including the mapped X. In conclusion, the VC dimension of $\mathcal{H}$ is $n + 2$.

◁

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. ◀

Answer. The VC dimension of $\mathcal{H}$ is n.

Shattering $2n+1$ points : Consider $\{x_i\}_n \subset \{0,1\}^n$ s.t. $x_{ij} = 0 \Leftrightarrow i = j$. Then for any label I where only the corresponding point in the index set A is 1, $wetake$S = [n]$\backslash A$, and $\{\boldsymbol{x} \in \{0,1\}^n \mid f_S(\boldsymbol{x}) = -1\}$ realizes I.

Attempting to shatter $n+1$ points : Note that $n+1$ points means 2^{n+1} labels, but $|\mathcal{H}_n| = \sum_{S \subseteq \{1,2,\cdots,n\}} 1 = 2^n < 2^{n+1}$. Thus $n+1$ points are impossible to be shattered

◁

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, $\pi_{\mathcal{H}}(m) \leq \sum_{i=0}^{d} \binom{m}{i}$. Prove that $\sum_{i=0}^{d} \binom{m}{i} \leq \left(\frac{\mathrm{e}m}{d}\right)^d$ when $m \geq d$.

(2) (8') For a hypothesis class $\mathcal{H}$ with VC-dimension d, define the hypothesis class $\mathcal{H}^k$ $(k \geq 2)$ 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. $(\ln 2 \approx 0.693,\ \mathrm{e} \approx 2.718,\ \ln 7 \approx 1.946,\ \ln\ln 2 \approx -0.367)$

◀

Answer. (1) In fact,

$$\left(\frac{d}{m}\right)^d \sum_{i \in [d]} \binom{m}{i} \leq \sum_{i \in [d]} \binom{m}{i} \left(\frac{d}{m}\right)^i \leq \sum_{i \in [m]} \binom{m}{i} \left(\frac{d}{m}\right)^i = \left(1 + \frac{d}{m}\right)^m \leq e^d$$

(2) Let m be the VC dimension of $\mathcal{H}^k$, then we have

$$\left(\frac{em}{d}\right)^{dk} \geq \pi_{\mathcal{H}}(m)^k \geq \pi_{\mathcal{H}^k}(m) = 2^m$$

Thus $f(m) = d - d\ln d + d\ln m - \frac{m\ln 2}{k} \geq 0$. Taking the derivative, we find that f is monotonically decreasing on $[\frac{dk}{\ln 2}, +\infty)$. Given that $7dk\ln k > \frac{dk}{\ln 2}$ and

$$f(7dk\ln k) = d\,(1 + \ln 7 + (1 - 7\ln 2)\ln k + \ln\ln k)$$

Consider $g(k) = 1+\ln 7+(1-7\ln 2)\ln k+\ln\ln k$, $g'(k) \leq 0 \Leftrightarrow \frac{1}{k\ln k}+\frac{1-7\ln 2}{k} \leq 0 \Leftarrow \frac{1}{\ln 2} < 7\ln 2-1$. Thus $g(k) \leq g(2) = 1 + \ln 7 + (1 - 7\ln 2)\ln 2 + \ln\ln 2 \approx -0.1 < 0$.

Then $f(7dk\ln k) < 0, f(m) \geq 0$, which means $m < 7dk\ln k$.

$\triangleleft$

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 $\mathbf{1}_{r_t \geq 1/2}$.

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 $\Phi_t = \log_2(|C_t|)$.] ◀

Answer. Consider the expectation of making mistakes in each round, we have:

r_t	y_t	penalty	$\Phi_t - \Phi_{t+1}$
$\geq \frac{1}{2}$	1	0	$\log_2 r_t$
$< \frac{1}{2}$	1	$1 - \frac{1}{2}\log_2 \frac{1}{1-r_t}$	$\log_2 r_t$
$\geq \frac{1}{2}$	0	1	$\log_2(1 - r_t)$
$< \frac{1}{2}$	0	$\frac{1}{2}\log_2 \frac{1}{1-r_t}$	$\log_2(1 - r_t)$

where $\Phi_t = \log_2(|C_t|)$.

We find that for each round, $E(\text{pently}) = \frac{1}{2}$, $E(\Phi_t - \Phi_{t+1}) = \frac{\log_2 r_t + \log_2(1-r_t)}{2} \leq -1$. Thus we have,

$$E(\text{mistakes}) \leq \frac{1}{2}\Phi_0 = \frac{1}{2}\log_2 N$$

$\triangleleft$