Machine Learning Exercises _chapter 10

Mohammad Javad Abbas pour

10.1. Boosting the confidence:

Algorithm A: There exist some $\delta_0 \in (0,1)$ and a function $m_H : (0,1) \to \mathbb{N}$ such that for every $\epsilon \in (0,1)$, if $m \geq m_H(\epsilon)$. For every D with probability of at least $1-\delta_0$: $L_D(A(S)) \leq \min_H L_D(h) + \epsilon$. Suggest a procedure that relies on A and learns H in the usual agnestic PAC learnig model and has a sample complexity of:

$$m_H(\epsilon, \delta) \leq k\, m_H(\epsilon) + \left\lceil \frac{2\log(4k/\delta)}{\epsilon^2} \right\rceil$$ where $k = \lceil \log(\delta)/\log(\delta_0) \rceil$.

Solution: Let have $\epsilon, \delta \in (0,1)$. Choose k chunks of size $m_H(\epsilon/2)$. Apply A on each of these chunks. The probability that $\min_{i \in [k]} L_D(\hat{h}_i) \leq \min L_D(h) + \epsilon/2$ is at least $1 - \delta_0^k \geq 1 - \delta/2$. Now we apply an ERM over $\hat{H} := \{\hat{h}_1, \dots, \hat{h}_k\}$ with the training data of size $\left\lceil \frac{2\log(4k/\delta)}{\epsilon^2} \right\rceil$.

Now we use Corollary 4.6, we see that with probability $1-\delta/2$, $L_D(\hat{h}) \leq \min_{i \in [k]} L_D(h_i) + \frac{\epsilon}{2}$

At last we apply union bound (with probability of $1-\delta$):

$$L_D(h) \leq \min_{i \in [k]} L_D(h_i) + \frac{\epsilon}{2} \leq \min_{h \in H} L_D(h) + \epsilon.$$

$\square$

10.4.(a) Assume that $\mathcal{B}$ is the class of all $\underbrace{\text{functions}}_{\text{from } \mathcal{X} \text{ to } \{0,1\}}$, and $\mathcal{X}$ be a finite set of size n.

Then we have $L(\mathcal{B}, T) = \mathcal{B}$, for $\forall T \geq 1$:

$$\text{VCdim}(\mathcal{B}) = \text{VCdim}(L(\mathcal{B}, T)) = \log 2^n = n.$$

(b). Assume B is the class of decision stums in $\mathbb{R}^d$,

$B = \{h_{j,b,\theta} : j \in [d], b \in \{-1,1\}, \theta \in \mathbb{R}\}$ where $h_{j,b,\theta}(x) = b \cdot \text{sign}(\theta - x_j)$.

$\forall j \in [d]$, $B_j = \{h_{b,\theta} : b \in \{-1,1\}, \theta \in \mathbb{R}\}$ where $h_{b,\theta}(x) = b \cdot \text{sign}(\theta - x_j)$

Note $VCdim(B_j) = 2$. ($B = \bigcup_{j=1}^{d} B_j$). with Applying Exercise 11:

$VCdim(B) \leq 16 + 2 \log d$.

- Note that w.l.o.g that $d = 2^k$ for $k \in \mathbb{N}$. Let A is the matrix whose columns range over the set $\{0,1\}^k$.

$\forall i \in [k]$, $x_i = A_{i \to}$. We know that $C = \{x_1, \dots, x_k\}$ is shattered.

Let $I \subseteq [k]$. we can lable the instances in I. positively while $[k] \setminus I$ instances are labled negetively. At last we claim that:

$\exists j$ st $A_{ij} = x_{ij} = 1$ iff $i \in I$ so, $h_{j,-1,\frac{1}{2}}(x_i) = 1$ iff $i \in I$.

(c) Using the hint of 10.4, for $\forall i \in [Tk/2]$ s.t. $x_i = [i/k] A_{i \to}$, we know that C is shattered by $L(B_d, T)$. Let $I \subseteq [Tk/2]$.

$I = I_1 \cup \dots \cup I_{T/2}$ (I_t is a subset of $\{(t-1)k+1, \dots tk\}$. for each $t \in [\frac{I}{2}]$

Let j_t be the colums of A:

$h(x) = \text{sign}((h_{j_1,-1,\frac{1}{2}} + h_{j_1,1,\frac{3}{2}} + \dots + h_{j_{\frac{I}{2}},-1,\frac{I}{2}-\frac{1}{2}})(x))$

At last $h(x) = 1$ if $i \in I \longrightarrow h \in L(B_d, T)$.

□