CS60073: Advanced Machine Learning

End-semester Examination Spring 2018

Answer all THREE questions. Clearly write your steps of derivations. Answer all parts of a question together. Make suitable assumptions if necessary. Time: 3 hrs. Total marks: 60.

1(a). State Sauer Lemma. **[5]**

(b). Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two hypothesis classes with $d_1 = \text{VC-dim}(\mathcal{H}_1)$, and $d_2 = \text{VC-dim}(\mathcal{H}_2)$. Using Sauer Lemma prove that $\text{VC-dim}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2\max(d_1, d_2) + 1$. **[10]**

(c). For every $n \in \mathbb{N}$, let $\mathcal{H}_n$ be the class of polynomial classifiers of degree n; namely, $\mathcal{H}_n$ is the set of all classifiers of the form $h(x) = \text{sign}(p(x))$, where $p : \mathbb{R} \to \mathbb{R}$ is a polynomial of degree n. Define $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \mathcal{H}_n$. Is $\mathcal{H}$ PAC Learnable? Explain. Is $\mathcal{H}$ non-uniformly learnable? Explain. **[5]**

2(a). Define γ-weak learnability. **[5]**

(b). Consider a hypothesis class, $\mathcal{H} = \{h_{\theta_1,\theta_2,b} : \theta_1, \theta_2 \in \mathbb{R}, \theta_1 < \theta_2, b \in \{\pm 1\}\}$ be defined for every $x \in \mathbb{R}$ as follows:

$$h_{\theta_1,\theta_2,b}(x) = \begin{cases} +b, & \text{if } \ x < \theta_1, \ \text{ or } \ x > \theta_2. \\ -b, & \text{if } \ \theta_1 \leq x \leq \theta_2. \end{cases}$$

Let B be the class of all (left or right) threshold functions over domain $\mathbb{R}$, i.e.,

$$B = \{x \mapsto \text{sign}(x - \theta).b \ ; \ \theta \in \mathbb{R}, b \in \{\pm 1\}\}.$$

Show that ERM_B is a γ-weak learner for $\mathcal{H}$. **[5]**

(c). Consider a base hypothesis class B over domain $\mathbb{R}$ as follows:

$$B = \{x \mapsto \text{sign}(x - \theta).b \ ; \ \theta \in \mathbb{R}, b \in \{\pm 1\}\}.$$

Now, consider a more complex hypothesis class $L(B, T)$ consisting of linear compositions of this base class.

$$L(B, T) = \{x \mapsto \text{sign}(\sum_{t=1}^{T} w_t h_t(x)) : w_t \in \mathbb{R}, \forall t, h_t \in B\}$$

Each $h \in L(B, T)$ is parameterized by T base hypothesis from B and by a set of weights $w_t \in \mathbb{R}$. Show that VC-dimension of $L(B, T)$ is at least $T + 1$. **[10]**

3.(a). Let $\mathcal{H}$ be a finite hypothesis class. Assume that there is a hypothesis in $h^c \in \mathcal{H}$ that makes no mistakes, i.e., is consistent with the training sequence (realizability assumption). In this scenario, describe an online learning algorithm which enjoys a mistake bound $M(\mathcal{H}) \leq \log_2(|\mathcal{H}|)$. Explain your answer. **[5]**

(b). We now remove the realizability assumption and instead assume that there is a hypothesis $h^m \in \mathcal{H}$ which makes at most $m \geq 0$ mistakes on the training sequence. Propose a modification of the above online learning algorithm such that the modified algorithm has a mistake bound $M(\mathcal{H}) \leq O((m+1)\log_2(|\mathcal{H}|))$. **[10]**

(c). Define Littlestone dimension L-dim($\mathcal{H}$) of a hypothesis class $\mathcal{H}$. **[5]**

—- BEST WISHES —-