CS60018: Statistical Learning Theory

Mid-Autumn Semester Exam

Time: 2:00hrs | Answer all THREE questions. | Max marks: 50

1.A. Define agnostic PAC learnability. [5]

1.B. Let $[a, b]$ and $[c, d]$ be two intervals of the real line $\mathbb{R}$ with $a \leq b \leq c \leq d$. Let $\epsilon > 0$ and $\Pr_{\mathcal{D}}((b, c)) > \epsilon$, where $\mathcal{D}$ is the distribution from which input points are drawn. Show that the probability that none of the points in an i.i.d sample S_m of size m fall in the interval (b, c) is at most $e^{-m\epsilon}$. You can use the inequality $1-x \leq e^{-x}$. Consequently, show that the hypothesis class consisting of union of two closed intervals on the real line ($[a, b] \cup [c, d]$) is PAC learnable. [10]

1.C. We want to learn a target concept using a hypothesis class consisting of union of two non-overlapping axis-parallel rectangles in $\mathbb{R}^2$. Assume that the target concept is realizable by this hypothesis class. What is the minimum size of an i.i.d training set S_m, such that a learner can learn the target function with error less than ϵ and with confidence more than $1 - \delta$. Show your derivation. [10]

2.A. Show that $\text{VC}(\mathcal{H}) \leq \log_2 |\mathcal{H}|$, where $\mathcal{H}$ is a finite hypothesis space. [5]

2.B. Consider the input space $\mathcal{X} = \{1, 2, 3, \ldots, 999\}$ of integers less than thousand. Let $\mathcal{H}$ be a hypothesis space having ten hypotheses $h_0, \ldots, h_9$. For a number $n \in \mathcal{X}$, $h_i(n) = 1$, if the decimal representation of n contains the digit i; and $h_i(n) = 0$, otherwise. For example, for $n = 911$, $h_9(n) = 1$ and $h_1(n) = 1$. Show that $\text{VC}(\mathcal{H}) = 2$. [10].

3.A. Over the input space $x \in \mathbb{R}$, define $\mathcal{H} = \{h_\theta(x) = sign(\sin(\theta x), \theta \in \mathbb{R}\}$. The $sign()$ function returns 1 for positive inputs, and 0 otherwise. Argue that $\text{VC}(\mathcal{H}) = \infty$. [5]

3.B. Show that adding one function to a hypothesis class of binary classifiers can increase the VC-dimension by at most one. That is, for any binary hypothesis class $\mathcal{H}$ and a hypothesis h: $\text{VC}(\mathcal{H} \cup h) \leq \text{VC}(\mathcal{H}) + 1$. [5].

— END —