Computational Learning Theory February 2016

Warsaw Graduate School

James Worrell

1. (a) Let D be a distribution on $\mathbb{R}$ and (b, c) an interval with $\Pr_D((b, c)) > \varepsilon$ for some $\varepsilon > 0$. Show that the probability that m points drawn i.i.d. from D all fall outside (b, c) is at most $e^{-m\varepsilon}$.

   (b) Show that the hypothesis class formed by the collection of unions of two closed bounded intervals of reals of the form $[a, b] \cup [c, d]$, with $a \le b \le c \le d$, is efficiently PAC learnable in the realisable case.

2. Let $\oplus$ denote the binary exclusive-or function. Given $f, g : X \to \{0, 1\}$, define $f \oplus g : X \to \{0, 1\}$ by $(f \oplus g)(x) = f(x) \oplus g(x)$ for all $x \in X$. Given a collection H of functions from X to $\{0, 1\}$ and some fixed function $f : X \to \{0, 1\}$, show that H and $\{h \oplus f : h \in H\}$ have the same VC dimension.

3. (a) For each fixed k, what is the VC dimension of the class of subsets of the real line expressible as the union of k closed intervals? Justify your answer.

   (b) Prove that the class of hyper-rectangles in $\mathbb{R}^n$, of the form $[a_1, b_1] \times \ldots \times [a_n, b_n]$, has VC dimension $2n$.

   (c) Prove that for $n > 1$ the class of Boolean functions on $\{0, 1\}^n$ that can be expressed as conjunctions of literals involving propositional variables $p_1, \ldots, p_n$ has VC dimension n.

4. For $i = 1, \ldots, n$, define $\boldsymbol{x}_i \in \mathbb{R}^n$ and $y_i \in \{-1, 1\}$ by

$$\boldsymbol{x}_i = (\underbrace{(-1)^i, \ldots, (-1)^i, (-1)^{i+1}}_{i \text{ first components}}, 0, \ldots, 0) \quad \text{and} \quad y_i = (-1)^{i+1}\,.$$

   Suppose that the Perceptron algorithm is run repeatedly over the sequence $S = (\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ until it makes no more mistakes. Show that the total number of mistakes is at least 2^{n-3}.

   Hint. Let $\boldsymbol{w} = (w_1, \ldots, w_n) \in \mathbb{Z}^n$ be (a normal vector of) any linear separator. Give lower bounds on the magnitude of successive entries $w_1, w_2, \ldots$.