Assignment 2

COMP 599 Fall 2020 McGill University
Instructors: Adam Oberman and Prakash Panangaden

Due 2^{nd} October 2020

Question 1.[15 points] An axis-aligned hyper-rectangle in $\mathbb{R}^n$ is a set of the form $[a_1, b_1] \times \ldots \times [a_n, b_n] \subseteq \mathbb{R}^n$. Show that the hypothesis class of hyper-rectangles is efficiently PAC-learnable (in the realizable case) by generalizing the reasoning we did in class for the case $n = 2$.

Question 2.[20 points]

(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 $e^{-m\varepsilon}$. This is routine and is a preparation for the next part of the question.

(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 $a \leq b \leq c \leq d$, is efficiently PAC learnable in the realizable case.

Question 3.[15 points]

The two-distribution model is defined as follows. Let H be a hypothesis class on domain X and and let $c \in H$ be the target concept. The learning algorithm $\mathcal{A}$ requests *separately* a set of m_+ positive examples and a set of m_- negative examples. The positive examples are drawn according to some distribution D_+ on $\{x \in X : c(x) = 1\}$ and the negative examples are drawn according to some fixed distribution D_- on the set $\{x \in X : c(x) = 0\}$. If either of the above sets is empty the learning algorithm receives the empty set in response to its request.

Fixing the accuracy ε and confidence δ the learning algorithm must output a hypothesis h such that with probability at least $1 - \delta$:

$$\Pr_{x \sim D_+}[h(x) = 0] \leq \varepsilon \text{ and } \Pr_{x \sim D_-}[h(x) = 1] \leq \varepsilon.$$

We say that $\mathcal{A}$ PAC-learns H in the two-distribution model if the required sample sizes are bounded by a polynomial in $1/\delta$ and $1/\varepsilon$.

Show that if the hypothesis class H is efficiently PAC-learnable in the standard (one-distribution) model, then it is also efficiently PAC learnable in the two-distribution model.

Question 4.[20 points]

1. For each fixed k, what is the VC dimension of the class of subsets of the real line expressible as the union of k or fewer closed intervals? Justify your answer.
2. 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 $2n$.

Question 5.[30 points]

Let $A \subseteq \mathbb{R}^m$ be a finite set of vectors with $\|\vec{a}\| \leq 1$ for all $\vec{a} \in A$. Prove the following inequality

$$\mathbb{E}_{\vec{\sigma}} \left[\max_{\vec{a} \in A} \sum_{i=1}^{m} \sigma_i a_i \right] \leq \sqrt{2 \log |A|}$$

where $\vec{\sigma}$ represents the vector $(\sigma_1, \sigma_2, \ldots, \sigma_m)$ of independent random variables uniformly distributed over $\{+1, -1\}$ and $a_1, \ldots, a_m$ are the components of the vector $\vec{a}$. This lemma is a key step in showing error bounds for VC dimension using Rademacher complexity as the tool.

Hints: You should try to prove that

$$\mathbb{E}_{\vec{\sigma}} \left[\max_{\vec{a} \in A} \sum_{i=1}^{m} \sigma_i a_i \right] \leq \frac{\log |A|}{t} + \frac{t}{2}$$

for all $t > 0$. The result then follows immediately by a suitable choice of t. To prove this inequality one should use some basic inequalities:

$$\max x_i \leq \log(\sum_i \exp(x_i) \text{ and } \frac{e^x + e^{-x}}{2} \leq e^{x^2/2}).$$

You will also need Jensen's inequality.