National University of Singapore
School of Computing
CS3244 Machine Learning

Tutorial 8: Computational Learning Theory

Issue: April 1, 2020 Due: —

Important Instructions:

- *Your solutions for this tutorial must be TYPE-WRITTEN.*
- *Make TWO copies of your solutions: one for you and one to be SUBMITTED TO THE TUTOR IN CLASS. Your submission in your respective tutorial class will be used to indicate your CLASS ATTENDANCE. Late submission will NOT be entertained.*
- *You may discuss the content of the questions with your classmates. But everyone should work out and write up ALL the solutions by yourself.*

TM 7.2 Consider the class C of concepts of the form $(a \leq x \leq b) \wedge (c \leq y \leq d)$ where a, b, c, and d are integers in the interval $[0, 99]$. Note each concept in this class corresponds to a rectangle with integer-valued boundaries on a portion of the x, y plane.
Hint: Given a region in the plane bounded by the points $(0, 0)$ and $(n-1, n-1)$, the number of distinct rectangles with integer-valued boundaries within this region is

$$\left(\frac{n(n+1)}{2}\right)^2 .$$

(a) Determine the number of randomly drawn training examples sufficient to guarantee that, for any target concept $c \in C$, the probability that a consistent learner using $H = C$ outputs a consistent hypothesis $h \in VS_{H,D}$ with $error_Q(h) < 0.15$ is at least 95%.

(b) Now, suppose that the rectangle boundaries a, b, c, and d take on **real** values instead of integer values. Update your answer to the first part of this question.

Solution.

(a) $|D| = 134$.

(b) $|D| = 1516$.

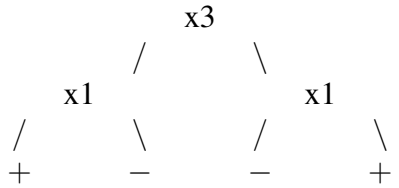
TM 7.5 Consider the instance space X corresponding to all points in the x, y plane. State the VC dimension of the hypothesis space H which is the set of all rectangles in the x, y plane. That is,

$$H = \{((a \leq x \leq b) \wedge (c \leq y \leq d)) \mid a, b, c, d \in \mathbb{R}\} .$$

You do not have to explain your answer.

Solution. $VC(H) = 4$.

TM 7.7 Consider the hypothesis class H_{rd2} of "regular, depth-2 decision trees" over n Boolean variables. A "regular, depth-2 decision tree" is a depth-2 decision tree (a tree with four leaves, all distance 2 from the root) in which the left and right child of the root are **required to contain the same variable**. For instance, the following tree is in H_{rd2}.



(a) As a function of n, how many syntactically distinct trees are there in H_{rd2}?

(b) Determine the number of training examples needed in the PAC model to learn H_{rd2} with error ϵ and confidence $1 - \delta$.

Solution.

(a) $|H_{rd2}| = n \times (n-1) \times 2^4 = 2^4 n(n-1)$.

(b) $|D| \geq (1/\epsilon)(\ln 2^4 n(n-1) + \ln(1/\delta)) = (1/\epsilon)(4 \ln 2 + \ln n + \ln(n-1) + \ln(1/\delta))$.