Computational Learning Theory
Michaelmas Term 2018
Week 5

Problem Sheet 3

Instructions: The problem sheets are designed to increase your understanding of the material taught in the lectures, as well as to prepare you for the final exam. You should attempt to solve the problems on your own after reading the lecture notes and other posted material, where applicable. Once you have given sufficient thought to a problem, if you are stuck, you are encouraged to discuss with others in the course and with the lecturer during office hours. Please avoid posting on Piazza until the Wednesday before the submission deadline. You are *not permitted* to search for solutions online. Questions marked with an asterisk (*) are optional.

1 Properties of AdaBoost

Consider the AdaBoost algorithm as described in the lecture notes.

1. Show that the error of h_t with respect to the distribution D_{t+1} is exactly 1/2.
2. What is the maximum possible value of $D_t(i)$ for some $1 \leq t \leq T$ and $1 \leq i \leq m$?
3. Fix some example, say i, let t_i be the first iteration such that $h_{t_i}(x_i) = y_i$. How large can t_i be?

2 Weak Learning CONJUNCTIONS and PARITIES

Consider the instance space $X_n = \{0,1\}^n$. Consider the following hypothesis class:

$$H_n = \{0, 1, x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n\}$$

The hypothesis class contains $2n+2$ functions. The functions "0" and "1" are constant and predict 0 and 1 on all instances in X_n. The function "x_i" evaluates to 1 on any $a \in \{0,1\}^n$ satisfying $a_i = 1$ and 0 otherwise. Likewise, the function "$\bar{x}_i$" evaluates to 1 on any $a \in \{0,1\}^n$ satisfying $a_i = 0$ and 0 otherwise. Thus a single bit of the input determines the value of these functions; for this reason these functions are sometimes referred to as *dictator* functions.

1. Show that the class CONJUNCTIONS is $\frac{1}{10n}$-weak learnable using H.
2. Let $\mathsf{CONJUNCTIONS}_k$ denote the class of conjunctions on at most k literals. Give an algorithm that PAC-learns $\mathsf{CONJUNCTIONS}_k$ and has sample complexity polynomial in k, $\log n$, $\frac{1}{\epsilon}$ and $\frac{1}{\delta}$. What would be the sample complexity if you had used the algorithm for learning CONJUNCTIONS discussed in the lectures (Lec 1)? *(Hint: First show that the weak learning algorithm in the previous part can be modified to be a $\frac{1}{10k}$-weak learner in this case.)*
3. Show that there does not exist a weak learning algorithm for PARITIES using H.

3 Teaching Dimension

We consider a new model of learning with the aid of a teacher. In this model, instead of receiving examples randomly from a distribution, a teacher can provide examples that are most *helpful* to a learner. Let X be a finite instance space. Given a concept class C and a target concept $c \in C$, we say that a sequence T of labelled examples is a *teaching sequence* for $c \in C$, if c is the only concept in C which is consistent with T. Let $T(c)$ be the set of all teaching sequences for $c \in C$. The *teaching dimension* of concept class C is then defined to be,

$$\mathsf{TD}(C) = \max_{c \in C} \min_{T \in T(c)} |T|$$

where $|T|$ denotes the number of examples in the sequence T.

1. Give an example of a concept class C for which $\mathsf{TD}(C) > \mathsf{VCD}(C)$.
2. Give an example of a concept class C for which $\mathsf{TD}(C) < \mathsf{VCD}(C)$.
3. Show that for any concept class C, $\mathsf{TD}(C) \leq |C| - 1$.
4. Show that for any concept class C, $\mathsf{TD}(C) \leq \mathsf{VCD}(C) + |C| - 2^{\mathsf{VCD}(C)}$.