Computational Learning Theory

Assignment 1

Deadline: December 8th.

1. Given the boolean function

$$f(x_1, x_2, x_3, x_4) = \overline{x_1 \vee ((x_4 \wedge \overline{x_2}) \oplus x_3 \oplus \overline{x_1})}$$

   where $\oplus$ is the Exclusive or.

   (a) Give a decision tree for the boolean function f.

   (b) What is the size of this decision tree?

   (c) What are the changes you need to make, for changing the decision tree of f to a decision tree for $\bar{f}$.

   (d) Prove that for any boolean function g, $size_{DT}(g) = size_{DT}(g^D)$.

   (e) Change the tree of f to DNF and CNF.

   (f) Change the tree to k-DL for some $k \leq 4$.

   (g) Can we write f as DL? Prove your answer.

   (h) Can we write f as 2-DL? Prove your answer.

   (i) Can we write f as 3-DL? Prove your answer.

   (j) Can we write f as a MDNF? Prove your answer.

2. Define the class XOR$= \{x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_j} \oplus \xi\}$ for all possible $1 \leq i_1 < i_2 < \cdots < i_j \leq n$, $0 \leq j \leq n$ and $\xi \in \{0, 1\}$. Define the class kXOR the set of all functions in XOR that contains at most k variables. Define XORk the set of all functions in XOR that contains at least $n - k$ variables. Define $\vee k$XOR the class that is the disjunction (or) of kXORs. In the same way we define $\vee$XORk, $\wedge k$XOR and $\wedge$XORk.

   (a) What is XOR^D, $\wedge k$XOR^D and $\vee$XORk^D.

   (b) Give an online learning algorithm for $\wedge k$XOR and $\vee$XORk for constant k that runs in polynomial time.

   (c) Given the sample $S = \{(000, 1), (001, 0), (010, 0), (011, 0), (100, 0), (110, 1), (111, 0)\}$. Give a $\wedge$2XOR that is consistent with S.

3. (a) Is the following true and in what conditions?
   If C is properly learnable from examples in polynomial time then the following problem is solvable in polynomial time:
   Given a sample S.
   Decide if there is a function $f \in C$ that is consistent with S.

   (b) Given the sample

$$S = \{(00000, 0), (10001, 1), (00010, 0), (10011, 1), (00100, 0), (10101, 1),$$

$$(00111, 0), (11000, 0), (01001, 0), (11010, 0), (11101, 1), (11110, 0), (01111, 1)\}.$$

   Find a consistent DL for S.

(c) Is there a term or a clause that is consistent with S? Prove your answers.

(d) Is there a $\wedge$2XOR function consistent with S? Prove your answer.

(e) Given the sample $S = \{(000,0),(001,1),(010,1),(011,0),(100,1),(110,0),(111,1)\}$. Is there a 2-CNF that is consistent with S? Is there a 2-DNF that is consistent with S? Prove your answers.

4. Define $\mathbf{BOX}_{a,b,c,d} : \Re^2 \rightarrow \Re$, where $\Re$ is the set of real numbers, and

$$\mathbf{BOX}_{a,b,c,d}(x_1,x_2) = [a \le x_1 \le b] \wedge [c \le x_2 \le d].$$

Define the class $\mathbf{BOX} = \{\mathbf{BOX}_{a,b,c,d} \mid a,b,c,d \in \Re \cup \{-\infty,\infty\}\}$.

(a) Show that $\mathbf{BOX}$ is properly learnable from examples.

(b) Show that the class

$$\wedge s - \mathbf{BOX} = \{\mathbf{BOX}_{a_1,b_1,c_1,d_1} \wedge \cdots \wedge \mathbf{BOX}_{a_s,b_s,c_s,d_s} \mid \mathbf{BOX}_{a_i,b_i,c_i,d_i} \in \mathbf{BOX}\}$$

is properly learnable from examples.

(c) Show that the class

$$\vee s - \mathbf{BOX} = \{\mathbf{BOX}_{a_1,b_1,c_1,d_1} \vee \cdots \vee \mathbf{BOX}_{a_s,b_s,c_s,d_s} \mid \mathbf{BOX}_{a_i,b_i,c_i,d_i} \in \mathbf{BOX}\}$$

is learnable from examples with a hypothesis of size $s \log m$. What is the time complexity?

(d) Show that although $\vee s - \mathbf{BOX} = s - \mathbf{MClause}(\mathbf{BOX})$ we cannot use the Composition Lemma for learning $\vee s - \mathbf{BOX}$.

(e) Show that for any boolean function $f : \{0,1\}^m \rightarrow \{0,1\}$,

$$f(\mathbf{BOX}_{a_1,b_1,c_1,d_1},\cdots,\mathbf{BOX}_{a_s,b_s,c_s,d_s}) \in \vee(2s+1)^2 - \mathbf{BOX}.$$

(f) Show that $\mathbf{FORMULA}(\mathbf{BOX})$ is learnable from examples. What is the hypothesis size and the time complexity?

(g) Define

$$\mathbf{BOX}^{(n)} = \{\mathbf{BOX}_{a,b,c,d} \mid a,b,c,d \in \{1,2,\ldots,n\} \cup \{-\infty,\infty\}\}.$$

Show that $\mathbf{BOX}^{(n)}$ is online learnable with Halving with $poly(\log n)$ mistakes.

(h) Show that $\mathbf{BOX}^{(n)}$ is online learnable with $poly(\log n)$ mistakes in $poly(n)$ time.
BONUS: $poly(\log n)$ time.

(i) Show that $\vee s - \mathbf{BOX}^{(n)}$ is online learnable with $poly(n,s)$ mistakes. Hint: See (4d).

(j) Show that $\mathbf{FORMULA}(\mathbf{BOX}^{(n)})$ is online learnable with $poly(n,s)$ mistakes.

(k) Show that $\mathbf{FORMULA}(\mathbf{BOX}^{(n)})$ is online learnable with $poly(\log n, s)$ mistakes and $poly(n,s)$ time. Hint: use WINNOW - Explain your answer.
BONUS: $poly(\log n, s)$ mistakes and time.

5. (a) Let $C_1,\ldots,C_t$ be classes of boolean functions, t constant. Suppose there is an algorithm $\mathcal{A}_i$ that online learns C_i in polynomial time with T_i mistakes. Prove that there is an algorithm that learns $C_1 \cup C_2 \cup \cdots \cup C_t$ in polynomial time with $O(\max_i T_i)$ mistakes.

(b) Let $C_1, \ldots, C_t$ be classes of boolean functions, t not constant. Suppose there is an algorithm $\mathcal{A}_i$ that online learns C_i in polynomial time with T_i mistakes. Prove that there is an algorithm that learns $C_1 \cup C_2 \cup \cdots \cup C_t$ in polynomial time with $O(\max_i T_i + \log t)$ mistakes.

(c) Suppose we run the perceptron algorithm with the update $w_{i+1} \leftarrow w_i + \lambda_i(y_i a_i)$ for $\lambda_i > 0$. What is the number of mistakes? Show that the minimal number of mistakes is achieved when all the λ_i are equal. Hint: Use Cauchy-Schwarz inequality.

(d) Consider the WINNOW algorithm with the following changes:
(3) If $f(z) = 1$ then for every $z_i = 1$ do $w_i \leftarrow \alpha w_i$.
(4) If $f(z) = 0$ then for every $z_i = 1$ do $w_i \leftarrow \beta w_i$.
Find the number of mistakes as a function of α and β. For what values of α and β the algorithm makes minimal number of mistakes.

(e) Consider the weighted majority algorithm with the update $w_{k,i+1} \leftarrow \alpha w_{k,i}$ in step 3.1. What is the number of mistakes of the algorithm?