Computational Learning Theory
Assignment 2

Let $[0,1] = \{x \mid 0 \le x \le 1\}$ and B be a set of all functions $g : [0,1] \to [0,1]$. For $g \in B$ we say that g is *monotone nondecreasing function* if for every $0 \le x \le y \le 1$ we have $g(x) \le g(y)$. We denote by B_M the set of all monotone nondecreasing functions in B. We say that g is *k-stairs function* if $g(x)$ takes at most k values. For example, $g(x) = \lfloor 10x \rfloor / 10$ is 11-stairs function. We denote by B_{kS} the set of all k-stairs functions in B. We denote $B_{MkS} = B_M \cap B_{kS}$.

For $g \in B$ Let $f_g : [0,1] \times [0,1] \to \{0,1\}$ where $g \in B$ and $f_g(x,y) = [y \ge g(x)]$. That is $f_g(x,y) = 1$ if $y \ge g(x)$ and $f_g(x,y) = 0$ otherwise. For the set B we denote by $C_e = \{f_g \mid g \in B\}$. In the same way we define $C_M = \{f_g \mid g \in B_M\}$, $C_{kS} = \{f_g \mid g \in B_{kS}\}$ and $C_{MkS} = \{f_g \mid g \in B_{MkS}\}$.

Let $Q_d = \{0, 1/d, 2/d, \ldots, (d-1)/d, 1\}$. The discritized sets DC_e, DC_M, DC_{kS} and DC_{MkS} are defined as above but for functions $g : Q_d \to Q_d$ and $f_g : Q_d \times Q_d \to \{0,1\}$.

1. Find $|DC_e|$, $|DC_M|$, $|DC_{kS}|$ and $|DC_{MkS}|$. A bound like $2^{\Theta(h(d,k))}$ suffices.
2. Use the first bound of OCCAM to find the number of examples needed to to PAC-learn DC_e, DC_M, DC_{kS} and DC_{MkS}.
3. Write a PAC learning algorithm for DC_e, DC_M, DC_{kS} and DC_{MkS}.
4. Prove that for any class of functions C we have $VCdim(C) \le \log |C|$ and find an upper bound for the $VCdim$ of the classes DC_e, DC_M, DC_{kS} and DC_{MkS}.
5. What is the connection between $VCdim(DC_X)$ and $VCdim(C_X)$ for $X \in \{e, M, kS, MkS\}$.
6. How many equivalence query we need in order to EXACT learn the classes DC_e, DC_M, DC_{kS} and DC_{MkS} using the HALVING algorithm.
7. Write a polynomial time EXACT learning algorithm for the classes $|DC_e|$ that uses number of equivalence queries that meets the number of equivalence queries in the HALVING algorithm. BONOS: For the classes DC_M, DC_{kS} and DC_{MkS}.
8. Show that VCdim of C_e, C_M and C_{kS} is ∞.
9. Show that C_e, C_M and C_{kS} are not PAC-learnable. Hint: Read Andrzej Ehrenfeucht, David Haussler, Michael J. Kearns, Leslie G. Valiant: A General Lower Bound on the Number of Examples Needed for Learning. Inf. Comput. (IANDC) 82(3):247-261 (1989).
10. Show that C_e, C_M and C_{kS} are not non-trivial EXAMPLE learnable.
11. Find $VCdim(C_{MkS})$.
12. Find $SC(C,m)$ (exactly) for C_e, C_M, C_{kS} and C_{MkS}.
13. How many examples do we need in order to properly PAC learn C_{MkS} using OCCAM.
14. Find an EXAMPLE learning algorithm for C_{MkS} using set cover. BONOS: Find a proper learning algorithm.

15. Find a PAC learning algorithm that uses $O((k/\epsilon)\log(k/\delta))$ examples. For what ϵ this bound is better that the bound in 13. Hint: as we did in ray define a stairs function with error $O(\epsilon/k)$ for each stair and find the probability of having a point between each two stairs.

16. Consider the class BOX of all rectangles in $[0,1]\times[0,1]$ and their complement. Show that for every $f\in C_{MkS}$ and any distribution D there is $h\in$BOX such that $\Pr[f=h]=1/2+1/(4k)$. Use Adaboost to PAC-learn C_{MkS}. What is the number of examples? What is the hypothesis? Here

$$\text{BOX}=\{f_{a,b,c,d}\mid f_{a,b,c,d}(x,y)=[a\le x\le b]\wedge[c\le y\le d],a,b,d,c\in[0,1]\}.$$

17. We say that an example $(x_0,y_0),\xi$ is γ-far from the target f_g if $|g(x_0)-y_0|\ge\gamma$. Show that if the examples are γ-far from the target in C_M then C_M is EXAMPLE learnable. Hint: It is consistent to C_{MkS} function for some k.

18. Define the following 16 squares: $[i/4,(i+1)/4]\times[j/4,(j+1)/4]$ for $0\le i,j\le 3$. Show that any function $g\in B_M$ can intersect at most 7 squares. Given an oracle that returns a random uniform $(x,y)\in[0,1]\times[0,1]$. How many calls to the oracle do we need to ensure that with probability at least $1-\delta$ we get at least one point in each square $[i/4,(i+1)/4]\times[j/4,(j+1)/4]$ for $0\le i,j\le 3$. Show that C_M can be learned under the uniform distribution (i.e., when the distribution is uniform) with error $\epsilon=7/16$ with a hypothesis that is a union of squares. How many examples do we need?

19. Using the above idea, write a PAC-learning algorithm that learns B_M under the uniform distribution for any ϵ and δ. How many examples do we need?

20. **Bonus:** Show that B_M is learnable under any product distribution (product distribution is a distribution D where $D(S\times R)=D_1(S)\times D_2(R)$ for some D_1 and D_2). Hint: take random points and use them to build the lattice. Then learn the function in each square.

21. In 18 we have seen a week PAC-learning algorithm. Can we change this week learning to a strong learning via boosting?

22. Show that B_{kS} is not PAC learnable under the uniform distribution.

23. The class $C=\bigoplus_d C_{MkS}$ is the set of xor of d functions in C_{MkS}. Give a bound on $VCdim(C)$.