Problem Sheet 2

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. Problems marked with an asterisk are optional. 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. You are *not permitted* to search for solutions online.

1 Learning Decision Lists

A k-decision list over n boolean variables $z_1, \ldots, z_n$, is defined by an ordered list

$$L = (t_1, b_1), (t_2, b_2), \ldots, (t_l, b_l),$$

and a bit b, where each t_i is a term (conjunction) of at most k literals (positive or negative) and each $b_i \in \{0, 1\}$. For $\mathbf{x} \in \{0, 1\}^n$ the value $L(\mathbf{x})$ is defined to be b_j, where j is the smallest index satisfying $t_j(\mathbf{x}) = 1$ and $L(\mathbf{x}) = b$ if no such index exists. Pictorially, a decision list can be depicted as shown below. As we move from left to right, the first time a term is satisfied, the corresponding b_j is output, if none of the terms is satisfied the default bit b is output.

$$\begin{tikzcd}
\boxed{z_1 \wedge \bar{z}_3} \arrow[r] \arrow[d] & \boxed{z_4} \arrow[r] \arrow[d] & \boxed{\bar{z}_2 \wedge z_3} \arrow[r] \arrow[d] & \boxed{z_1 \wedge z_5} \arrow[r] \arrow[d] & \boxed{\bar{z}_4 \wedge z_6} \arrow[r] \arrow[d] & \boxed{z_1 \wedge \bar{z}_6} \arrow[r] \arrow[d] & b \\
b_1 & b_2 & b_3 & b_4 & b_5 & b_6 &
\end{tikzcd}$$

Give an *efficient* PAC learning algorithm for the class of decision lists. As a first step, argue that it is enough to just consider the case where all the terms have length 1, *i.e.,* in fact they are just literals.

Hint: For the first part, use the same reduction as we considered in the lectures to reduce the problem of learning k-CNF to that of learning conjunctions.

For the second part, find an efficient consistent learning algorithm and count the number of functions that can be represented by decision lists. Then apply Occam's Razor.

2 Learning Convex Sets in $[0, 1]^2$

In this question we will consider the learnability of convex sets. Let us consider the domain to be $X = [0, 1]^2$, the unit square in the plane. For $S \subset X$ a convex set, let $c_S : X \to \{0, 1\}$, where $c_S(x) = 1$ if $x \in S$ and 0 otherwise. Let $C = \{c_S \mid S \text{ convex subset of } X\}$ be the concept class defined by convex sets of $[0, 1]^2$.

1. Show that the VC dimension of C is ∞. This shows that the concept class of convex sets of $[0,1]^2$ cannot be learnt by an algorithm (efficiently or otherwise) that uses a sample whose size is bounded by a polynomial in $1/\epsilon$ and $1/\delta$ alone.

2. We will consider a restriction of PAC-learning where the learning algorithm is only required to work for a specific distribution D over X. Show that if D is the uniform distribution over $[0,1]^2$, then the concept class of convex sets is *efficiently* PAC-learnable in this restricted sense, where efficiency means running time (and sample complexity) bounded by a polynomial in $\frac{1}{\epsilon}$ and $\frac{1}{\delta}$.

Hints:

1. Consider a suitable geometric arrangement of points, where no matter which subset of points are labelled 1, you can always find a convex set that encloses the positively labelled points and excludes all the negatively labelled ones. Make sure that this arrangement can be made with d points, for an arbitrary natural number d.

2. Consider the algorithm that simply outputs the convex hull of positive points as the output hypothesis. You may assume without proof that the convex hull of a set of points in the plane can be computed in polynomial time. You may use the fact that the perimeter of any convex set in the unit square can be at most 4.

   For the analysis you may find it helpful to form a $k \times k$ grid of the unit square $[0,1]^2$, where you should choose k as a function of ϵ carefully. Then argue that the boundary of the convex set lies in $\Theta(k)$ squares.

3 Growth Function

Prove that for any $d \in \mathbb{N}$, there is a concept class C such that $\mathsf{VCD}(C) = d$, and that for any $m \in \mathbb{N}$, $\Pi_C(m) = \Phi(m, d)$.

Hint: Recall that $\Phi(m,d) = \sum_{i=0}^{d} \binom{m}{i}$. Combinatorially, think about what each of the terms in the summation represent and then make sure that all of those can be represented by concepts in your concept class.