CMSC5724: Exercise List 4

Problem 1. A *rectangular classifier* h in $\mathbb{R}^2$ is described by an axis-parallel rectangle $r = [x_1, x_2] \times [y_1, y_2]$. Function h maps all the points covered by r to label 1, and all the points outside r to label -1. Give a set of 4 points $\mathbb{R}^2$ that can be shattered by the class of rectangular classifiers.

Solution. $A = (0, 1), B = (1, 0), C = (-1, 0), D = (0, -1)$.

Problem 2. Prove: there does not exist any set of 5 points in $\mathbb{R}^2$ that can be shattered by the class of rectangular classifiers.

Solution. Take any 5 points P in $\mathbb{R}^2$. Identify a subset $S \subseteq P$ as follows.

- Initially, S is empty.
- Add to S a point in P with the minimum x-coordinate.
- Add to S a point in P with the maximum x-coordinate.
- Add to S a point in P with the minimum y-coordinate.
- Add to S a point in P with the maximum y-coordinate.

The size of S is at most 4.

Any axis-parallel rectangle covering S must cover the entire P and, hence, must also cover all the points in $P \setminus S$. Consider the label assignment where the points in S have label 1, and those in $P \setminus S$ have label -1. No rectangular classifier can produce these labels.

Problem 3. Let $\mathcal{P}$ be a set of points in $\mathbb{R}^d$ for some integer $d > 0$. Let $\mathcal{H}$ be a set of classifiers each of which maps $\mathbb{R}^d$ to $\{-1, 1\}$. Prove: for any $\mathcal{H}' \subseteq \mathcal{H}$, it holds that VC-dim$(\mathcal{P}, \mathcal{H}') \le$ VC-dim$(\mathcal{P}, \mathcal{H})$.

Solution. Let $\lambda =$ VC-dim$(\mathcal{P}, \mathcal{H})$. It suffices to prove that $\mathcal{P}$ does not contain a subset P of size $\lambda + 1$ that can be shattered by $\mathcal{H}'$. This is obvious because such a P can be shattered by $\mathcal{H}$ as well, which contradicts VC-dim$(\mathcal{P}, \mathcal{H}) = \lambda$.

Problem 4. Denote by $\mathcal{X} = \mathbb{R}^d$ (where d is an integer) the instance space and by $\mathcal{Y} = \{-1, 1\}$ the label space. Recall that a classifier is a function $h : \mathcal{X} \to \mathcal{Y}$. Given a classifier h, define its *complement* as the function $\bar{h} : \mathcal{X} \to \mathcal{Y}$ which, given an instance $x \in \mathcal{X}$, outputs 1 if $h(x) = -1$, or -1 otherwise. Let $\mathcal{H}$ be a set of classifiers. Define another set of classifiers as follows: $\bar{\mathcal{H}} = \{\bar{h} \mid h \in \mathcal{H}\}$. Prove: $(\mathcal{X}, \mathcal{H})$ and $(\mathcal{X}, \bar{\mathcal{H}})$ have the same VC dimension.

Solution. It suffices to prove that $\mathcal{H}$ can shatter a set $S \subseteq \mathcal{X}$ if and only of $\bar{\mathcal{H}}$ can shatter S. Due to symmetry, it suffices to prove that if $\mathcal{H}$ can shatter S, so can $\bar{\mathcal{H}}$. Consider an arbitrary subset $T \subseteq S$. We will show that $\bar{\mathcal{H}}$ has a function g such that $g(x) = 1$ for every $x \in T$, and $g(x) = -1$ for every $x \in S \setminus T$. This will imply that $\bar{\mathcal{H}}$ can shatter S.

Because $\mathcal{H}$ can shatter S, there must exist a function $h \in H$ such that $h(x) = 1$ for every $x \in S \setminus T$ and $h(x) = -1$ for every $x \in T$. Therefore, $\bar{h}$ is the function g we are looking for.

Problem 5*. In this problem, we will see that deciding *whether* a set of points is linearly separable can be cast as an instance of linear programming.

In the *linear programming* (LP) problem, we are given n constraints of the form:

$$\boldsymbol{\alpha}_i \cdot \boldsymbol{x} \geq 0$$

where $i \in [1, n]$, $\boldsymbol{\alpha}_i$ is a constant d-dimensional vector (i.e., $\boldsymbol{\alpha}_i$ is explicitly given), and $\boldsymbol{x}$ is a d-dimensional vector we search for. Let $\boldsymbol{\beta}$ be another constant d-dimensional vector. Denote by S the set of vectors $\boldsymbol{x}$ satisfying all the n constraints. The objective is to

- either find the best $\boldsymbol{x} \in S$ that maximizes the *objective function* $\boldsymbol{\beta} \cdot \boldsymbol{x}$ — in this case we say that the LP instance is *feasible*;
- or declare that S is empty — in this case we say that the instance is *infeasible*.

Suppose that we have an algorithm $\mathcal{A}$ for solving LP in at most $f(n, d)$ time. Let P be a set of n points in $\mathbb{R}^d$, each given a label that is either 1 or -1. Explain how to use $\mathcal{A}$ to decide in $O(nd) + f(n, d+1)$ time whether P is linearly separable, i.e., whether there exists a vector $\boldsymbol{w}$ such that:

- $\boldsymbol{w} \cdot \boldsymbol{p} > 0$ for each $\boldsymbol{p} \in P$ of label 1;
- $\boldsymbol{w} \cdot \boldsymbol{p} < 0$ for each $\boldsymbol{p} \in P$ of label -1.

Note that the inequalities in the above two bullets are strict, while the inequality in LP involves equality.

Solution. Construct an instance of $(d + 1)$-dimensional LP as follows. For each $p \in P$ with label 1, create a constraint

$$\boldsymbol{p} \cdot \boldsymbol{x} \geq t$$

and for each point $p \in P$ with label -1, create:

$$\boldsymbol{p} \cdot \boldsymbol{x} \leq -t$$

We want to find $\boldsymbol{x}$ and t to satisfy all the n constraints, and in the meantime, maximize t.

To see that this is indeed a $(d+1)$-dimensional LP, define $\boldsymbol{y}$ as the $(d+1)$-dimensional vector that concatenates $\boldsymbol{x}$ and $-t$, namely, the first d components of $\boldsymbol{y}$ constitute x, and the last component of $\boldsymbol{y}$ is $-t$. Accordingly, for each point $\boldsymbol{p} \in P$ of label 1, define $\boldsymbol{p'}$ as the concatenation of $\boldsymbol{p}$ and 1; for each point $\boldsymbol{p} \in P$ of label -1, define $\boldsymbol{p'}$ as the concatenation of $\boldsymbol{p}$ and -1. Then, the constraint of a label-1 point p can be rewritten as

$$\boldsymbol{p'} \cdot \boldsymbol{y} \geq 0$$

while that of a label-(-1) point p as

$$\boldsymbol{p'} \cdot \boldsymbol{y} \leq 0.$$

The objective is to maximize $(0, ..., 0, -1) \cdot \boldsymbol{y} = t$.

The above LP instance can be constructed in $O(nd)$ time. We now deploy the algorithm $\mathcal{A}$ to solve the instance in $f(n, d + 1)$ time. Let t^* be the returned value for the objective function (note that the instance is always feasible). If $t > 0$, we claim that P is linearly separable; otherwise, we claim that P is not.