Problem 1. A rectangular classifier $M$ in $\mathbb{R}^2$ is specified by an axis-parallel rectangle $r = [x_1, x_2] \times [y_1, y_2]$. Given a point $p \in \mathbb{R}^2$, $M(p)$ equals 1 if $p$ is covered by $r$, or $-1$ otherwise. Give a set of 4 points $\mathbb{R}^2$ that can be shattered by $M$.

Problem 2. A rectangular classifier $M$ in $\mathbb{R}^2$ is specified by an axis-parallel rectangle $r = [x_1, x_2] \times [y_1, y_2]$. Given a point $p \in \mathbb{R}^2$, $M(p)$ equals 1 if $p$ is covered by $r$, or $-1$ otherwise. Prove: there does not exist any set of 5 points in $\mathbb{R}^2$ that can be shattered by $M$.

Problem 3. Let $P$ be a set of points in $\mathbb{R}^d$, and $R$ the maximum distance from the origin to a point in $P$. Construct an alternative dataset $P'$ as follows: for each point $p \in P$, add to $P'$ a point $p'$ satisfying $p'[i] = p[i]/R$ for each $i \in [1, d]$. Prove:

- The maximum distance from the origin to a point in $P'$ is 1.
- If there exists a canonical linear classifier $M$ on $P$ that has margin $\gamma$, there exists a canonical linear classifier $M'$ on $P'$ that has margin $\gamma/R$.

Problem 4*. 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, each of which has the form:

$$\alpha_i \cdot \vec{x} \geq 0$$

where $i$ ranges from 1 to $n$, $\alpha_i$ is a constant $d$-dimensional vector (i.e., $\alpha_i$ is explicitly given), and $\vec{x}$ is a $d$-dimensional vector we search for. Let $\vec{\beta}$ be another constant $d$-dimensional vector. Let $S$ be the set of vectors $\vec{x}$ that satisfy all the $n$ constraints. The objective is to

- either find the best $\vec{x}$ in $S$ that maximizes the objective function $\vec{\beta} \cdot \vec{x}$ — 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 $A$ for solving the LP problem in at most $f(n, d)$ time. Let $P$ be a set of $n$ points in $\mathbb{R}^d$, each of which is associated with a label that is either 1 or $-1$. Explain how to use $A$ to decide in $O(nd) + f(n + 1, d + 1)$ time whether $P$ is linearly separable, i.e., whether there exists a vector $\vec{c}$ such that:

- For each label-1 point $p \in P$, it holds that $\vec{c} \cdot \vec{p} > 0$;
- For each label-$(-1)$ point $p \in P$, it holds that $\vec{c} \cdot \vec{p} < 0$.

Note: The inequalities in the above two bullets are strict, while the inequality in each constraint of LP involves equality.