MathTools HW 11

1. **Vertices and extremal points.**

   Definition: Let P be a polytope. We say that a point $x \in P$ is **extremal** if for every vector $u \neq 0$, either $x + u \notin P$ or $x - u \notin P$ (or both).

   (a) Show that if $x \in P$ is a vertex, then it is extremal.

   (b) Let $P = \{y \, : \, Ay \leq b\}$ be a polyhedron in $\mathbb{R}^n$. Denote by $a_1, \ldots, a_m$ the rows of A, so that in other words $P = \bigcap_{i=1}^{m} \{y \, : \, a_i^\top y \leq b_i\}$. For a point $x \in P$, denote by $I(x) = \{i \in [m] \, : \, a_i^\top x = b_i\}$, the inequalities on which x is tight.

   Prove that $x \in P$ is a vertex **if and only if** $\operatorname{span}\{a_i \, : \, i \in I(x)\} = \mathbb{R}^n$.

   (c) Suppose that $x \in P$ is an extremal point. Prove that it is a vertex.

   Remark: An extremal point is in some ways a generalization of the notion of a vertex, that applies to convex sets in general – for polyhedra these two notions coincide, as you have just proved. The celebrated Krein-Milman theorem says that every compact convex body is the convex hull of its extremal points (this holds in a *very* general setting) – just as a polytope is the convex hull of its vertices.

2. **VC dimension of linear separators.** Let $\mathcal{H}$ be a set of functions from $\mathbb{R}^d$ to $\{\pm 1\}$; that is, every $f \in \mathbb{R}^d$ is a *labeling*, assigning each point $x \in \mathbb{R}^d$ either the label[1] $f(x) = +1$ or the label $f(x) = -1$. We say that a set $S = \{x_1, \ldots, x_k\}$ is *shattered* if $\mathcal{H}$ induces on S all the possible labelings: that is, for every labeling $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k) \in \{\pm 1\}^k$, there is a function $f \in \mathcal{H}$ such that $f(x_i) = \varepsilon_i$ for all $i = 1, \ldots, k$. The **VC dimension** of $\mathcal{H}$ is the size of the largest shattered set.

   Denote by $\mathcal{H}_{lin}$ the class of **linear separators**: functions of the form $h_{w,b}(x) = \operatorname{sgn}(w^\top x + b)$ for any $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$. Here $\operatorname{sgn}(\alpha) \in \{\pm 1\}$ is the sign of $\alpha \in \mathbb{R}$, where we define $\operatorname{sgn}(0) = +1$.

   Prove that $\operatorname{VCdim}(\mathcal{H}_{lin}) = d + 1$.

   Hint: It might be useful to use Radon's theorem.

   Remark: The VC dimension is an extremely important notion in machine learning. I refer you to e.g. wikipedia for more information (that is, if you haven't taken a course in ML yet).

3. **Helly's theorem.** Recall the following theorem from the recitation:

   Theorem 1 (Helly) *Let $A_1, \ldots, A_m \subset \mathbb{R}^d$ be convex sets where $m \geq d + 1$. Suppose that the intersection of every $d + 1$ sets is non-empty, in other words, $A_{i_1} \cap \ldots \cap A_{i_{d+1}} \neq \emptyset$ for all $i_1, \ldots, i_{d+1} \in [m]$. Then $\bigcap_{i=1}^{m} A_i \neq \emptyset$.*

[1] For example, x represents a picture (RGB values for every pixel), and f assigns $f(x) = +1$ if and only if it contains a cat.

Prove Helly's theorem. You may follow these steps:

(a) Explain why it suffices to show the following: Let $m \geq d+2$, and let $A_1, \ldots, A_m \subset \mathbb{R}^d$ be any convex sets such that the intersection of every $m-1$ sets is non-empty. Then the intersection of all m sets is non-empty.

(b) We now prove (a). For every j, let $x_j \in \bigcap_{i \in [m] : i \neq j} A_i$; that is, x_j belongs to all sets A_i except for, possibly, $i = j$. Apply Radon's theorem on $S = \{x_1, \ldots, x_m\}$ to find a non-trivial partition $(T, S \setminus T)$ such that $\mathrm{conv}(T) \cap \mathrm{conv}(S \setminus T) \neq \emptyset$. What can you say about any point in this intersection?

4. **Fractional relaxation of a hard problem.** Let $G = (V, E)$ be a graph. Recall that a *triangle* in G is a triplet $a, b, c \in V$ such that $ab, bc, ac \in E$. We would like to remove the **least** amount of edges from G so to make it triangle free.

We can write this an an ILP (linear program with integer constraints). Denote by $w : E \to \{0,1\}$ so to signify which edges we delete. From every triangle, we need to erase at least one edge; a different way of writing this is that $w(ab) + w(bc) + w(ac) \geq 1$ for all triangles abc. The goal is to minimize $\sum_{e \in E} w(e)$ over mappings $w : E \to \{0,1\}$ satisfying this constraint. Let m be this minimum.

(a) Consider the following fractional relaxation of this ILP: instead of the hard constraint $w(e) \in \{0,1\}$, we allow for $w(e) \in [0,1]$; that is, we minimize over mappings $w : E \to [0,1]$. The resulting optimization problem is clearly an LP.

Let m^* be the value of the resulting LP; prove that $m^* \leq m$.

(b) Describe an efficient algorithm that takes G and removes at most $3m$ edges so to make it triangle free.

Hint: "Round" the solution of the fractional relaxation appropriately.