Assignment 1

1. Lower Bounds on Approximate Caratheodory (10 points)

We will show that the Approximate Caratheory's theorem can not be improved in high dimensions. For any dimension d, construct a set $X \subseteq \mathbb{R}^d$ contained in $\mathbb{B}_2 = \{x : \|x\|_2 \leq 1\}$ and a point y in the convex hull of X so that for all k we have

$$\min_{\substack{\{x_1,\dots x_k\} \in X}} \min_{\substack{\alpha_i \geq 0 \\ \sum_{i=1}^k \alpha_i = 1}} \left\| y - \sum_{i=1}^n \alpha_i x_i \right\|_2 \geq \sqrt{\frac{1}{k} - \frac{1}{d}}.$$

2. Pairwise Independence (10 points)

The purpose of this exercise is to demonstrate a specific sense in which pairwise independence is a much weaker condition than full independence. Concretely, given access to n independent Rademacher variables, provide a deterministic recipe to construct at least e^{cn} pairwise independent Rademacher variables for any $c > 0$ of your choice.

3. Random MaxCut (10 points)

Consider a random graph on n vertices where each pair of vertices has a edge with probability $1/2$, independently of other pairs. Every subset of vertices is associated with a cut, namely, edges with exactly one endpoint in a subset; the size of the cut is the number of such edges. Prove that with probability at least 0.999, the maximum cut of such a graph is within $f(n) \pm g(n)$. Try to find the best such $f(n)$ and $g(n)$, but do not worry about the leading constant in $g(n)$.

4. VC Dimension Examples (10 points)

Calculate the VC dimensions of:

1. The set of halfspaces (i.e., $\mathbf{1}_{a^\top x + b \geq 0 : a \in \mathbb{R}^n, b \in \mathbb{R}}$) in n-dimensions
2. The set of n-dimensional spheres.

Lastly, prove for any $\mathcal{H}_1$ and $\mathcal{H}_2$ defined over the same feature space that

$$\mathrm{VC}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq \mathrm{VC}(\mathcal{H}_1) + \mathrm{VC}\mathcal{H}_2) + 1.$$

Hint: Radon's theorem can be useful for the first part.

5. Sparse Predictors (10 points)

Consider $\mathcal{X} = \mathbb{R}^n$, and $\mathcal{H}_k = \{\mathbf{1}_{a^\top x + b \geq 0} : a \in \mathbb{R}^n, b \in \mathbb{R}, \|a\|_0 \leq k\}$ be the class of linear predictors that depend on at most k coordinates. Prove that:

1. $\mathrm{VC}(\mathcal{H}_1)$ is at least $\Omega(\log n)$.

2. $\mathrm{VC}(\mathcal{H}_k)$ is at most $O\!\left(k \log \frac{n}{k}\right)$.

6. Learning by Asking is Faster (10 points)

The point of this exercise to show that if you are allowed to ask for the correct labels for points of your choice, you can learn with far less labeling/supervision. Concretely, construct a hypothesis class $\mathcal{H} \subseteq \{0,1\}^{[0,1]}$ such that both conditions below are satisfied simultaneously:

- $\mathcal{H}$ is (realizable) PAC learnable with sample complexity $\Omega(\log(1/\delta)/\varepsilon)$.

- Consider a model where the learner can (A) *query* the correct label for any given feature vector, and (B) additionally draw an infinite number of samples from the marginal distribution of $\mathcal{D}$ on the feature space (without labels). It must be that, for any distribution $\mathcal{D}$, as long as $\mathrm{err}_{\mathcal{D}}(h) = 0$ for some $h \in \mathcal{H}$, a learning algorithm can make $O(\log(1/\varepsilon))$ such *queries* to produce a $h_{\mathcal{A}} \in \mathcal{H}$ such that $\mathrm{err}_{\mathcal{D}}(h_{\mathcal{A}}) \leq \varepsilon$.