Homework 3, due Tuesday 11/14

1. In the agnostic PAC learning model, it is not assumed that the distribution \( \mathcal{D} \) is separable by \( \mathcal{H} \); that is, there may not exist \( h^* \in \mathcal{H} \) with \( \text{err}_\mathcal{D}(h^*) = 0 \). In this case, the goal is to return \( h \in \mathcal{H} \) such that

\[
\text{err}_\mathcal{D}(h) \leq \epsilon + \inf_{h^* \in \mathcal{H}} \text{err}_\mathcal{D}(h^*)
\]

(everything else remains the same as in the definition of PAC learning).

Show that the class of axis-parallel rectangles in \( \mathbb{R}^2 \) is efficiently learnable in the agnostic PAC model.

2. Show that \( \text{VC}(\mathcal{H}_1 \cup \mathcal{H}_2) = O(\max(\text{VC}(\mathcal{H}_1), \text{VC}(\mathcal{H}_2))) \).

3. What upper and lower bounds can you give on the VC dimension of Boolean conjunctions over \( \{0,1\}^n \)?

4. In this problem, we will get bounds on the VC dimension of the class of balls in \( \mathbb{R}^d \), that is, \( \mathcal{B} = \{ B_{\mu,r} : \mu \in \mathbb{R}^d, r > 0 \} \) where

\[
B_{\mu,r}(x) = \begin{cases} 1 & \text{if } \|x - \mu\|^2 \leq r^2 \\ 0 & \text{otherwise} \end{cases}
\]

(a) Consider the mapping \( \phi : \mathbb{R}^d \to \mathbb{R}^{d+1} \) defined by \( \phi(x) = (x, \|x\|^2) \). Show that if \( x_1, \ldots, x_m \) are shattered by \( \mathcal{B} \), then \( \phi(x_1), \ldots, \phi(x_m) \) are shattered by the class of linear separators in \( \mathbb{R}^{d+1} \).

What does this tell us about \( \text{VC}(\mathcal{B}) \)?

(b) Exhibit a set of \( d + 1 \) points in \( \mathbb{R}^d \) that are shattered by \( \mathcal{B} \).

Footnote. In fact, \( \text{VC}(\mathcal{B}) = d + 1 \). One way to arrive at the upper bound is by a classic result in geometry called Radon’s theorem. It says that any \( d + 2 \) points in \( \mathbb{R}^d \) can be partitioned into two disjoint sets \( S, T \) such that the convex hull of \( S \) intersects the convex hull of \( T \). (Try it with four points in \( \mathbb{R}^2 \).) To get the VC result, pick any set of \( d + 2 \) points in \( \mathbb{R}^d \), and let \( S, T \) be the two subsets of the Radon partition. If the points can be shattered, then there exists a ball containing \( S \) but excluding \( T \), and another ball containing \( T \) but excluding \( S \). But this must mean that there is a hyperplane separating \( S \) from \( T \) (do you see why?), which is a contradiction since the convex hulls of \( S \) and \( T \) intersect.

5. An \( \epsilon \)-cover of a metric space \( (X,d) \) is a subset \( C \subset X \) such that for any \( x \in X \), there exists \( y \in C \) with \( d(x,y) \leq \epsilon \).

Consider \( S^{d-1} \), the surface of the unit sphere in \( \mathbb{R}^d \), imbued with the \( l_2 \) (Euclidean) metric. In this problem, we’ll consider two ways of getting an \( \epsilon \)-cover of \( S^{d-1} \).

(a) Consider a maximal set of points \( x_1, \ldots, x_M \in S^{d-1} \) such that the balls \( B(x_i, \epsilon/2) \) are disjoint.

i. Show that \( x_1, \ldots, x_M \) constitute an \( \epsilon \)-cover of \( S^{d-1} \).

ii. By comparing the volume of the balls \( B(x_i, \epsilon/2) \) to that of a larger ball that encloses all of them, show that

\[
M \leq \left(1 + \frac{2}{\epsilon} \right)^d
\]

(b) Here’s an even easier way to construct an \( \epsilon \)-cover: pick \( N \) points at random from the uniform distribution \( \mu \) over \( S^{d-1} \). How large should \( N \) be, so that the points are an \( \epsilon \)-cover with probability at least \( 1/2^N \)? (Hint: Use problem 4. Also, you may assume that \( \epsilon \leq 1/2 \), and that for any \( x \in S^{d-1} \), the spherical cap of radius \( \epsilon \) around \( x \) has probability mass \( \mu(B(x,\epsilon) \cap S^{d-1}) \geq \epsilon^{d-1}/(6\sqrt{d}) \).)