2.1 VC-dimension of Halfspaces

A *halfspace* in $\mathbb{R}^n$ is the set of all points satisfying

$$\{x \in \mathbb{R}^n : w^\intercal x + b > 0\},$$

for some $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ [$w^\intercal x = \langle w, x\rangle$ is the standard inner product]. Let $\mathcal{H}_n$ be the collection of all half-spaces over $\mathbb{R}^n$. Formally, $f_{w,b} \in \mathcal{H}_n$ is defined by

$$f_{w,b}(x) = \operatorname{sgn}(w^\intercal x + b).$$

Show that the VC-dimension of $\mathcal{H}_n$ is $n+1$.

Hint: you may use the following theorem of Radon:

Theorem If S is a set of $n+2$ points in $\mathbb{R}^n$ then S can be partitioned into 2 disjoint sets whose convex hulls intersect.

[For a finite $A \subset \mathbb{R}^n$, its convex hull is defined by

$$\operatorname{conv}(A) = \left\{ x \in \mathbb{R}^n : x = \sum_{a\in A} \lambda_a a \text{ where } \sum_{a\in A} \lambda_a = 1 \text{ and } \lambda_a \geq 0 \right\};$$

in words, it's the set of all points expressible as convex combinations of elements of A. [As an optional exercise, prove Radon's theorem.]

2.2 VC-dimension vs. Number of Parameters

One might be tempted to conjecture that the VC-dimension of a concept class is somehow related to the number of parameters needed to specify a concept. That turns out not to be the case. Consider the concept class $\mathcal{C}$ defined over the instance space $\mathcal{X} = \mathbb{R}$ as follows. For $\alpha \in \mathbb{R}$, define the concept

$$f_\alpha(x) \quad = \quad \operatorname{sgn}(\sin(\alpha x))$$

and let $\mathcal{C} = \{f_\alpha : \alpha \in \mathbb{R}\}$. Observe that concept in $\mathcal{C}$ is parametrized by a single real number, and prove that the VC-dimension of $\mathcal{C}$ is infinite.

2.3 Tightness of VC

Let $\mathcal{C}$ be a concept class over the instance space $\mathcal{X}$. Recall our definition of a *projection*: for any finite $S \subset \mathcal{X}$, the projection of $\mathcal{C}$ onto $S = \{s_1, s_2, \ldots, s_m\}$ is the collection of vectors

$$\mathcal{C}(S) \quad = \quad \{(f(s_1), f(s_2), \ldots, f(s_m)) : f \in \mathcal{C}\} \subseteq \{0,1\}^m.$$

The Sauer-Shelah-Vapnik-Chervonenkis lemma shows that for any concept class $\mathcal{C}$ with VC-dimension d and any $S \subset \mathcal{X}$ with $|S| = m$, we have

$$|\mathcal{C}(S)| \quad \leq \quad \Phi_d(m), \tag{2.3.1}$$

where $\Phi_d(m)$ is defined inductively as $\Phi_d(0) = \Phi_0(m) = 1$ and

$$\Phi_d(m) = \Phi_d(m-1) + \Phi_{d-1}(m-1).$$

Prove that the bound in (2.3.1) is tight: for any d, construct a concept class $\mathcal{C}$ with VC-dim d such that for all sets S of size m, we have $|\mathcal{C}(S)| = \Phi_d(m)$.

2.4 The VC dimension of the primes

For a prime p, define the function $f_p : \mathbb{N} \to \{0, 1\}$ to map $n \in \mathbb{N}$ to 0 if p divides n and to 1 otherwise. Let $P \subset \mathbb{N}$ be some set of prime numbers and define the function class $\mathcal{F}_P = \{f_p : p \in P\}$.

(a) Prove that for finite P, we have $\mathrm{VCdim}(\mathcal{F}_P) = \lfloor \log_2 |P| \rfloor$.

(b) Conclude that when P is infinite (e.g., P is all the primes), $\mathrm{VCdim}(\mathcal{F}_P) = \infty$.

2.5 Noise magnitude for count queries

Let $\bar{x} \in \{0, 1\}^n$ and consider the function $f(\bar{x}) = \sum_{i=1}^n x_i$.

1. We saw that the randomized algorithm $\hat{f}(\bar{x}) = f(\bar{x}) + Y$ where $Y \sim \mathrm{Lap}(1/\epsilon)$ is ϵ-differentially private. Show that for all $\bar{x} \in \{0, 1\}^n$,

$$\Pr\left[\left|\hat{f}(\bar{x}) - f(\bar{x})\right| \geq \frac{\ln(\frac{1}{\delta})}{\epsilon}\right] \leq \delta.$$

2. Prove that for any ϵ-differentially private (approximation) algorithm $\hat{f}$

$$\Pr\left[\left|\hat{f}(\bar{x}) - f(\bar{x})\right| > \frac{\ln(\frac{1-\delta}{\delta})}{2\epsilon}\right] > \delta$$

holds for some $\bar{x} \in \{0, 1\}^n$. The probability is taken over the randomness of the approximation algorithm, $\hat{f}()$.

Hint: consider instances $\bar{x}, \bar{x}'$ that are at Hamming Distance $\frac{\ln(\frac{1-\delta}{\delta})}{\epsilon}$ apart.

2.6 Private learning of *discrete* intervals

Let $\mathcal{C}_d = \{I_{[a,b]}\}_{0 \leq a \leq b < 2^d}$ where a, b are integers and $I_{[a,b]}$ is the indicator function for the discrete interval $[a, b] = \{a, a+1, a+2, \ldots, b\}$. Show that $\mathcal{C} = \{\mathcal{C}_d\}_{d \in \mathbb{N}}$ is privately learnable properly and efficiently, where sample complexity depends linearly on d.

2.7 Private learning of *continuous* intervals

Let $\mathcal{C} = \{I_{[a,b]}\}_{[a,b] \subseteq [0,1]}$ where $I_{[a,b]}$ is the indicator function for the (continuous) interval $[a, b]$. Show that $\mathcal{C}$ is *not* privately and properly learnable.

Hint: Given parameters α, β, ϵ construct a large number of (distribution, concept) pairs for which any hypothesis that exhibits generalization error α on one (distribution, concept) pair fails to do so for all other (distribution, concept) pairs.