Homework 7

Due: Thursday, March 13, 2025 at 12:00pm **(Noon)**

Problem 1: VC Dimension

(20 points)
For any hypothesis class $\mathcal{H}$ on domain $\mathcal{X}$, to show that the VC Dimension of $\mathcal{H}$ is d, you should prove each of the following:

- There exists a set $C \subset \mathcal{X}$ of size d such that $\mathcal{H}$ shatters C. (Recall that a set is a collection of unique elements.)
- There does not exist a set $C' \subset \mathcal{X}$ of size $d+1$ such that $\mathcal{H}$ shatters C'.

1. Compute and prove the VC dimension for the following hypothesis classes:

   a. The class of signed intervals in $\mathbb{R}$, $\mathcal{H} = \{h_{a,b,s} : a \leq b, s \in \{-1, 1\}\}$ where:

   $$h_{a,b,s} = \begin{cases} s & \text{if } x \in [a, b] \\ -s & \text{if } x \notin [a, b] \end{cases}$$

   b. The class of origin-centered spheres in $\mathbb{R}^d$, $\mathcal{H} = \{h_{a,s} : s \in \{-1, 1\}, a \in \mathbb{R}\}$ where:

   $$h_{a,s} = \begin{cases} s & \text{if } x \text{ is within or on the origin centered sphere of radius } a \\ -s & \text{if } x \text{ is outside the origin centered sphere of radius } a \end{cases}$$

2. Consider two hypothesis classes $\mathcal{H}_1$, $\mathcal{H}_2$ such that $\mathcal{H}_1 \subset \mathcal{H}_2$. Prove that the VC Dimension of $\mathcal{H}_2$ is at least as large as the VC Dimension of $\mathcal{H}_1$.

Solution

1. a. The VC Dimension of $\mathcal{H}$ is 3.

   Let $\mathcal{H}$ be the hypothesis class of intervals. For any set of three points C, all labelings of the three points: (s, s, s), $(s, -s, s)$, $(-s, s, s)$, $(-s, -s, s)$ ... can be realized by hypothesis class $\mathcal{H}$ by placing $[a, b]$ to include all continuous points with the same label. We only need to consider these four assignments out of the total eight assignments as they are identical to the other four possible assignments as we can switch our choice of s. Thus $VCDim(\mathcal{H}_i) \geq 3$ as $\mathcal{H}$ shatters C. *Note that it is sufficient to show a concrete example of 3 points and enumerate each possible labeling.*

   Now consider an arbitrary set C' of four points. Since this is a set, we know that each point is distinct. Without loss of generality, we can order these points as $\{x_1, x_2, x_3, x_4\}$ with $x_1 < x_2 < x_3 < x_4$. $\mathcal{H}$ cannot realize the labeling $(s, -s, s, -s)$ assignment to this set. Any hypothesis $h_{a,b,s} \in \mathcal{H}$ that assigns the label s to points x_1, x_3 must also assign s to x_2 since we know that $a \leq x_1 < x_2 < x_3 \leq b$. Thus, there does not exist any set C' of size 4 such that $\mathcal{H}$ shatters C'.

   This implies that $VCDim(\mathcal{H}) = 3$.

b. The VC Dimension of $\mathcal{H}$ is 2.

Let $\mathcal{H}$ be the hypothesis class of origin-centered spheres. For a set C of size 2 with points having different L2 norms (i.e. $\{(1,1),(2,2)\}$), we can choose a value of the radius to realize the potential labelings $(s,-s)$ and (s,s). Similar to above, we only need to consider these two cases as they are equivalent to the other two cases by flipping our choice of s. Thus, we know that $\mathcal{H}$ shatters C and that $VCDim(\mathcal{H}) \geq 2$. *Again, it is sufficient to consider a concrete example.*

Now consider an arbitrary set of three points C'. We can order these points by their L2 norm as some points $\{x_1, x_2, x_3\}$ where $||x_1||_2 \leq ||x_2||_2 \leq ||x_3||_2$. *Note that we must use $\leq$ here as distinct points may have the same distance to the origin.* Now, consider the labeling of points $(s,-s,s)$ or that the closest to the origin has label s, the furthest point also has label s, and the point in between has label $-s$. In this case, there is no such hypothesis $h \in \mathcal{H}_s$ that can realize this labelling as any hypothesis that labels x_1, x_3 as s must also label x_2 as s. Thus, there does not exist any set C' of size 3 such that $\mathcal{H}$ shatters C'. This gives us that $VCDim(\mathcal{H}) < 3$, and combining the inequalities, we have that $VCDim(\mathcal{H}) = 2$.

2. Let $VC(\mathcal{H}_1) = k$. Then, we know that $\exists$ a set C of size k such that $\mathcal{H}_1$ shatters C. This means that for each possible labeling of the points C, $\exists h \in \mathcal{H}_1$ that realizes the labeling of those points. Since $\mathcal{H}_1 \subset \mathcal{H}_2$, this same hypothesis $h \in \mathcal{H}_2$. This implies that $\mathcal{H}_2$ can also shatter C. Thus, $VCDim(\mathcal{H}_2) \geq k = VCDim(\mathcal{H}_1)$.

Problem 2: VC Dimension and PAC Learning

(20 points)

Decision trees can split on data with binary features ($\mathcal{X} = \{0,1\}^d$) or continuous features ($\mathcal{X} = \mathbb{R}^d$). Assume that the nodes of a continuous decision tree have splitting rules that threshold the value of a single feature. *Note that for continuous decision trees, multiple splits can be made on the same feature. For binary decision trees, only a single split can be made on a feature.*

Consider the following hypothesis classes:

$$\mathcal{H}_1 = \{h : h \text{ is a decision tree for data with only binary features}\}$$
$$\mathcal{H}_2 = \{h : h \text{ is a decision tree for data with only continuous features}\}$$

1. Compute and prove the VC dimension of $\mathcal{H}_1$.
2. Show that the VC dimension is infinite for $\mathcal{H}_2$.
3. Is $\mathcal{H}_1$ PAC learnable? How about $\mathcal{H}_2$? Explain.

Solution

1. The VC dimension is 2^d.

With d binary attributes, there are 2^d possible input values. The set of size 2^d with each possible input can be shattered by splitting on every feature to isolate each example to its own leaf. Then, any possible labeling can be realized as we can choose the label assigned to each leaf. This is sufficient for the set to be shattered.

There is no set of size 2^d+1, thus 2^d is the largest size set that can be shattered and $VCDim(\mathcal{H}_1) = 2^d$.

(((**The following reasoning is actually wrong (there's no set of size 2^d+1) and should get partial credit:** Any larger set will have two examples with the same attribute values. The tree must then always predict the same label for these examples so the set cannot be shattered. As a set of size 2^d can be shattered but not a set of size 2^d+1, the VC dimension is 2^d.)))

2. The VC dimension is infinite if for any size m there is a set C that is shattered by $\mathcal{H}_2$. Consider a set C of size m, where all points have a unique value for attribute x_0 and all other attributes are 0. (*This effectively reduces the problem into a single dimension* $\mathbb{R}$). Then, repeatedly choosing nodes that create splits in between each of the points' attribute values of x_0 isolates each example to its own leaf, so $\mathcal{H}_2$ shatters C. This holds for a set C for arbitrary size m, so the VC dimension is infinite.

3. By the Fundamental Theorem of Statistical Learning, a hypothesis class if PAC learnable if and only if it has a finite VC dimension. Therefore, we can conclude that H_1 is PAC learnable but H_2 is not.

Problem 3: Uniform Convergence

(10 points)

In class and in Corollary 4.6 in the textbook, we proved that finite hypothesis classes enjoy uniform convergence, and therefore are agnostically PAC learnable. In those proofs, we assumed that the loss function has a range of $[0, 1]$. Prove that if the range of the loss function is instead $[a, b]$, then the sample complexity satisfies:

$$m_{\mathcal{H}}(\epsilon, \delta) \leq m_{\mathcal{H}}^{\text{UC}}(\epsilon/2, \delta) \leq \left\lceil \frac{2\log(2|\mathcal{H}|/\delta)(b-a)^2}{\epsilon^2} \right\rceil .$$

You may not use the result from in class or Corollary 4.6 as an argument for your proof, but you may (and it is recommended!) use the same proof steps in guiding your approach.

Solution: The first inequality

$$m_{\mathcal{H}}(\epsilon, \delta) \leq m_{\mathcal{H}}^{\text{UC}}(\epsilon/2, \delta)$$

is a restatement of Corollary 4.4. To show the second inequality, we want to prove that for any ϵ, δ

$$\mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \frac{\epsilon}{2}\}) \leq \delta .$$

Using the Union Bound, we have that

$$\mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \frac{\epsilon}{2}\}) \leq \sum_{h \in \mathcal{H}} \mathcal{D}^m(\{S : |L_S(h) - L_D(h)| > \frac{\epsilon}{2}\}) .$$

To use Hoeffding's inequaility, we need to formulate our problem correctly in terms of assumptions and random variables. We can consider our random variable $\theta_i = l(h, z_i)$ or the loss that a hypothesis h achieves on a sample data point z_i. Then, we have that each θ_i is iid because each of the points z_i are sampled iid. Finally, we have that $L_S(h) = \sum_{i=1}^m \frac{1}{m}\theta_i$ and $L_D(h) = \mu$. Now, we can invoke Hoeffding's Inequality to get that:

$$\sum_{h \in \mathcal{H}} \mathcal{D}^m(\{S : |L_S(h) - L_D(h)| > \frac{\epsilon}{2}\}) \leq \sum_{h \in \mathcal{H}} 2\exp(-2m\epsilon^2/(4(b-a)^2))$$

$$= 2|\mathcal{H}|\exp(-2m\epsilon^2/(4(b-a)^2))$$

since each term is identical. Now, we can choose a specific m (to solve for δ) to find our sample complexity. Thus, we take

$$m = \left\lceil \frac{2\log(2|\mathcal{H}|/\delta)(b-a)^2}{\epsilon^2} \right\rceil .$$

Substituting this definition into the previous equations:

$$\begin{aligned}
\mathcal{D}^m(\{S : \exists h \in \mathcal{H}, |L_S(h) - L_{\mathcal{D}}(h)| > \frac{\epsilon}{2}\}) &\leq 2|\mathcal{H}| \exp\left(-2\epsilon^2 \left(\frac{2\log(2|\mathcal{H}|/\delta)(b-a)^2}{4\epsilon^2(b-a)^2}\right)\right) \\
&= 2|\mathcal{H}| \exp\left(\frac{-\log(2|\mathcal{H}|/\delta)(b-a)^2}{(b-a)^2}\right) \\
&= \delta
\end{aligned}$$

Problem 4: Socially Responsible Computing: Data Privacy

(5 points)

In 2012, Minerva High School, a public school in Pittsburgh, PA with nearly 3,000 students, hit a record student dropout rate of nine percent.[1] The school principal and board decided to put the extensive data the school had already collected about its students' behavior to use. These datasets included demographic information, academic performance, disciplinary and attendance records, and teacher statistics (i.e. percent of students failing per class, years of teaching). The school also tracked students' internet use and monitored their movements throughout the campus.

The board members suggested that developments in machine learning could be applied to this information to understand what causes students to drop out so that new incentive structures for teachers and students could be created. They contracted a local data science company, Hephaestats, to provide them with their existing databases and gave them access to new data as it was collected. Given the urgency of the situation, the principal proceeded quickly, without time to notify students and parents of this agreement, nor giving them the opportunity to opt out. They justified that this decision was supported by the school board and fell within the general mandate to promote positive educational outcomes for all.

1. In this case, the decision to adopt AI technologies came from above—a suggestion from the school board, implemented by the principal. Who are the other relevant stakeholders, and how could they have been involved? Should they have been involved in the decision to use Hephaestats?

2. Review the introduction of this section on Data Privacy in the Blueprint for an AI Bill of Rights. According to the blueprint, did the school violate the privacy of its students by sharing their data with Hephaestats? If you were the principal, what would you have done?

Grading Breakdown

The grading breakdown for the assignment is as follows:

Problem 1	40%
Problem 2	40%
Problem 3	15%
Problem 4	5%
Total	100%

[1] Case study by Princeton Dialogues on AI and Ethics licensed under CC Attribution 4.0 International

Handing In

You will turn in your final handin via Gradescope, as detailed in the email sent to the course. If you have questions on how to set up or use Gradescope, ask on Edstem! For this assignment, you should have written answers for Questions 1, 2, 3, and 4.

Anonymous Grading

You need to be graded anonymously, so do not write your name anywhere on your handin.

Obligatory Note on Academic Integrity

Plagiarism — don't do it.

As outlined in the Brown Academic Code, attempting to pass off another's work as your own can result in failing the assignment, failing this course, or even dismissal or expulsion from Brown. More than that, you will be missing out on the goal of your education, which is the cultivation of your own mind, thoughts, and abilities. Please review this course's collaboration policy and if you have any questions, please contact a member of the course staff.