Department of Computer Science
Ashoka University

Introduction to Machine Learning

Assignment 2B

Collaborators: None **Name: Rushil Gupta**

Question 1

Done on the jupyter notebook.

Question 2

Done on the jupyter notebook.

Question 3

Done on the jupyter notebook.

Question 4

1. $h(x) = \mathbb{1}(a < x), a \in \mathbb{R}$

   Claim: The VC dimension of this hypothesis class is 1.

   To shatter 1 point: Given a single point x_1, we can achieve both labelings by setting a appropriately:

   - For $(+)$: Choose $a < x_1$
   - For $(-)$: Choose $a > x_1$

   Impossible to shatter 2 points: For points $x_1 < x_2$, the labeling $(+, -)$ cannot be achieved with any choice of threshold a, as any a that makes x_1 positive (i.e., $a < x_1$) would also make x_2 positive since $x_1 < x_2$. This threshold function can only create a single decision boundary, resulting in all points to the right being positive and all points to the left being negative.

2. $h(x) = \mathbb{1}(a < x < b), a, b \in \mathbb{R}$

   Claim: The VC dimension of this hypothesis class is 2.

   To shatter 2 points: Given points $x_1 < x_2$, all four labelings can be achieved by setting a and b appropriately:

   - For $(+, +)$: Choose $a < x_1$ and $b > x_2$
   - For $(-, -)$: Choose $a > x_2$ or $b < x_1$
   - For $(-, +)$: Choose $a < x_2 < b$ with $a > x_1$
   - For $(+, -)$: Choose $a < x_1 < b < x_2$

   Cannot shatter 3 points: For points $x_1 < x_2 < x_3$, the labeling $(+, -, +)$ is impossible with this hypothesis class. If x_1 and x_3 are within the interval (positive), then x_2 must also be within the interval (positive) since the hypothesis represents a single connected interval. This shows that any arrangement of 3 points on a line cannot be shattered by this hypothesis class.

3. $h(x) = \mathbb{1}(a \sin x > 0), a \in \mathbb{R}$

   Claim: The VC dimension of this hypothesis class is 1.

   To shatter 1 point: For any single point x_1 with $\sin x_1 \neq 0$, both labelings can be achieved:

   - For $(+)$: Choose a with the same sign as $\sin x_1$, making $a \sin x_1 > 0$
   - For $(-)$: Choose $a = 0$ or a with the opposite sign as $\sin x_1$, making $a \sin x_1 \leq 0$

   Impossible to shatter 2 points: For any two points x_1, x_2:

   - If $\sin x_1$ and $\sin x_2$ have opposite signs, the labeling $(+, +)$ is impossible.
   - If they have the same sign, the labelings $(+, -)$ and $(-, +)$ are impossible.

   This is because the parameter a can only control the sign of the hypothesis, giving us at most three distinct functions: positive when $\sin x > 0$ (when $a > 0$), positive when $\sin x < 0$ (when $a < 0$), or identically 0 (when $a = 0$). Therefore, the VC dimension is 1.

4. $h(x) = \mathbb{1}(\sin(x + a) > 0), a \in \mathbb{R}$

   Claim: The VC dimension of this hypothesis class is 2.

   To shatter 2 points: For any two points x_1, x_2 whose distance is less than π (e.g., $x_2 = x_1 + \frac{\pi}{2}$), all four labelings can be achieved by selecting an appropriate shift a:

   - For $(+, +)$: Choose a so that both $x_1 + a$ and $x_2 + a$ lie within a region where sin is positive
   - For $(-, -)$: Choose a so that both $x_1 + a$ and $x_2 + a$ lie within a region where sin is negative
   - For $(+, -)$: Choose a so that $x_1 + a$ is in a positive region and $x_2 + a$ is in a negative region
   - For $(-, +)$: Choose a so that $x_1 + a$ is in a negative region and $x_2 + a$ is in a positive region

   Cannot shatter 3 points: For any three points $x_1 < x_2 < x_3$, the labeling $(+, -, +)$ is impossible. If $\sin(x_1 + a) > 0$ and $\sin(x_3 + a) > 0$, then $x_1 + a$ and $x_3 + a$ must lie in the same positive interval of length π. Since this interval is contiguous, the entire segment between them, including $x_2 + a$, must also lie in this positive region, forcing $h_a(x_2) = 1$. This contradiction shows that no set of three points can be shattered, so the VC dimension is 2.

Question 5

Question 5.1

1. **(a)** Can be shattered.
2. **(b)** Cannot be shattered.
3. **(c)** Can be shattered.

Question 5.2

Since the hypothesis class could shatter a set of 4 points, it has a VC dim of at least 4.