IML Assignment 2B

Aman K. Foujdar

May 12, 2025

Question 4

Let the input domain of a learning problem be X= R. We have to give the the VC dimension for each of the following classes of hypotheses.

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

This is a threshold function on the real line. It classifies all $x > a$ as 1 and all others as 0.

- It can shatter 1 point: For a point x_1, choosing $a < x_1$ gives label 1; choosing $a > x_1$ gives label 0.
- It cannot shatter 2 points: For $x_1 < x_2$, the labeling $\{1, 0\}$ is not realizable by any threshold a.

VC dimension: 1

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

This represents a single open interval.

- It can shatter 2 points: For $x_1 < x_2$, all four labelings $\{00, 01, 10, 11\}$ are possible by selecting appropriate interval (a, b).
- It cannot shatter 3 points: For $x_1 < x_2 < x_3$, the labeling $\{1, 0, 1\}$ is not possible with a single interval.

VC dimension: 2

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

This function depends on the sign of $a \sin x$. The effect of a is to flip the sign of $\sin x$, so the class consists essentially of three functions:

- $1\{\sin x > 0\}$
- $1\{\sin x < 0\}$
- Constant zero function (if $a = 0$)

This gives very limited labelings.

- It cannot shatter 2 arbitrary points: For example, if $\sin x_1 > 0$ and $\sin x_2 < 0$, then labeling $\{1, 1\}$ is not realizable.

VC dimension: 1

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

Here, we shift the sine wave horizontally via a. This allows more flexibility than scaling.

- It can shatter 2 points: By choosing appropriate phase shifts a, we can achieve any labeling on 2 points.
- It cannot shatter 3 points: The sine function has at most two sign changes per period. Thus, for labelings such as $\{1, 0, 1\}$, we cannot find a phase shift a that realizes it.

VC dimension: 2

Hypothesis Class	**Description**	**VC Dimension**
$1\{a < x\}$	Threshold function	1
$1\{a < x < b\}$	Single interval	2
$1\{a \sin x > 0\}$	Scaled sine	1
$1\{\sin(x + a) > 0\}$	Shifted sine	2

Question 5

Suppose we are using semicircles in the 2D plane to classify a collection of 2-dimensional data points. The diameters of these semicircles need not be aligned with the coordinate axes. Each semicircle classifier labels the points **inside** the semicircle as 0, and the points **outside** as 1. Let this collection of classifiers be denoted by $\mathcal{H}$.

(i) Which of the following point sets can $\mathcal{H}$ shatter?

- **Set (a):** It is not possible to shatter 4 points arranged this way using a single semicircle. For example, trying to realize an alternating label pattern like $\{0, 1, 0, 1\}$ around the circle cannot be done since a semicircle defines a single, contiguous 0-labeled region. Therefore, **set (a) cannot be shattered**.
- **Set (b):** Similar to (a), some labelings such as $\{1, 0, 1, 0, 1\}$ cannot be realized by any semicircle. The classifier cannot separate non-contiguous subsets. Thus, **set (b) cannot be shattered**.
- **Set (c):** This class again can't realize all the labelings of these 4 points can be realized by a semicircle. Hence, **set (c) can't be shattered**.

(ii) What does this tell you about the VC dimension of $\mathcal{H}$?

By definition, the VC dimension is the size of the largest set that can be shattered by the hypothesis class. A semicircle can be positioned (rotated and scaled) to include any subset of three non-colinear points — so all 8 labelings are realizable. Hence,$\mathcal{H}$ can shatter 3 points).

Therefore, the VC dimension of $\mathcal{H}$ is 3.

Acknowledgment

I would like to acknowledge that I used ChatGPT(for latex erros and syntax correction), Desmos(for graphs), Prof Juneja's class notes and Andrew Ng notes for reference and help during this work.