Advanced Machine Learning

Alex Preda

Exercise 1

a)Finite hypothesis class $\mathcal{H}$ with $VC_{\mathbf{dim}}(H) = 2024$

The example that I have chosen consists of a hypothesis class $\mathcal{H}$ consisting of all possible binary labelings of a set of 2024 points.

$\mathcal{H}_{\text{finite}} = \{h : \{x_1, x_2, ..., x_{2024}\} \to \{0, 1\}\}$

There are 2^{2024} such functions.

To demonstrate the $VC_{\text{dim}}(H)$ we will find a general pattern.

$VC_{\mathbf{dim}}$ of 1:

Single Point: There 2 possible labelings (0 or 1), and the hypothesis class can represent both. Therefore, it can shatter 1 point.

$VC_{\mathbf{dim}}$ of 2:

Single Point: There 2^2 possible labelings (00, 01, 10 or 11), and the hypothesis class can represent all. Therefore, it can shatter 2 point.

$VC_{\mathbf{dim}}$ of 3:

Single Point: There 2^3 possible labelings (000, 001, 010, 011, 100, 101, 110, 111), and the hypothesis class can represent all. Therefore, it can shatter 3 point.

General Pattern:

For a set of n points, there are 2^n possible binary labelings, since $\mathcal{H}_{\text{finite}}$ contains all the possible binary labelings, for any subset of 2024 points, it can represent all possible labelings. But, if there are 2^{2025} possible labelings, since $\mathcal{H}_{\text{finite}}$ contains only 2^{2024} hypotheses, it cannot represent all specific points and it cannot shatter 2025 points.

In conclusion, the finite hypothesis class $\mathcal{H}_{\text{finite}}$ with 2^{2024} hypothesis has a $\underline{VC_{\text{dim}} = 2024}$ because it can shatter any set of 2024 points, but it cannot shatter any set of 2025 points.

b)Infinite hypothesis class $\mathcal{H}$ with $VC_{\textbf{dim}}(\mathcal{H}) = 2024$

The example that I have chosen is a set of polynomials of degree up to 2023. To demonstrate that this example has the $VC_{\text{dim}}(\mathcal{H}) = 2024$, I did the following approach:

Polynomials of degree 1:

$\mathcal{H}_1 = \{h(x) = ax + b | a, b \in \mathbb{R}\}$

$VC_{\text{dim}}(\mathcal{H}_1) = 2$ because a linear polynomial can shatter any set of 2 points but cannot shatter any set of 3 points.

Polynomials of degree 2:

$\mathcal{H}_2 = \{h(x) = ax^2 + bx + c | a, b, c \in \mathbb{R}\}$

$VC_{\text{dim}}(\mathcal{H}_2) = 3$ because a quadratic polynomial can shatter any set of 3 points but cannot shatter any set of 4 points.

Polynomials of degree 3:

$\mathcal{H}_3 = \{h(x) = ax^3 + bx^2 + cx + d | a, b, c, d \in \mathbb{R}\}$

$VC_{\text{dim}}(\mathcal{H}_2) = 4$ because a cubic polynomial can shatter any set of 4 points but cannot shatter any set of 5 points.

Generalization:

For a polynomial of degree n, the hypothesis class can be written as:

$\mathcal{H}_n = \{h(x) = a_n x^n + a_{\text{n-1}} x^{\text{n-1}} + ... + a_1 x + a_0 | a_i \in \mathbb{R}, 0 \leq i \leq d\}$

$VC_{\text{dim}}(\mathcal{H}_n) = n + 1$

To match our VCdim, we choose the following hypothesis:

$\mathcal{H}_{2023} = \{h(x) = a_2 023 x^2 023 + a_{\text{n-1}} x^{\text{n-1}} + ... + a_1 x + a_0 | a_{0...2023} \in \mathbb{R}\}$

which has the $\underline{VC_{\text{dim}}(\mathcal{H}2023) = 2024}$

Exercise 3

To compute VCdim($\mathcal{H}$) we will first determine the VCdim of $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$.

VC Dimension of $\mathcal{H}_1$:
$\mathcal{H}_1$ represents a thresholded function $h_a(x) = 1$ if $X \leq 1$ and 0 otherwise. The $\underline{VC_{\text{dim}}(\mathcal{H}_1) = 1}$ because:

- For one point we can always find a point that shatters it
- For two points we cannot find a threshold a that will shattter all possile combinations (00, 01, 10, 11).

This was also demonstrated in Lecture 6 page 26.

VC Dimension of $\mathcal{H}_2$:
Similarly, $\mathcal{H}_2$ can shatter any single point set $\{x_1\}$, but cannot shatter any set of two points where one point is labeled 1 and the other 0 in the order $x_1 < x_2$, then $\underline{VC\text{dim}(\mathcal{H}_2) = 1}$.

VC Dimension of $\mathcal{H}_3$:
$\mathcal{H}_3$ represents a interval function $h_{\text{c, d}}(x) = 1$ if $c \leq x \leq d$ and 0 otherwise. The $\underline{VC_{\text{dim}}(\mathcal{H}_3) = 2}$ because:

- For two points we can always find intervals $[c, d]$ that can shatter them
- For three points, we cannot find intervals $[c, d]$ to label all eight possible combinations (000, 001, 010, 011, 100, 101, 110, 111).

This was also demonstrated in Lecture 6 page 27.

VC Dimension of H:
Since $\mathcal{H}_3$ has the highest $VC_{\text{dim}} = 2$ among $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$, and considering that any point shatterable by $\mathcal{H}_1$ or $\mathcal{H}_2$ is also shatterable by $\mathcal{H}_3$, the VC dimension of $\mathcal{H}$ is the maximum VC dimension of its components. Therefore, the $\underline{\text{VC dimension of } \mathcal{H} \text{ is } 2}$.

Exercise 4

a)

First, we know that the $VC_{\text{dim}}(\mathcal{H})(x) = 3$ based on the demonstration on Seminar 3, Exercise 4.

We also know that the $\tau_{\mathcal{H}}(m)$ for $m \leq 3$ is 2^m based on Lecture 7.

We need to find a general formula that matches the shattering coefficient for $m \geq 3$, to do that we know that all the labeling for $m \leq 2$ are unique, but for $m \geq 3$ there are some labels that coincide and we have to remove them. This general formula should work on the general case: $C = \{c_1, c_2, ..., c_m\}, m > d$.

From Lecture 7 we know that for a interval, there are $m+1$ possible choices because a or b can be in the following case: $a_1 < c_1 < a_2 < c_2 < ... < c_m < a_{m+1}$

To generalize the formula we take $L \in [1, d]$ as the length of the interval and find for each L how many different intervals we find.

$L(0)$: $a < b < c_1$ so there is only 1 solution $L(0) = 1$.

$L(1)$: The first interval is $a < c_1 < b < c_2$ and the last interval is $c_{m-1} < a < c_m < b$, so $L(1) = m$.

$L(m)$: This is the last case where $a < c_1 < c_m < b$.

The general formula for adding all Ls is $\frac{m(m+1)}{2} * 2 + 2 = m(m+1) + 2$. We multiply by 2 because we have singed intervals and it can be 1 or -1 and +2 comes from $L(0)$ and $L(m)$.

We can observe that we have some duplicates that arise from assigning both values for intervals where one endpoints is less that the other because, for a given set of labels, the elements at the end of the list will pe part of the interval on the left side, and the remaining elements will form another interval on the right. This results in overlaps for intervals of length L minus some m.

To ilustrate such example, if we compute this for $d = 5$ with $L = 1$, it has this corresponding sequence: $[1, -1, -1, -1]$ and $[-1, 1, 1, 1$ and for $L = 4$, it has this corresponding sequence: $[-1, 1, 1, 1$ and $[1, -1, -1, -1]$. We can observe that some of these labeling pairs need to be remove so to do that we can remove the first interval out of every length L. By doing this we get the following shattering coefficient:

$\tau_{\mathcal{H}}(m) = m^2 - m + 2$

b)

Based on the $\tau_{\mathcal{H}}(m)$ calculated on previous exercise, we can compute the upper bound using the Sauer's lemma.

We know that $VC_{dim} = 3 => d = 3$ then the upper bound is:

$C_m^0 + C_m^1 + C_m^2 + C_m^3 =$
$1 + \frac{m!}{1!(m-1)!} + \frac{m!}{2!(m-2)!} + \frac{m!}{3!(m-3)!} =$

$1 + m + \frac{m(m-1)}{2} + \frac{m(m-1)(m-2)}{6} =$
$\frac{6+6m+3m^2-3m+m^3-3m^2+2m}{6} =$
$\frac{m^3+5m+6}{6}$

Based on this upper bound, we can compare it with the shuttering coefficient:

$\frac{m^3+5m+6}{6} \geq m^2 - m + 2$
$m^3 + 5m + 6 \geq 6m^2 - 6m + 12$
$m^3 - 6m^2 + 11m - 6 \geq 0$
$m^3 - 5m^2 + 6m - (m^2 - 5m + 6) \geq 0$
$m(m^2 - 3m - 2m + 6) - (m^2 - 3m - 2m + 6) \leq 0$
$(m-1)(m(m-3) - 2(m-3)) \geq 0$
$(m-1)(m-2)(m-3) \geq 0$

We know that $m \geq 0$, then our formula is equal to 0 for $m \in \{1, 2, 3\}$ and ≥ 0 for others.

c)

Here's a rephrased version of the text:

We can see that $\pi_H(m)$ increases exponentially and matches the upper bound for all values $m \leq d$. Therefore, we can select a hypothesis class H where the shattering coefficient reaches the same general upper bound, effectively making it infinite. An example of such a class is in Lecture 7:

$$H_{\sin} = \{h_\theta : \mathbb{R} \to \{0, 1\} \mid h_\theta(x) = [\sin(\theta x)], \theta \in \mathbb{R}\}$$

Exercise 5

The function $h_\theta(x)$ takes the value 1 if x is in the interval $[\theta, \theta+1]$ or in the interval $[\theta+2, \infty]$. Now let's analyze how many points can be shattered by this hypothesys.

One point shattering:
Any single point x_1 can or cannot be included by choosing θ. For example:

- To include x_1, we can choose θ such that $x_1 \in [\theta, \theta+1]$ or $x_1 \in [\theta+2, \infty]$.
- To exclude x_1, we can choose θ such that $x_1 \notin [\theta, \theta+1]$ and $x_1 \notin [\theta+2, \infty]$

Considering this, we can tell that $VC_{\text{dim}} \geq 1$.

Two point shattering:
Having two point x_1 and x_2, we will check if we can achieve all possible labeling 00, 01, 10, 11:

- For 00: θ such that both points are not in $[\theta, \theta+1] \cup [\theta+2, \infty)$.
- For 01: θ such that $x_1 < [\theta, \theta+1] \cup [\theta+2, \infty)$ and $x_2 \in [\theta, \theta+1]$.
- For 10: We need θ such that $x_1 \in [\theta, \theta+1]$, then $x_2 \in (\theta+1, \theta+2)$ if $1 < x_2 - x_1 < 2$.
- For 11: θ such that both points are in $[\theta, \theta+1] \cup [\theta+2, \infty)$.

Considering this, we can tell that $VC_{\text{dim}} \geq 2$.

Three point shattering:
Having three points x_1, x_2, x_3, we will check if we can achieve all possible labeling 000, 001, 010, 011, 100, 101, 110, 111. To do that we choose $x_1 = 3, x_2 = 4, x_3 = 4.2$.

- for 000: θ such that none of the points are in $[\theta, \theta+1] \cup [\theta+2, \infty)$.
- for 001: θ such that $x_1, x_2 \notin [\theta, \theta+1] \cup [\theta+2, \infty)$ and $x_3 \in [\theta, \theta+1]$.
- for 010: θ such that $x_1 < \theta$, $x_2 \in [\theta, \theta+1]$ and $x_3 \in (\theta+1, \theta+2)$.
- for 011: θ such that $x_1 \in (\theta+1, \theta+2)$ and $x_2, x_3 \in [\theta+2, \infty)$.
- for 100: θ such that $x_1 \in (\theta, \theta+1)$ and $x_2, x_3 \in (\theta+1, \theta+2)$.
- for 101: θ such that $x_1 \in [\theta, \theta+1]$ and $x_2 \in (\theta+1, \theta+2)$ and $x_3 \in [\theta+2, \infty)$.
- for 110: θ such that $x_1, x_2 \in [\theta, \theta+1]$ and $x_3 \in [\theta+2, \infty)$.
- for 111: θ such that $x_1, x_2, x_3 \in [\theta, \infty)$.

Considering this, we can tell that $VC_{\text{dim}} \geq 3$.

Four point shattering:

We can tell for the label example 0101, that it is impossible to label as that because of the interval restrictions, so $\underline{HC_{\text{dim}} = 3}$