Advanced Machine Learning Seminar 4

Exercise 1 $\mathcal{H}^d_{\text{mcon}}$ = class of monotone Boolean conjunctions over $\{0,1\}^d$.

$$\mathcal{H}^d_{\text{mcon}} = \left\{ h\colon \{0,1\}^d \to \{0,1\},\, h_{(x_1,x_2,\dots,x_d)} = \bigwedge_{i=1}^{d} l(x_i) \right\} \cup \{ \underset{\substack{h^-(x_1,\dots,x_d)=0 \\ \text{always}}}{\underset{\downarrow}{h^-}} \}$$

$$l(x_i) \in \{\underset{\substack{\text{positive}\\ \text{literal}}}{\underset{\downarrow}{x_i}}, \underset{\substack{\text{missing}\\ \text{literal}}}{\underset{\downarrow}{1}}\}$$

So $|\mathcal{H}^d_{\text{mcon}}| = 2^d + 1$.
Examples:

$$d = 2 \qquad \mathcal{H}^2_{\text{mcon}} = \{ \underset{h^-}{\underset{\downarrow}{0}}, \underset{h_{empty}}{\underset{\downarrow}{1}}, x_1, x_2, x_1 \wedge x_2\}$$

$$d = 3 \qquad \mathcal{H}^3_{\text{mcon}} = \{0, 1, x_1, x_2, x_3, x_1 \wedge x_2, x_1 \wedge x_3, x_2 \wedge x_3, x_1 \wedge x_2 \wedge x_3\}$$

We need to show that $VC\dim(\mathcal{H}^d_{\text{mcon}}) = d$.

Proof. We know that $|\mathcal{H}^d_{\text{mcon}}| = 2^d + 1$, so $VC\dim(\mathcal{H}^d_{\text{mcon}}) \le \lfloor \log_2(|\mathcal{H}^d_{\text{mcon}}|)\rfloor$, which in turn means

$$VC\dim(\mathcal{H}^d_{\text{mcon}}) \le \lfloor \log_2(2^d + 1)\rfloor = d$$

We only need to find a set $C \subseteq \{0,1\}^d$ with d points that is shattered by $\mathcal{H}^d_{\text{mcon}}$.

Usually, taking $C = \{e_1, e_2, \dots, e_d\}, e_i = (0, \dots, 0, 1, 0, \dots, 0)$ works, but not for this $\mathcal{H} = \mathcal{H}^d_{\text{mcon}}$. You cannot have a conjunction that will have $h(e_1) = h(e_2) = 1$ and $h(e_3) = \dots = h(e_d) = 0$.

Instead, we choose $C = \{(0,1,1,\dots,1), (1,0,1,1,\dots,1), \dots, (1,1,\dots,\underset{i}{0},1,1)\}$ set of vectors of the form $c_i = (1,1,\dots,1) - e_i, i = \overline{1,d}$.

We want to show that, for each possible labeling $(y_1, y_2, \dots, y_d)$ of the points $c_i = (1,1,\dots,1) - e_i$, there exists a function $h \in \mathcal{H}^d_{\text{mcon}}$ such that $h(c_i) = y_i, \forall i = \overline{1,d}$.

Consider $(y_1, y_2, \dots, y_d)$ a labeling and take $\mathcal{J} = \{j \mid y_j = 1\}$.

If $\mathcal{J} = \emptyset$, then h^- realizes the labeling $(0, 0, \dots, 0)$.

If $\mathcal{J} = \{1, 2, \dots, d\}$, then $h_{empty} = 1$ (all literals are missing) realizes the labeling $(1, 1, \dots, 1)$.

If $1 \le |\mathcal{J}| \le d - 1$, then consider $h_{\mathcal{J}}(x_1, x_2, \dots, x_d) = \bigwedge_{j \notin \mathcal{J}} x_j$.

For example, if $d = 4$ and $\mathcal{J} = \{2, 3\}$, $h_{\mathcal{J}}(x_1, x_2, x_3, x_4) = x_1 \wedge x_4$:

$$\begin{aligned} h_{\mathcal{J}}(c_1) &= h_{\mathcal{J}}(0,1,1,1) = 0 \\ h_{\mathcal{J}}(c_2) &= h_{\mathcal{J}}(1,0,1,1) = 1 \\ h_{\mathcal{J}}(c_3) &= h_{\mathcal{J}}(1,1,0,1) = 1 \\ h_{\mathcal{J}}(c_4) &= h_{\mathcal{J}}(1,1,1,0) = 0 \end{aligned}$$

We have that $h_{\mathcal{J}}(c_i) = 1$ if $i \in \mathcal{J}$ and $h_{\mathcal{J}}(c_i) = 0$ if $i \notin \mathcal{J}$.

For all indices $i \in \mathcal{J}$, c_i will have value 0 on the position i and 1 in rest, but variable x_i is not considered in the conjunction. So $h_{\mathcal{J}}(c_i) = 1$.

For all indices $i \notin \mathcal{J}$, c_i will have value 0 on the position i and, because the conjunction contains the literal x_i, then we have that $h_{\mathcal{J}}(c_i) = 0$. $\square$

Exercise 2 $\mathcal{X} = \{0,1\}^n$

$$\mathcal{H}_{\text{n-parity}} = \left\{ h_I \mid I \subseteq \{1,2,\dots,n\},\ h_I(x_1,x_2,\dots,x_n) = \left(\sum_{i \in I} x_i\right) \mod 2 \right\}$$

What is $VC\dim(\mathcal{H}_{\text{n-parity}})$?

Proof. For each subset $I \subseteq \{1,2,\dots,n\}$ we have a h_I, so $|\mathcal{H}_{\text{n-parity}}| = 2^n$.

We know that $VC\dim(\mathcal{H}_{\text{n-parity}}) \le \lfloor \log_2 2^n \rfloor = n$.

So $VC\dim(\mathcal{H}_{\text{n-parity}}) \le n$.

Can we find a set C with n points that is shattered by $\mathcal{H}_{\text{n-parity}}$?

Let's try the "usual" set of unit vectors $C = \{e_1, e_2, \dots, e_n\}, e_i = (0,\dots,0,\underset{i}{1},0,\dots,0)$.

We want to show that, for each possible labeling $(y_1,y_2,\dots,y_n)$ of $(e_1,e_2,\dots,e_n)$, you can find a corresponding h such that $h(e_i) = y_i, \forall i = \overline{1,n}$.

Consider $(y_1,y_2,\dots,y_n)$ such a labeling and take $I = \{i \mid y_i = 1\}$.

Then we have

$$h_I(e_i) = \left(\sum_{i \in I} x_i\right) \mod 2 = \begin{cases} 1, & \text{if } i \in I \\ 0, & \text{otherwise} \end{cases}$$

So $VC\dim(\mathcal{H}_{\text{n-parity}}) = n$. □

Exercise 3 $\mathcal{X}$ – finite domain, $|\mathcal{X}| = n < \infty$, $k \le |\mathcal{X}|$

3.1. $\mathcal{H}^{\mathcal{X}}_{=k} = \left\{ h \in \{0,1\}^{\mathcal{X}} \,\middle|\, |\{x \colon h(x) = 1\}| = k \right\}$
= set of all functions that assign the value 1 to exactly k elements of $\mathcal{X}$

$VC\dim(\mathcal{H}^{\mathcal{X}}_{=k}) =?$

Proof.

If $k = 0 \Rightarrow \mathcal{H}^{\mathcal{X}}_{=0} = \{h^-\}$, all points get the value 0

If $k = 1 \Rightarrow \mathcal{H}^{\mathcal{X}}_{=1}$ has $|\mathcal{X}|$ functions $= n$ functions

$\mathcal{X} = \{x_1, x_2, \dots, x_n\}, n = |\mathcal{X}|$

$$h_i \colon \{x_1,\dots,x_n\} \to \{0,1\} \quad h_i(x_j) = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases}$$

If $k = 2 \Rightarrow \mathcal{H}^{\mathcal{X}}_{=2}$ has C_n^2 elements

$\vdots$

If $k = n-1 \Rightarrow \mathcal{H}^{\mathcal{X}}_{=n-1}$ has n elements

If $k = n \Rightarrow \mathcal{H}^{\mathcal{X}}_{=n}$ has 1 element $h^+(x_i) = 1\ \forall i = \overline{1,n}$

We first show that $VC\dim(\mathcal{H}^{\mathcal{X}}_{=k}) \le \min(k, n-k) = VC\dim(\mathcal{H})$.

Case 1) if $n \ge 2k$, in this case, $\min(k, n-k) = k,\ k \le \frac{n}{2}$

$\mathcal{H}$ will consist of functions h that label exactly k elements of $\mathcal{X}$ with label 1. So any set C with more than k points cannot be shattered because the labeling with all 1's $(1,1,1,\dots,1)$ cannot be realized by any $h \in \mathcal{H}$.

Case 2) if $n < 2k$, in this case $\min(k, n-k) = n-k,\ k > \frac{n}{2}$

$\mathcal{H}$ will consist of functions h that labels k elements of $\mathcal{X}$ with label 1, and $n-k$ points of $\mathcal{X}$ with label 0. So any set with more than $n-k+1$ points cannot be shattered by $\mathcal{H}$ as the labeling with all 0's $(0,0,\dots,0)$ cannot be realized by any $h \in \mathcal{H}$.

So we have that $VC\dim(\mathcal{H}) \le \min(k, n-k)$.

We will prove that $VC\dim(\mathcal{H}) = \min(k, n-k)$.

Consider $k' = \min(k, n-k)$.

We need to show that there exists a set of points $A = \{x_{i1}, x_{i2}, \dots, x_{ik'}\} \subseteq \mathcal{X}$ that is shattered by $\mathcal{H}$. This means that, for each subset $B \subseteq A$, we can find $h_B \in \mathcal{H}$ such that

$$h_B(x) = \begin{cases} 1, & x \in B \\ 0, & x \in A \setminus B \end{cases}$$

We choose a set of $k-|B|$ points $B' = \{b_1, b_2, \dots, b_{k-|B|}\} \subseteqq \mathcal{X} \setminus A$.

We can make the choice since $k - |B| \leq |\mathcal{X} \setminus A|$

$$\begin{aligned} k - |B| &\leq n - k' \\ k' - |B| &\leq n - k \\ k' - |B| &\leq k' \leq n - k \text{ this is true} \end{aligned}$$

So $B \subseteq A$ has $|B|$ elements

$B' \subseteq \mathcal{X} \setminus A$ has $k - |B|$ elements

So $|B \cup B'| = |B| + |B'|$ (as $B \cap B' = \emptyset$) $= k$

So, if we consider the characteristic function of the set $B \cup B'$, we have

$$\mathbb{1}_{B \cup B'}(x) = \begin{cases} 1, & x \in B \cup B' \\ 0, & \text{otherwise} \end{cases}$$

What is more important, $\mathbb{1}_{B \cup B'}$ takes value 1 for exactly k points, so it is a member of $\mathcal{H}$.

So, in this case, we take $h_B = \mathbb{1}_{B \cup B'}$.

h_B will have the desired property that $h_B(x) = \begin{cases} 1, & x \in B \\ 0, & x \in A \setminus B \end{cases}$

So any set A of $k' = min(k, n-k)$ points can be shattered by $\mathcal{H}$.

So $VC \dim(\mathcal{H}) = k' = \min(k, n-k)$. $\square$

3.2. $\mathcal{H}_{\text{at-most-}k} = \left\{h \in \{0,1\}^{\mathcal{X}} \colon |\{x \colon h(x) = 1\}| \leq k \text{ or } |\{x \colon h(x) = 0\}| \leq k\right\}$

Proof.

If $k = 0$ $\quad \mathcal{H}_{\text{at-most-0}} = \{h^-, h^+\}$ $\quad$ where $|\{x \colon h^-(x) = 1\}| \leq 0$ and $|\{x \colon h^+(x) = 0\}| \leq 0$

If $k = 1$ $\quad \mathcal{H}_{\text{at-most-1}} = \{h^-, h^+\} \cup \underbrace{\{\text{functions } h \text{ which label just one point with label } 1\}}_{\mathcal{H}^{\mathcal{X}}_{=1}}$

$\cup \{\text{functions } h \text{ which label just 1 point with label } 0\}$

<u>Case 1</u> $\quad$ If $n = |\mathcal{X}| \leq 2k+1$, then we have that $\mathcal{H}_{\text{at-most-}k} = \{0,1\}^{\mathcal{X}} = \{h \colon \mathcal{X} \to \{0,1\}\}$

This is true because any function $h \colon \mathcal{X} \to \{0,1\}$ will have either at most k points labeled with 0 or at most k points labeled with 1.

$\Downarrow$ $h \in \mathcal{H}_{\text{at-most-}k}$

Example (see Table 1): Take $n = 7,\ k = 4$ $\qquad \mathcal{X} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$

	x	x_1	x_2	x_3	x_4	x_5	x_6	x_7
at most 4 1's or 4 0's $\leftarrow$	$h(x)$	1	1	0	0	1	0	1
at most 4 1's $\leftarrow$	$h(x)$	0	0	1	0	0	0	0
at most 4 1's $\leftarrow$	$h(x)$	0	1	0	0	1	0	0

Table 1

In this case, $VC\dim(\mathcal{H}_{\text{at-most-}k}) = VC\dim(\{0,1\})^{\mathcal{X}} = n = |\mathcal{X}|$.

<u>Case 2</u> $\quad$ If $n = |\mathcal{X}| \geq 2k+2$

We first show that $VC\dim(\underset{\substack{\| \\ \mathcal{H}}}{\mathcal{H}_{\text{at-most-}k}}) = VC\dim(\mathcal{H}) \geq 2k+1$

Consider any set A of $2k+1$ points in $\mathcal{X}$: $A = \{a_1, a_2, \ldots, a_{2k+1}\}$.

We will show that A is shattered by $\mathcal{H}$. It is enough to show that, for each possible labeling $(y_1, y_2, \ldots, y_{2k+1})$ of the points $(a_1, a_2, \ldots, a_{2k+1})$, we can find an $h \in \mathcal{H}$ such that $h(y_i) = a_i$.

Take $\mathcal{J} = \{j \mid y_j = 1\}$, and take $B_{\mathcal{J}} = \{a_j \in A \mid \underset{j \in \mathcal{J}}{y_j = 1}\}$

If $|\mathcal{J}| \leq k$ $\quad$ we know that $\mathbb{1}_{B_{\mathcal{J}}} \in \mathcal{H}_{\text{at-most-}k}$, so we take

$$h = \mathbb{1}_{B_{\mathcal{J}}} \quad h(a_j) = \begin{cases} 1, & y_j = 1 \\ 0, & \text{otherwise} \end{cases}$$

If $|\mathcal{J}| > k$ $\quad$ take $C_{\mathcal{J}} = \{a_{\mathcal{J}} \in A \mid \underset{j \notin \mathcal{J}}{y_j = 0}\}$ has at most k elements, so we take in this case

$$h = \mathbb{1}_{A \setminus C_{\mathcal{J}}} \quad h(a_j) = \begin{cases} 1, & y_j = 1 \\ 0, & y_j = 0 \end{cases}$$

So, we have that $VC\dim(\mathcal{H}) \geq 2k+1$.

We show now that $VC\dim(\mathcal{H}) < 2k+2$.

Consider any set A of $2k+2$ points $\quad A = \{a_1, a_2, \ldots, a_{2k+2}\}$.

There is no $h \in \mathcal{H}$ that will label the first $k+1$ points with 1 and the rest $k+1$ points with 0.

So, in conclusion, $VC\dim(\mathcal{H}_{\text{at-most-}k}) = \min(|\mathcal{X}|, 2k+1)$. $\square$