Learning theory

Machine Learning II
2021-2022 - UMONS
Souhaib Ben Taieb

1 Exercise 1

Solve Problem 2.8 in LFD.

Problem 2.8 Which of the following are possible growth functions $m_{\mathcal{H}}(N)$ for some hypothesis set:

$$1+N;\ 1+N+\frac{N(N-1)}{2};\ 2^N;\ 2^{\lfloor \sqrt{N} \rfloor};\ 2^{\lfloor N/2 \rfloor};\ 1+N+\frac{N(N-1)(N-2)}{6}.$$

Figure 1: Source: Abu-Mostafa et al. Learning from data. AMLbook.

Solution

There are two cases for the growth function:

- $d_{VC} = \infty$ and $m_{\mathcal{H}}(N) = 2^N$ for all N.
- d_{VC} is finite and $m_{\mathcal{H}}(N) \leq N^{d_{VC}} + 1$

Now let's see the growth functions:

(i) $m_{\mathcal{H}}(N) = 1 + N$
Here, $d_{VC} = 1$ as $m_{\mathcal{H}}(2) = 3 < 2^2$. Therefore, $m_{\mathcal{H}}(N)$ must be bounded by $N^1 + 1$ which is true. Hence, $m_{\mathcal{H}}(N) = 1 + N$ is a possible growth function.

(ii) $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)}{2}$
Here, $d_{VC} = 2$ as $m_{\mathcal{H}}(3) = 7 < 2^3$. Therefore, $m_{\mathcal{H}}(N)$ must be bounded by $N^2 + 1$ for all N, which is true here. Hence, $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)}{2}$ is a possible growth function.

(iii) $m_{\mathcal{H}}(N) = 2^N$
Here, $d_{VC} = \infty$. Hence, $m_{\mathcal{H}}(N) = 2^N$ is a possible growth function.

(iv) $m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}$
Here, $d_{VC} = 1$ as $m_{\mathcal{H}}(2) = 2 < 2^2$. Therefore, $m_{\mathcal{H}}(N)$ must be bounded by $N^1 + 1$ for all N. Consider an example $N = 25$, $m_{\mathcal{H}}(25) = 32 \geq 25 + 1$. So, the bound $N^1 + 1$ for all N is not true here. Hence, $m_{\mathcal{H}}(N) = 2^{\lfloor \sqrt{N} \rfloor}$ is **not** a possible growth function.

(v) $m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor}$
Here, $d_{VC} = 0$ as $m_{\mathcal{H}}(1) = 1 < 2^1$. Therefore, $m_{\mathcal{H}}(N)$ must be bounded by $N^0 + 1 = 2$ for all N. Consider an example $N = 4$, $m_{\mathcal{H}}(4) = 4 \geq 2$. So, the bound $N^0 + 1$ for all N is not true here. Hence, $m_{\mathcal{H}}(N) = 2^{\lfloor N/2 \rfloor}$ is **not** a possible growth function.

(vi) $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$
Here, $d_{VC} = 1$ as $m_{\mathcal{H}}(2) = 3 < 2^2$. Therefore, $m_{\mathcal{H}}(N)$ must be bounded by $N^1 + 1$ for all N. Consider an example $N = 3$, $m_{\mathcal{H}}(3) = 5 \geq 3^1 + 1$. So, the bound $N^0 + 1$ for all N is not true here. Hence, $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)(N-2)}{6}$ is **not** a possible growth function.

2 Exercise 2

Solve Problem 2.3 in LFD.

Problem 2.3 Compute the maximum number of dichotomies, $m_{\mathcal{H}}(N)$, for these learning models, and consequently compute d_{VC}, the VC dimension.

(a) Positive or negative ray: $\mathcal{H}$ contains the functions which are $+1$ on $[a, \infty)$ (for some a) together with those that are $+1$ on $(-\infty, a]$ (for some a).

(b) Positive or negative interval: $\mathcal{H}$ contains the functions which are $+1$ on an interval $[a, b]$ and -1 elsewhere or -1 on an interval $[a, b]$ and $+1$ elsewhere.

(c) Two concentric spheres in $\mathbb{R}^d$: $\mathcal{H}$ contains the functions which are $+1$ for $a \leq \sqrt{x_1^2 + \ldots + x_d^2} \leq b$.

Figure 2: Source: Abu-Mostafa et al. Learning from data. AMLbook.

Solution

(a) **Positive or negative ray:** The growth function for positive rays is $N+1$. Now, for negative rays we get $N-1$ numbers of new dichotomies (the opposite of the ones from positive rays - two dichotomies where all points are +1 and where all are -1. Hence,

$$m_{\mathcal{H}}(N) = N + 1 + N - 1 = 2N$$
$$d_{VC} = 2 \qquad [\text{As } m_{\mathcal{H}}(3) = 6 < 2^3]$$

(b) **Positive or negative interval:** The growth function for positive interval is $\binom{N+1}{2} + 1$ as we can place interval ends in two of $N+1$ spots. Similarly, growth function for negative interval is $\binom{N+1}{2} + 1$. Now, we need to subtract the number of dichotomies covered by both the positive and negative interval. The overlap is $2N$ where all the positive and negative points are grouped separately (Consider for $N = 3$, 6 dichotomies are covered by both). Hence,

$$m_{\mathcal{H}}(N) = \binom{N+1}{2} + 1 + \binom{N+1}{2} + 1 - 2N = N^2 - N + 2$$
$$d_{VC} = 3 \qquad [\text{As } m_{\mathcal{H}}(4) = 14 < 2^4]$$

Now, we need to add the number of dichotomies for negative interval. For negative interval, we can place interval ends in two of $N-1$ spots as we need to subtract two spots where all points are +1 and all are -1. Therefore, for negative interval, the number of dichotomies is $\binom{N-1}{2}$. Hence,

$$m_{\mathcal{H}}(N) = \binom{N+1}{2} + 1 + \binom{N-1}{2} = N^2 - N + 2$$
$$d_{VC} = 3 \qquad [\text{As } m_{\mathcal{H}}(4) = 14 < 2^4]$$

(c) **Two concentric spheres in $\mathbb{R}^d$:** We can map two concentric sphere from $\mathbb{R}^d$ to $[0, \infty)$ by using the function below:

$$f : (x_1, x_2, \ldots, x_d) \mapsto r = \sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}$$

Now, the problem of two concentric circles in $\mathbb{R}^d$ is equivalent to the problem of positive interval. Hence,

$$
\begin{aligned}
m_{\mathcal{H}}(N) &= \binom{N+1}{2} = \frac{N^2}{2} + \frac{N}{2} + 1 \\
d_{VC} &= 2 \qquad [\text{As } m_{\mathcal{H}}(3) = 7 < 2^3]
\end{aligned}
$$

3 Exercise 3

Solve Problem 2.4 in LFD.

Problem 2.4 Show that $B(N, k) = \sum_{i=0}^{k-1} \binom{N}{i}$ by showing the other direction to Lemma 2.3, namely that

$$B(N, k) \geq \sum_{i=0}^{k-1} \binom{N}{i} .$$

To do so, construct a specific set of $\sum_{i=0}^{k-1} \binom{N}{i}$ dichotomies that does not shatter any subset of k variables. *[Hint: Try limiting the number of -1's in each dichotomy.]*

Figure 3: Source: Abu-Mostafa et al. Learning from data. AMLbook.

Solution

Let's assume that we have N points and $m_{\mathcal{H}}(N) = 2^N$. Now, we focus on the dichotomies that contains $(k-1)$ -1's. These dichotomies are:

- Number of dichotomies that doesn't contain -1 : $\binom{N}{0} = 1$.
- Number of dichotomies that contain one -1 : $\binom{N}{1} = N$.
- Number of dichotomies that contain two -1's : $\binom{N}{2}$.
- Number of dichotomies that contain three -1's : $\binom{N}{3}$.
- ...
- Number of dichotomies that contain $(k-1)$ -1's : $\binom{N}{k-1}$.

In total there are $\sum_{i=0}^{k-1} \binom{N}{i}$ such dichotomies. Moreover, these dichotomies do not shatter any subset of k variables and the set does not have any dichotomy which contains k (-1)s. Hence,

$$B(N, k) \geq \sum_{i=0}^{k-1} \binom{N}{i}$$

Now, Sauer's lemma is:

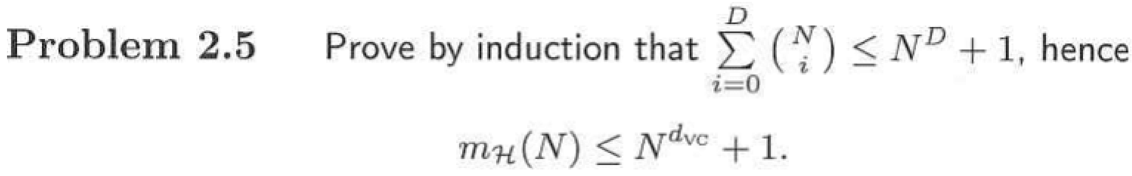
$$B(N, k) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

Therefore, we can conclude that,

$$B(N, k) = \sum_{i=0}^{k-1} \binom{N}{i}$$

4 Exercise 4

Solve Problem 2.5 in LFD.



Problem 2.5 Prove by induction that $\sum_{i=0}^{D} \binom{N}{i} \leq N^D + 1$, hence

$$m_{\mathcal{H}}(N) \leq N^{d_{\mathrm{vc}}} + 1.$$

Figure 4: Source: Abu-Mostafa et al. Learning from data. AMLbook.

Solution

To prove the inequality, let's look at the following cases:

- For $D = 0$,

$$1 = \binom{N}{0} \leq N^0 + 1$$

- Consider the inequality is true for D ($D \geq 1$),

$$\sum_{i=0}^{D} \binom{N}{i} \leq N^D + 1$$

- Now, we need to prove that it is true for $D + 1$,

$$\begin{aligned}
\sum_{i=0}^{D+1} \binom{N}{i} &= \sum_{i=0}^{D} \binom{N}{i} + \binom{N}{D+1} \\
&\leq N^D + 1 + \binom{N}{D+1} \\
&\leq N^D + 1 + \frac{N!}{(D+1)!(N-D-1)!}
\end{aligned}$$

  Now, we want to prove that $\frac{N!}{(N-D-1)!} \leq N^{D+1}$,

$$\frac{N!}{(N-D-1)!} = \prod_{i=0}^{D} (N-i) \leq N^{D+1}$$

  Therefore, we can write,

$$\begin{aligned}
\sum_{i=0}^{D+1} \binom{N}{i} &\leq N^D + 1 + \frac{N^{D+1}}{(D+1)!} \\
&\leq N^D + 1 + \frac{N^{D+1}}{2} \qquad [\text{As } D \geq 1, \text{ we have } (D+1)! \geq 2 \iff \frac{1}{(D+1)!} \leq \frac{1}{2}]
\end{aligned}$$

  Moreover, we have assumed $N \geq D + 1$ (otherwise, $\binom{N}{D+1} = 0$). So, $N \geq 2$ and consequently

$$\frac{1}{N} \leq \frac{1}{2} \iff \frac{N^D}{N^{D+1}} \leq \frac{1}{2} \iff N^D \leq \frac{N^{D+1}}{2}$$

Hence we can write,

$$\begin{aligned}\sum_{i=0}^{D+1}\binom{N}{i} &\le N^D + 1 + \frac{N^{D+1}}{2}\\ &\le \frac{N^{D+1}}{2} + 1 + \frac{N^{D+1}}{2}\\ &\le N^{D+1} + 1\end{aligned}$$

So, we have proved $\sum_{i=0}^{D}\binom{N}{i} \le N^D + 1$. Now,

$$\begin{aligned}m_{\mathcal{H}}(N) &\le \sum_{i=0}^{d_{VC}}\binom{N}{i}\\ &\le N^{d_{VC}} + 1\end{aligned}$$

5 Exercise 5

Solve Problem 2.13 in LFD..

Problem 2.13

(a) Let $\mathcal{H} = \{h_1, h_2, \ldots, h_M\}$ with some finite M. Prove that $d_{\text{vc}}(\mathcal{H}) \leq \log_2 M$.

(b) For hypothesis sets $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_K$ with finite VC dimensions $d_{\text{vc}}(\mathcal{H}_k)$, derive and prove the tightest upper and lower bound that you can get on $d_{\text{vc}}\left(\cap_{k=1}^{K}\mathcal{H}_k\right)$.

(c) For hypothesis sets $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_K$ with finite VC dimensions $d_{\text{vc}}(\mathcal{H}_k)$, derive and prove the tightest upper and lower bounds that you can get on $d_{\text{vc}}\left(\cup_{k=1}^{K}\mathcal{H}_k\right)$.

Figure 5: Source: Abu-Mostafa et al. Learning from data. AMLbook.

Solution

(a) Let, $d_{VC} = d$, then, $m_H(d) = 2^d$ (by definition). Now,

$$\begin{aligned} m_{\mathcal{H}}(d) &= \max_{x_1, x_2, \ldots, x_d} |\mathcal{H}(x_1, x_2, \ldots, x_d)| \\ &= \max_{x_1, x_2, \ldots, x_d} |\{(h(x_1), h(x_2), \ldots, h(x_d)) : h \in \mathcal{H}\}| \\ &= \max_{x_1, x_2, \ldots, x_d} |\{(h(x_1), h(x_2), \ldots, h(x_d)) : h \in \{h_1, h_2, \ldots, h_M\}\}| \\ &\leq |\mathcal{H}| = M \end{aligned}$$

Therefore, we can write,

$$\begin{aligned} & 2^d \leq M \\ \iff & d \leq \log_2(M) \end{aligned}$$

(b) At worst, we have $\cap_{k=1}^{K}\mathcal{H}_k = \{h\}$. Here, trivially VC dimension is 0 as $m_{\mathcal{H}}(N) = 1$ for all N. So, we can write $d_{VC}(\cap_{k=1}^{K}\mathcal{H}_k) \geq 0$.
Now, we will prove that,

$$d_{VC}(\cap_{k=1}^{K}\mathcal{H}_k) \leq \min_{1 \leq k \leq K} d_{VC}(\mathcal{H}_k)$$

To prove that let's assume,

$$d_{VC}(\cap_{k=1}^{K}\mathcal{H}_k) > \min_{1 \leq k \leq K} d_{VC}(\mathcal{H}_k) = d$$

It means that $\cap_{k=1}^{K}\mathcal{H}_k$ can shatter $d+1$ points, let $x_1, \ldots, x_{d+1}$ be those points. Now, we may write,

$$\begin{aligned} \{-1, +1\}^{d+1} &= \cap_{k=1}^{K}\mathcal{H}_k(x_1, \ldots, x_{d+1}) \\ &= \{(h(x_1), \cdots, h(x_{d+1})) : h \in \cap_{k=1}^{K}\mathcal{H}_k\} \\ &\subseteq \{(h(x_1), \cdots, h(x_{d+1})) : h \in \mathcal{H}_k\} \quad \text{for all } k = 1, \cdots, K \end{aligned}$$

If we compute the cardinality of these sets, we obtain,

$$\begin{aligned}
&2^{d+1} \leq |\{(h(x_1), \cdots, h(x_{d+1})) : h \in \mathcal{H}_k\}| \leq 2^{d+1} \quad \text{for all } k = 1, \cdots, K \\
&\Rightarrow |\{(h(x_1), \cdots, h(x_{d+1})) : h \in \mathcal{H}_k\}| = 2^{d+1} \quad \text{for all } k = 1, \cdots, K
\end{aligned}$$

Therefore, any $\mathcal{H}_k$ can shatter $d+1$ points.
Now, let $\min_{1\leq k\leq K} d_{VC}(\mathcal{H}_k) = d_{VC}(\mathcal{H}_{k0})$. Then we have,

$$d = d_{VC}(\mathcal{H}_{k0}) \geq d+1$$

which is not possible. Hence,

$$0 \leq d_{VC}(\cap_{k=1}^K \mathcal{H}_k) \leq \min_{1\leq k\leq K} d_{VC}(\mathcal{H}_k)$$

(c) Let, $d_{VC}(H_k) = d_k$ for all $k = 1, \cdots, K$. This implies $\mathcal{H}_k$ shatters d_k points $x_1, \cdots, x_{dk}$,

$$\begin{aligned}
\{-1,+1\}^{d_k} &= \{(h(x_1), \cdots, h(x_{d_k})) : h \in \mathcal{H}_k\} \\
&\subset \{(h(x_1), \cdots, h(x_{d_k})) : h \in \cup_{k=1}^K \mathcal{H}_k\}
\end{aligned}$$

Now if we compute the cardinality of these sets, we obtain

$$\begin{aligned}
&2^{d_k} \leq |\{(h(x_1), \cdots, h(x_{d_k})) : h \in \cup_{k=1}^K \mathcal{H}_k\}| \leq 2^{d_k} \\
&\Rightarrow |\{(h(x_1), \cdots, h(x_{d_k})) : h \in \cup_{k=1}^K \mathcal{H}_k\}| = 2^{d_k} \quad \text{for all } k = 1, \cdots, K
\end{aligned}$$

More simply we can write,

$$\begin{aligned}
&m_{\cup_{k=1}^K \mathcal{H}_k}(d_k) = 2^{d_k} \quad \forall k \\
&\Rightarrow d_{VC}(\cup_{k=1}^K \mathcal{H}_k) \geq d_k \quad \forall k \\
&\Rightarrow d_{VC}(\cup_{k=1}^K \mathcal{H}_k) \geq \max_{1\leq k\leq K} d_k = \max_{1\leq k\leq K} d_{VC}(\mathcal{H}_k)
\end{aligned}$$

Now, consider $K = 2$ and $d_{VC}(\mathcal{H}_1) = d_1$ and $d_{VC}(\mathcal{H}_2) = d_2$. The number of dichotomies generated by $\mathcal{H}_1 \cup \mathcal{H}_2$ is at most the sum of the dichotomies generated by $\mathcal{H}_1$ and by $\mathcal{H}_2$. Therefore,

$$\begin{aligned}
m_{\mathcal{H}_1\cup\mathcal{H}_2} &\leq m_{\mathcal{H}_1} + m_{\mathcal{H}_2} \\
&\leq \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=0}^{d_2} \binom{N}{i} \\
&\leq \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=0}^{d_2} \binom{N}{N-i} \\
&\leq \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=N-d_2}^{d_2} \binom{N}{i} \\
&< \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=d_1+1}^{N-d_2-1} \binom{N}{i} + \sum_{i=N-d_2}^{d_2} \binom{N}{i} = \sum_{i=0}^{N} \binom{N}{i} = 2^N
\end{aligned}$$

$\forall N$ such that $d_1 + 1 \leq N - d_2 - 1 \iff N \geq d_1 + d_2 + 1$. So, we can deduce that

$$d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq d_1 + d_2 + 1$$

Now, we will prove by induction,

$$d_{VC}(\cup_{k=1}^K \mathcal{H}_k) \leq K - 1 + \sum_{k=1}^K d_{VC}(\mathcal{H}_k)$$

- For $K = 2$,

$$d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 1 + \sum_{k=1}^{2} d_{VC}(\mathcal{H}_k) \quad \text{[Already proven]}$$

- Consider it is true for $K-1$,

$$d_{VC}(\cup_{k=1}^{K-1}\mathcal{H}_k) \leq K - 2 + \sum_{k=1}^{K-1} d_{VC}(\mathcal{H}_k)$$

- For K,

$$\begin{aligned}
d_{VC}(\cup_{k=1}^{K}\mathcal{H}_k) &= d_{VC}((\cup_{k=1}^{K-1}\mathcal{H}_k) \cup \mathcal{H}_K) \\
&\leq 1 + d_{VC}(\cup_{k=1}^{K-1}\mathcal{H}_k) + d_{VC}(\mathcal{H}_K) \\
&\leq 1 + K - 2 + \sum_{k=1}^{K-1} d_{VC}(\mathcal{H}_k) + d_{VC}(\mathcal{H}_K) \\
&\leq K - 1 + \sum_{k=1}^{K} d_{VC}(\mathcal{H}_k)
\end{aligned}$$

Finally, we obtain,

$$\max_{1 \leq k \leq K} d_{VC}(\mathcal{H}_k) \leq d_{VC}(\cup_{k=1}^{K}\mathcal{H}_k) \leq K - 1 + \sum_{k=1}^{K} d_{VC}(\mathcal{H}_k)$$