# Machine Learning, 2024 Spring
# Assignment 5

**Name: Zhou Shouchen**

Student ID: 2021533042

### Notice

Plagiarizer will get 0 points.
LaTeXis highly recommended. Otherwise you should write as legibly as possible.

Problem 1

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

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

Solution

1. $1+N$:
   Since $1+1=2^1$, and $1+2<2^2$, so $k=2$ is the breakpoint.
   $$\sum_{i=0}^{k-1}\binom{N}{i}=\sum_{i=0}^{1}\binom{N}{i}=1+N$$
   Since
   $$m_{\mathcal{H}}(N)=1+N\leq\sum_{i=0}^{k-1}\binom{N}{i}$$
   So $m_{\mathcal{H}}(N)=1+N$ is a possible growth function.
2. $1+N+\dfrac{N(N-1)}{2}$:
   Since $1+2+\dfrac{2*1}{2}=2^2$, and $1+3+\dfrac{3*2}{2}<2^3$, so $k=3$ is the breakpoint.
   $$\sum_{i=0}^{k-1}\binom{N}{i}=\sum_{i=0}^{2}\binom{N}{i}=1+N+\frac{N(N-1)}{2}$$
   So
   $$m_{\mathcal{H}}(N)=1+N+\frac{N(N-1)}{2}\leq\sum_{i=0}^{k-1}\binom{N}{i}$$
   So $m_{\mathcal{H}}(N)=1+N+\dfrac{N(N-1)}{2}$ is a possible growth function.
3. $2^N$:
   $\forall k=1,2,\cdots,N$, we have
   $$m_{\mathcal{H}}(N)=2^N$$
   So the breakpoint is $k=\infty$.
   $$\sum_{i=0}^{k-1}\binom{N}{i}=2^N$$
   So
   $$m_{\mathcal{H}}(N)=2^N\leq\sum_{i=0}^{k-1}\binom{N}{i}$$
   So $m_{\mathcal{H}}(N)=2^N$ is a possible growth function.
4. $2^{\lfloor\sqrt{N}\rfloor}$:
   Since $2^{\lfloor\sqrt{1}\rfloor}=2^1$ and $2^{\lfloor\sqrt{2}\rfloor}<2^2$, so $k=2$ is the breakpoint.
   $$\sum_{i=0}^{k-1}\binom{N}{i}=\sum_{i=0}^{1}\binom{N}{i}=1+N$$
   But $m_{\mathcal{H}}(N)=2^{\lfloor\sqrt{N}\rfloor}$ is close to a exponential function, and the $1+N$ is polynormial function, so it must has a $N$, such as $N=25$:
   $$m_{\mathcal{H}}(N)=32>1+N=26$$
   So $m_{\mathcal{H}}(N)=2^{\lfloor\sqrt{N}\rfloor}$ is not a possible growth function.

5. $2^{\lfloor \frac{N}{2} \rfloor}$:
Since $2^{\lfloor \frac{0}{2} \rfloor} = 2^0$ and $2^{\lfloor \frac{1}{2} \rfloor} = 1 < 2^1$, so $k = 1$ is the breakpoint.

$$\sum_{i=0}^{k-1} \binom{N}{i} = \sum_{i=0}^{0} \binom{N}{i} = 1$$

But $m_{\mathcal{H}}(N) = 2^{\lfloor \frac{N}{2} \rfloor}$ is an increasing function, and the 1 is a constant function, so $\forall N \geq 2$

$$m_{\mathcal{H}}(N) = 2^{\lfloor \frac{N}{2} \rfloor} > 1$$

So $m_{\mathcal{H}}(N) = 2^{\lfloor \frac{N}{2} \rfloor}$ is not a possible growth function.

6. $1 + N + \dfrac{N(N-1)(N-2)}{6}$:
Since $1 + 1 + \dfrac{1(1-1)(1-2)}{6} = 2^1$ and $1 + 2 + \dfrac{2(2-1)(2-2)}{6} = 3 < 2^2$, so $k = 2$ is the breakpoint.

$$\sum_{i=0}^{k-1} \binom{N}{i} = \sum_{i=2}^{0} \binom{N}{i} = 1 + N$$

But when $\dfrac{N(N-1)(N-2)}{6} \neq 0$, i.e. $N \geq 3$, we have

$$m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)(N-2)}{6} > 1 + N$$

So $m_{\mathcal{H}}(N) = 1 + N + \dfrac{N(N-1)(N-2)}{6}$ is not a possible growth function.

So above all, the possible growth functions $m_{\mathcal{H}}(N)$ are

$$1 + N; 1 + N + \frac{N(N-1)}{2}; 2^N$$

And the followings are not possible growth functions.

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

## Problem 2

For an $\mathcal{H}$ with $d_{\text{vc}} = 10$, what sample size do you need (as prescribed by the generalization bound) to have a $95\%$ confidence that your generalization error is at most 0.05 ?

## Solution

From the VC-inequality, we have

$$\mathbb{P}\left[\sup_{h\in\mathcal{H}} |E_{\text{in}}(h) - E_{\text{out}}(h)| > \epsilon\right] \leq 4m_{\mathcal{H}}(2N)e^{-\frac{1}{8}\epsilon^2 N}$$

i.e.

$$E_{\text{generalization}}(h) = |E_{\text{out}}(h) - E_{\text{in}}(h)| \leq \sqrt{\frac{8}{N}\ln\left(\frac{4((2N)^{\text{d}_{\text{vc}}}+1)}{\delta}\right)}$$

where $\text{d}_{\text{vc}}$ is the VC-dimension of $\mathcal{H}$, and $\delta$ is the confidence level.
Since we want the confidence to be $95\%$, so $\delta = 0.05$.
So we want to find when $d_{\text{vc}} = 10, \delta = 0.05$, what is the minimum sample size $N$ such that $E_{\text{generalization}}(h) \leq 0.05$.
Which is hard to solve analytically, so we can use the following python code to find the minimum sample size $N$.

p2.py

```python
from math import sqrt, log

def general_error(N, dvc=10, delta=0.05):
    return sqrt(8 / N * log(4 * ((2 * N) ** dvc + 1) / delta))

left, right = int(1), int(1e8)
ans = 1

while left <= right:
    mid = (left + right) // 2
    if general_error(mid) <= 0.05:
        ans = mid
        right = mid - 1
    else:
        left = mid + 1

print('The minimum sample size is N = ', ans)
print(f'N = {ans - 1}, generalization error = {general_error(ans - 1)}')
print(f'N = {ans}, generalization error = {general_error(ans)}')
```

With the code, we can get that:

$$N = 452956, E_{\text{generalization}}(h) = 0.05000004435489733 > 0.05$$
$$N = 452957, E_{\text{generalization}}(h) = 0.04999999306115058 < 0.05$$

The minimum sample size is $N = 452957$.
So above all, the sample size $N$ should be at least $452957$ to have a $95\%$ confidence that the generalization error is at most $0.05$.

Problem 3

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

Solution

Since there are totally $M$ hypothesis in $\mathcal{H}$, so we can generate $M$ results with these $M$ hypothesis.
And since $d_{\text{vc}}(\mathcal{H})$ means that for $d_{\text{vc}}(\mathcal{H})$ groups of data, the hypothesis $\mathcal{H}$ can seperate all $2^{d_{\text{vc}}(\mathcal{H})}$ cases.
Since the maximum number of cases that can be seperated by the hypothesis $\mathcal{H}$ is $M$, so we have

$$2^{d_{\text{vc}}(\mathcal{H})} \leq M$$

i.e.

$$d_{\text{vc}}(\mathcal{H}) \leq \log_2 M$$

So above all, we have proved that $d_{\text{vc}}(\mathcal{H}) \leq \log_2 M$.

Problem 4

Let $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_K$ be $K$ hypothesis sets with finite VC dimension $d_{\text{vc}}$. Let $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \ldots \cup \mathcal{H}_K$ be the union of these models. Show that $d_{\text{vc}}(\mathcal{H}) < K\,(d_{\text{vc}} + 1)$.

Solution

To be more general, suppose that $d_{\text{vc}}(\mathcal{H}_i) = d_i$ for $i = 1, 2, \ldots, K$.
And we can prove that

$$d_{\text{vc}}(\mathcal{H}) \leq K - 1 + \sum_{i=1}^{K} d_i$$

1. If $K = 1$, then

$$d_{\text{vc}}(\mathcal{H}) = d_{\text{vc}}(\mathcal{H}_1) = d_1 \leq 0 + d_1$$

2. If $K = 2$, then we need to prove that

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

We can prove this with contradiction.
If we assume that $d_{\text{vc}}(\mathcal{H}) = d_{\text{vc}}(\mathcal{H}_1 \cup \mathcal{H}_2) \geq d_1 + d_2 + 2$.
Then we have

$$m_{\mathcal{H}}(d_1 + d_2 + 2) \geq 2^{d_1 + d_2 + 2}$$

But we also have:

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

Which is a contradiction.
Therefore, we have $d_{\text{vc}}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 1 + d_1 + d_2$.

3. Then we use induction to prove that $d_{\text{vc}}(\mathcal{H}) \leq K - 1 + \sum_{i=1}^{K} d_i$.

Suppose that when $K = k$, we have $d_{\text{vc}}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \ldots \cup \mathcal{H}_k) \leq k - 1 + \sum_{i=1}^{k} d_i$.

When $K = k + 1$, we have

$$\begin{aligned}
d_{\text{vc}}(\mathcal{H}) &= d_{\text{vc}}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \ldots \cup \mathcal{H}_{k+1}) \\
&= d_{\text{vc}}\left[(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \ldots \cup \mathcal{H}_k) \cup \mathcal{H}_{k+1}\right] \\
&\leq 1 + d_{\text{vc}}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \ldots \cup \mathcal{H}_k) + d_{\text{vc}}(\mathcal{H}_{k+1}) \\
&\leq 1 + \left( k - 1 + \sum_{i=1}^{k} d_i \right) + d_{k+1} \\
&= k + \sum_{i=1}^{k+1} d_i
\end{aligned}$$

So we have proved that when $K = k + 1$,

$$d_{\text{vc}}(\mathcal{H}) \leq (k + 1) - 1 + \sum_{i=1}^{k+1} d_i$$

4. By induction, we can prove that $d_{\text{vc}}(\mathcal{H}) \leq K - 1 + \sum\limits_{i=1}^{K} d_i$.

Since we have $d_{\text{vc}}(\mathcal{H}_i) = d_i = d_{\text{vc}},\ \forall i = 1, 2, \ldots, K$, we can get that

$$\begin{aligned} d_{\text{vc}}(\mathcal{H}) &= d_{\text{vc}}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \ldots \cup \mathcal{H}_K) \\ &\leq K - 1 + \sum_{i=1}^{K} d_i \\ &= K - 1 + K d_{\text{vc}} \\ &< K + K d_{\text{vc}} \\ &= K(d_{\text{vc}} + 1) \end{aligned}$$

So above all, we have proved that

$$d_{\text{vc}}(\mathcal{H}) < K(d_{\text{vc}} + 1)$$

Problem 5

In this part, you need to complete some mathematical proofs about VC dimension. Suppose the hypothesis set

$$\mathcal{H} = \{f(x, \alpha) = \text{sign}(\sin(\alpha x)) \mid, \alpha \in \mathbb{R}\}$$

where $x$ and $f$ are feature and label, respectively.

- Show that $\mathcal{H}$ cannot shatter the points $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$.

(Key: Mathematically, you need to show that there exists $y_1, y_2, y_3, y_4$, for any $\alpha \in \mathbb{R}, f(x_i) \neq y_i, i = 1, 2, 3, 4$, for example, $+1, +1, -1, +1$ )

- Show that the VC dimension of $\mathcal{H}$ is $\infty$. (Note the difference between it and the first question)

(Key: Mathematically, you have to prove that for any label sets $y_1, \cdots, y_m, m \in \mathbb{N}$, there exists $\alpha \in \mathbb{R}$ and $x_i, i = 1, 2, \cdots, m$ such that $f(x; \alpha)$ can generate this set of labels. Consider the points $x_i = 10^{-i} \ldots$)

Solution

(1) For the case $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$ with label $y_1 = +1$, $y_2 = +1$, $y_3 = -1$, $y_4 = +1$, we can prove that $\forall \alpha \in \mathbb{R}, \exists i \in \{1, 2, 3, 4\}, f(x_i) \neq y_i$.

- 1. For $x_1 = 1$
  Since $y_1 = +1$, we have $f(x_1) = \text{sign}(\sin \alpha) = +1 \Rightarrow \sin \alpha > 0$
  i.e. we have $\sin \alpha > 0$.
- 2. For $x_2 = 2$
  Since $y_2 = +1$, we have $f(x_2) = \text{sign}(\sin 2\alpha) = +1 \Rightarrow \sin 2\alpha > 0$
  i.e. we have $\sin 2\alpha = 2 \sin \alpha \cos \alpha > 0$.
  From 1. we have known that $\sin \alpha > 0$, so we must have $\cos \alpha > 0$.
  i.e. we have $\cos \alpha > 0$ and $\sin 2\alpha > 0$.
- For $x_3 = 3$
  Since $y_3 = -1$, we have $f(x_3) = \text{sign}(\sin 3\alpha) = -1 \Rightarrow \sin 3\alpha < 0$
  i.e. we have $\sin 3\alpha = \sin \alpha \cos 2\alpha + \sin 2\alpha \cos \alpha < 0$.
  From 1. we have $\sin \alpha > 0$, and from 2. we have $\cos \alpha > 0$, $\sin 2\alpha > 0$,
  so we must have $\cos 2\alpha < 0$.
  i.e. we have $\cos 2\alpha < 0$.
- For $x_4 = 4$
  Since $y_4 = +1$, we have $f(x_4) = \text{sign}(\sin 4\alpha) = +1 \Rightarrow \sin 4\alpha > 0$
  i.e. we have $\sin(4\alpha) = 2\sin(2\alpha)\cos(2\alpha) > 0$
  But from 2. we have $\sin 2\alpha > 0$, from 3. we have $\cos 2\alpha < 0$,
  so we must have $\sin(4\alpha) < 0$, which is a contradiction.

So we have prove that for $y_1 = +1$, $y_2 = +1$, $y_3 = -1$, $y_4 = +1$,
$\forall \alpha \in \mathbb{R}$, if $\forall i \in \{1, 2, 3\}, f(x_i) = y_i$, then $f(x_4) \neq y_4$.
So above all, $\mathcal{H}$ cannot shatter the points $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4,\ \forall \alpha \in \mathbb{R}$.

(2) We can construct that:
Let $\alpha = 1$, $x_i = \pi - \dfrac{\pi}{2} \cdot y_i + 2\pi i, i = 1, 2, \cdots, m$

- For $y_i = +1$, we have
$$f(x_i) = \text{sign}(\sin(\pi - \frac{\pi}{2} \cdot y_i + 2\pi i)) = \text{sign}(\sin(\frac{\pi}{2} + 2\pi i)) = +1$$
- For $y_i = -1$, we have
$$f(x_i) = \text{sign}(\sin(\pi - \frac{\pi}{2} \cdot y_i + 2\pi i)) = \text{sign}(\sin(\frac{3\pi}{2} + 2\pi i)) = -1$$

So above all, we have proved that for any label sets $y_1, \cdots, y_m, m \in \mathbb{N}$,
there exists $\alpha \in \mathbb{R}$ and $x_i, i = 1, 2, \cdots, m$ such that $f(x; \alpha)$ can generate this set of labels.
i.e. the VC dimension of $\mathcal{H}$ is $\infty$.