Machine Learning, 2024 Spring
Assignment 5

Name: Zhou Shouchen

Student ID: 2021533042

Notice

Plagiarizer will get 0 points.
LaTeX is 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
   $$m_{\mathcal{H}}(N) = 1 + N \le \sum_{i=0}^{k-1} \binom{N}{i}$$
   So $m_{\mathcal{H}}(N) = 1 + N$ is a possible growth function.

2. $1 + N + \frac{N(N-1)}{2}$:
   Since $1 + 2 + \frac{2*1}{2} = 2^2$, and $1 + 3 + \frac{3*2}{2} < 2^3$, 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
   $$m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)}{2} \le \sum_{i=0}^{k-1} \binom{N}{i}$$
   So $m_{\mathcal{H}}(N) = 1 + N + \frac{N(N-1)}{2}$ 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
   $$m_{\mathcal{H}}(N) = 2^N \le \sum_{i=0}^{k-1} \binom{N}{i}$$
   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 d_{vc} is the VC-dimension of $\mathcal{H}$, and δ 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 $d_{\text{vc}}(\mathcal{H}) \le \log_2 M$.

Solution

Since there are totally M hypotheses in $\mathcal{H}$, we can distinguish M cases. $d_{\text{vc}}(\mathcal{H})$ means that for $n = d_{\text{vc}}(\mathcal{H})$ groups of data, these M hypotheses can distinguish all $2^{d_{\text{vc}}(\mathcal{H})}$ cases, and a total of at most M cases can be distinguished, so

$$2^{d_{\text{vc}}(\mathcal{H})} \le M \; d_{\text{vc}}(\mathcal{H}) \le \log_2(M)$$

Problem 4

Let $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_K$ be K hypothesis sets with finite VC dimension d_{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 $d_{\text{vc}}(\mathcal{H}) < K\left(d_{\text{vc}} + 1\right)$.

Solution

Problem 5

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

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

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 ∞. (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)

$$\sin(2\alpha) = 2\sin\alpha\cos\alpha$$
$$\sin(4\alpha) = 2\sin(2\alpha)\cos(2\alpha)$$

(2) $\alpha = 1$, $x_i = \pi - \frac{\pi}{2} \cdot y_i + 2\pi i$