Learning from Data
Homework # 4

Khoi Pham

Generalization Error

My code for answering problems $1 - 3$ is uploaded here.

1. Answer: [d]

2. Answer: [d]

3. Answer: [c]

Bias and Variance

My code for running the experiments to answer problems $4 - 7$ is uploaded here.

4. Answer: [d]

5. Answer: [b]

6. Answer: [a]

7. Answer: [b]

VC Dimension

8. Assume $q \geq 1$ is an integer and let $m_{\mathcal{H}}(1) = 2$. What is the VC dimension of a hypothesis set whose growth function satisfies: $m_{\mathcal{H}}(N+1) = 2m_{\mathcal{H}}(N) - \binom{N}{q}$? Recall that $\binom{M}{m} = 0$ when $m > M$.

Let d_{VC} be the VC dimension of the hypothesis set. According to the recurrence relation of the growth function, we have:

$$\begin{aligned} m_{\mathcal{H}}(d_{VC}) &= 2 \times m_{\mathcal{H}}(d_{VC} - 1) - \binom{d_{VC} - 1}{q} \\ &= 2 \times 2^{d_{VC}-1} - \binom{d_{VC} - 1}{q} \\ &= 2^{d_{VC}} - \binom{d_{VC} - 1}{q} \end{aligned}$$

Because d_{VC} is the VC dimension of the hypothesis set, $m_{\mathcal{H}}(d_{VC}) = 2^{d_{VC}}$. Thus $\binom{d_{VC}-1}{q} = 0 \Leftrightarrow q > d_{VC} - 1 \Leftrightarrow q \geq d_{VC}$ (1).

As $d_{VC} + 1$ is the break point of the hypothesis set, $m_{\mathcal{H}}(d_{VC} + 1) < 2^{d_{VC}}$. According to the recurrence relation, we have:

$$\begin{aligned} m_{\mathcal{H}}(d_{VC} + 1) &= 2 \times m_{\mathcal{H}}(d_{VC}) - \binom{d_{VC}}{q} \\ &= 2 \times 2^{d_{VC}} - \binom{d_{VC}}{q} \\ &= 2^{d_{VC}+1} - \binom{d_{VC}}{q} \end{aligned}$$

As shown above, because $m_{\mathcal{H}}(d_{VC} + 1) < 2^{d_{VC}}$, $\binom{d_{VC}}{q} > 0 \Leftrightarrow q \leq d_{VC}$ (2). From (1) and (2), we have proved that $d_{VC} = q$.

Answer: [c]

9. For hypothesis sets $\mathcal{H}_1$, $\mathcal{H}_2$, ..., $\mathcal{H}_K$ with finite, positive VC dimensions $d_{VC}(\mathcal{H}_k)$, some of the following bounds are correct and some are not. Which among the correct ones is the tightest bound (the smallest range of values) on the VC dimension of the **intersection** of the sets: $d_{VC}(\bigcap_{k=1}^{K} \mathcal{H}_k)$?

First, suppose all of the hypothesis sets are disjoint, then the intersection of them is an empty set, whose VC dimension is 0. Thus, the left bound of the VC dimension of their intersection is $0 \leq d_{VC}(\bigcap_{k=1}^{K} \mathcal{H}_k)$.

Second, let $p = \min\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K}$, and $\mathcal{H}_j$ be the hypothesis set whose VC dimension is equal to p . We assume that the maximum possible value of $d_{VC}(\bigcap_{k=1}^{K} \mathcal{H}_k) > p$. Thus the intersection of all K hypothesis sets are able to shatter more than p data points. Because all of the hypotheses in this intersection also belong to $\mathcal{H}_j$ (because this set is the **intersection**), $\mathcal{H}_j$ is also able to shatter more than p points, which contradicts the fact that $d_{VC}(\mathcal{H}_j) = p$. Thus we find the right bound:

$d_{VC}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq min\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K}$.

In conclusion, we have: $0 \leq d_{VC}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \min\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K}$.

Answer: [b]

10. Similar to problem 9, but ask about the tightest bound on the VC dimension of the **union** of the sets: $d_{VC}(\bigcup_{k=1}^{K} \mathcal{H}_k)$.

Let $\mathcal{H} = \bigcup_{k=1}^{K} \mathcal{H}_k$, $p = \max\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K}$, and $\mathcal{H}_j$ be the hypothesis set whose VC dimension is equal to p. It is obvious that $\mathcal{H}$ contains $\mathcal{H}_j$ because $\mathcal{H}$ is a union, thus $d_{VC}(\mathcal{H}) \geq p$. The equality happens when $\mathcal{H}_j$ contains all other hypotheses in its set, thus making the VC dimension of the union equal to p. Hence: $\max\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{VC}(\bigcup_{k=1}^{K} \mathcal{H}_k)$.

Suppose we have 2 disjoint hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_2$ which can be used to shatter d_1 and d_2 points respectively. Thus, by combining these 2 together, we can shatter at least $d_1 + d_2$ points. It is observed that by giving $\mathcal{H}_1$ an additional $(d_1 + 1)$-th point P, $\mathcal{H}_1$ will not be able to shatter it along with its first d_1 points. As a result, P will be classified as either black or white by $\mathcal{H}_1$. The same thing occurs with $\mathcal{H}_2$ if we give P to it as an additional $(d_2 + 1)$-th point, but by some chances (if we are lucky), $\mathcal{H}_2$ will classify it differently from $\mathcal{H}_1$. Thus, P might potentially become an additional data point that the union of $\mathcal{H}_1$ and $\mathcal{H}_2$ are able to shatter. Thus, the maximum possible value of the VC dimension of the union of 2 hypothesis sets is $d_1 + d_2 + 1$. Thus, $d_{VC}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq K - 1 + \sum_{k=1}^{K} d_{VC}(\mathcal{H}_k)$.

In conclusion, we have: $\max\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{VC}(\bigcup_{k=1}^{K} \mathcal{H}_k) \leq K - 1 + \sum_{k=1}^{K} d_{VC}(\mathcal{H}_k)$.

Answer: [e]