Problem Set 4

Claire Goeckner-Wald

October 24, 2016

Generalization Error

1. D. 460,000

$$
\begin{aligned}
\epsilon &= \sqrt{\frac{8}{N} \cdot \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}} \\
0.95 &\geq 1 - \delta \\
\delta &\leq 0.05 \\
\epsilon &= 0.05 \\
d_{vc} &= 10 \\
m_{\mathcal{H}}(2N) &= (2N)^{d_{vc}} = (2N)^{10} \\
0.05 &= \sqrt{\frac{8}{N} \cdot \ln \frac{4(2N)^{10}}{0.05}} \\
(0.05)^2 &= \frac{8}{N} \cdot \ln \frac{4(2N)^{10}}{0.05} \\
\frac{N(0.05)^2}{8} &= \ln \frac{4(2N)^{10}}{0.05} \\
N &= -0.323, 0.323, 452957
\end{aligned}
$$

2. D. Devroye

$$
\begin{aligned}
\delta &= 0.05 \\
d_{vc} &= 50 \\
m_{\mathcal{H}}(N) &= N^{d_{vc}} = N^{50}
\end{aligned}
$$

[a]

$$
\begin{aligned}
\epsilon &\le \sqrt{\frac{8}{N} \cdot \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}} \\
\epsilon &\le \sqrt{\frac{8}{N} \cdot \ln \frac{4(2N)^{50}}{0.05}} \\
\epsilon &\le 0.632
\end{aligned}
$$

[b]

$$
\begin{aligned}
\epsilon &\le \sqrt{\frac{2\ln(2Nm_{\mathcal{H}}(N))}{N}} + \sqrt{\frac{2}{N}\ln\frac{1}{\delta}} + \frac{1}{N} \\
\epsilon &\le \sqrt{\frac{2\ln(2NN^{50})}{N}} + \sqrt{\frac{2}{N}\ln\frac{1}{0.05}} + \frac{1}{N} \\
\epsilon &\le 0.331
\end{aligned}
$$

[c]

$$
\begin{aligned}
\epsilon &\le \sqrt{\frac{1}{N}(2\epsilon + \ln\frac{6m_{\mathcal{H}}(2N)}{\delta})} \\
\epsilon &\le \sqrt{\frac{1}{N}(2\epsilon + \ln\frac{6(2N)^{50}}{0.05})} \\
100\epsilon &\le \sqrt{2\epsilon + 499.962} \\
\epsilon &\le 0.223
\end{aligned}
$$

[d]

$$
\begin{aligned}
\epsilon &\le \sqrt{\frac{1}{2N}(4\epsilon(1+\epsilon) + \ln\frac{4m_{\mathcal{H}}(N^2)}{\delta})} \\
\epsilon &\le \sqrt{\frac{1}{2N}(4\epsilon(1+\epsilon) + \ln\frac{4(N^2)^{50}}{0.05})} \\
\epsilon &\le \sqrt{(4\epsilon(\epsilon+1) + 925.416)/(100\sqrt{2}} \\
\epsilon &\le 0.215
\end{aligned}
$$

3. C. Parrondo and Van den Broek

[a]

$$
\begin{aligned}
\epsilon &\le \sqrt{\frac{8}{N} \cdot \ln \frac{4(2N)^{50}}{0.05}} \\
\epsilon &\le 13.82
\end{aligned}
$$

[b]

$$
\begin{aligned}
\epsilon &\leq \sqrt{\frac{2\ln(2NN^{50})}{N}} + \sqrt{\frac{2}{N}\ln\frac{1}{0.05}} + \frac{1}{N} \\
\epsilon &\leq 7.05
\end{aligned}
$$

[c]

$$
\begin{aligned}
\epsilon &\leq \sqrt{\frac{1}{N}(2\epsilon + \ln\frac{6(2N)^{50}}{0.05})} \\
\epsilon &\leq 0.632456\sqrt{\epsilon + 59.9584} \\
\epsilon &\leq 5.10136
\end{aligned}
$$

[d]

$$
\begin{aligned}
\epsilon &\leq \sqrt{\frac{1}{2N}(4\epsilon(1+\epsilon) + \ln\frac{4(N^2)^{50}}{0.05})} \\
\epsilon &\leq 0.632456\sqrt{\epsilon^2 + \epsilon + 41.3315} \\
\epsilon &\leq 5.59313
\end{aligned}
$$

Bias and Variance

4. E. None of the above

The average slope b of the chosen function g, averaging over one million times, was 1.42. See the code attached for an experimental solution.

5. B. 0.3

The average bias of the chosen function g was 0.262. See the code attached for an experimental solution.

6. A. 0.2

The variance of the chosen function g, averaging over one million times, was 0.237. See the code attached for an experimental solution.

7. D or E.

[a] $h(x) = b$

$E_{out} = \text{bias} + \text{variance} = 0.5 + 0.25 = 0.75$

[b] $h(x) = ax$

$E_{out} = \text{bias} + \text{variance} = 0.5 + 0.25 = 0.75$

[c] $h(x) = ax + b$

$E_{out} = \text{bias} + \text{variance} = 0.2 + 1.7 = 1.9$

[d] $h(x) = ax^2$

$E_{out} = \text{bias } + \text{ variance} =$

[e] $h(x) = ax^2 + b$

$E_{out} = \text{bias } + \text{ variance} =$

VC Dimension

8. C. q

$$\begin{aligned}
m_{\mathcal{H}}(N+1) &= m_{\mathcal{H}}(N) - \binom{N-1}{q} \\
m_{\mathcal{H}}(1 \leq q) &= 2 = 2^1 \\
m_{\mathcal{H}}(2 \leq q) &= 2 \cdot 2 = 2^2 \\
m_{\mathcal{H}}(3 \leq q) &= 2 \cdot 2 \cdot 2 = 2^3 \\
m_{\mathcal{H}}(p < q) &= 2^p \\
m_{\mathcal{H}}(p = q) &= 2^p - \binom{p-1}{q} = 2^p - 0 = 2^p \\
m_{\mathcal{H}}(p = q+1) &= 2^p - \binom{q}{q} = 2^p - 1
\end{aligned}$$

Thus, at $q+1$, the growth function returns a value less than 2^{q+1}. Therefore, the breakpoint $k = q+1$ and the VC dimension $d_{vc} = q$.

9. B.

In the case that $\bigcap_{k=1}^{K} \mathcal{H}_k = \emptyset$, then $d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) = 0$. Thus, the lower bound of 0 is valid. However, if $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$ is not zero as well, then it is an invalid lower bound in this scenario. It is not a valid lower bound because there is no reason that d_{vc} for some $\mathcal{H}_k$ in the set is necessarily zero. Thus, we can eliminate choices **[d]** and **[e]** for invalid lower bounds.

In the case that $\bigcap_{k=1}^{K} \mathcal{H}_k$ is such that all the hypotheses amongst the sets are in common, then each $\mathcal{H}_k$ in the set of all $\mathcal{H}$ is equivalent. Then, $d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) = d_{vc}(\mathcal{H})$, where $\mathcal{H} = \mathcal{H}_k, \forall k$. In this case,

$$d_{vc}(\mathcal{H}) = \min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} = \max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \leq \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k).$$

Thus, we can eliminate **[a]** for being a looser fit on the upper bound.

We can also eliminate **[c]** for being a looser fit than **[b]**. Since $d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k)$ returns the greatest subset of all of the $\mathcal{H}$, the minimum function is valid. For example, if $\mathcal{H}_1 \subset \mathcal{H}_2$, then $\mathcal{H}_1 \cap \mathcal{H}_2 = \mathcal{H}_1$. Then, $d_{vc}(\mathcal{H}_1)$ is the proper d_{vc}. However, since $\mathcal{H}_1 \cap \mathcal{H}_2 = \mathcal{H}_1$, then $d_{vc}(\mathcal{H}_1) \leq d_{vc}(\mathcal{H}_2)$.

[a] $0 \le d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \le \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The lower bound of 0 is valid. The upper bound of $\sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$ is valid but looser than $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$ and $\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$.

[b] $0 \le d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \le \min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$

The lower bound of 0 is valid.

[c] $0 \le d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \le \max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$

The lower bound of 0 is valid.

[d] $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \le d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \le \max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$

The lower bound of $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$ is invalid.

[e] $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \le d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) \le \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The lower bound of $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$ is invalid.

10. E.

Given some set of $\mathcal{H}$, each $\mathcal{H}_k$ has some d_{vc}. Then, if we take the intersection of $\mathcal{H}_\cup = \bigcup_{k=1}^{K} \mathcal{H}_k$, the super-set $\mathcal{H}_\cup$ contains all hypotheses that led to each d_{vc} before. Thus, a lower bound of $\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$ is valid.

$\sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$ is not a valid upper bound. Thus, we can eliminate **[a]**, **[c]**, and **[d]**. For a counter-example of its validity, consider the hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_2$, given below, for an input space composed of three labels $[\pm, \pm, \pm]$. For $\mathcal{H}_1$, $d_{vc} = 1$ because [-, -] cannot be shattered. For $\mathcal{H}_2$, $d_{vc} = 1$ because [+, +] cannot be shattered. However, if we take $\mathcal{H}_1 \bigcap \mathcal{H}_2$, as $\bigcup_{k=1}^{K} \mathcal{H}_k$ does, then the $d_{vc}(\mathcal{H}_1 \bigcap \mathcal{H}_2) = 3 > d_{vc}(\mathcal{H}_1) + d_{vc}(\mathcal{H}_2)$. Thus, $\sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$ is not a valid upper bound. Then, the only other option $K - 1 + \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$ must be a valid upper bound.

[a] $0 \le d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \le \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The upper bound is invalid.

[b] $0 \le d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \le K - 1 + \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The upper bound is valid. The lower bound is valid, but more loose than $\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K}$.

[c] $\min\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \le d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \le \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The upper bound is invalid.

[d] $\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \le d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \le \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The upper bound is invalid.

[e] $\max\{d_{vc}(\mathcal{H}_k)\}_{k=1}^{K} \le d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k) \le K - 1 + \sum_{k=1}^{K} d_{vc}(\mathcal{H}_k)$

The upper bound is valid. The lower bound is valid, and more tight than 0.