Homework 5

Karan Sarkar
sarkak2@rpi.edu

October 8, 2019

Exercise 2.8.

(a) We generate many datasets. $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$. From these we construct final hypotheses $g_1, g_2, \ldots, g_K$. Therefore, $\bar{g}(x) = \frac{1}{K}\sum_{i=0}^{K} g_K(x)$. Therefore, $\bar{g}$ is a linear combination of elements g_i of $\mathcal{H}$. Because $\mathcal{H}$ is closed under linear combination, it follows that $\bar{g} \in \mathcal{H}$.

(b) We define a model to be the majority algorithm. If the data has a majority -1 we return -1 for all values. If the data has a majority $+1$ we return $+1$ for all values. The hypothesis set is all $+1$ and all -1. If randomly generated datasets have an equal predisposition for -1 as $+1$, we have that $\bar{g}(x) = 0.5$. This $\bar{g}(x)$ is not one of the two hypotheses in the hypothesis set.

(c) In general, we do not expect $\bar{g}$ to be a binary function. We would expect fractional values indicating the uncertain nature of learning.

Exercise 2.14.

(a) Consider two hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_2$ with VC-dimensions d_1 and d_2 respectively. Note that if $d_1 + 1 \leq N - d_2 - 1$ i.e. $N \geq d_1 + d_2 + 2$ it follws

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

Therefore, if N is large enough $d_{\mathrm{VC}}(\mathcal{H}_1 \cup \mathcal{H}_2) < N$. Thus, $d_{\mathrm{VC}}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq d_1 + d_2 + 1$. Now suppose that the VC-dimension of $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_K$ is d_{VC}. We will prove by induction that $d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_K) \leq K d_{\mathrm{VC}} + K - 1$. We will do induction on K. For the base case, let $K = 2$. We already have proven that $d_{\mathrm{VC}}(\mathcal{H}_1 \cup \mathcal{H}_2) \leq 2 d_{\mathrm{VC}} + 1$.

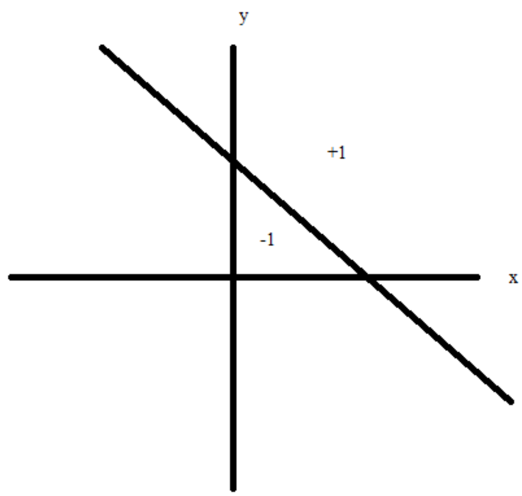
For the inductive step, assume that for some k we have that $d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_k) \leq k d_{\mathrm{VC}} + k - 1$. Note that:

$$\begin{aligned} d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_k \cup \mathcal{H}_{k+1}) &= d_{VC}((\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_k) \cup \mathcal{H}_{k+1}) \\ &\leq k d_{\mathrm{VC}} + k - 1 + d_{\mathrm{VC}} + 1 \\ &= (k+1) d_{\mathrm{VC}} + k \end{aligned}$$

Thus it holds for all positive K that $d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_K) \leq K d_{\mathrm{VC}} + K - 1$. Thus, $d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_K) < K(d_{\mathrm{VC}} + 1)$. q

(b) Note that $m_{\mathcal{H}_k}(l) \leq l^{d_{\mathrm{VC}}} + 1$. Therefore, by the union bound, $m_{\mathcal{H}}(l) \leq K l^{d_{\mathrm{VC}}} + K$. Note that when $l^{d_{\mathrm{VC}}} \geq 1$, it follows that $K l^{d_{\mathrm{VC}}} + K \leq 2 K l^{d_{\mathrm{VC}}}$. Therefore,$m_{\mathcal{H}}(l) \leq 2 K l^{d_{\mathrm{VC}}}$. Therefore, from the hypothesis it follows that $m_{\mathcal{H}}(l) < 2^l$. Thus, $d_{\mathrm{VC}}(\mathcal{H}) \leq l$.

Exercise 2.15.



(a)

(b) We can use a thresholded function with an arbitrary number of steps. This allows us to crreate an arbitary number of positive and negative intervals. Therefore, $m_{\mathcal{H}}(N) = 2^N$ and the VC-dimension is infinite.

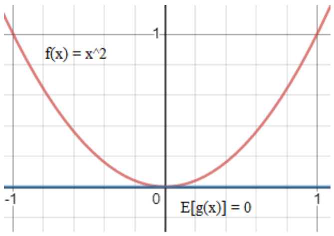
Exercise 2.24.

(a)

$$\begin{aligned} \bar{g}(x) &= \mathbb{E}_{\mathcal{D}}[g(x)] \\ &= \frac{1}{4} \int_{-1}^{1} \int_{-1}^{1} \frac{y_2 - y_1}{x_2 - x_1} x + \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2} dx_1 dx_2 \\ &= \frac{1}{4} \int_{-1}^{1} \int_{-1}^{1} \frac{x_2^2 - x_1^2}{x_2 - x_1} x + \frac{x_1 x_2^2 - x_2 x_1^2}{x_1 - x_2} dx_1 dx_2 \\ &= \frac{1}{4} \int_{-1}^{1} \int_{-1}^{1} (x_1 + x_2) x - x_1 x_2 dx_1 dx_2 \\ &= 0 \end{aligned}$$

(b) We can randomly generate data sets. Then we can fit the final hypothesis for each data set. Lastly we can test random values to approximate the expected out of sample error, bias and variance.

(c) We found that $\mathbb{E}[E_{\text{out}}] = 0.54$. We got a bias of 0.2 and a variance of 0.34. The bias variance decomposition of error holds. We plotted $\bar{g}$ vs f.



(d) From part (a), we know that $g(x) = (x_1 + x_2)x - x_1x_2$. Therefore, we have that:

$$
\begin{aligned}
\mathbb{E}[E_{\text{out}}] &= \mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_x[(g(x) - x^2)^2]\right] \\
\mathbb{E}[E_{\text{out}}] &= \frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}((x_1 + x_2)x - x_1x_2 - x^2)^2 dx dx_1 dx_2 \\
&= \frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}(x_1x + x_2x - x_1x_2 - x^2)^2 dx dx_1 dx_2 \\
&= \frac{8}{15}
\end{aligned}
$$

We can now compute the bias.

$$
\begin{aligned}
\text{bias} &= \mathbb{E}_x[(\bar{g}(x) - f(x))^2] \\
&= \frac{1}{2}\int_{-1}^{1} x^4 dx = \frac{1}{5}.
\end{aligned}
$$

We can now compute the variance.

$$
\begin{aligned}
\text{variance} &= \mathbb{E}_{\mathcal{D}}[\mathbb{E}_x[(\bar{g}(x) - g(x))^2]] \\
&= \frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}((x_1 + x_2)x - x_1x_2)^2 dx dx_1 dx_2 \\
&= \frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}(x_1x + x_2x - x_1x_2)^2 dx dx_1 dx_2 \\
&= \frac{1}{3}
\end{aligned}
$$

Thus, we see the analytically computed values are similar to the numerically computed values.