CSE 417T - HW2

467261 - Yifu Wang

2018 - 09 - 30

1.12 **(a)** Denote h_{mean} as $\bar{y}$. To prove $\bar{y}$ makes $E_{in}(h)$ minimal. Basically we only need to show $E_{in}(h) \geq E_{in}(\bar{y})$. Which is obviously since $E'_{in}(h) = 2\sum_{n=1}^{N}(h - y_n) = 2N(h - \bar{y})$, when $h < \bar{y}$ $E' < 0$, when $h > \bar{y}$ $E' > 0$, thus E reach its minimal value at $\bar{y}$. $\square$

(b) We try to calculate the derivative of E as we did in **(a)** but since E is sum of absolutes, we can redefine E to get rid of absolute.

$$E_{in}(h) = \begin{cases} \sum_{n=1}^{N}(y_n - h) & h \leq y_1 \\ \sum_{n=1}^{i}(h - y_n) + \sum_{n=i+1}^{N}(y_n - h) & y_i \leq h \leq y_{i+1}\ ,\ i \in [1, n-1] \\ \sum_{n=1}^{N}(h - y_n) & h \geq y_N \end{cases}$$

Thus we have

$$E'_{in}(h) = \begin{cases} -N & h \leq y_1 \\ 2i - N & y_i \leq h \leq y_{i+1}\ ,\ i \in [1, n-1] \\ N & h \geq y_N \end{cases}$$

E' is a monotonically decreasing function, and by the definition of h_{med}, E' will reach its zero point at h_{med}. Namely, E will reach its minimal value at h_{med}. $\square$

And notice that h_{med} will be a single value only if N is odd.

(c) h_{mean} will tend to ∞ in all condition.

h_{med} will tend to ∞ if $N \leq 2$, will remain same otherwise.

2.3 **(a)** Giving N points, $\mathbb{R}$ will be divide into $N + 1$ interval (assume no coincided since we just need the max value). The result of positive or negative ray is totaly depends on which interval contains a. Considering $a \in (-\infty, X_1) \cup (X_N, \infty)$, only two results will be produced all $+1$ and all -1. Thus $m_{\mathcal{H}}(N) = 2(N - 1) + 2 = 2N$, $d_{VC} = 2$. $\square$

(b) Firstly consider positive interval only, if at one X_i locate inside the interval $[a, b]$ then a and b must locate in different interval divided by X, which produce $(N+1)N/2$ results. Plus the all -1 condition that should be the max number of dichotomies of positive intervals. Take negative intervals into consider, the only thing we need to pay attention is the following situation. Denote a positive interval as $[a_1, b_1]$ and a negative interval as $[a_2, b_2]$. If $a_1 < X_1$ $b_2 > X_N$ and $b_1, a_1 \in [X_i, X_{i+1}]$. These two interval will produce same results. Thus $m_{\mathcal{H}}(N) = (2 + N)(N - 1) - 2(N - 1) + 2 = N^2 - N + 2$, $d_{VC} = 3$. $\square$

(c) This is equivalent to positive interval with $\mathcal{X} = [0, \infty)$ and $a, b \geq 0$, which won't affect the result. Thus $m_{\mathcal{H}}(N) = N(N + 1)/2 + 1$, $d_{VC} = 2$. $\square$

2.8 A valid growth function $m_{\mathcal{H}}(N)$ should firstly satisfy $\forall N \in \mathbb{N}^+, m_{\mathcal{H}}(N) \leq 2^N$, then $\forall k$ satisfy $m_{\mathcal{H}}(k) < 2^k, m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} C_N^i \forall N \in \mathbb{N}^+$. Thus once we find the d_{VC} we find the lower bound.

$\mathbf{1 + N}$ Good: The d_{VC} is 1.

$\mathbf{1 + N + \frac{N(N-1)}{2}}$ Good: The d_{VC} is 2.

$\mathbf{2^N}$ Good: No break point.

$2^{\lfloor\sqrt{N}\rfloor}$ Bad: The d_{VC} is 1, break its bound when $N \to \infty$.

$2^{\lfloor N/2 \rfloor}$ Bad: The d_{VC} is 0, break its bound when $N \to \infty$.

$1 + N + \frac{N(N-1)(N-2)}{6}$ Bad: The d_{VC} is 1, break its bound when $N \to \infty$.

2.10 This is actually obvious and almost involе no math. And I thought this derive from a more general formula

$$m_{\mathcal{H}}(N_1 + N_2) \le m_{\mathcal{H}}(N_1) m_{\mathcal{H}}(N_2)$$

When taking $N_1 = N_2 = N$ you get the $m_{\mathcal{H}}(2N) \le m_{\mathcal{H}}(N)^2$.
Since you could always devide a $N_1 + N_2$ size sample into two part size of N_1 and N_2 if the two parts is independent, you will get $m_{\mathcal{H}}(N_1 + N_2) = m_{\mathcal{H}}(N_1) m_{\mathcal{H}}(N_2)$ else you will get $m_{\mathcal{H}}(N_1 + N_2) < m_{\mathcal{H}}(N_1) m_{\mathcal{H}}(N_2)$. □

2.13 **(a)** Since $m_{\mathcal{H}}(N) \le M$, we have $2^{d_{VC}} \le M$, namely $d_{VC} \le \log_2 M$. □

(b) $0 \le d_{VC}(\cap_{k=1}^{K} \mathcal{H}_k) \le \max\limits_{1<k<K} d_{VC}(\mathcal{H}_k)$
This result is obvious by the definition of intersection of sets. And when there is no intersection the result will reach the lower bound, when all sets are equale the result will reach the upper bound. □

(c) $\max\limits_{1<k<K} d_{VC}(\mathcal{H}_k) \le d_{VC}(\cap_{k=1}^{K} \mathcal{H}_k) \le \sum\limits_{1<k<K} d_{VC}(\mathcal{H}_k)$
This result is obvious by the definition of union of sets. And when all sets are equale the result will reach the upper bound, when all sets are paire-wise unintersected the result will reach upper bound. □

2.22

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}}[E_{\text{out}}(g^{(\mathcal{D})})] =\ & \mathbb{E}_{\mathcal{D},x,y}[(g^{(\mathcal{D})}(x) - y(x))^2] \\
=\ & \mathbb{E}_{\mathcal{D},x}[(g^{(\mathcal{D})}(x))^2 - 2g^{(\mathcal{D})}(x)\mathbb{E}_y[y(x)] + \mathbb{E}_y[(y(x))^2]] \\
=\ & \mathbb{E}_{\mathcal{D},x}[(g^{(\mathcal{D})}(x))^2 - 2g^{(\mathcal{D})}(x)\mathbb{E}_\epsilon[f(x) + \epsilon] + \mathbb{E}_\epsilon[(f(x))^2 + 2\epsilon f(x) + \epsilon^2]] \\
=\ & \mathbb{E}_{\mathcal{D},x}[(g^{(\mathcal{D})}(x))^2 - 2g^{(\mathcal{D})}(x)f(x) + (f(x))^2 + \sigma^2] \\
=\ & \sigma^2 + \text{bias} + \text{var}
\end{aligned}$$

□

2.24 **(a)**

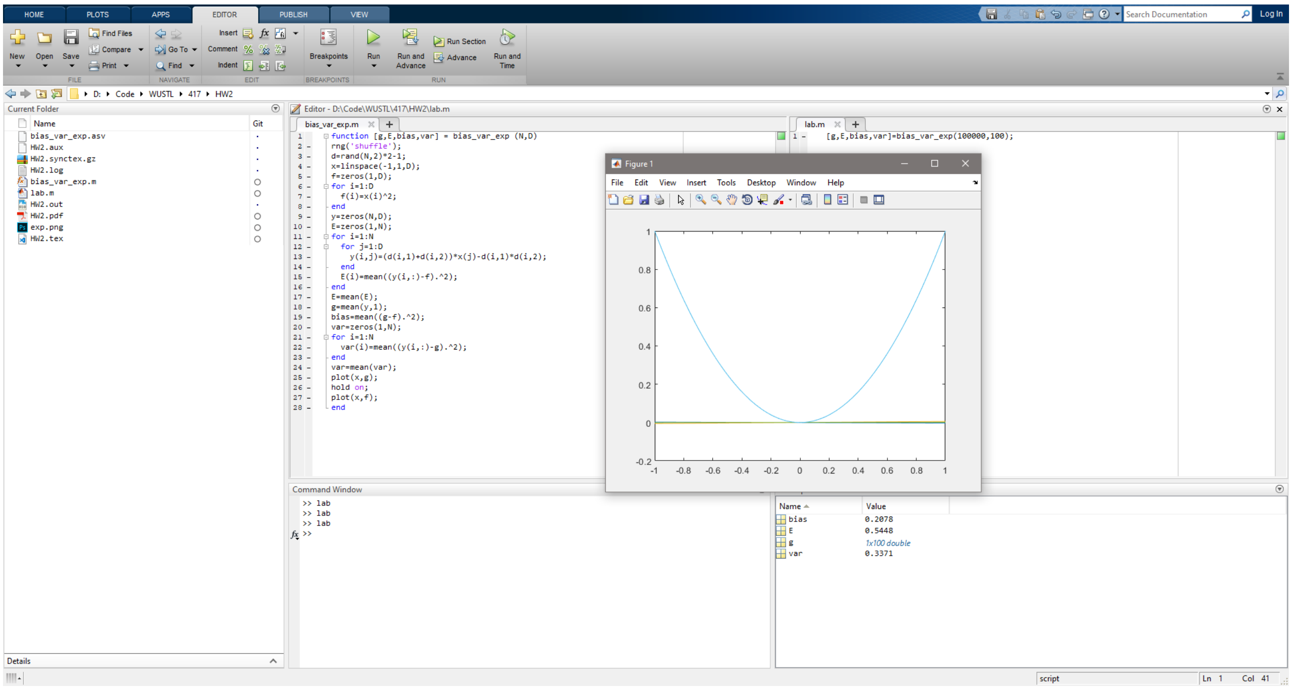
$$\begin{aligned}
\bar{g}(x) =\ & \mathbb{E}_{\mathcal{D}}[g^{(\mathcal{D})}(x)] \\
=\ & \mathbb{E}_{\{(x_1,x_1^2),(x_2,x_2^2)\}}[(x_1 + x_2)x - x_1 x_2] \\
=\ & 0
\end{aligned}$$

(b) Pick up many data sets $\mathcal{D}_i$, calculate $g^{(\mathcal{D})}(x)$. Then calculate $\bar{g}$ then bias and var then E_{out}.

(c) I repeated the experiment for 100,000 times, got

$$\begin{aligned}
\text{bias} =\ & 0.2078 \\
\text{var} =\ & 0.3371 \\
E_{\text{out}} =\ & 0.5448 \ \approx \ \text{bias} + \text{var} \\
|g(x)| \le\ & 10^{-3}
\end{aligned}$$

Here is all the result I got in one screenshot, code included. And I upload it to image host, click here if you feel the image below is low resolution, or you can zoom in.



(d) Bias

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

Var

$$
\begin{aligned}
var &= \mathbb{E}_{x,\mathcal{D}}[(g^{(\mathcal{D})}(x) - \bar{g}(x))^2] \\
&= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \frac{1}{8}\frac{3}{2}((y+z)x - yz)^2 dz dy dx \\
&= \frac{1}{3}
\end{aligned}
$$

$\mathbb{E}[E_{\text{out}}]$

$$
\begin{aligned}
\mathbb{E}[E_{\text{out}}] &= bias + var \\
&= \frac{8}{15}
\end{aligned}
$$