Assignment 3

Ashraful Islam
RIN: 661854753

1 Exercise 1.13

1.1 a

$$
\begin{aligned}
&P(\text{error that h makes in approximating y}) && (1)\\
&= P(h(x) \neq y) && (2)\\
&= P(h(x) \neq y | y = f(x))P(y = f(x)) + P(h(x) \neq y | y \neq f(x))P(y \neq f(x)) && (3)\\
&= \mu\lambda + (1-\mu)(1-\lambda) && (4)\\
&= \mu(2\lambda - 1) + 1 - \lambda && (5)
\end{aligned}
$$

1.2 b

From Eqn. 5, we get when $\lambda = 0.5$, $P(\text{error that h makes in approximating y})$ will be independent of μ.

2 Exercise 2.1

1. Positive rays: break points $k = 2$, as 2 points cannot be shattered by the ray. $m_H(k) = k + 1 = 3 < 2^k = 4$
2. Positive intervals: break points $k = 3$, and $m_H(k) = \binom{4}{2} + 1 = 7 < 2^3 = 8$.
3. Convex sets: There's no break point, as the convex set can shatter any N points.

3 Exercise 2.2

3.1 a

1. Positive rays: $m_H(N) = N + 1$. And $\sum_{i=0}^{k-1} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} = 1 + N$. So, $m_H(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$.
2. Positive intervals: $m_H(N) = \binom{N+1}{2} + 1 = \frac{N(N+1)}{2} + 1$. And $\sum_{i=0}^{k-1} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = 1 + N + \frac{N(N-1)}{2} = 1 + \frac{N(N+1)}{2}$. So, $m_H(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$.

3. Here, $k = \inf$, and $\sum_{i=0}^{k-1} \binom{N}{i} = \inf$. So, we still $m_H(N) = 2^N \leq \sum_{i=0}^{k-1} \binom{N}{i}$.

3.2 b

$m_H(N) = N + 2^{\lfloor \frac{N}{2} \rfloor}$. Now, $m_H(2) = 4 \not< 2^N = 4$, $m_H(3) = 5 < 2^3 = 8$. Hence, break point $k = 3$.

Now, we know, for all N, $m_H(N) \leq \sum_{i=0}^{k-1} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = 1 + N + \frac{N(N-1)}{2} = 1 + \frac{N(N+1)}{2}$. However, $N + 2^{\lfloor \frac{N}{2} \rfloor} < 1 + \frac{N(N+1)}{2}$ does not hold for all N, as the left side is exponential and right side is polynomial. Hence, the hypothesis cannot exist.

4 Exercise 2.3

We know, $d_{VC} = k - 1$. So,

1. Positive rays: $d_{vc} = 1$
2. Positive intervals: $d_{vc} = 2$
3. Convex sets: $d_{vc} = \inf$

5 Exercise 2.6

5.1 a

We know, $E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \log \frac{2m_H(N)}{\delta}}$, where $\sqrt{\frac{1}{2N} \log \frac{2m_H(N)}{\delta}}$ is the error bar. For training cases, $M = 1000$ and $N = 400$, training error bar $= 0.115$. For testing, $M = 1$ and $N = 200$, testing error bar $= 0.096$. So, training has higher error bar.

5.2 b

If we reserve more data for testing, then we will have less data for training. In that case, the training error bar will be larger, and we might not have a good hypothesis. Hence, even though the error bar will be smaller, $E_{\text{test}}(g)$ might be larger from where $E_{\text{in}}(g)$ might be larger too.

6 Problem 1.11

$$
\begin{aligned}
E_{\text{in}}(h) &= \frac{1}{N} \sum_{i=1}^{N} e(h(x_n), f(x_n)) \\
&= \frac{1}{N} (\sum_{f(x_n)=1} e(h(x_n), 1) + \sum_{f(x_n)=-1} e(h(x_n), -1))
\end{aligned}
$$

Now, for Supermarket,

$$
\begin{aligned}
E_{\text{in}}(h) &= \frac{1}{N}\Big(\sum_{f(x_n)=1} e(h(x_n),1) + \sum_{f(x_n)=-1} e(h(x_n),-1)\Big) \\
&= \frac{1}{N}\Big(\sum_{f(x_n)=1} 10[h(x_n \neq 1)] + \sum_{f(x_n)=-1} [h(x_n \neq -1)]
\end{aligned}
$$

Now, for CIA,

$$
\begin{aligned}
E_{\text{in}}(h) &= \frac{1}{N}\Big(\sum_{f(x_n)=1} e(h(x_n),1) + \sum_{f(x_n)=-1} e(h(x_n),-1)\Big) \\
&= \frac{1}{N}\Big(\sum_{f(x_n)=1} [h(x_n \neq 1)] + \sum_{f(x_n)=-1} 1000[h(x_n \neq -1)]
\end{aligned}
$$

Here, $[a \neq b] = 1$, if $a \neq b$, else it is 0.

7 Problem 1.12

7.1 a

Here, $E_{\text{in}}(h) = \sum_{n=1}^{N}(h - y_n)^2$. To minimize $E_{\text{in}}(h)$,

$$
\begin{aligned}
\frac{dE_{\text{in}}(h)}{dh} &= 0 \\
\implies 2\sum_{n=1}^{N}(h - y_n) &= 0 \\
\implies h = \frac{1}{N}\sum_{n=1}^{N} y_n &= h_{\text{mean}}
\end{aligned}
$$

7.2 b

$$
\begin{aligned}
\frac{dE_{\text{in}}(h)}{dh} &= 0 \\
\frac{d}{dh}\sum_{i=1}^{N}|h - y_n| &= 0
\end{aligned}
$$

We know that $\frac{d}{dx}|x| = +1$, if $x > 0$, otherwise -1. Hence, there should be equal number of $+1$ and -1 in the left side of the above equation, or h should be median of $\{y_n\}$, so that there will be equal number of $y_n > h$ as there is $y_n < h$.

7.3 c

If $y_N = y_N + \epsilon$, where $\epsilon \to \inf$, then From (a), $h_{\text{mean}} \to \inf$. However, h_{median} will remain the same. So (a) is sensitive to outlier whereas (b) is not.