Assignment 3

Name: Mohaiminul Al Nahian
RIN: 662026703
Course no.: CSCI 6100

1 Exercise 1.13

(a) The probability of h making error is

$$\begin{aligned} P(error) &= P(h \neq y|y = f(x)P(y = f(x)) + P(h \neq y|y \neq f(x)p(y \neq f(x)) \\ &= \mu * \lambda + (1 - \mu) * (1 - \lambda) \\ &= 2 * \mu * \lambda - \mu - \lambda + 1 \end{aligned}$$

(b) To make the performance of h independent of μ, we have to make the above equation independent of μ. So,

$$\begin{aligned} 2\mu\lambda - \mu &= 0 \\ or, 2\lambda - 1 &= 0 \\ or, \lambda &= 0.5 \end{aligned}$$

so, $\lambda = 0.5$ will make error independent of μ

2 Exercise 2.1

FOR POSITIVE RAYS: $m_{H(N)} = N + 1$

$m_H(1) = 1 + 1 = 2 = 2^1$
$m_H(2) = 2 + 1 = 3 < 2^2$ so, n=2 is a break point.

FOR POSITIVE INTERVAL: $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

$m_H(2) = 2 + 1 + 1 = 4 = 2^2$
$m_H(3) = 1/2 * 3^2 + 3/2 + 1 = 7 < 2^3$. So, n=3 is a break point for H.

FOR CONVEX SET: $m_H(N) = 2^N = 2^N$, so no break point exists for H

3 Exercise 2.2

(a) FOR POSITIVE RAYS: break point n=2 from previous exercise and $m_H(N) = N + 1$
$And, \sum_{i=0}^{n-1} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} = N + 1$
So, $m_H(N) \leq \sum_{i=0}^{n-1} \binom{N}{i}$ holds.

FOR POSITIVE INTERVAL: break point n=3 from previous exercise and $m_H(N) = \binom{N+1}{2} = \frac{N(N+1)}{2} + 1$
And, $\sum_{i=0}^{n-1} \binom{N}{i} = \binom{N}{0} + \binom{N}{1} + \binom{N}{2} = 1 + N + \frac{N(N-1)}{2} = \frac{N(N+1)}{2} + 1$
So, $m_H(N) \leq \sum_{i=0}^{n-1} \binom{N}{i}$ holds.

FOR CONVEX CASE: Since no break point, so Theorem 2.4 does not hold

(b) $m_H(N) = N + 2^{\lfloor N/2 \rfloor}$
$m_H(3) = 5 < 2^3$, so n=3 is a break point.
Now, we know that for all N, $m_H(N) \leq \sum_{i=0}^{n-1} \binom{N}{i}$

For, n=3, $\sum_{i=0}^{3-1}\binom{N}{i} = 1 + \frac{N(N+1)}{2}$
However, N+ $2^{\lfloor N/2 \rfloor} \leq 1 + \frac{N(N+1)}{2}$ does not hold for all N, because LHS increases exponentially with N, whereas RHS follows polynomial increase.
So, no such hypothesis exists.

4 Exercise 2.3

We know $d_{VC} = k - 1$

POSITIVE RAY: breakpoint n=2 from exercise 2.1, so, $d_{VC} = 2 - 1 = 1$

POSITIVE INTERVAL: breakpoint n=3 from exercise 2.1, so, $d_{VC} = 3 - 1 = 2$

CONVEX: breakpoint $n = \infty$ from exercise 2.1, so, $d_{VC} = \infty - 1 = \infty$

5 Exercise 2.6

(a) We know that, $E_{out}(g) \leq E_{in}(g) + \sqrt{\frac{1}{2N} ln \frac{2M}{\delta}}$
Here, error bar=$\sqrt{\frac{1}{2N} ln \frac{2M}{\delta}}$
Here, $\delta = 0.05$, For training, M=1000, N=400 and for testing, M=1, N=200

so, for training, error bar=$\sqrt{\frac{1}{2*400} ln \frac{2*1000}{0.05}} = 0.1151$

for testing, error bar=$\sqrt{\frac{1}{2*200} ln \frac{2*1}{0.05}} = 0.096$

so, training has higher error bar.

(b) If we reserve more examples for testing, then even though the error bar on testing may become less, but we shall have even less example for training. In that case we might not even find a good hypothesis g to begin with and our training may fail.

6 Problem 1.11

The matrix is given below:

		f: $+1$	f: -1
h	$+1$	0	1
	-1	10	0

Supermarket

		f: $+1$	f: -1
h	$+1$	0	1000
	-1	1	0

CIA

For Supermarket, $E_{in-supermarket}(h) = \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n))$
where, $e(h(x_n) = +1, f(x_n) = +1) = 0$, $e(h(x_n) = -1, f(x_n) = -1) = 0$, $e(h(x_n) = +1, f(x_n) = -1) = 1$ and $e(h(x_n) = -1, f(x_n) = +1) = 10$

For CIA, $E_{in-CIA}(h) = \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n))$
where, $e(h(x_n) = +1, f(x_n) = +1) = 0$, $e(h(x_n) = -1, f(x_n) = -1) = 0$, $e(h(x_n) = +1, f(x_n) = -1) = 1000$ and $e(h(x_n) = -1, f(x_n) = +1) = 1$

7 Problem 1.12

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

$$\frac{d}{dh}E_{in}(h) = 0$$
$$=> 2\sum_{n=1}^{N}(h - y_n) = 0$$
$$=> Nh - \sum_{n=1}^{N} y_n = 0$$
$$So, h = \frac{1}{N}\sum_{n=1}^{N} y_n = h_{mean}$$

(b) $E_{in}(h) = \sum_{n=1}^{N} |h - y_n|$. To minimize it,

$$\frac{d}{dh}E_{in}(h) = 0$$
$$=> \frac{d}{dh}\sum_{n=1}^{N} |h - y_n| = 0$$
$$=> \sum_{n=1}^{N} sign(h - y_n) = 0$$

In order for the LHS of the above equation to be equal to zero, there must be equal number of $y_n > h$ and $y_n < h$. So, h must be the median value.
Hence, $h = h_{med}$

(c) If y_N is perturbed to become $y_N + \epsilon$ where $\epsilon \to \infty$ then $h_{mean} \to \infty$, as the value of y_N is used to calculate h_{mean}.
However, h_{med} will remain unchanged as median is determined by the order of sample values, not by the value of any sample.