Exercise 1.13

a)

There are two situations where error could happen:

P[$h(x) \neq f(x)$ and $y = f(x)$] = $\mu \times \lambda$

P[$h(x) = f(x)$ and $y \neq f(x)$] = $(1 - \mu) \times (1 - \lambda)$

We have P[error] = $\mu \times \lambda + (1 - \mu) \times (1 - \lambda) = 2\mu\lambda - \lambda - \mu + 1$

b)

$\lambda = 0.5$

P[error] will always be 0.5 no matter how μ is changed; hence, h will be totally independent of μ.

Exercise 2.1

1.

Positive rays:

$m_H(N) = N + 1$

The break point is N = 2 since $m_H(2) < 2^2$, which is $3 < 4$

2.

Positive intervals:

$m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

The break point is N = 3 since $m_H(3) < 2^3$, which is $7 < 8$

3.

Convex sets:

$m_H(N) = 2^N$

There is no break point since $m_H(N)$ is always equal to the 2^N.

Exercise 2.2

a)

i)

The break point is 2 for the positive ray.

We have $m_H(N) \leq \sum_{i=0}^{1} \binom{N}{i}$

For positive ray, LHS = N + 1

RHS = $\frac{N!}{0!\times N!} + \frac{N!}{1!\times(N-1)!} = 1 + N$

LHS <= RHS for all N.

Verified theorem 2.4.

ii)

The break point is 3 for the positive interval.

We have $m_H(N) \leq \sum_{i=0}^{2} \binom{N}{i}$

LHS = $\frac{1}{2}N^2 + \frac{1}{2}N + 1$

RHS =

$\frac{N!}{0!\times N!} + \frac{N!}{1!\times(N-1)!} + \frac{N!}{2!\times(N-2)!} = 1 + N + \frac{N(N-1)}{2} = 1 + N + \frac{1}{2}N^2 - \frac{1}{2}N = 1 + \frac{1}{2}N^2 + \frac{1}{2}N$

LHS <= RHS for all N.

Verified theorem 2.4.

iii)

There is no break point for convex sets.

k = N+1

We have $m_H(N) \leq \sum_{i=0}^{N} \binom{N}{i}$, since $\sum_{i=0}^{k-1} \binom{N}{i} \leq N^{k-1} + 1$

LHS = 2^N

RHS $\leq N^N + 1$

LHS must be smaller than or equal to the RHS for all N.

Verified theorem 2.4

b)

NO.

According to the theorem 2.4, we got $N + 2^{N/2} \leq \sum_{i=0}^{k-1} \binom{N}{i}$ for all N. Since $\sum_{i=0}^{k-1} \binom{N}{i} \leq N^{k-1} + 1$, we could deduce that $N + 2^{N/2} \leq N^{k-1} + 1$.

The above statement can not be true since $2^{N/2}$ of LHS grows exponentially, while N^{k-1} of RHS grows polynomially.
This hypothesis does not obey the bound stated in theorem 2.4. Hence, this hypothesis does not exist.

Exercise 2.3
i) Positive rays
Since the break point is N = 2, we have
$d_{vc}(H) = 2 - 1 = 1$

ii) Positive intervals
Since the break point is N = 3, we have
$d_{vc}(H) = 3 - 1 = 2$

iii) Convex sets
Since $m_H(N) = 2^N$ for all N, we have
$d_{vc}(H) = \infty$

Exercise 2.6

a)

Apply $\sqrt{\frac{1}{2N}ln(\frac{2M}{\delta})}$ from (2.1)

Training examples:

$\sqrt{\frac{1}{2\times400}ln(\frac{2\times1000}{0.05})} \approx 0.115$

Testing examples:

$\sqrt{\frac{1}{2\times200}ln(\frac{2\times1}{0.05})} \approx 0.096$

Training examples has a higher 'error bar'.

b)

If we reverse more testing examples, we will have less examples for training. The error bar for E_{out} will definitely increase; the error bar for testing will decrease. We do not get the hypothesis with the performance as well as on our previous training examples; it is getting far away from our estimation of E_{out} which is E_{test}.

Problem 1.11

Supermarket:

$$E_{in}(h) = \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n))$$

$$= \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), 1) + e(h(x_n), -1)$$

$$= \frac{1}{N}\sum_{n=1}^{N} [h(x_n) \neq 1] \times 10 + [h(x_n) \neq -1] \times 1$$

CIA:

$$E_{in}(h) = \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), f(x_n))$$

$$= \frac{1}{N}\sum_{n=1}^{N} e(h(x_n), 1) + e(h(x_n), -1)$$

$$= \frac{1}{N}\sum_{n=1}^{N} [h(x_n) \neq 1] \times 1 + [h(x_n) \neq -1] \times 1000$$

Problem 1.12

a)

$$E_{in}(h) = \sum_{n=1}^{N} (h - y_n)^2 = \sum_{n=1}^{N} h^2 - 2hy_n + y_n^2$$

$$= \sum_{n=1}^{N} h^2 - \sum_{n=1}^{N} 2hy_n + \sum_{n=1}^{N} y_n^2$$

$$= Nh^2 - 2h \times \sum_{n=1}^{N} y_n + \sum_{n=1}^{N} y_n^2$$

Compute the derivative of $E_{in}(h)$

$$E'_{in}(h) = 2Nh - 2\sum_{n=1}^{N} y_n$$

If we want to compute the minimum, let $E'_{in}(h) = 0$, we have,

$$2Nh - 2\sum_{n=1}^{N} y_n = 0$$

$$2Nh = 2\sum_{n=1}^{N} y_n$$

$$h = \frac{2\sum_{n=1}^{N} y_n}{2N} = \frac{1}{N}\sum_{n=1}^{N} y_n = h_{mean}$$

b)

$$E_{in}(h) = \sum_{n=1}^{N} \left|h - y_n\right|$$

In order to minimize the $E_{in}(h)$, we need $E'_{in}(h) = 0$.

In order to satisfy this, we need have half of the y_n at most h while half of y_n at least h. In this case, all h in the sum can be cancelled out with rest of term y_ns.

$E'_{in}(h) = 0$ under the above situation.

Hence, we can conclude that the estimate is the h_{med}.

c)

h_{mean} will become ∞ while h_{med} will not change.

Since y_N is being perturbed to ∞, $\sum_{n=1}^{N} y_n$ will become ∞. Hence, the h_{mean} will be ∞. As for h_{med}, its position remains unchanged as y_N approaches ∞; hence the value of h_{med} will not change.