1. Exercise 1.13

    (a) two types of error, false reject ($f(x) = +1$, $h(x) = -1$) and false accept ($f(x) = -1$, $h(x) = 1$), when correctly defined, no errors.
    False reject $f = y$, $h \neq y = \lambda(\mu)$
    False accept $f \neq y$, $h = y = (1-\lambda)(1-\mu)$

    (b) 0.5, then $P(y|x) = 0.5$ regardless of $y = f(x)$ or $y \neq f(x)$, it becomes completely random.

2. Exercise 2.1

    (a) $m_H(N) = N + 1$, try $k = 2$, $m_H(2) = 3$, $2^k = 4$, $m_H(k) < 2^k$, breakpoint is at $k = 2$.

    (b) $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$, try $k = 3$, $m_H(3) = \frac{9}{2} + \frac{3}{2} + 1 = 7$, $2^3 = 8$, $m_H(3) < 2^3$, so breakpoint is at $k = 3$.

    (c) no breakpoint for convex sets as $m_H(N) = 2^N$.

3. Exercise 2.2

    (a) $N + 1 \leq \sum_{i=0}^{k-1} \binom{N}{i}$, let $N = 2$ using our previous breakpoint, then $3 \leq \binom{2}{0} + \binom{2}{1}$, and $3 \leq 3$, which is true. We can try again for $N = 3$, where we get the result of $4 \leq 7$, which is also true.

    (b) $\frac{1}{2}N^2 + \frac{1}{2}N + 1 \leq \sum_{i=0}^{k-1} \binom{N}{i}$, let $N = 3$ and solving, from (2b), we know it's equal to 7, solving $\sum_{i=0}^{2} \binom{3}{i} = 7$, if we solve for $N = 4$, we will get the result of $11 \leq 15$, which is also true.

    (c) Theorem doesn't apply since $m_H(N) = 2^N$, and theorem requires that $m_H(N) < 2^N$

4. Exercise 2.3

    (a) From 2.1, we know $k = 1$ is the largest value of N where $m_H(N) = 2^N$, therefore $d_{vc} = 1$ and breakpoint $k = 1 + 1 = 2$.

    (b) From 2.1 we know $k = 2$ is the largest value of N where $m_H(N) = 2^N$, therefore $d_{vc} = 2$ and breakpoint $k = 2 + 1 = 3$.

    (c) $d_{vc} = \infty$ since $m_H(N) = 2^N$ for all N and there is no breakpoint.

5. Exercise 2.6

    (a) We can use formula

$$E_{out} \leq E_{in} + \sqrt{\frac{1}{2N} ln(\frac{2M}{\delta})}$$

$$E_{out} = \sqrt{\frac{1}{2(400)} ln(\frac{2(1000)}{0.05})} = 0.115$$

$$E_{test} = \sqrt{\frac{1}{2(200)} ln(\frac{2(1)}{0.05})} = 0.096$$

    E_{out} has the higher error bar

(b) Samples used for testing cannot be used in E_{out}, this will improve E_{test}, but worsen E_{out}, which is meaningless in the end.

6. Problem 1.11

We can multiply the result by their weights corresponding in the form of

$$\frac{1}{N}\sum_{i=1}^{N} w_n \times [h(x_n) \neq f(x_n)]$$

Supermarket

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

CIA

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

7. Problem 1.12

(a) We want to minimize $E_{in}(h)$, so we can consider taking its derivative, and when it equals 0, it is the minimal value.

$$\begin{aligned}
E_{in}(h) &= \sum_{n=1}^{N}(h - y_n)^2 \\
\frac{\partial E_{in}(h)}{\partial h} &= 2\sum_{n=1}^{N}(h - y_n)\frac{\partial(h - y_n)}{\partial h} \text{ chain rule} \\
&= 2\sum_{n=1}^{N}(h - y_n) \\
&= 2(\sum_{n=1}^{N} h - \sum_{n=1}^{N} y_n) \\
&= 2(Nh - \sum_{n=1}^{N} y_n) \\
&= 2[N(h - \frac{1}{N}\sum_{n=1}^{N} y_n)]
\end{aligned}$$

When $h = \frac{1}{N}\sum_{n=1}^{N} y_n$, the equation is minimized since it equals 0. since $2[N(0)] = 0$

(b) **TO DO**

(c) h_{mean} becomes ∞ since mean is calculated by the average, when there exists an infinitely large number, the mean gets pulled up to infinity.

h_{med} will likely not change because median is calculated by the middle "ordered" value, when there exists an infinitely large number, the median will not change.