Exercise 1.13:
Part 1:
Case 1: h(x) = f(x) and y != f(x): $(1 - \mu)(1 - \lambda)$
Case 2: h(x) != f(x) and y = f(x): $\mu\lambda$
So the probability is: $(1 - \mu)(1 - \lambda) + \mu\lambda$

Part 2:
$(1 - \mu)(1 - \lambda) + \mu\lambda$
$=1 - \mu - \lambda + 2\mu\lambda$
$=1 - \lambda + \mu(2\lambda - 1)$
So $\mu(2\lambda - 1)$ needs to be equal to 0
$\lambda = 0.5$

Exercise 2.1:
Positive rays:
mH(N) = N+1
mH(1) = 1 + 1 = 2 = 2^1
mH(2) = 2 + 1 < 2^2
Break point at k = 2

Positive interval:
mH(N) = ½N^2 + ½N + 1
mH(1) = ½ + ½ + 1 = 2^1 = 2
mH(2) = ½ * 4 + ½ * 2 + 1 = 2^2 = 4
mH(3) = ½ *9 +½ * 3 + 1 = 7 < 2^3 = 8
Break point at k = 3

Convex set:
No Breakpoint, mH(N) = 2^N

Exercise 2.2:
Part a:
$\sum_{i=0}^{k-1} (N,\ i)$
Positive rays:
At break point k = 2
mH(2) = N + 1 <= $\sum_{i=0}^{2-1} (N,\ i)$ = (N, 0) + (N, 1) <= 1 + N

Positive interval::
At break point k = 3
mH(2) = ½N^2 + ½N + 1 <= $\sum_{i=0}^{3-1} (N,\ i)$ = (N, 0) + (N, 1) + (N, 2)

½N^2 + ½N + 1 <= 1 + N + N(N - 1)/2
½N^2 + ½N + 1 <= 1 + ½N + ½N^2

Convex set:
From exercice 2.1 we no we don’t have a breakpoint for Convex set, so the bond is not exist

Part b:
mH(N) = N + 2^$\left[\frac{N}{2}\right]$
mH(1) = 1 + 2^$\left[\frac{1}{2}\right]$ = 1 + 1 = 2 = 2^1
mH(2) = 2 + 2^$\left[\frac{2}{2}\right]$ = 2 + 2 = 4 = 2^2
mH(3) = 3 + 2^$\left[\frac{3}{2}\right]$ = 3 + 2 = 5 < 2^3 = 9
The break point is at k = 3
Chack the bond:
$\sum_{i=0}^{3-1}(N,\ i)$ = (N, 0) + (N, 1) + (N, 2) = 1 + ½N + ½N^2
mH(3) = N + 2^$\left[\frac{N}{2}\right]$

Since 2^N is growing faster than N^2, so mH(N) > $\sum_{i}^{k-1}(N,\ i)$ at the break point where k = 3

Answer: No

Exercise 2.3:
From definition on the textbook:
If mH(N) = 2^N, for all N $d_{vc} = \infty$, else k = d_{vc} + 1
Positive ring:
From exercise 2.1 we know k for positive ring is 2,
d_{vc} = k - 1 = 2 - 1 = 1.
d_{vc} = 1
Positive interval:
From exercise 2.1 we know k for positive interval is 3
d_{vc} = 3 - 1 = 3 - 1 = 2.
d_{vc} = 2
Convex set:
From exercise 2.1 we know k for convex set is ∞
$d_{vc} = \infty$

Exercise 2.6:
Part a:
For E_{out}:

δ = 0.05, M = 1000, N = 400,
Plug in the formula we get E_{out} = 0.115

For E_{test}:
δ = 0.05, M = 1000, N = 400,
Plug in the formula we get E_{out} = 0.096

So E_{out} has a higher error bar

Part b:
Because if we use more data to do testing, we will have less data to be used in training, as the result of the g we find from the training set will be less accurate.

Problem 1.11
Supermarket:

$E_{in} = \frac{1}{N}\sum_{n=1}^{N}(10*[h(x_n) = -1, f(x_n) = 1] + [h(x_n) = 1, f(x_n) = -1])$

CIA:

$E_{in} = \frac{1}{N}\sum_{n=1}^{N}([h(x_n) = -1, f(x_n) = 1] + 1000 * [h(x_n) = 1, f(x_n) = -1])$

Problem 1.12:
Part a:

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

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

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

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

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

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

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

$\frac{1}{N}\sum_{n=1}^{N} y_n$ and $\sum_{n=1}^{N} y_n^2$ is constant, when $E_{in}(h)$ reaches min, $N(h - \frac{1}{N}\sum_{n=1}^{N} y_n)^2 = 0$

h - $\frac{1}{N}\sum_{n=1}^{N} y_n$ = 0

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

Part b:

Assume there is j number that is less or equal to h_{med}:

We can rewrite the equation as:

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

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

If we take derivative of E_{in} on h, since $\sum_{n=1}^{j} y_n$ and $\sum_{n=j+1}^{N} y_n$ are constant, so the derivative of them is 0.

E_{in}' = j - (N - j)

E_{in}' = -N + 2j.

E_{in} is minimum when E_{in}' = 0, so j = ½N when E_{in} is minimum, so there is j number of data point that is smaller than h_{med}, which is half of total data point, which means there will also be half of the data point that is greater than h_{med}

Part c:

h_{mean} will be growing the infinite, since the sum will be infinite. h_{med} will be unchanged because the largest number will not affect the median number unless N <= 2.