Exercise 2.4
Part a

$$X = \begin{Bmatrix} x_{00} & x_{01} & x_{02} & x_{03} & \cdots & x_{0d+1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{d+1\,0} & x_{d+1\,1} & x_{d+1\,2} & x_{d+1\,3} & \cdots & x_{d+1\,d+1} \end{Bmatrix}$$

y = [y1, y2, y3 …. yd+1]
We have xw = y, since X is invertible so we can also get w = yx^-1. We can use equation "w = yx^-1" to get all ws, so there must be a w for each y. Since we have d+1 ys, means will have d+1 w, so there is at least d+1 data point can be shattered
Part b:
We know y = sign((w^T)*x)
assume that x_{d+2} can be linearly represented as its d + 2 points in a d + 1 dimension.
So we will have:

$x_{d+2} = \sum_{i=1}^{d+1} c_i x_i$

x_{d+2}w^T = $\sum_{i=1}^{d+1} c_i x_i$w^T

sign(x_{d+2}w^T) = sign($\sum_{i=1}^{d+1} c_i x_i$w^T)

So the result of sign(x_{d+2}w^T) won't be affected by whatever input we have for x_{d+2}, it will only depend on the first d+1 data. Which means there are some dichotomies that won't be able to be implemented.

Problem 2.3:
Part a:
With N points, the line can split into N + 1 regions for each of the positive and negative rays.
mH(N) = 2N+2, consider there is situation where points are all -1 or +1, so mH(N) = 2N
mH(2) = 2*2 = 4 = 2^2
mH(3) = 2*3 = 6 < 2^3
d_{vc}= 2

Part b:
With N points, we need choose 2 spot that is in between data points, so there is N + 1 spot can be choose from, but if we choose the 2 same spot will be the same as choose from the start of

the data set to the end of the data set(data points are all +1 or all -1), so we need to subtract 2N case, also the end regions fall in the same religion, so need to add 2. We have:
$2(\left(\frac{N+1}{2}\right))$ - 2N +2 (can't do (N+1) choose 2 on docs so just show like this)

$\left(\frac{N+1}{2}\right)$ = (N+1)N -2N +2 = N^2 - N + 2,
mH(3) = 9 - 3 + 2 = 8 = 2^3
mH(4) = 4*4 - 4 + 2 = 14 < 2^4
d_{vc} = 3

Part c:
Similar to part c, but in part c interval can be +1 or -1, but for this one it will always be -1 in the interval.
mH(2) = (2^2-2+2)/2 =4 = 2^2
mH(3) = (3^3 -3+2)/2 <2^3
d_{vc} = 2
Problem 2.8:

$$m_{\mathcal{H}}(N) \leq N^{d_{\mathrm{vc}}} + 1.$$

We have

$$2^{\lfloor \sqrt{N} \rfloor};\ 2^{\lfloor N/2 \rfloor}$$

For , since they are not polynomials about N and they are not 2^N, so they can't be possible growth functions.

1 + N: mH(1) = 2 = 2^1, mH(2) = 3 < 2^2
So d_{vc} = 1, mH(N) <= N^d_{vc} +1 <= N +1 true.
it is growing function

1 + N + N(N-1)/2: mH(1) = 2 = 2^1, mH(2) = 4 = 2^2, mH(3) = 7 < 2^3
So d_{vc} = 2, mH(N) <= N^d_{vc} +1 <= N^2 +1 true.
it is growing function

2^N will always = 2^N
So d_{vc} = infinite, it is growing function

$$2^{\lfloor \sqrt{N} \rfloor}$$

: mH(1) = 2 = 2^1, mH(2) = 2 < 2^2
So d_{vc} = 1, but mH(N) !<= N^d_{vc},
So it can't be a growth function

$2^{\lfloor N/2 \rfloor}$

: mH(1) = 1 < 2^1

So d_{vc}= 0, it can't be a growth function

$$1+N+\frac{N(N-1)(N-2)}{6}.$$

mH(1)>2^1, mH(2) < 2^2

So d_{vc}= 1, but 1 + N + N(N-1)(N-2)/6 > N^d_{vc} +1= N+1

So it can't be a growth function.

Therefore 1 + N, 1 + N + N(N-1)/2, 2^N are the possible growth function

Problem 2.10

We can separate 2N points into 2 separated data sets each with N points. The max dichotomy for each N points data set is mH(N), so the combination of 2N points will be max mH(N)^2.

So, mH(2N) <= mH(n)^2

Generalization bond: $E_{out}(g) <= E_{in}(g) + \sqrt{\frac{8}{N} ln\frac{4m_H(N)^2}{\delta}}$

Problem 2.12

Form:

Theorem 2.5 (VC generalization bound). For any tolerance $\delta > 0$,

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

with probability $\geq 1 - \delta$.

So we have $\delta = 0.05$, and d_{vc} = 10,

$$m_{\mathcal{H}}(N) \leq N^{d_{\text{vc}}} + 1.$$

Also

So P($E_{out}(g) <= E_{in}(g) + \sqrt{\frac{8}{N} ln\frac{4m_H(2N)^2}{\delta}}$) >=0.95

P($E_{out}(g) <= E_{in}(g) + \sqrt{\frac{8}{N} ln\frac{4((20)^2+1)}{0.05}}$) >=0.95

After iterator in python(code submitted), I got N = 452957