Exercise 2.8:
Part a:
A: $\overline{g}$ represent the average of the function g($g_1, g_2, g_3 ... g_n$), so $\overline{g}$ is linear combination of $g_1, g_2, g_3 ... g_n$. Since H is closed under linear combination, and $g_1, g_2, g_3 ... g_n$ belongs to H, so $\overline{g}$ has to be in the H.
Part b:
A: H = {g_1, g_2}
g_1 = 0
g_2 = 1
$\overline{g}$ in this example is not in H.
Part c:
$\overline{g}$ can be a binary function but it won't always be a binary function. As part b as example, $\overline{g}$ is not a binary function, but if there's another g_3 in H and g_3 = -1. Now $\overline{g}$ is a binary function.

Problem 2.14
Part a:
So we know d_{vc}(H) <= $\sum_{1}^{K} d_{vc}(H_1)$
Each Hi has a VC dimension of dvc
$\sum_{i=1}^{K} d_{vc}(H_i)$ = K * d_{vc}
d_{vc}(H) <= K * d_{vc}
Since K is a positive number, if we add K to the right side the less or equal sign will become less sign.
So: d_{vc}(H) < K * d_{vc} + K
d_{vc}(H) < K * (d_{vc}+ 1)

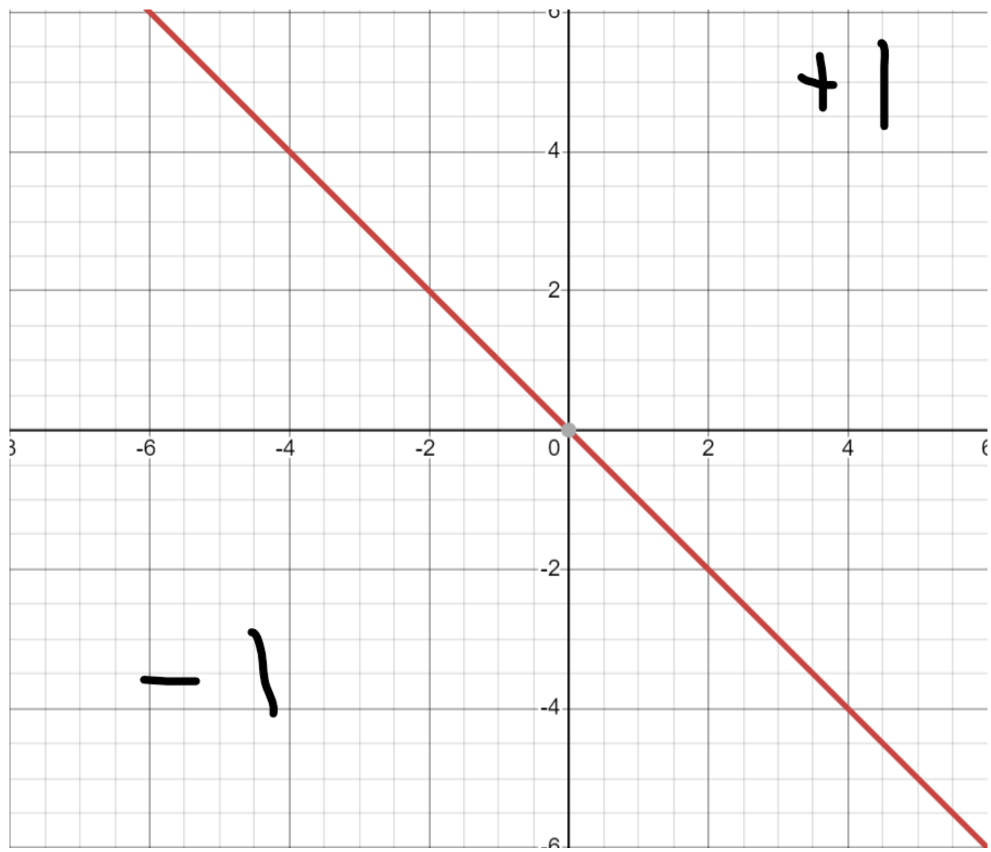
Part b:
We know: $m_H(l) <= l^{d_{vc}} + 1$
Since K and $l^{d_{vc}}$ are both positive number
$l^{d_{vc}} + 1$ <= K($l^{d_{vc}} + 1$)
$m_H(l) <= Kl^{d_{vc}} + K$
$Kl^{d_{vc}} >= K$, so $Kl^{d_{vc}} + K$ < = 2$Kl^{d_{vc}}$
From the question we know 2^l > 2$Kl^{d_{vc}}$
So $m_H(l)$ < 2^l

Since $m_H(l)$ <= 2^d_{vc}(H)
So d_{vc}(H) <= l

Part c:
From part a we know d_{vc}(H) < K * (d_{vc}+ 1), so only need to prove d_{vc}(H) <= 7(d_{vc}+K)$log_2 d_{vc}$K
We set l from part b to 7(d_{vc}+K)$log_2 d_{vc}$K.
Plug into 2^l > $2Kl^{d_{vc}}$
We have 2^(7(d_{vc}+K)$log_2 d_{vc}$K) > 2K$(7(d_{vc} + K)log_2 d_{vc} K)^{d_{vc}}$
Take log from each side

7d_{vc} $log_2 d_{vc}$K+7K$log_2 d_{vc}$K > 1 + log_2K + 7$log_2 d_{vc}$ + $log_2 d_{vc}$(d_{vc}+ K) + $log_2 log_2 d_{vc} K$
Since d_{vc} $log_2 d_{vc}$K> 1
d_{vc} $log_2 d_{vc}$K > log_2K
d_{vc} $log_2 d_{vc}$K > $log_2 log_2 d_{vc} K$
d_{vc} $log_2 d_{vc}$K > 7$log_2 d_{vc}$
7K$log_2 d_{vc}$K > 7$log_2 d_{vc}$
So left side must be greater than right side, so 2^l > 2Kl^d_{vc} when l = 7(d_{vc}+K)$log_2 d_{vc}$K, so we can use conclusion from part b d_{vc}(H) <= l, at l = 7(d_{vc}+K)$log_2 d_{vc}$K, so d_{vc}(H) <= 7(d_{vc}+K)$log_2 d_{vc}$
K
So we proved d_{vc}(H) < K * (d_{vc}+ 1) from part 1, and d_{vc}(H) <= 7(d_{vc}+K)$log_2 d_{vc}$K from part b.

Problem 2.15
Part a:
h(x_1, x_2) = sign($x_1 + x_2$)



Part b:

From the hint, assume we generated the first point x_a and generating the second point by increase x_1 of x_a, and decreasing x_2 of x_a, and continue doing this after generating N points. Any two points in those N points won't have any relation. Thus, m(H) = 2^N, $d_{vc} = \infty$

Problem 2.24

Part a:

Give $\mathcal{D} = \{(x_1, x_1^2), (x_2, x_2^2)\}$.

And g(x) = ax + b

x_1^2= ax_1+ b

x_2^2= ax_2+ b

There we know a = x_1 + x_2

b = -$x_1 x_2$

g(x) = (x_1 + x_2)x -$x_1 x_2$

$\overline{g}(x) = \frac{1}{2} * \frac{1}{2}\int_{-1}^{1}\int_{-1}^{1}(x_1 + x_2)x - x_1x_2\, dx_1 dx_2 = 0$

Part b:

$$E_{\text{out}}(g^{(\mathcal{D})}) = \mathbb{E}_{\mathbf{x}}\left[(g^{(\mathcal{D})}(\mathbf{x}) - f(\mathbf{x}))^2\right]$$

$$\text{bias}(\mathbf{x}) = (\bar{g}(\mathbf{x}) - f(\mathbf{x}))^2,$$

$$\text{var}(\mathbf{x}) = \mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(\mathbf{x}) - \bar{g}(\mathbf{x}))^2],$$

Iterate N time, each time generate two random points in range [-1,1] each time. Get gout in iteration, and compute the average after iteration. calculate $\bar{g}$(x) by getting average of all g(x), compute bias by using $\bar{g}$(x) and given f(x), compute var by compare $\bar{g}$(x) to each g(x)
Part d:

$$E_{\text{out}}(g^{(\mathcal{D})}) = \mathbb{E}_{\mathbf{x}}\left[(g^{(\mathcal{D})}(\mathbf{x}) - f(\mathbf{x}))^2\right]$$

g(x) = (x1+x2)x - x1x2
f(x) = x^2
Eout = E[((x1+x2)x - x1x2 -x^2)^2]
= $\frac{1}{2}\frac{1}{2}\frac{1}{2}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}$((x1+x2)x - x1x2 -x^2)^2 dx1dx2 dx
= 0.533

$$\text{bias}(\mathbf{x}) = (\bar{g}(\mathbf{x}) - f(\mathbf{x}))^2,$$

$\bar{g}$(x) - f(x) =(0-x^2)
Bias = $\frac{1}{2}\int_{-1}^{1}$x^2^2
=0.2

$$\text{var}(\mathbf{x}) = \mathbb{E}_{\mathcal{D}}[(g^{(\mathcal{D})}(\mathbf{x}) - \bar{g}(\mathbf{x}))^2],$$

Var = $\frac{1}{2}\frac{1}{2}\frac{1}{2}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}$(x1+x2)x - x1x2)^2 dx1dx2 dx
Var = 0.333