Problem 1.12

a) to find the minimum

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

$$\frac{\partial^2 E_{in}(h)}{\partial h} = 2N \quad > 0$$

thus setting $\frac{\partial E_{in}(h)}{\partial h} = 0$ we find the minimum

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

b)

Suppose there are totally $n$ points.

If we set up a starting point to be between point #a and #b that $a, b < \frac{n}{2}$

then if we move the point $h$ to right by $\Delta$

$$E_{in}(h)^{new} = E_{in}(h)_{old} + a\cdot\Delta - (n-a)\Delta \downarrow$$

thus if we keeping move $h$ toward the mid point, $E_{in}(h)$ could get its minimum

for odd $n$, it lies exactly on point $\frac{n-1}{2}+1$

for even $n$ it lies between point $\frac{n}{2}, \frac{n}{2}+1$

c) for $\epsilon \to \infty$, $\bar{y} \to \infty$. Which means $h_{mean}$ would have a drastic change but, with $h_{med}$ does not change much.

problem 2.3

a)

We know for positive ray we have $m_H(N) = N+1$

for negative-ray we also have $m_H(N) = N+1$

However, the case for all + and all − are overlaped

thus totally we have $2(N+1) - 2 = 2N$

the largest value for $m_H(N) = 2^N$ is 2, thus $d_{vc} = 2$

b) we know for positive intervals we have $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$

for negative intervals we have $m_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1$ also

However there are two case are counted.

which are $\{\underbrace{+, + \cdots +}_{a}\ \underbrace{-, \cdots, -}_{b}\}$ and $\{\underbrace{-, \cdots, -}_{a}\ \underbrace{+, \cdots +}_{b}\}$

$a+b = N$, thus total $2N$ cases over counted

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

with $d_{vc} = 3$

c)

It can be seen as a high dimensional positive inter, with a $\mathbb{R}^d$ circle to cut a $\mathbb{R}^d$ space and inside is +1 out side -1

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

with $d_{vc} = 2$

Problem 2.8.

growth function , equal to $2^N$ or bounded by $N^{d_{vc}}+1$

① $1+N$ : $d_{vc}=1$. is bounded by $N^1+1$

thus is a possible growth function

② $1+N+\frac{N(N-1)}{2}$ , $d_{vc}=2$. is bounded by $N^2+1$

thus $1+N+\frac{N(N-1)}{2}$ is possible growth function

③ $2^N$ is the growth function of convex sets.

④ $2^{\lfloor\sqrt{N}\rfloor}$ $\quad 2^{\lfloor\sqrt{1}\rfloor}=2^1$ $\quad 2^{\lfloor\sqrt{2}\rfloor}=2<2^2$

thus $d_{vc}=1$

However when N is large it can't be bound by $N+1$

($N=25$ as an example) it is not a growth function

⑤ $2^{\lfloor\frac{N}{2}\rfloor}$ $\quad 2^{\lfloor\frac{1}{2}\rfloor}=2^0<2^1$ thus $d_{vc}=0$

However when N is large it cannot be bound by $1+1$

it is not a growth function

⑥ $1+N+\frac{N(N-1)(N-2)}{6}$

$1+1+0=2=2^1$

$1+2+0=3<2^2$. thus $d_{vc}=2$

it can be bound by $N^2+1$ Thus is possible.

problem 2.10

Dichotomy of 2N points can be viewed as a dichomy of first N points and second N points.

Denote $m_H(N) = p$
then for the 2N case both first N and second N has the upper bound p, And dichotomy of 2N would be cross product of the elements for the first and second N,
which means

$$m_H(2N) \le m_H(N)^2$$

VC-bound

$$E_{out}(g) \le E_{in}(g) + \sqrt{\frac{8}{N}\ln\frac{4m_H(2N)}{\delta}}$$

$$\le E_{in}(g) + \sqrt{\frac{8}{N}\ln\frac{4m_H(N)^2}{\delta}}$$

( $\ln(\cdot)$ and $\sqrt{\cdot}$ monotonic increasing )

problem 2.13

a) Denoted $d_{vc}(H) = k$, then $m_H(k) \le 2^k$

then we have

$$2^k = m_H(k) = \max_{x_1 \cdots x} \left| H(x_1 \cdots x_m) \right|$$

$$= \max \left| \{ h(x_1) \cdots h(x_m) ,\ h \in H \} \right|$$

$$\le |H| = M$$

thus $d_{vc}(H) < \log_2 M$

b)

The lower bound is obvious 0, if we have one case for $m_H(N) = 1$. ($d_{vc} = 0$)

for the upper bound, we proof the guess by contradiction.

Denote $k_1 = d_{vc}(\cap_{k=1}^{K} H_k)$, $k_2 = \min\{(d_{vc}(H_k))\}_{k=1}^{K}$

if $k_1 \ge k_2 + 1$, which means $\cap_{k=1}^{K} H_k$ can shatter $k_2+1$ elements, Thus at lease $\exists H_i$ $i \in 1 \cdots k$, that $H_i$ shatters $k+1$ element, which is impossible, because $\min\{(d_{vc} H_k)\}_{k=1}^{K} = k_2$

thus $0 \le d_{vc}(\cap_{k=1}^{K} H_k) \le \min\{ d_{vc}(H_k) \}_{k=1}^{K}$

c)

lower bound.

if $H_i \subset H_j$ then $d_{vc}(H_i) \leq d_{vc}(H_j)$

thus $\forall H_i \subset \bigcup_{k=1}^{K} H_k$, $d_{vc}(H_i) \leq \cdots \leq \max\{d_{vc}(H_k)\}_{k=1}^{K}$
$\leq \bigcup_{k=1}^{K} H_k$.

upper bound:

when $k=2$

$$d_{vc}(H_1 \cup H_2) \leq m_{H_1}(N) + m_{H_2}(N) + 1$$
$$= 2-1+\sum_{k=1}^{2} d_{vc}(H_k)$$

for all elements in $H_1 \cup H_2$, the must be fixed in $H_1$ or $H_2$.
thus $H_1 \cup H_2 \leq H_1 + H_2$
Assume the inequality hold for $k \leq n$.
when $k=n+1$

$$d_{vc}(\bigcup_{k=1}^{k+1} H_k) = d_{vc}((\bigcup_{k=1}^{k} H_k \cup H_{k+1})$$

$$= 1 + d_{vc}(\bigcup_{k=1}^{k} H_k) + d_{vc}(H_{k+1})$$

$$= 1 + k-1 + \sum_{i=1}^{k} d_{vc}(H_k) + d_{vc}(H_{k+1})$$

$$= (k+1)-1 \sum_{i=1}^{k+1} d_{vc}(H_k)$$

Thus $\max\{dvc\, H_k\}_{k=1}^{K} \le dvc\left(\bigcup_{k=1}^{K} H_k\right) \le K-1+\sum_{k=1}^{K} dvc(H_k)$

Problem 2.22

$$E_D[E_{out}(g^{(D)})] = E_D[E_{x,y}[(g^{(D)}(x) - y(x))^2]]$$

$$= E_D[E_{x,y}[(g^{(D)}(x) - f(x) - \varepsilon)^2]]$$

$$= E_D[E_{xy}[\underbrace{(g^{(D)}(x) - f(x))^2}_{1} + \underbrace{E_x(\varepsilon^2)}_{2} -$$

$$\underbrace{2E_x[\varepsilon \cdot (g^{(D)}(x) - f(x))]}_{3}$$

$$E_D[1] = E_x[\underbrace{E_D(g^{(D)}(x)^2 - \bar{g}(x)^2}_{g^{(D)}x^2 - \bar{g}(x^2)} + \underbrace{\bar{g}(x)^2 - 2\bar{g}(x)f(x) + f(x)^2}_{(\bar{g}(x) - f(x))^2}]$$

bias

Var

$$E_D[2] = E_D[E_x(\varepsilon^2)] = E_D(\delta)^2 = \delta^2$$

$$E_D[3] = E_D[-2E_x(\varepsilon)[g^{(D)}(x) - f(x)]]$$

$$= E_D[0 \cdot [g^{(D)}(x) - f(x)]] = 0$$

thus $E_D[E_{out}(g^{(D)})] = \sigma^2 + \text{bias} + \text{var}$

problem 2.24

a)

$$\begin{cases} x_2^2 = ax_2 + b \\ x_1^2 = ax_1 + b \end{cases} \Rightarrow \begin{array}{l} a = (x_1 + x_2) \\ b = -x_1 x_2 \end{array}$$

$$\bar{g}(x) = \frac{1}{k}\sum_{i=1}^{k} g_i(x) \approx \left(\frac{1}{2}\right)^2 \int_{-1}^{1}\int_{-1}^{1} (x_1+x_2)x + -x_1x_2 \, dx_1 dx_2$$

As k large enough it is approximately a continuous uniform distribution

$$= \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} (x_1+x_2)\, dx_1 dx_2 \quad \frac{1}{4}\iint_{-1}^{1} (x_1 x_2)\, dx_1 dx_2$$

$$= 0$$

thus $\bar{g}(x) \approx 0$

d)

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

$$= \mathbb{E}[x^4] - 2a\,\mathbb{E}_x[x^3] + (a^2 - 2b)\,\mathbb{E}_x[x^2] + 2ab[\mathbb{E}_x[x]] + b^2$$

$$= \int_{-1}^{1} \frac{1}{2}x^4 dx - 2a\int_{-1}^{1} \frac{1}{2}x^3 dx + (a^2-2b)\int_{-1}^{1} \frac{1}{2}x^2$$

$$+ 2ab\int_{-1}^{1} x\, dx + b^2$$

$$= \frac{8}{15}$$

$$bias(x) = (\bar{g}(x) - f(x))^2 = (-f(x))^2 = x^4$$

$$bias = \int_{-1}^{1} \frac{1}{2}x^4 dx = \frac{1}{5}$$

$$\mathrm{Var}(x) = E_D\left[(g(x) - \bar{g}(x))^2\right]$$

$$= E_D[a^2]\cdot x^2 + 2E_D[ab]\cdot x + E_D[b^2]$$

$$= E_D[(x_1+x_2)^2]x^2 + 2E_D[(x_1+x_2)(-x_1x_2)]\cdot x + E_D[(-x_1x_2)^2]$$

$$= \left(\tfrac{1}{2}\right)^2\int_{-1}^{1}\int_{-1}^{1}(x_1+x_2)^2dx_1dx_2\cdot x^2 + 2\cdot\left(\tfrac{1}{2}\right)^2\int_{-1}^{1}\int_{-1}^{1}(x_1+x_2)(-x_1x_2)dx_1dx_2\, x$$

$$+\left(\tfrac{1}{2}\right)^2\int_{-1}^{1}\int_{-1}^{1}(-x_1x_2)^2dx_1dx_2$$

$$= \frac{2}{3}\cdot x^2 + \frac{1}{9}$$

$$\mathrm{Var} = E_x\left[\frac{2}{3}x^2 + \frac{1}{9}\right] = \frac{2}{3}\int_{-1}^{1}\frac{1}{2}x^2dx + \frac{1}{9} = \frac{1}{3}$$

$$\frac{1}{3} + \frac{1}{5} = \frac{8}{15} \Leftrightarrow E_{out} = \mathrm{bias} + \mathrm{Var}$$