# Problems

1. LFD Problem 1.12
2. LFD Problem 2.3
3. LFD Problem 2.8
4. LFD Problem 2.10
5. LFD Problem 2.13
6. LFD Problem 2.22
7. LFD Problem 2.24

# Q1: LFD Problem 1.12

## (a)

**Proof:** Given that:

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

To minimize $E_{in}(h)$, let $\frac{d}{dh}E_{in}(h) = 0$, then:

$$\sum_{n=1}^{N} 2h - 2y_n = 0$$
$$\sum_{n=1}^{N} y_n = Nh$$
$$h = \frac{1}{N}\sum_{n=1}^{N} y_n$$

□

## (b)

**Proof:**

$$E_{in}(h) = \sum_{n=1}^{N} |h - y_n|$$
$$\frac{d}{dh}E_{in}(h) = sign(h - y_n)$$

To minimize $E_{in}(h)$, let $\frac{d}{dh}E_{in}(h) = 0$. This equals to zero only when the number of positive items equals the number of negative which happens when h=median$\{y_1, y_2, ...y_n\}$

□

## (c)

In this case $h_{mean}$ will become very large ($h_{mean} \to \infty$), while $h_{median}$ will stay constant.

## Q2: LFD Problem 2.3

### (a)

The positive ray will generate N + 1 dichotomy, and the negative ray will generate N-1 unique dichotomies ( all dichotomies opposite from dichotomies generated by positive ray minus the two repeated ones - all N points are negative and all N points are positive). So the total dichotomies are **2N**.

The VC dimension can be solved by find the maximum N such that $2^N = 2N$, so $d_{vc} = 2$.

### (b)

The positive interval will generate $\frac{n(n+1)}{2} + 1$ dichotomy, and the negative ray will generate $\frac{n(n+1)}{2} + 1 - 2N$ unique dichotomies (all dichotomies opposite from dichotomies generated by positive interval minus the 2N repeated ones - all N points are negative and all N points are positive plus all dicotomies including the minimum and maximum). So total dichotomies are **N(N+1) + 2 - 2N**.

The VC dimension can be solved by find the maximum N such that $N(N+1) + 2 - 2N = 2^N$, so $d_{vc} = 3$

### (c)

The result will be the same as positive interval (set d = 1 then the concentric spheres will be exactly the positive interval).

So the maximum number of dichotomies is $\frac{n(n+1)}{2} + 1$ and $d_{vc} = 2$

## Q3: LFD Problem 2.8

We know that if $m_{\mathcal{H}}(k) < 2^k$, then $\forall$ N, the growth function must have a polynomial bound:

$$m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$$

or in other word, there are 2 cases:

$$\begin{cases} d_{vc} = +\infty, m_{\mathcal{H}}(k) = 2^K \\ d_{vc} \text{is finite}, m_{\mathcal{H}}(k) \text{ is bounded by} N^{d_{vc}} + 1 \end{cases}$$

Then:

- (N+1:) is possible growth function:
  $d_{vc} = 1$, N + 1 is bounded by $N + 1$
- $(1 + N + \frac{N(N-1)}{2})$ is possible growth function:
  $d_{vc} = 2$, $1 + N + \frac{N(N-1)}{2}$ is bounded by $N^2 + 1$ for all $N \geq 1$
- $(2^N)$ is possible growth function:
  $d_{vc} = +\infty$, and $m_H(k) = 2^K$
- $(2^{\sqrt{N}})$ is not a possible growth function:
  $d_{vc} = 1$, but $2^{\sqrt{N}}$ is not bounded by N+1 (for N = 25, $2^{\sqrt{25}} = 32 > 25 + 1 = 26$).
- $(2^{\frac{N}{2}})$ is not a possible growth function:
  $d_{vc} = 0$, but $2^{\sqrt{N}}$ is not bounded by 2 (for N = 4, $2^{\frac{4}{2}} = 4 > 2$).
- $(1 + N + \frac{N(N-1)(N-2)}{6})$ is not a possible growth function:
  $d_{vc} = 1$, but $1 + N + \frac{N(N-1)(N-2)}{6}$ is not bounded by N +1
  (for N = 3, $1 + N + \frac{N(N-1)(N-2)}{6} = 5 > 1 + 3 = 4$).

## Q4: LFD Problem 2.10

Partition the 2N points to 2 fractions: $\{x_1, x_2, ....x_N\}$ and $\{x_{N+1}, x_{N+2}, ....x_{2N}\}$, each with N points. For each fraction there are $m_{\mathcal{H}}(N)$ ways to dichotomies N points, so for 2N points, there are at most $m_{\mathcal{H}}(N) * m_{\mathcal{H}}(N)$ dichotomies. Which suggests $m_{\mathcal{H}}(2N) \leq m_{\mathcal{H}}(N)^2$.

Therefore, we can rewrite the VC generalization bound as follow:

$$\begin{aligned} E_{out}(g) &\leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}} \\ &\leq E_{in}(g) + \sqrt{\frac{8}{N} \ln \frac{4(m_{\mathcal{H}}(N))^2}{\delta}} \end{aligned}$$

# Q5: LFD Problem 2.13

## (a)

Since $d_{vc}$ is defined as the largest value of N for which $m_{\mathcal{H}}(N) = 2^N$, while $m_{\mathcal{H}}(N)$ is the maximum number of possible partition on N points. Since there are k hypothesis, the maximum number of partitions can not exceed k for any N, which produce an upper bound of the growth function $m_{\mathcal{H}}(N) \leq M$ for any N including the value of VC dimension. Therefore :

$$\begin{aligned} m_{\mathcal{H}}(d_{vc}) = 2^{d_{vc}} \leq M \\ d_{vc} \leq \log_2 M \end{aligned}$$

## (b)

**Lower bound:** 0
It could be possible that the intersection of all hypothesis set is the empty set. In this case, $d_{vc} = 0$ since there are only 1 possible dichotomy for any N.

**Upper bound:** $\min\limits_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$
First, show that $\min\limits_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$ is an upper bound.

Let $\min\limits_{1\leq k\leq K} d_{vc}(\mathcal{H}_k) = d$. Assume $\bigcap\limits_{k=1}^{K} \mathcal{H}_k > d$.

Then by definition, $\bigcap\limits_{k=1}^{K} \mathcal{H}_k$ must shutter d+1 points, which suggest for all $1 \leq k \leq K$, $\mathcal{H}_k$ also shutter d+1 points. This contradicts with the fact there exist some k such that $1 \leq k \leq K$ and $d_{vc}(\mathcal{H}_k) = d$. Therefore d must be an upper bound of $d_{vc}(\bigcap\limits_{k=1}^{K} \mathcal{H}_k)$.

Then, show $\min\limits_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$ is the tightest upper bound. Given the case that K = 1, then:

$$d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) = d_{vc}(\mathcal{H}_1) = d_{vc} \min_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$$

Then $\min\limits_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$ is a maximum (an achievable upper bound), which is the tightest possible upper bound.

**(c)**

**Lower Bound:** $\max_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$

First, show that $\max_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$ is a lower bound.

Let $\max_{1\leq k\leq K} d_{vc}(\mathcal{H}_k) = d$. Assume $\bigcap_{k=1}^{K} \mathcal{H}_k < d$.

Then by definition, $\bigcup_{k=1}^{K} \mathcal{H}_k$ does not shutter d points, which suggest for all $1 \leq k \leq K$, there does not exist a $\mathcal{H}_k$ that shutter d points. This contradicts with the fact there exist some k such that $1 \leq k \leq K$ and $d_{vc}(\mathcal{H}_k) = d$. Therefore d must be a lower bound of $d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k)$.

Then, show $\min_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$ is the tightest upper bound. Given the case that K = 1, then:

$$d_{vc}(\bigcap_{k=1}^{K} \mathcal{H}_k) = d_{vc}(\mathcal{H}_1) = d_{vc} \max_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$$

Then $\max_{1\leq k\leq K} d_{vc}(\mathcal{H}_k)$ is a minimum (an achievable lower bound), which is the tightest possible lower bound.

**Upper Bound:** $\sum_{k=1}^{K} d_{vc}(\mathcal{H}_k) + K - 1$

The maximum number of partition for union of two hypothesis set can be deduced as follow:
Since the maximum number of dichotomy generate by two hypothesis set is smaller than the sum of number of partition generated by each set, the following inequality must hold:

$$\begin{aligned} m_{\mathcal{H}_1}(N) + m_{\mathcal{H}_2}(N) &\leq \sum_{i=0}^{d1} \binom{N}{i} + \sum_{i=0}^{d2} \binom{N}{i} \\ &= \sum_{i=0}^{d1} \binom{N}{i} + \sum_{i=0}^{d2} \binom{N}{N-i} \end{aligned}$$

For all N satisfy $d_1 + d_2 < N - 1$,

$$\sum_{i=0}^{d1} \binom{N}{i} + \sum_{i=0}^{d2} \binom{N}{N-i} = \sum_{i=0}^{N} \binom{N}{i} - \sum_{i=d_1}^{d_2} \binom{N}{i}$$

$$\sum_{i=0}^{d1} \binom{N}{i} + \sum_{i=0}^{d2} \binom{N}{N-i} + \sum_{i=d_1}^{d_2} \binom{N}{i} = 2^N$$

$$\sum_{i=0}^{d1} \binom{N}{i} + \sum_{i=0}^{d2} \binom{N}{N-i} < 2^N$$

Then for all $N > d_1 + d_2 + 1$:

$$
\begin{aligned}
m_{\mathcal{H}_1}(N) + m_{\mathcal{H}_2}(N) &\leq \sum_{i=0}^{d1} \binom{N}{i} + \sum_{i=0}^{d2} \binom{N}{i} \\
&< 2^N
\end{aligned}
$$

which is equivalent to $\mathcal{H}_1 \cup \mathcal{H}_2$ can not shutter $d_1 + d_2 + 1$points. Thus, $d_1 + d_2 + 1$ is an upper bound of $d_vc$ for $\mathcal{H}_1 \cup \mathcal{H}_2$.
Using similar logic for the K union of hypothesis, $\bigcup_{k=1}^{K} \mathcal{H}_k$, we can conclude that the upper bound of $d_{vc}(\bigcup_{k=1}^{K} \mathcal{H}_k)$ is $\sum\limits_{k=1}^{K} d_{vc}(\mathcal{H}_k) + K - 1$.

# Q6: LFD Problem 2.22

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}[E_{out}(g^{(D)}] &= \mathbb{E}_{\mathcal{D}}[E_{x,y}[(g^{(D)}(x) - y(x))^2]] \\
&= \mathbb{E}_{x,y}[E_D[(g^{(D)}(x) - y(x))^2]] \\
&= \mathbb{E}_{x,y}[E_D[(g^{(D)}(x))^2] - 2E_D[(g^{(D)}y(x)] + y(x)^2] \\
&= \mathbb{E}_{x,y}[E_D[(g^{(D)}(x))^2] - 2E_D[(g^{(D)}(f(x) + \epsilon)] + (f(x) + \epsilon)^2] \\
&= \mathbb{E}_{x}[E_D[(g^{(D)}(x))^2] - 2E_D[(g^{(D)}f(x)] + (f(x) + \epsilon)^2] + \mathbb{E}_{x,y}[2f(x)\epsilon + \epsilon^2 - 2\overline{g}(x)\epsilon] \\
&= var + bias + \mathbb{E}_{x,y}[2\epsilon(f(x) - \overline{g}(x)) + \epsilon^2] \\
&= var + bias + \mathbb{E}[\epsilon^2] + 2\,\mathbb{E}[\epsilon]\,\mathbb{E}_x[f(x) - \overline{g}(x)]
\end{aligned}
$$

Since $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[\epsilon^2] = Var[\epsilon] - (\mathbb{E}[\epsilon])^2 = \sigma^2$, then:

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}[E_{out}(g^{(D)}] &= var + bias + \mathbb{E}[\epsilon^2] + 2\,\mathbb{E}[\epsilon]\,\mathbb{E}_x[f(x) - \overline{g}(x)] \\
&= var + bias + \sigma^2 + 2 * 0 * \mathbb{E}_x[f(x) - \overline{g}(x)] \\
&= var + bias + \sigma^2
\end{aligned}
$$

# Q7: LFD Problem 2.24

## (a)

$$\begin{aligned}
\overline{g}(x) &= \mathbb{E}_D[g^{(D)}(x)] \\
&= \mathbb{E}_D[\frac{x_2^2 - x_1^2}{x2 - x1}x - x_1x_2] \\
&= \mathbb{E}_{x_1,x_2}[(x_1 + x_2)x - x_1x_2] \\
&= \frac{1}{4}x(\int_{-1}^{1}\int_{-1}^{1} x_1 + x_2 dx_1dx_2) - \frac{1}{4}(\int_{-1}^{1}\int_{-1}^{1} x_1x_2dx_1dx_2) \\
&= \frac{1}{4}x(0) + 0 \\
&= 0
\end{aligned}$$

## (b)

1. Generate a K pair of $x_1, x_2$ randomly with uniform distribution from [-1, 1], for each pair $1 \leq k \leq K$ fit $g_k(x) = a_kx + b$.
2. Compute $\overline{g}(x)$ using $g(x)$ for all pair of $x_1, x_2$ generated such that:

$$\overline{g}_(x) = \frac{\sum_{i=1}^{K} a_k}{K}x + \frac{\sum_{i=1}^{K} b_k}{K}$$

3. Generate K x randomly ($\{x_1, x_2, x_3, ...x_K\}$) with uniform distribution from [-1, 1], using the following formula to compute $E_{out}$, bias and var directly.

$$\mathbb{E}_{\mathcal{D}}[E_{out}] = \frac{\sum\limits_{j=0}^{K}\sum\limits_{i=0}^{K}(g_i(x_j) - x_j^2)^2}{K^2}$$

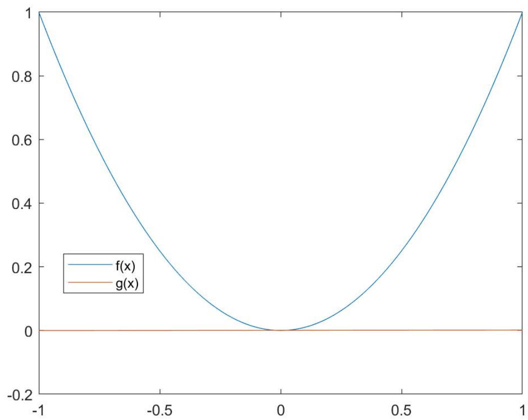
$$var = \frac{\sum\limits_{j=0}^{K}\sum\limits_{i=0}^{K}(g_i(x_j) - \overline{g}(x_j))^2}{K^2}$$

$$bias = \frac{\sum\limits_{i=0}^{K}(\overline{g}(x_k) - x_k^2)^2}{K}$$

## (c)

bias = 0.20
var = 0.33

$E_{out} = 0.53 = bias + var$




### (d)

First compute $E_{out}$:

$$
\begin{aligned}
E_{out} &= E_x[(g(x) - f(x))^2] \\
&= E_x[(ax + b - f(x))^2] \\
&= E_x[(x^2 - ax - b)^2] \\
&= E_x[a^2x^2 + 2abx - 2ax^3 + b^2 - 2bx^2 + x^4)] \\
&= \frac{1}{2}\int_{-1}^{1} a^2x^2 + 2abx - 2ax^3 + b^2 - 2bx^2 + x^4 dx \\
&= \frac{2/5 + (2a^2)/3 - (4b)/3 + 2b^2}{2} \\
&= \frac{1}{5} + \frac{2a^2}{3} - \frac{2b}{3} + b^2
\end{aligned}
$$

Then $\mathbb{E}_{\mathcal{D}}[E_{out}]$ :

$$
\begin{aligned}
\mathbb{E}_{\mathcal{D}}[E_{out}] &= \mathbb{E}_{x_1,x_2}[E_{out}] \\
&= \mathbb{E}_{x_1,x_2}[\frac{1}{5} + \frac{(x_1+x_2)^2}{3} + \frac{2x_1x_2}{3} + (x_1x_2)^2] \\
&= \frac{1}{4}(\int_{-1}^{1}\int_{-1}^{1} \frac{1}{5} + \frac{(x_1+x_2)^2}{3} + \frac{2x_1x_2}{3} + (x_1x_2)^2 dx_1dx_2) \\
&= \frac{8}{15}
\end{aligned}
$$

Similarly:

$$\begin{aligned}
bias(x) &= (\overline{g} - f(x))^2 \\
&= f(x)^2 \\
&= x^4 \\
bias &= \mathbb{E}_x[x^4] \\
&= \frac{1}{2}\int_{-1}^{1} x^4 dx \\
&= \frac{1}{5}
\end{aligned}$$

And since by definition, $\mathbb{E}_{\mathcal{D}}[E_{out}] = bias + var$, $var = \frac{8}{15} - \frac{1}{5} = \frac{1}{3}$.