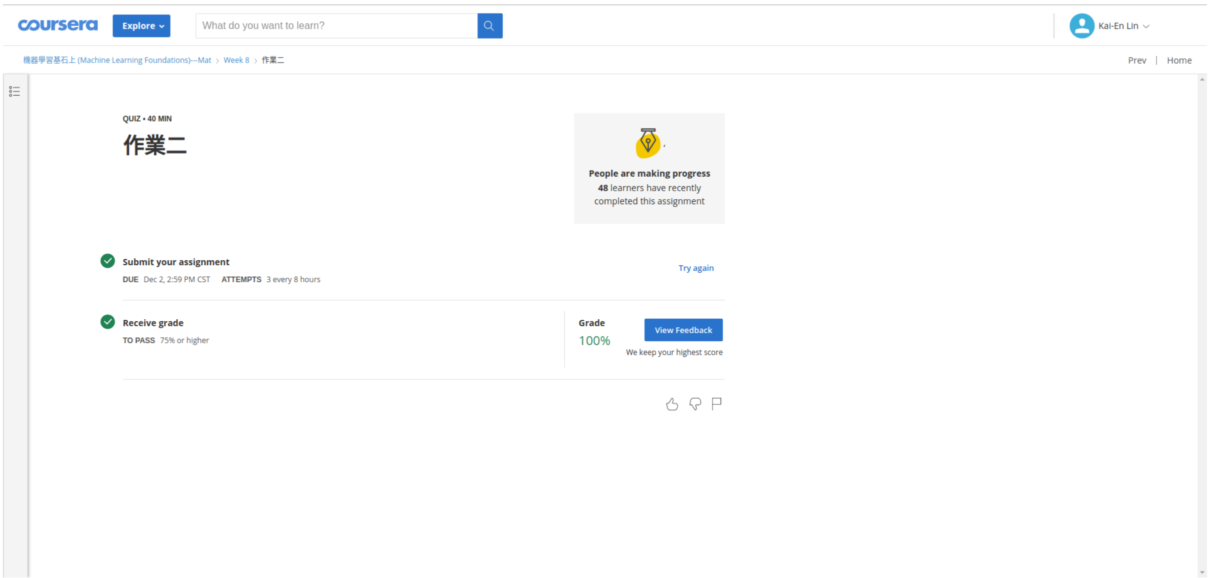
Homework #2

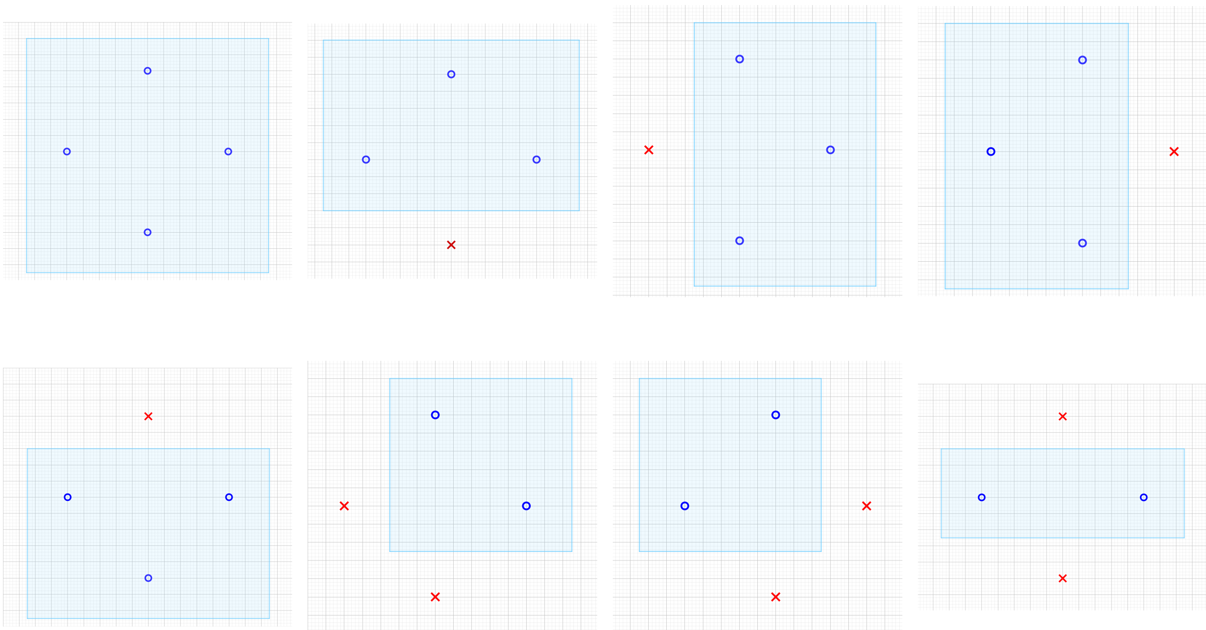
Student Name: 林楷恩
Student ID: b07902075

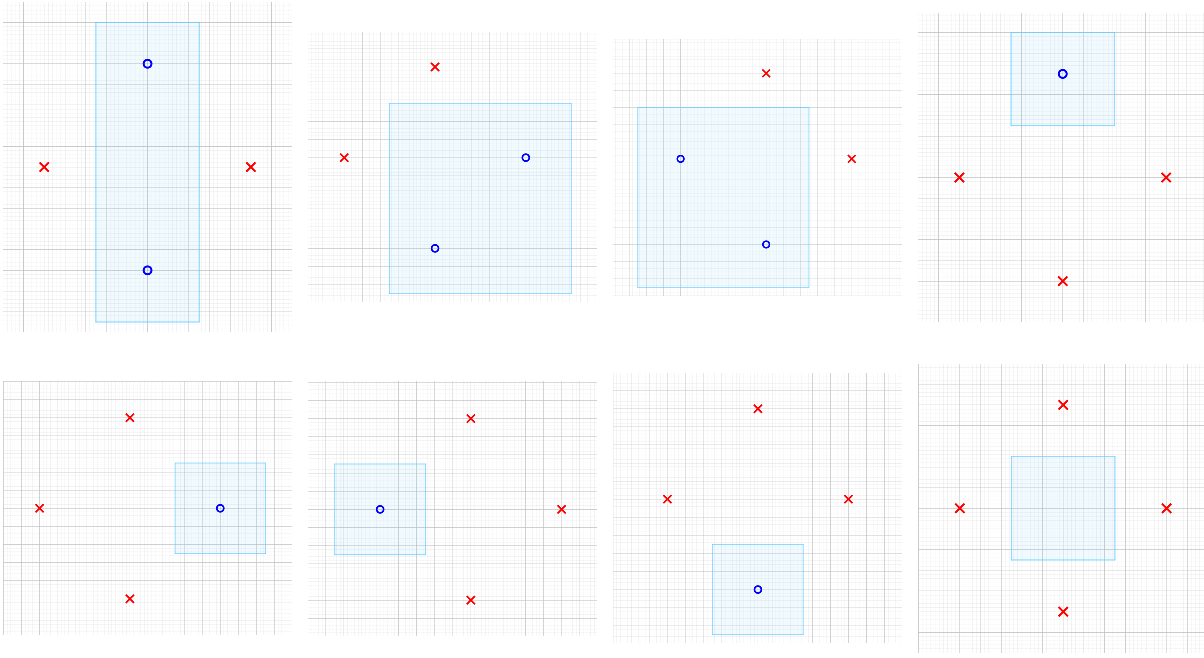
1



2

To prove that d_{vc} of this hypothesis set ≥ 4, I show that there exists some inputs of size 4 that can be shattered by the hypothesis set. The arangement of the 4 points and the total $2^4 = 16$ dichotomies are as follow:





3

Assume $sign(0) = -1$

1. First, to simplify the problem, I map (αx) mod 4 directly to 1 & -1 directly by observation:

$$h_\alpha(x) = \left\{ \begin{array}{ll} 1 & if\ (\alpha x)\ mod\ 4 \in [0,\ 1) \cup (3, 4) \\ -1 & if\ (\alpha x)\ mod\ 4 \in [1, 3] \end{array} \right.$$

2. If we represent αx in base-4 number, which looks like $\ldots a4^2 + b4^1 + c4^0 + d4^{-1} \ldots$, then we can find that (αx) mod 4 will truncate those terms which become 0 when being mod 4. That is, the 4^i terms where $i \geq 1$ can be ignored.

3. By above observation, we can encode y into α by manipulating the coefficients in the base-4 representation such that:

   - if $y_i = 1$, then (αx_i) mod $4 \in [0, 1) \in [0,\ 1) \cup (3, 4) \Rightarrow h_\alpha(x_i) = 1$
   - if $y_i = -1$, then (αx_i) mod $4 \in [1, 2) \in [1, 3] \Rightarrow h_\alpha(x_i) = -1$

   Thus, if $y_i = 1$, let the coefficient of the 4^i term in α be 0, or if $y_i = -1$, let the coefficient of the 4^i term in α be 1. And let $x_i = 4^{-i}$ for all i. Then we can get:

   - if $y_i = 1$, (αx_i) mod $4 = 0 \times 4^0 + \ldots$
   - if $y_i = -1$, (αx_i) mod $4 = 1 \times 4^0 + \ldots$

   For both cases, the terms below 4^0 does not make a difference ≥ 1, because $\sum_{k=1}^{\infty} 4^{-k} = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3} < 1$ even if all coefficient is 1.

4. Therefore, for any $N \geq 1$, let $x_i = 4^{-i}$, for $i \in \{1, 2, \ldots, N\}$, then α can be constructed as follow for all 2^N different dichotomies:

$$\alpha = \sum_{i=1}^{N} c\,4^i,\ \ c = \left\{ \begin{array}{ll} 1 & if\ y_i = 1 \\ 0 & if\ y_i = -1 \end{array} \right.$$

   By above approach, for any N, there exists some inputs we can shatter, so $d_{vc} = \infty$

4

Since $\mathcal{H}_1 \cap \mathcal{H}_2 \subseteq \mathcal{H}_1$, for any inputs, if $\mathcal{H}_1 \cap \mathcal{H}_2$ can shatter, then $\mathcal{H}_1$ must be able to shatter. However, it is possible that there are some inputs such that $\mathcal{H}_1$ can shatter but $\mathcal{H}_1 \cap \mathcal{H}_2$ cannot. That is:

$$\mathcal{H}_1 \cap \mathcal{H}_2 \text{ can shatter } \implies \mathcal{H}_1 \text{ can shatter}$$

Thus, we can *prove by contradiction.* Assume $d_{vc}(\mathcal{H}_1) < d_{vc}(\mathcal{H}_1 \cap \mathcal{H}_2) = x \implies \mathcal{H}_1$ cannot shatter any inputs of size x, but there exists a input of size x such that $\mathcal{H}_1 \cap \mathcal{H}_2$ can shatter, which implies $\mathcal{H}_1$ can shatter such an input $\implies$ *contradiction*, $d_{vc}(\mathcal{H}_1) \geq d_{vc}(\mathcal{H}_1 \cap \mathcal{H}_2)$ proved.

5

The only intersection of $\mathcal{H}_1$ and $\mathcal{H}_2$ are "all positive" and "all negative " hypotheses for all possible input size N. Thus:

$$m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) = m_{\mathcal{H}_1}(N) + m_{\mathcal{H}_2}(N) - 2 = (N+1) + (N+1) - 2 = 2N$$

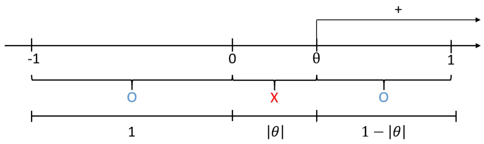
	$m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N)$	2^N
$N=1$	2	2
$N=2$	4	4
$N=3$	6	8

From the table, we know 3 is the first break point, so $d_{vc}(\mathcal{H}_1 \cup \mathcal{H}_2) = 3 - 1 = 2$ #

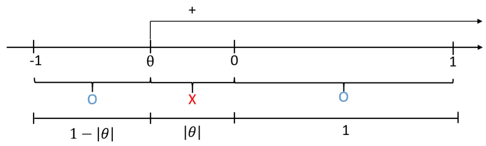
6

- $s = 1$ (positive ray):

  * $\theta > 0$

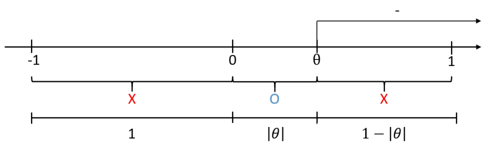


  * $\theta < 0$

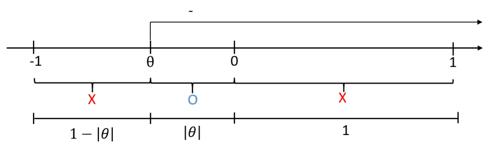


- $s = -1$ (negative ray):

  * $\theta > 0$



  * $\theta < 0$



Let μ be the probability that the hypothesis $h_{s,\theta}(x) = s \cdot sign(x - \theta)$ makes an error. Observing the pattern of when an error occurs in the case of $s = 1$ and $s = -1$, we can conclude that:

$$\begin{aligned}
\mu &= \begin{cases} \frac{|\theta|}{2} & if\ \ s = 1 \text{ (positive ray)} \\ \frac{2-|\theta|}{2} & if\ \ s = -1 \text{ (negative ray)} \end{cases} \\
&= \frac{1+s}{2}\left(\frac{|\theta|}{2}\right) + \frac{1-s}{2}\left(\frac{2-|\theta|}{2}\right) \\
&= \frac{|\theta|}{4} + \frac{s|\theta|}{4} + \frac{2 - |\theta| - 2s + s|\theta|}{4} \\
&= \frac{2s|\theta| - 2s + 2}{4} \\
&= \frac{s(|\theta| - 1) + 1}{2}
\end{aligned}$$

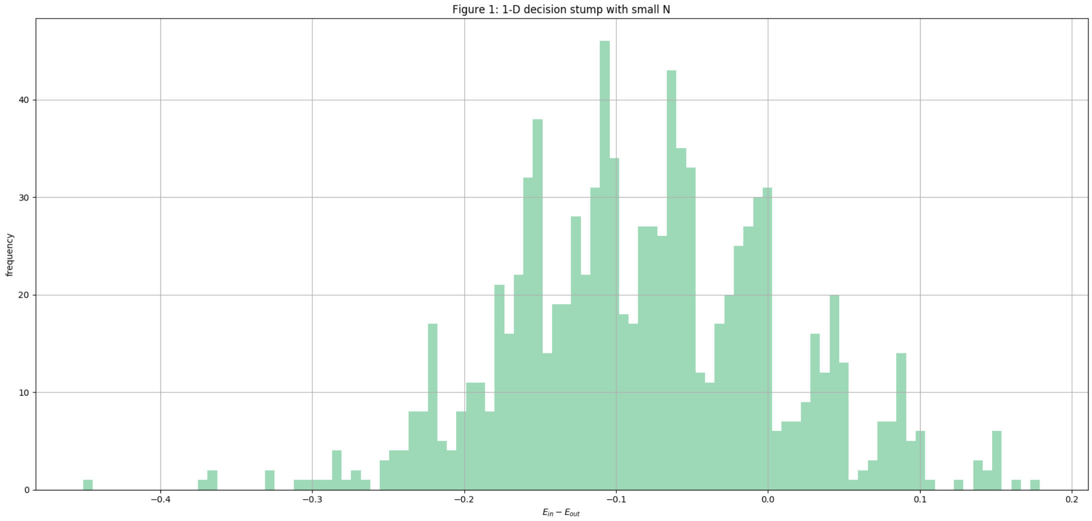
And since the probability of noise is 20%, E_{out} is calculated as follow:

$$\begin{aligned}
E_{out} &= \mu \cdot 0.8 + (1 - \mu) \cdot 0.2 \\
&= \frac{s(|\theta| - 1) + 1}{2} \cdot 0.8 + \frac{1 - s(|\theta| - 1)}{2} \cdot 0.2 \\
&= 0.4s(|\theta| - 1) + 0.4 + 0.1 - 0.1s(|\theta| - 1) \\
&= 0.3s(|\theta| - 1) + 0.5 \ \#
\end{aligned}$$

7

Result: $E_{in} = 0.17270,\ E_{out} = 0.25373$

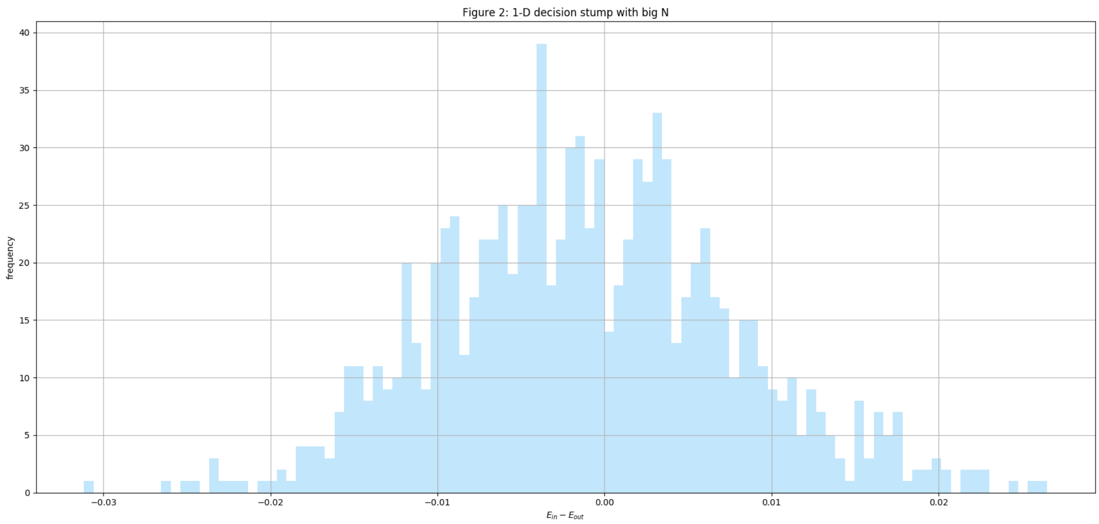
I notice that the difference between E_{in} and E_{out} is quite large(about 0.08), and I think it can be explained by the *Hoeffding's Inequality*: $\mathbb{P}[|E_{in}(h) - E_{out}(h)| > \epsilon] \leq 2exp(-2\epsilon^2 N)$. Since the N in this experiment is rather small, the probability of "BAD things" becomes large, so $E_{in} \approx E_{out}$ is not quite correct here.



8

Result: $E_{in} = 0.19950,\ E_{out} = 0.20058$

The difference between becomes a lot smaller here(about 0.001), since the N is large enough to make the right hand side of *Hoeffding's Inequality* small enough to bound the probability of "BAD things".



9

- $d_{vc} \geq 2^d$: I show that there exists inputs of size 2^d that can be shattered:
  1. Let $t = (0, 0, 0, \ldots, 0)$
  2. Let $\mathcal{X} = \{-1, 1\}^d \implies |\mathcal{X}| = 2^d$
  3. Then $\forall\, \hat{x}_i \in \mathcal{X}$, the corresponded $\hat{v}_i$ is unique (all $\hat{x}_i$ lie in different hyper-rectangular regions)
  4. Let $\hat{v}_i \in S$ if and only if $y_i = 1$ (assign each hyper-rectangular regions the corresponded y_i)
- $d_{vc} \leq 2^d$: I show that we cannot shatter any set of inputs of size $2^d + 1$:

  Proof: By induction on d:

  - **Base case**: $d = 1$: the hypothesis set we have is the same as the union of "positive rays" and "negative rays", so by works in problem 5, we cannot shatter inputs of size $3 = 2^d + 1$.
  - **Inductive Step**: $d = k$: we arbitrarily choose a dimension and apply a threshold on it, then the points on the hyperplane are divided into 2 groups. If there are $2^k + 1$ points, we can observe that one of the groups' size must $\geq 2^{k-1} + 1$. Because we have use one of the dimension, so the 2 groups can be regarded as only have $d - 1$ dimensions. By induction hypothesis, we cannot shatter the group whose size is $\geq 2^{k-1} + 1$, so we cannot shatter the whole set of points, too.

By the two facts, $d_{vc} = 2^d$, the statement gets proved.