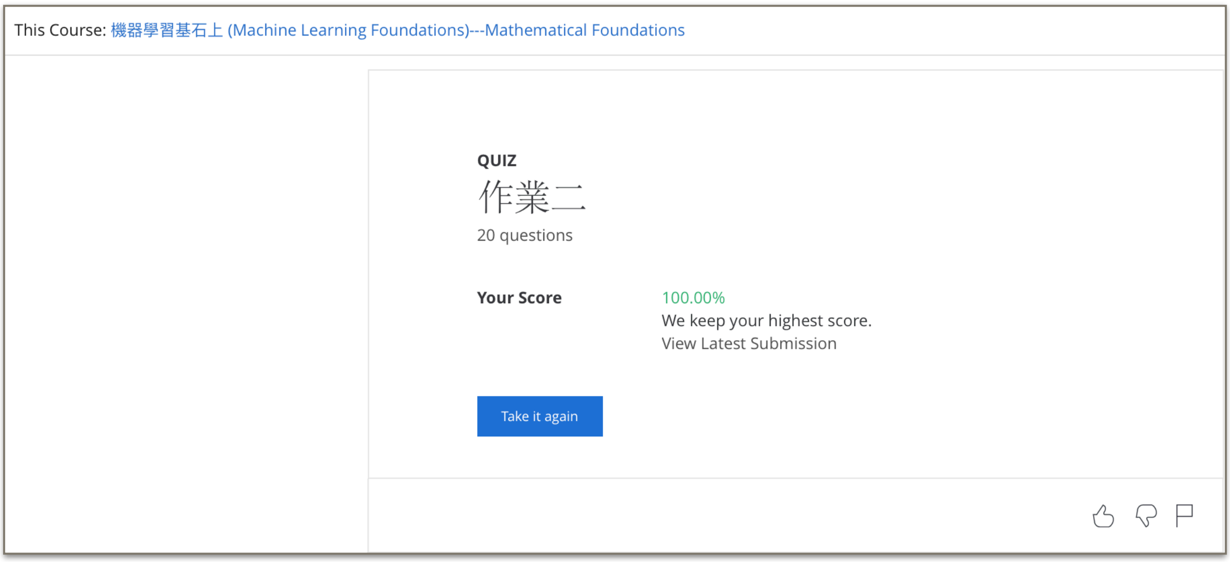
Machine Learning Foundations - *Homework #2*

B05902118 資工三 陳盈如

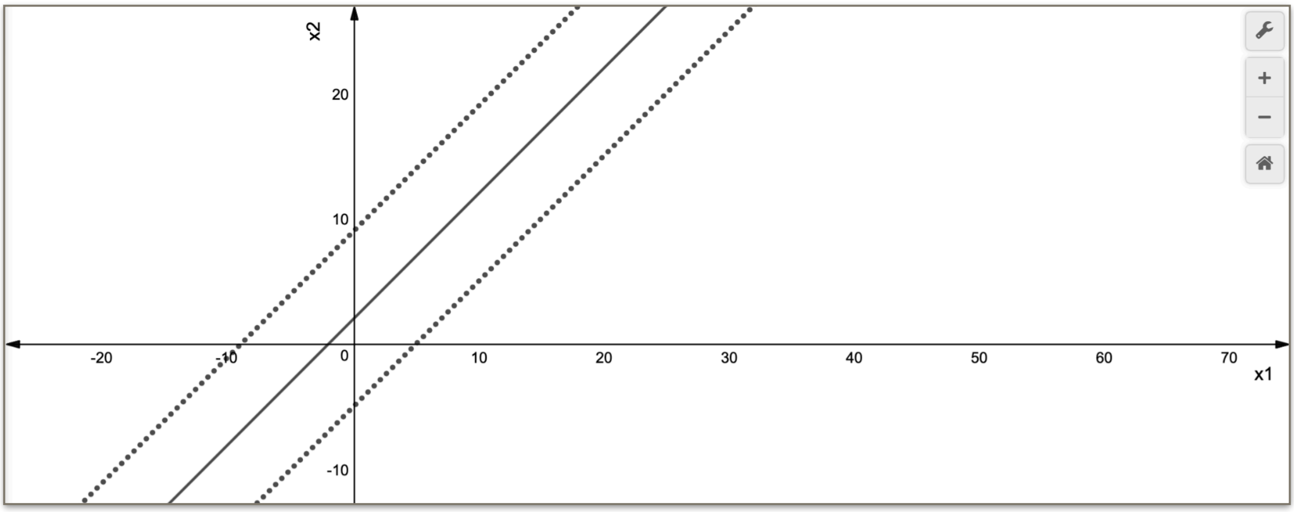
Problem 1.



Problem 2.

Prove it by proving that the hypothesis set must shatter inputs when N = 1, 2, 3.

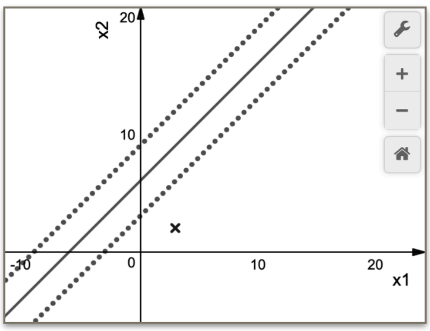
We can draw the function of $|w_0 + w_1x_1 + w_2x_2| \le \theta$. The solid line represents the function, $w_0 + w_1x_1 + w_2x_2 = 0$, and it could be different with different value of $w_0,\ w_1,\ w_2$. Two dotted lines have the same m as the solid line, and the distance between the dotted line and the solid line depends on the value of θ.

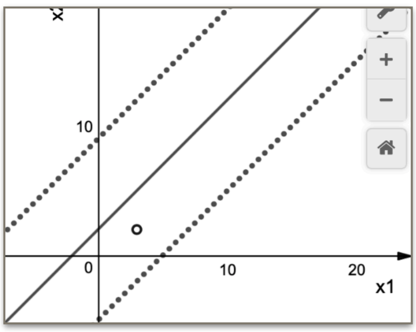


- N = 1 (o→1, x→-1)

  It can be shattered.

$(x_1,\ x_2)$
o
x

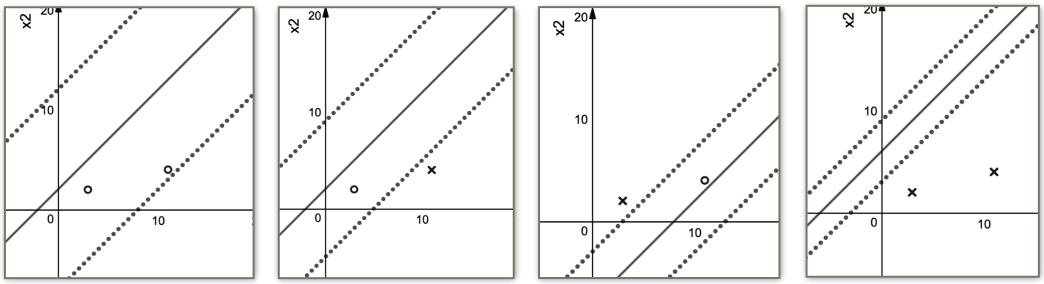




- N = 2

  It can also be shattered.

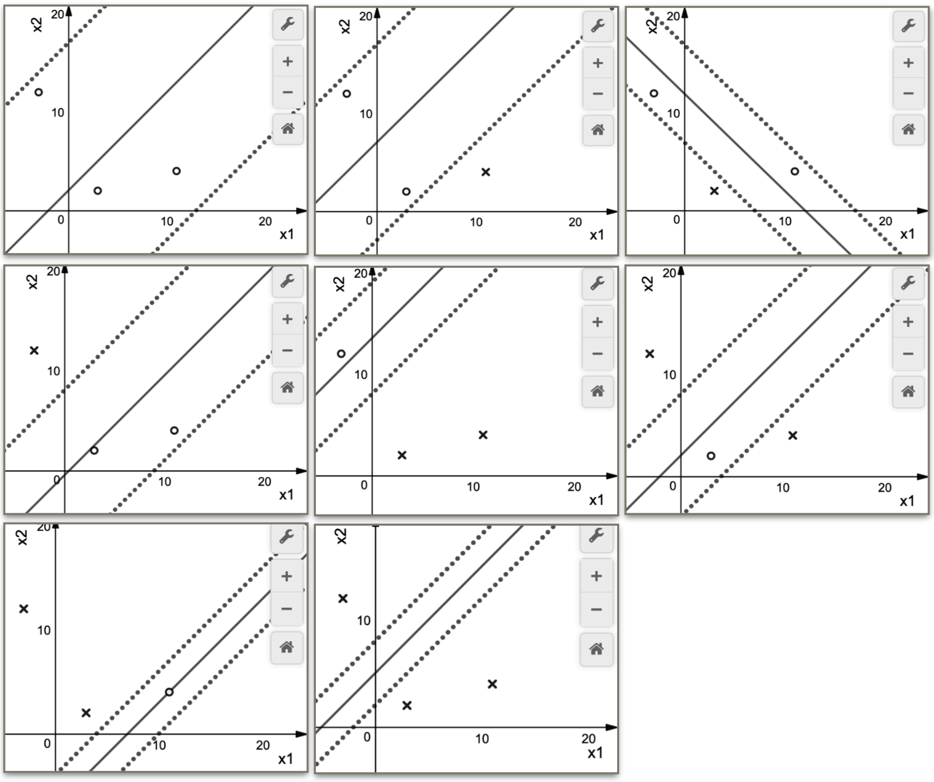
(x_{11}, x_{21})	(x_{12}, x_{22})
o	o
o	x
x	o
x	x



- N = 3

  It still can be shattered.

(x_{11}, x_{21})	(x_{12}, x_{22})	(x_{13}, x_{23})
o	o	o
o	o	x
o	x	o
x	o	o
o	x	x
x	o	x
x	x	o
x	x	x



Therefore, we can know that VC-Dimension of the hypothesis set is no less than 4.

Problem 3.

The VC-Dimension of such an H is ∞.

Suppose that $x = 4^1, 4^2, \dots, 4^i, \dots, 4^N$

$$\alpha = (\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \dots + (\frac{1}{4})^i a_i + \dots + (\frac{1}{4})^N a_N, \qquad a_i \in \{0, 2\}$$

If $y_i = +1$, $a_i = 0$. And if $y_i = -1$, $a_i = 2$

$$H = \{h_\alpha |\ h_\alpha(x) = sign(|\alpha x \mod 4 - 2| - 1),\ \alpha \in R\}$$

If there are N inputs:

(1) $x_i = 4^i$

(2) $\alpha = (\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \dots + (\frac{1}{4})^i a_i + \dots + (\frac{1}{4})^N a_N$

Case 1: $y_i = +1$

$$
\begin{aligned}
\alpha x &= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^i a_i + \ldots + (\frac{1}{4})^N a_N] \times 4^i \\
&= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^{i-1} a_{i-1} + (\frac{1}{4})^i a_i + (\frac{1}{4})^{i+1} a_{i+1} + \ldots + (\frac{1}{4})^N a_N] \times 4^i \\
&= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^{i-1} a_{i-1}] \times 4^i + [(\frac{1}{4})^i a_i] \times 4^i + [(\frac{1}{4})^{i+1} a_{i+1} + \ldots + (\frac{1}{4})^N a_N] \times 4^i
\end{aligned}
$$

$$
\begin{aligned}
\alpha x \mod 4 &= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^{i-1} a_{i-1}] \times 4^i \mod 4 \\
&\quad + [(\frac{1}{4})^i a_i] \times 4^i \mod 4 \\
&\quad + [(\frac{1}{4})^{i+1} a_{i+1} + \ldots + (\frac{1}{4})^N a_N] \times 4^i \mod 4 \\
&= 0 + (a_i \mod 4) + ((\frac{1}{4})a_{i+1} + \ldots + (\frac{1}{4})^{N-i} a_N \mod 4) \\
&= 0 + (0 \mod 4) + ((\frac{1}{4})a_{i+1} + \ldots + (\frac{1}{4})^{N-i} a_N \mod 4)
\end{aligned}
$$

$\because 0 < (\frac{1}{4})a_{i+1} + \ldots + (\frac{1}{4})^{N-i} a_N < 1$

$\therefore 0 < \alpha x \mod 4 < 1$

$\Rightarrow |\alpha x \mod 4 - 2| > 1$

$\underline{\Rightarrow sign(|\alpha x \mod 4 - 2| - 1) = +1}$#

Case 2: $y_i = -1$

$$
\begin{aligned}
\alpha x &= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^i a_i + \ldots + (\frac{1}{4})^N a_N] \times 4^i \\
&= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^{i-1} a_{i-1} + (\frac{1}{4})^i a_i + (\frac{1}{4})^{i+1} a_{i+1} + \ldots + (\frac{1}{4})^N a_N] \times 4^i \\
&= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^{i-1} a_{i-1}] \times 4^i + [(\frac{1}{4})^i a_i] \times 4^i + [(\frac{1}{4})^{i+1} a_{i+1} + \ldots + (\frac{1}{4})^N a_N] \times 4^i
\end{aligned}
$$

$$
\begin{aligned}
\alpha x \mod 4 &= [(\frac{1}{4})a_1 + (\frac{1}{4})^2 a_2 + \ldots + (\frac{1}{4})^{i-1} a_{i-1}] \times 4^i \mod 4 \\
&\quad + [(\frac{1}{4})^i a_i] \times 4^i \mod 4 \\
&\quad + [(\frac{1}{4})^{i+1} a_{i+1} + \ldots + (\frac{1}{4})^N a_N] \times 4^i \mod 4 \\
&= 0 + (a_i \mod 4) + ((\frac{1}{4})a_{i+1} + \ldots + (\frac{1}{4})^{N-i} a_N \mod 4) \\
&= 0 + (2 \mod 4) + ((\frac{1}{4})a_{i+1} + \ldots + (\frac{1}{4})^{N-i} a_N \mod 4)
\end{aligned}
$$

$\because 0 < (\frac{1}{4})a_{i+1} + \ldots + (\frac{1}{4})^{N-i} a_N < 1$

$\therefore 2 < \alpha x \mod 4 < 3$

$\Rightarrow |\alpha x \mod 4 - 2| < 1$

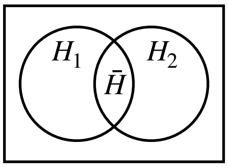
$\underline{\Rightarrow sign(|\alpha x \mod 4 - 2| - 1) = -1}$#

In conclusion, i can be any number $\in N$ and makes N inputs shattered. Therefore, the VC-Dimension of such an H is ∞.

Problem 4.

Assume that $\bar{H} = H_1 \cap H_2$

$\because \bar{H} \subseteq H_2$



$\therefore$ if N inputs can be shattered by $\bar{H}$, they must be shattered by H_2, too.

$\therefore$ The VC-Dimension of such a hypothesis set H_2 is no less than the VC-Dimension of $\bar{H}$.

Problem 5.

(1) $\underline{m_{H_1 \cup H_2}(N) = 2N\#}$

(2) N = 1: $m_{H_1 \cup H_2}(1) = 2 = 2^1$

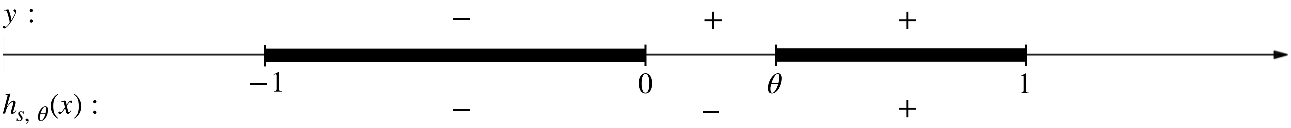
N = 2: $m_{H_1 \cup H_2}(2) = 4 = 2^2$

N = 3: $m_{H_1 \cup H_2}(3) = 6 < 2^3$

$\underline{d_{vc}(H_1 \cup H_2) = 3\#}$

Problem 6.

s = +1:



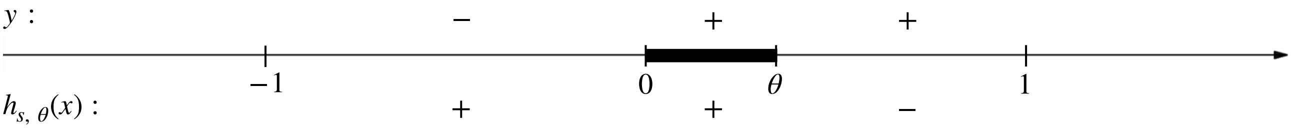
$$E_{out_1}(h_{s,\theta}) = \frac{(1-|\theta|)+1}{2} \times 20\% + \frac{|\theta|}{2} \times 80\%$$

$$= (1 - \frac{|\theta|}{2}) \times 20\% + \frac{|\theta|}{2} \times 80\%$$

$$= 0.2 - 0.1 \times |\theta| + 0.4 \times |\theta|$$

$$= 0.2 + 0.3|\theta|$$

s = -1:



$$E_{out_2}(h_{s,\theta}) = \frac{(1-|\theta|)+1}{2} \times 80\% + \frac{|\theta|}{2} \times 20\%$$

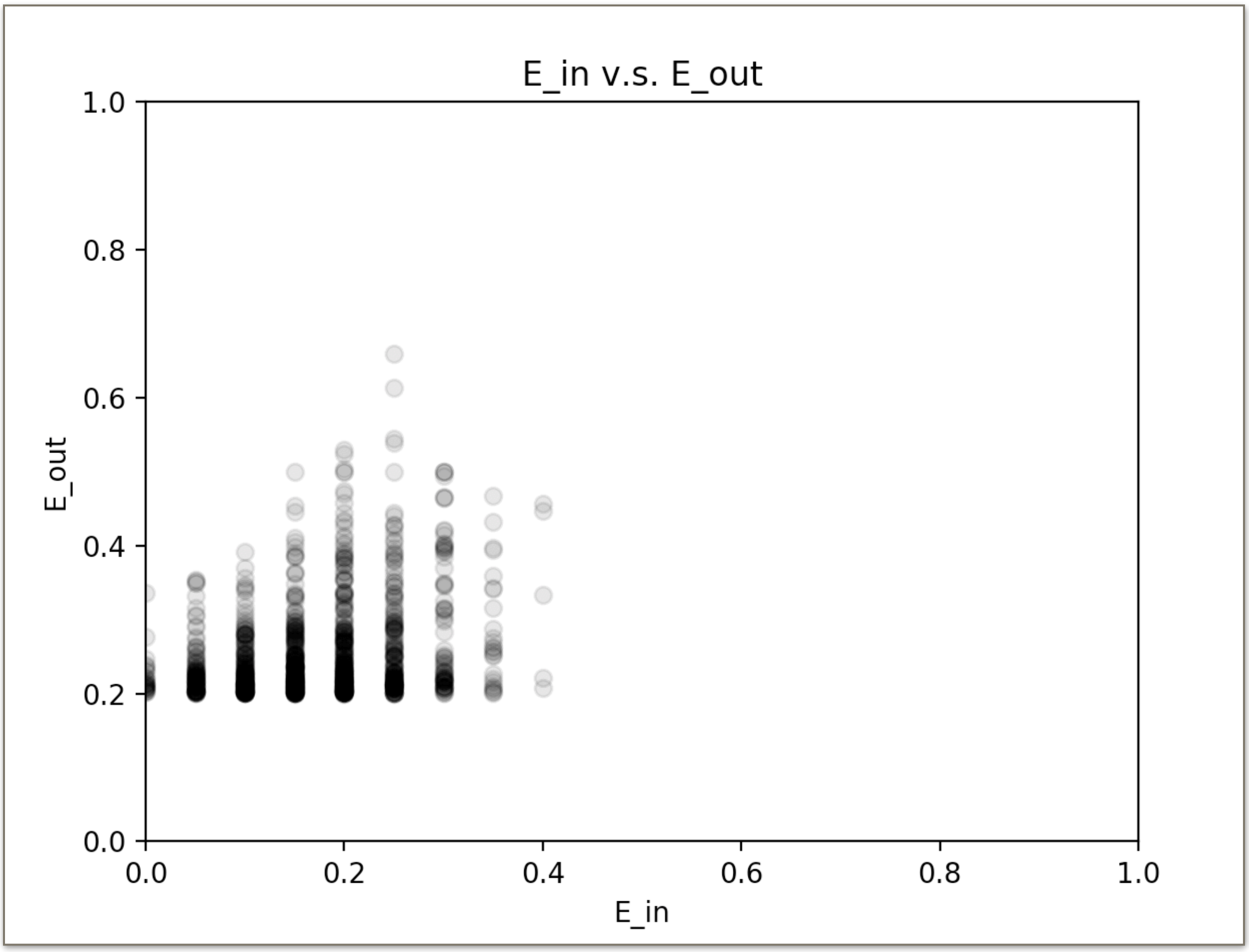
$$= (1 - \frac{|\theta|}{2}) \times 80\% + \frac{|\theta|}{2} \times 20\%$$

$$= 0.8 - 0.4 \times |\theta| + 0.1 \times |\theta|$$

$$= 0.8 - 0.3|\theta|$$

$\Rightarrow \underline{E_{out}(h_{s,\theta}) = 0.5 + 0.3(1 - |\theta|)s\#}$

Problem 7.



Problem 8.

$$B(N, k) \leq \sum_{i=0}^{k-1} C_i^N$$

Assume input is a vector with N dimensions, input ∈ {o, x}^{N}

k: minimum break point

There are C_0^N possibilities that the vector has 0 "o".

There are C_1^N possibilities that the vector has 1 "o".

There are C_2^N possibilities that the vector has 2 "o".

.
.
.
.
.

There are C_{k-1}^N possibilities that the vector has k-1 "o".

Therefore, there are $\sum_{i=0}^{k-1} C_i^N$ possibilities.

If we want to shatter k inputs in N, there must be 2^k permutations. However, what we calculate above doesn't contain the possibility of k "o" in the vector. That is, k inputs in N cannot be shattered. "k" is the minimum break point. And then, $B(N, k) = \sum_{i=0}^{k-1} C_i^N$ now.