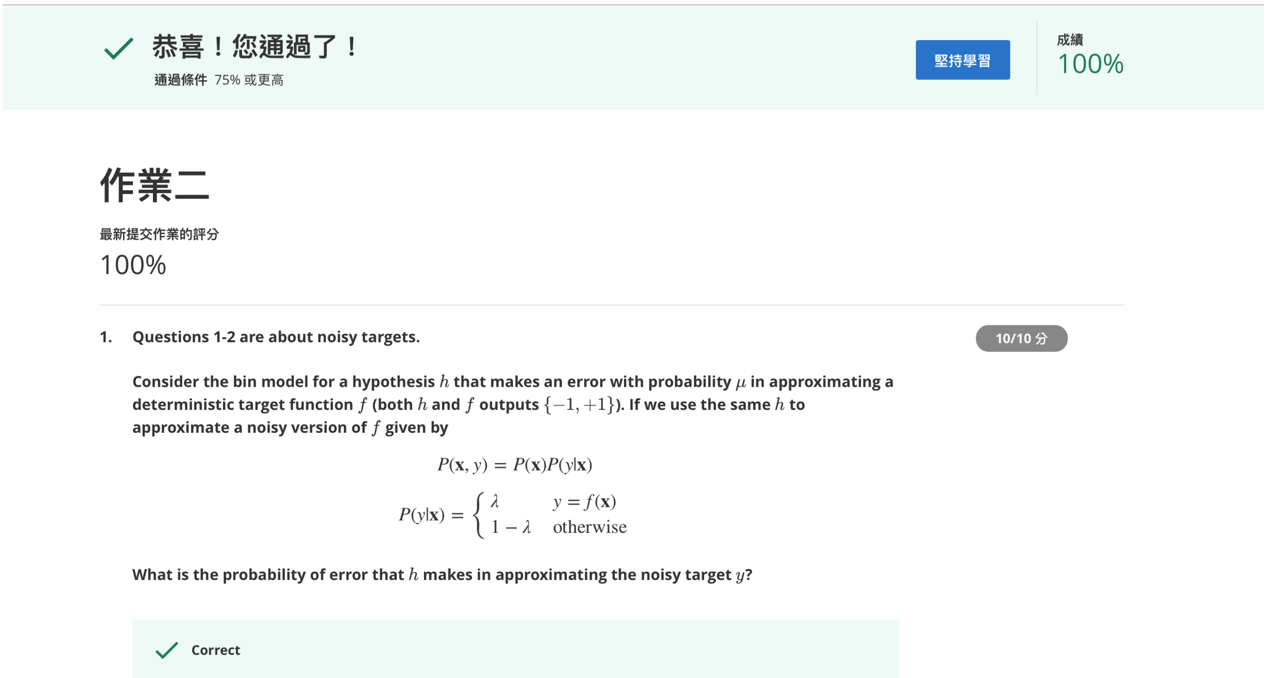
Machine Learning Foundations HW2

B06705028 資管三 朱紹瑜

1.



恭喜！您通過了！

通過條件 75% 或更高

堅持學習

成績

100%

作業二

最新提交作業的評分

100%

1. Questions 1-2 are about noisy targets.

10/10 分

Consider the bin model for a hypothesis h that makes an error with probability μ in approximating a deterministic target function f (both h and f outputs $\{-1,+1\}$). If we use the same h to approximate a noisy version of f given by

$$P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$$

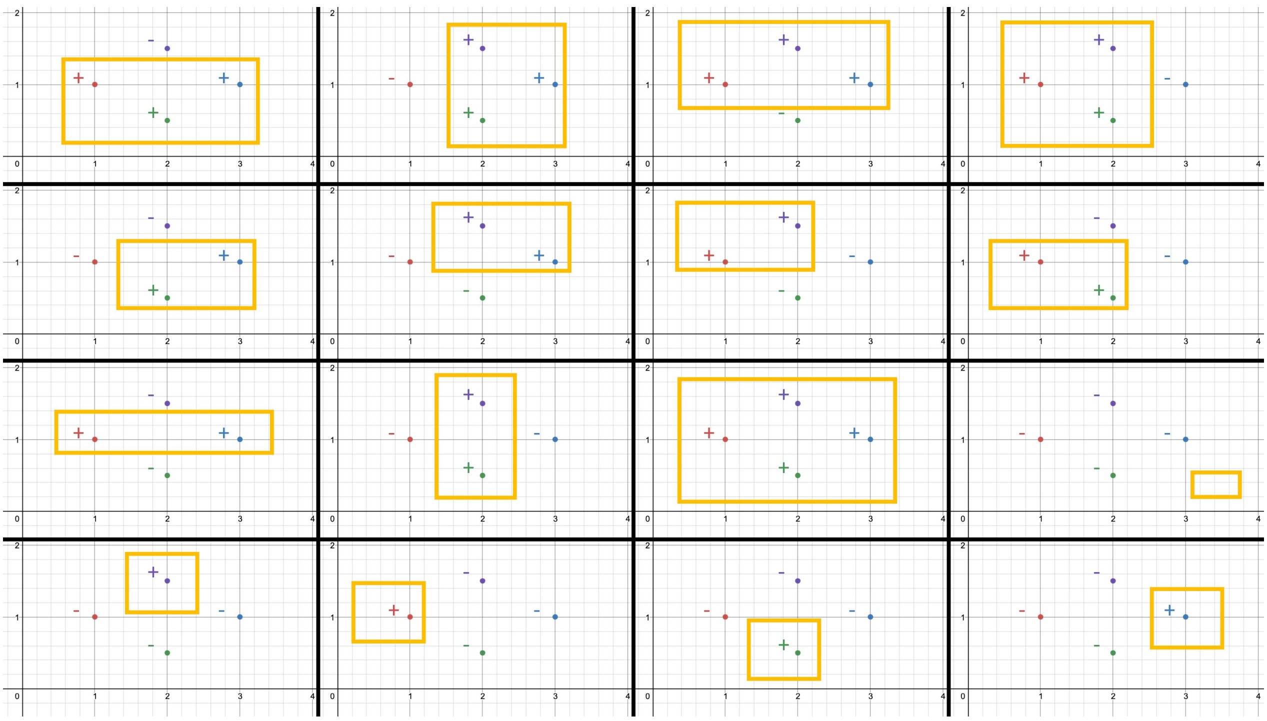
$$P(y|\mathbf{x}) = \begin{cases} \lambda & y = f(\mathbf{x}) \\ 1-\lambda & \text{otherwise} \end{cases}$$

What is the probability of error that h makes in approximating the noisy target y?

Correct

2.

To prove the VC-Dimension of the "positive rectangle" hypothesis set $\mathcal{H}$ is no less than 4, we try to show that there exists an input of size 4 that can be shattered by $\mathcal{H}$.



The image above shows one such input. From the images we can see, for all 16 dicotomies, there is at least one hypothesis (the yellow rectangle) from $\mathcal{H}$ that satisfies the specific dicotomy. This is equalivent to say $\mathcal{H}$ shatters such input.

Since there exist input of size 4 that can be shattered by $\mathcal{H}$, the VC-Dimension of the "positive rectangle" hypothesis set $\mathcal{H}$ is no less than 4.

3.

$d_{VC}(\mathcal{H}) = \infty$

Proof We prove $d_{VC}(\mathcal{H}) = \infty$ by showing that $\mathcal{H}$ shatters input of any size.

Let $X = \{x_1, x_2, \ldots, x_N\}$ be an input of size N, where $x_i = 2^i$ and N is an arbitrary positive integer. Let $\alpha_0 \in \mathbb{R}$ be an arbitrary number less than 4. We can express α_0 in binary as $\alpha_0 = (b_1 b_0 . b_{-1} b_{-2} \ldots)_2$, where $b_i \in \{0, 1\}$, which means $\alpha_0 = b_1 \cdot 2^1 + b_0 \cdot 2^0 + b_{-1} \cdot 2^{-1} + b_{-2} \cdot 2^{-2} + \ldots$

For any $x_i = 2^i \in X$,

$$\begin{aligned} h_{\alpha_0}(x_i) &= sign(|\alpha_0 x_i \bmod 4 - 2| - 1) \\ &= sign(|((b_1 \cdot 2^1 + b_0 \cdot 2^0 + b_{-1} \cdot 2^{-1} + b_{-2} \cdot 2^{-2} + \ldots) \cdot 2^i) \bmod 4 - 2| - 1) \\ &= sign(|(b_1 \cdot 2^{1+i} + b_0 \cdot 2^{0+i} + b_{-1} \cdot 2^{-1+i} + b_{-2} \cdot 2^{-2+i} + \ldots) \bmod 4 - 2| - 1) \\ &= \begin{cases} +1 & \text{if } b_{1-i} = 1 \\ -1 & \text{if } b_{1-i} = 0 \end{cases} \end{aligned}$$

From the equations above, we can see that $y_i = h_{\alpha_0}(x_i)$ can be decided by a single bit in α_0. To fit any dicotomy $H \in \{+1, -1\}^N$, we can build α' by 'flipping' b_{1-i} from α_0 for all i that satisfy $h_{\alpha_0}(x_i) \neq H_i$. Thus, $\mathcal{H}$ shatters X.

Since $\mathcal{H}$ shatters input of any size, $d_{VC}(\mathcal{H}) = \infty$.

4.

Claim If hypothesis sets $\mathcal{A}$ and $\mathcal{B}$ have the property $\mathcal{A} \subseteq \mathcal{B}$, then $d_{VC}(\mathcal{A}) \leq d_{VC}(\mathcal{B})$.

(Proof)

Assume $d_{VC}(\mathcal{A}) = N$, which means there exists at least one input of size N that can be shattered by $\mathcal{A}$. Let $\{x_1, x_2, \ldots, x_N\}$ be such an input and let $\{h_1, h_2, \ldots, h_N\} \subseteq \mathcal{A}$ be the set of hypotheses that shatters $\{x_1, x_2, \ldots, x_N\}$.

Since $\{h_1, h_2, \ldots, h_N\} \subseteq \mathcal{A}$ and $\mathcal{A} \subseteq \mathcal{B}$, $\{h_1, h_2, \ldots, h_N\} \subseteq \mathcal{B}$. It is guaranteed that $\mathcal{B}$ can shatter $\{x_1, x_2, \ldots, x_N\}$. Thus, $d_{VC}(\mathcal{B}) \geq N = d_{VC}(\mathcal{A})$.

Proof Since $(\mathcal{H}_1 \cap \mathcal{H}_2) \subseteq \mathcal{H}_1$, by the claim we proved above, we can easily get $d_{VC}(\mathcal{H}_1 \cap \mathcal{H}_2) \leq d_{VC}(\mathcal{H}_1)$.

5.

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

Then, we try to find $d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2)$ with the growth function. $N = 1 : m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) = 2 = 2^1$ $N = 2 : m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) = 4 = 2^2$ $N = 3 : m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) = 6 < 2^3$

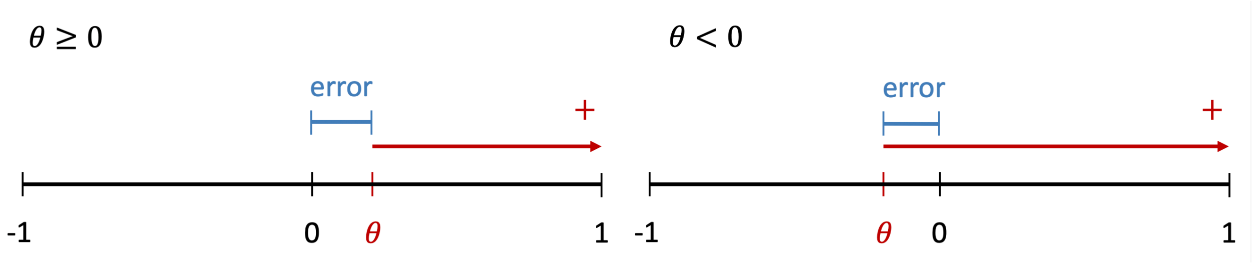
From the three equations and inequalities above, we get $d_{VC}(\mathcal{H}_1 \cup \mathcal{H}_2) = 2$.

6.

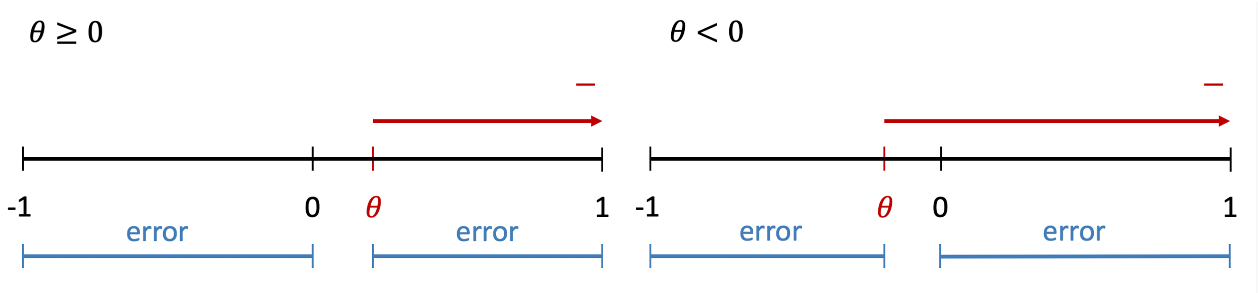
Assume that $h_{s,\theta}$ makes an error with probability μ in approximating the non-noise version of f.

$E_{out}(h_{\theta,s}) = (1 - 0.2)\mu + 0.2(1 - \mu) = 0.6\mu + 0.2$

Case 1: $s = +1$



Case 2: $s = -1$



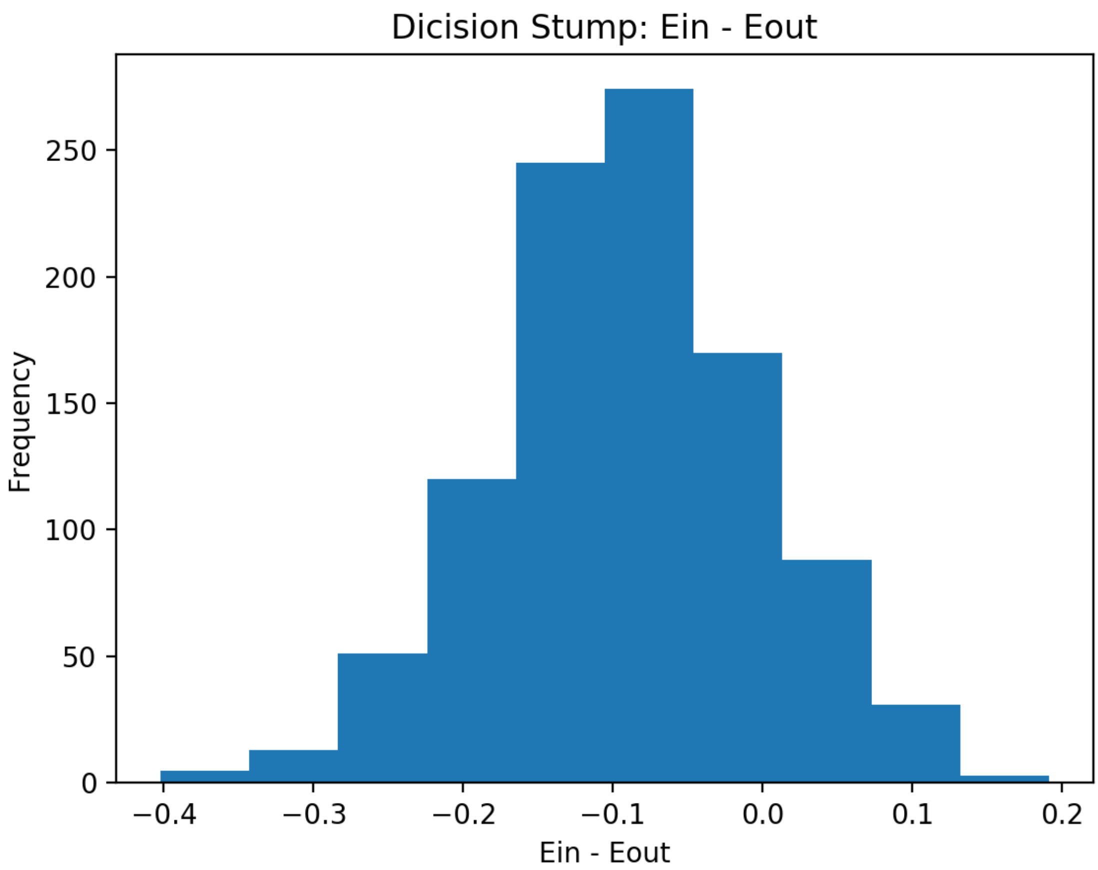
From the two cases above, we can see that $\mu = \begin{cases} \frac{|\theta|}{2} & \text{if } s = +1 \\ 1 - \frac{|\theta|}{2} & \text{if } s = -1 \end{cases}$

Then we combine the two cases and substitute μ in E_{out} with the value we obtained.

$\mu = \frac{s+1}{2} \cdot \frac{|\theta|}{2} - \frac{s-1}{2}(1 - \frac{|\theta|}{2}) = \frac{1}{2} + \frac{s}{2}(|\theta| - 1)$
$E_{out} = 0.6\mu + 0.2 = 0.6(\frac{1}{2} + \frac{s}{2}(|\theta| - 1)) + 0.2 = 0.5 + 0.3(|\theta - 1|)$

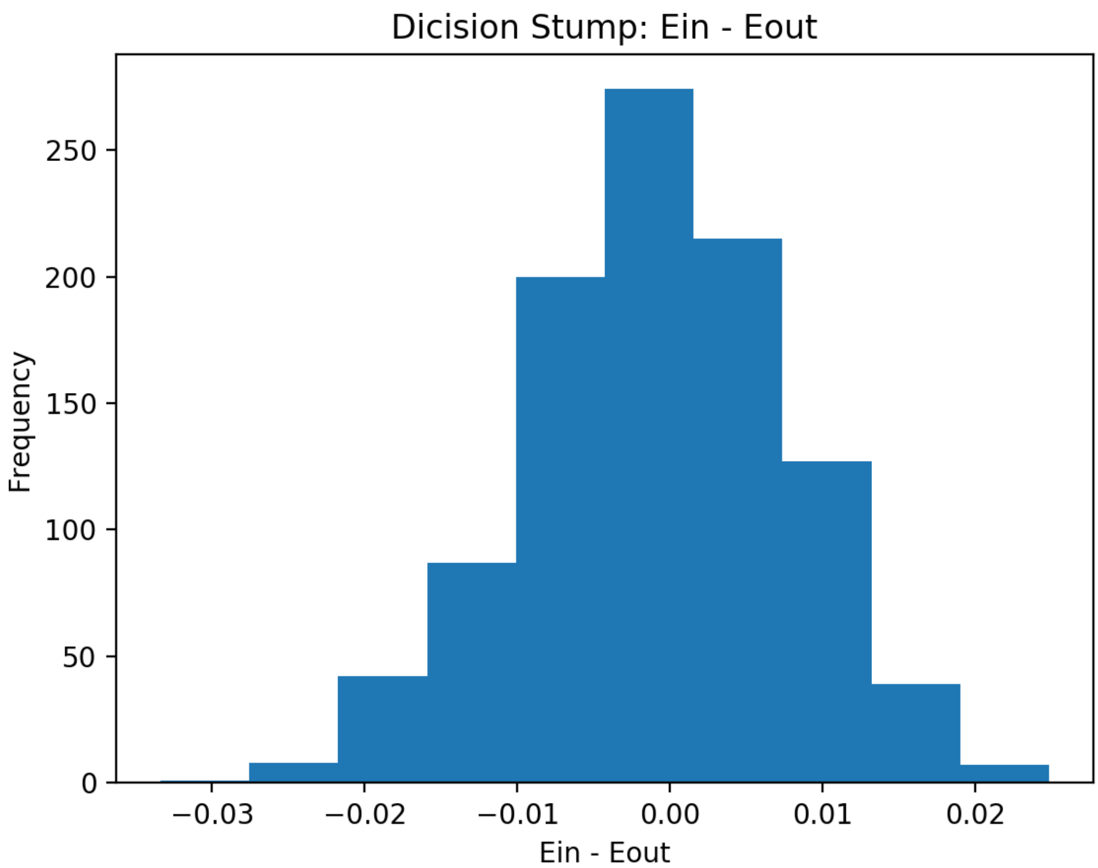
Thus, we get $E_{out}(h_{\theta,s}) = 0.5 + 0.3(|\theta - 1|)$.

7.



From the histogram, we can see that the distribution of $E_{in} - E_{out}$ is unimodel, and the mode class falls near -0.1. We can infer that E_{in} and E_{out} are often close and E_{in} is often slightly smaller than E_{out}.

8.



From the histogram, we can see that the distribution of $E_{in} - E_{out}$ is still unimodel, but the mode class gets closer to 0, comparing to what we've got in the previous problem. In this case, the data size is larger, dicotomies are distributed more evenly, causing a smaller difference between E_{in} and E_{out}.

9.

$$\mathcal{H} = \{h_{t,S} \mid h_{t,S}(\mathbf{x}) = 2[[\mathbf{v} \in S]] - 1, \text{where } \mathbf{v}_i = [[x_i > t_i]], \\ S \text{ is a collection of vectors in } \{0,1\}^d, t \in \mathbb{R}^d\}$$

To prove $d_{VC}(\mathcal{H}) = 2^d$, we try to show $2^d \leq d_{VC}(\mathcal{H})$ and $d_{VC}(\mathcal{H}) \leq 2^d$.

Claim 1: $2^d \leq d_{VC}(\mathcal{H})$

Proof of Claim 1 $2^d \leq d_{VC}(\mathcal{H})$ is equivalent to the statement "there exists input of size 2^d that $\mathcal{H}$ can shatter".

We construct the input X of size 2^d as the following. Get an arbitrary $t \in \mathbb{R}^d$. Let

$$X = \{t_1 - 1, t_1 + 1\} \times \{t_2 - 1, t_2 + 1\} \times \ldots \times \{t_d - 1, t_d + 1\} \\ = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{2^d}\}.$$

Let $V = \{v_1, v_2, \ldots, v_{2^d}\}$, where $v_i \in \{0,1\}^d$ is the result of thresholding $\mathbf{x}_i$ with t. According to our construction of X, $v_i \neq v_j$ if $i \neq j$. For any $y \in \{0,1\}^d$, we can find a hypothesis in $\mathcal{H}$ by selecting $S = \{v_i \mid i \in \{0,1,\ldots,2^d\} \text{ and } y_i = +1\}$. Thus, $\mathcal{H}$ can shatter X, so $2^d \leq d_{VC}(\mathcal{H})$.

Claim 2: $d_{VC}(\mathcal{H}) \leq 2^d$

Proof of Claim 2 Prove by contradiction. Assume $d_{VC}(\mathcal{H}) > 2^d$, which means there exists input of size $2^d + 1$ that $\mathcal{H}$ can shatter.

Let such input be $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{2^d}, \mathbf{x}_{2^d+1}\}$. Let $v_i = [[x_i > t_i]] \in \{0, 1\}^d$ for every $i \in \{1, 2, \ldots, 2^d + 1\}$. Since $|X| = 2^d + 1 > 2^d = |\{0, 1\}^d|$, according to pigeonhole priciple, there must exist i, j, where $i \neq j$ and $v_i = v_j$. Thus, no matter how S is selected, $y_i = y_j$. It is equivalent to say that $\mathcal{H}$ cannot shatter any input of size $2^d + 1$, which contradicts our assumption.

By contradiction, we proved that $d_{VC}(\mathcal{H}) \leq 2^d$.

Proof

By combining **Claim 1** and **Claim 2**, we get $2^d \leq d_{VC}(\mathcal{H}) \leq 2^d$, which indicates $d_{VC}(\mathcal{H}) = 2^d$.