R08944035 呂翊愷

Machine Learning Foundations
Homework #2

1.

作業二
評分測驗 • 40 min

截止時間 12月15日 22:59 PST

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

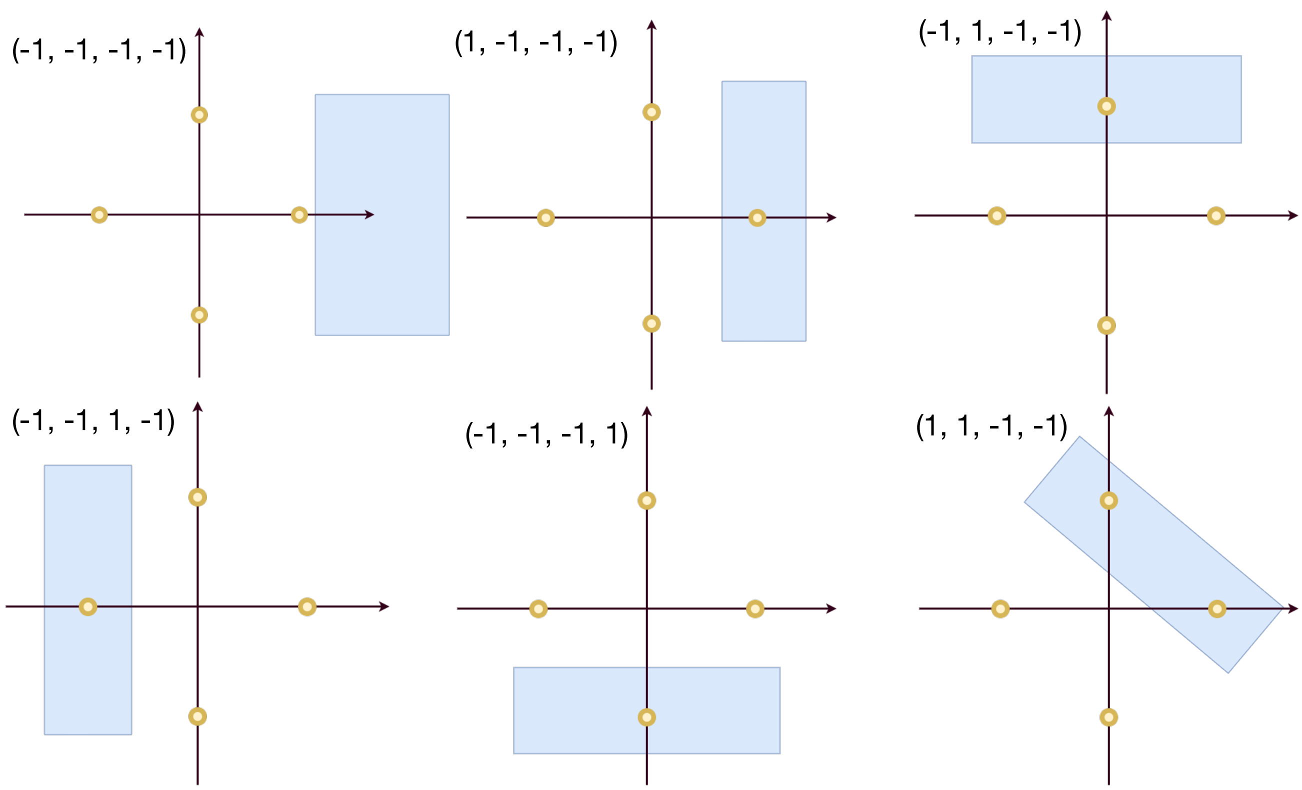
2. Following Question 1, with what value of λ will the performance of h be independent of μ? 10/10 分

2.

設$H=\{$positive rectangle in $R^2\}$，欲證明$d_{vc}(H) \geq 4$

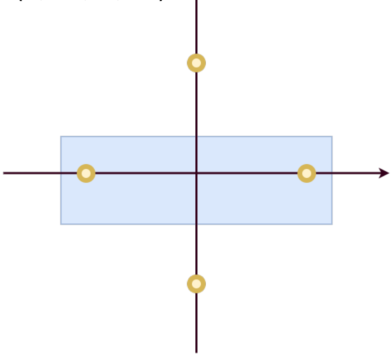
⇒即可找到一組4個點皆可被H所shatter

若4個點分別為$(1,0),(0,1),(-1,0),(0,-1)$，以下列出這4個點被$H$ shatter的16種dichotomies

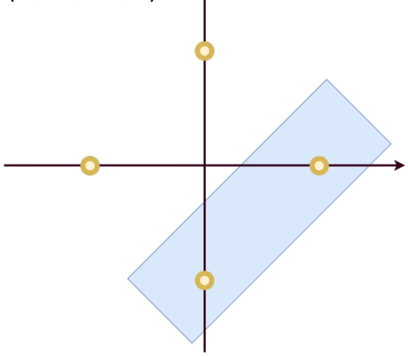


R08944035 呂翊愷

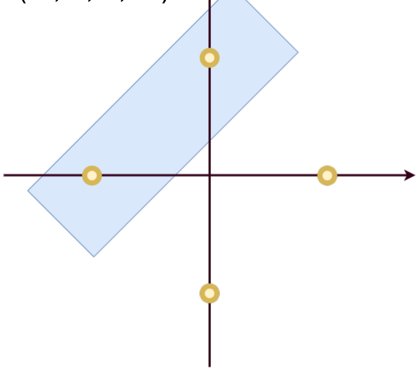
(1, -1, 1, -1)



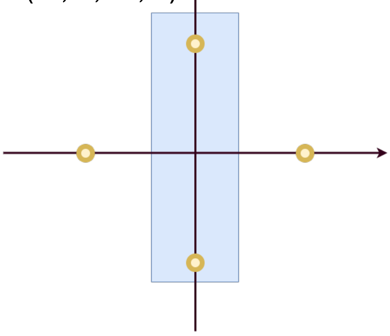
(1, -1, -1, 1)



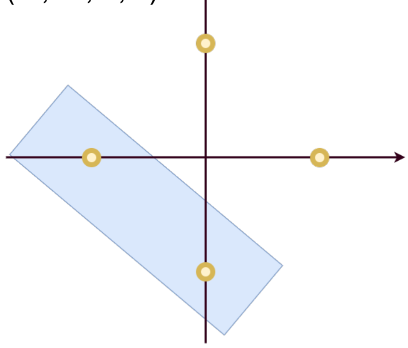
(-1, 1, 1, -1)



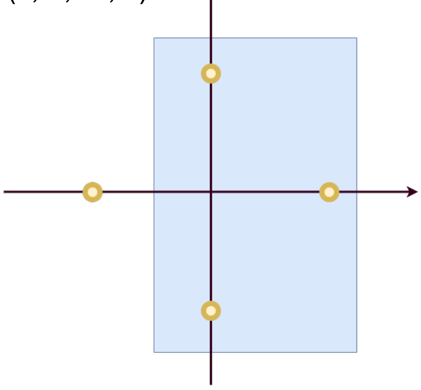
(-1, 1, -1, 1)



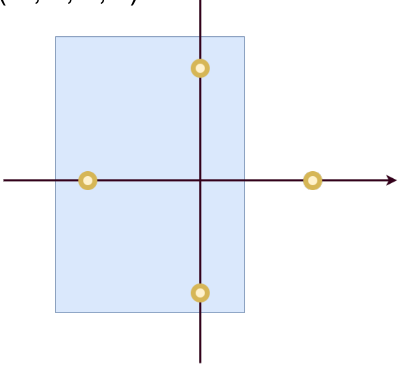
(-1, -1, 1, 1)



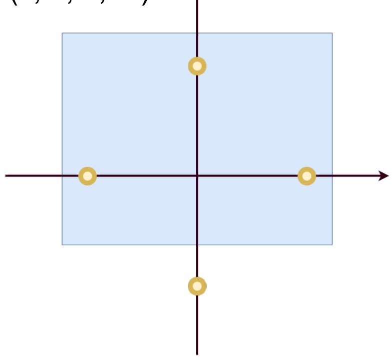
(1, 1, -1, 1)



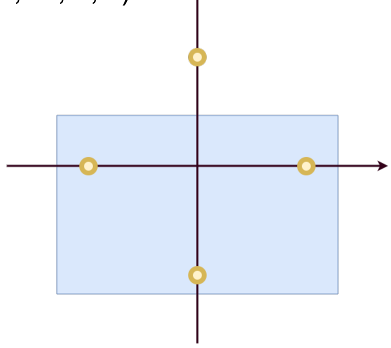
(-1, 1, 1, 1)



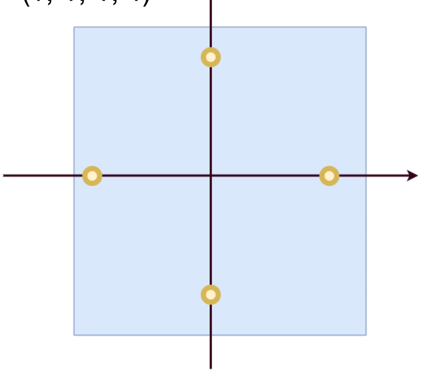
(1, 1, 1, -1)



(1, -1, 1, 1)



(1, 1, 1, 1)



因此得證$d_{vc}(H) \geq 4$

R08944035 呂翊愷

3.
假設X有N個點，分別為$x_1, x_2, ..., x_N$
令$x_i = 2^i$，$i = 1, 2, ..., N$
欲證明有N個X點時，能被2^N種dichotomies shatter，其中每種dichotomy為：
$\dfrac{i}{2^N} < \alpha < \dfrac{i+1}{2^N}, i = 0, 1, 2, ..., 2^N - 1$
根據數學歸納法，

① $N = 1$時，X有1個點$x_1 = 2$，則
$$\begin{cases} 0 < \alpha < \frac{1}{2}\text{時，}H(x_1) = +1 \\ \frac{1}{2} < \alpha < 1\text{時，}H(x_1) = -1 \end{cases}$$
可shatter出1個點，
因此有1個X點時，能被$2 = 2^1$種dichotomies shatter

② $N = 2$時，X有2個點$x_1 = 2, x_2 = 4$，則
$$\begin{cases} 0 < \alpha < \frac{1}{2^2}\text{時，}H(x_1) = +1, H(x_2) = +1 \\ \frac{1}{2^2} < \alpha < \frac{2}{2^2}\text{時，}H(x_1) = +1, H(x_2) = -1 \\ \frac{2}{2^2} < \alpha < \frac{3}{2^2}\text{時，}H(x_1) = -1, H(x_2) = +1 \\ \frac{3}{2^2} < \alpha < 1\text{時，}H(x_1) = -1, H(x_2) = -1 \end{cases}$$
可shatter出2個點，
因此有2個X點時，能被$4 = 2^2$種dichotomies shatter

③ 令$N = k$時，有k個點$x_1, x_2, ..., x_k$，
則$\dfrac{i}{2^k} < \alpha < \dfrac{i+1}{2^k}, i = 0, 1, 2, ..., 2^k - 1$時，
可shatter出k個點

④ 令$N = k + 1$，有k+1個點$x_1, x_2, ..., x_k, x_{k+1}$，
則將$N = k$時α的每段範圍再切一半，使得x_{k+1}可以被shatter，
而切一半後得到的範圍即為$\dfrac{i}{2^{k+1}} < \alpha < \dfrac{i+1}{2^{k+1}}, i = 0,1,2,...,2^{k+1} - 1$
可shatter出$k + 1$個點，
因此有$k + 1$個X點時，能被2^{k+1}種dichotomies shatter

因此推廣到N個點即可證明X的任一點都可以被2^N種dichotomies shatter，所以
$d_{vc}(H) = \infty$

4.
因為$H_1 \cap H_2$可以shatter $d_{vc}(H_1 \cap H_2)$個點，又$(H_1 \cap H_2) \subseteq H_1$，
所以代表H_1也可以shatter至少$d_{vc}(H_1 \cap H_2)$個點，
$\therefore d_{vc}(H_1 \cap H_2) \leq d_{vc}(H_1)$

5.

已知$m_{H_1}(N) = N + 1$且$m_{H_2}(N) = N + 1$，欲求$m_{H_1 \cup H_2}(N)$

$m_{H_1 \cup H_2}(N) = m_{H_1}(N) + m_{H_2}(N) - m_{H_1 \cap H_2}(N)$

而H_1和H_2的交集情況是全為1和全為-1兩種情況，

所以$m_{H_1 \cup H_2}(N) = m_{H_1}(N) + m_{H_2}(N) - m_{H_1 \cap H_2}(N) = N + 1 + N + 1 - 2 = 2N$

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

$N = 2 \Rightarrow m_{H_1 \cup H_2}(N) = 2N = 2 * 2 = 2^2$

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

$\therefore d_{vc}(H_1 \cup H_2) = 2$

6.

假設hypothesis h有μ的機率使得$h(x) \neq f(x)$，而λ為沒有噪音的機率，則

$$\begin{cases} y = f(x)\text{時，錯誤的機率為}\lambda\mu \\ y \neq f(x)\text{時，錯誤的機率為}(1-\lambda)(1-\mu) \end{cases}$$

$\Rightarrow E_{out} = \lambda\mu + (1-\lambda)(1-\mu)$

已知$\lambda = 0.8, x = [-1,1], h(x) = s * sign(x - \theta)$

(1) 若$s = 1, \theta > 0$，h在$x \in [0,\theta]$時會預測錯誤，

使$h(x) \neq f(x)$，而$x \in [0,\theta]$的機率為$\dfrac{\theta}{2}$

(2) 若$s = 1, \theta < 0$，h在$x \in [-\theta,0]$時會預測錯誤，

使$h(x) \neq f(x)$，而$x \in [-\theta,0]$的機率為$\dfrac{|\theta|}{2}$

(3) 若$s = -1, \theta > 0$，h在$x \notin [0,\theta]$時會預測錯誤，

使$h(x) \neq f(x)$，而$x \notin [0,\theta]$的機率為$\dfrac{2-\theta}{2}$

(4) 若$s = -1, \theta < 0$，h在$x \notin [-\theta,0]$時會預測錯誤，

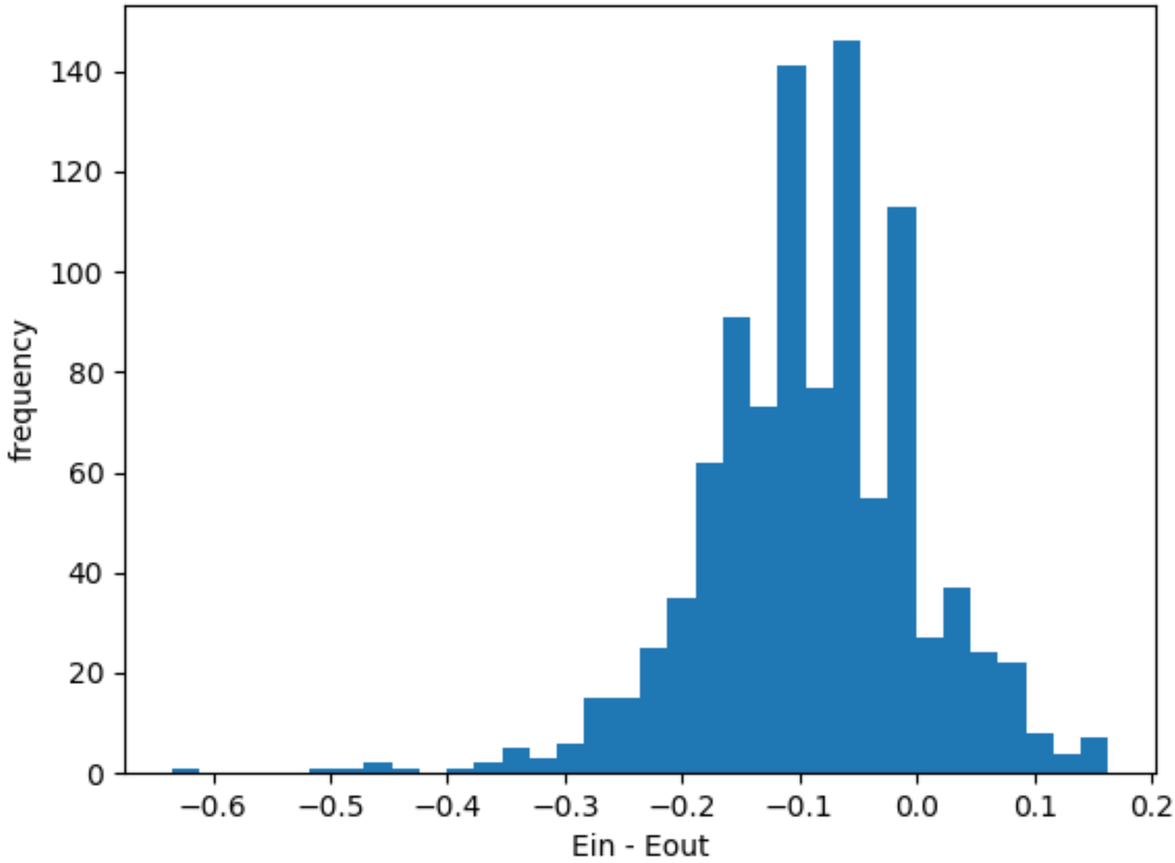
使$h(x) \neq f(x)$，而$x \notin [-\theta,0]$的機率為$\dfrac{2-|\theta|}{2}$

$$\therefore \mu = \left(\frac{s+1}{2}\right)\left(\frac{|\theta|}{2}\right) - \left(\frac{s-1}{2}\right)\left(\frac{2-|\theta|}{2}\right)$$

$E_{out} = \lambda\mu + (1-\lambda)(1-\mu), \lambda = 0.8$, 代入上式的$\mu$

即可得到$E_{out} = 0.5 + 0.3 * s * (|\theta| - 1)$

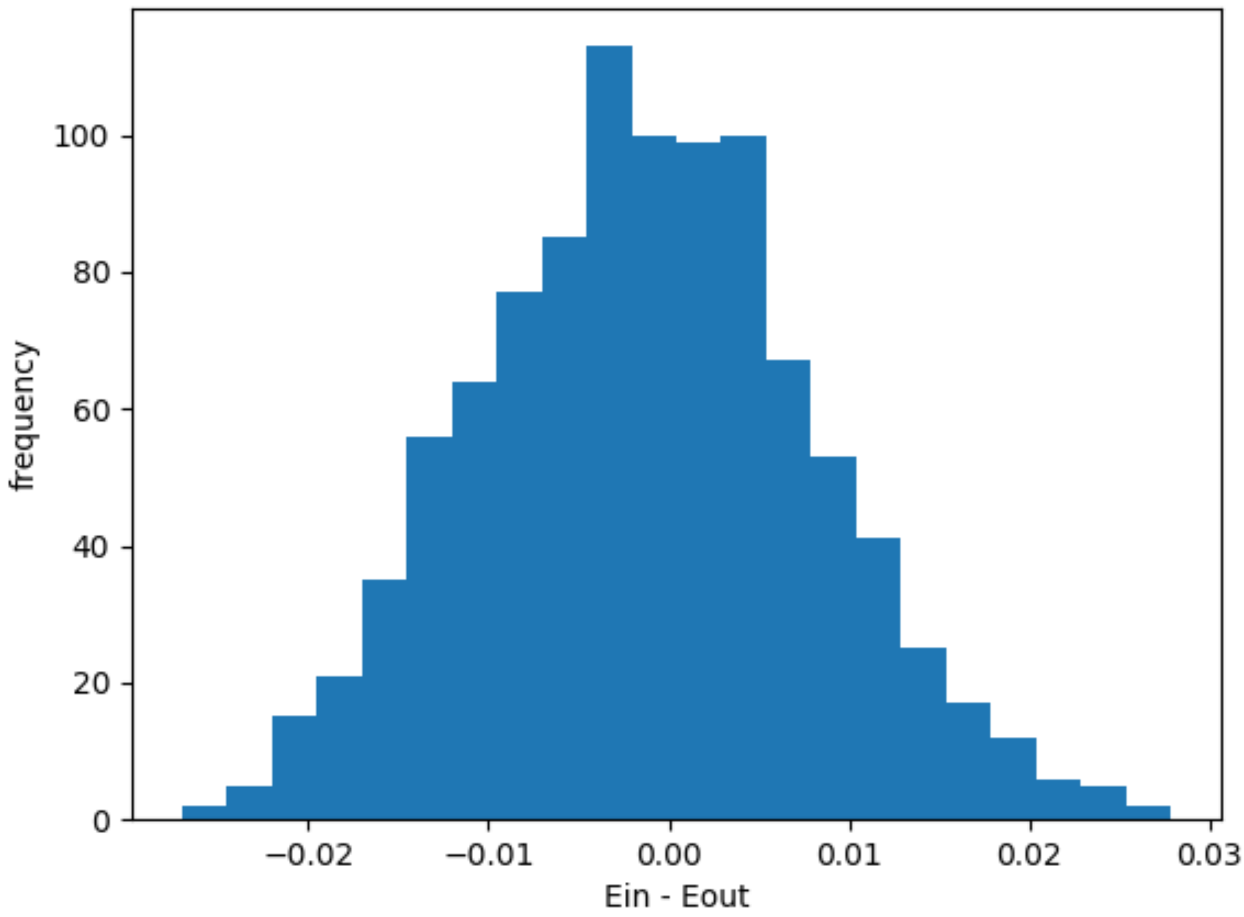
7.



average $E_{in} = 0.17070000000000002$
average $E_{out} = 0.2604758866102603$

由$E_{in} - E_{out}$的histogram可以發現，大部分的$E_{in} - E_{out}$落在-0.3～0之間，代表$E_{in} > E_{out}$且差距可達0.3，可見在datasize = 20的情況下，找到的hypothesis仍無法太準確預測data。

8.



average $E_{in} = 0.19930999999999968$
average $E_{out} = 0.20060542381074067$

由$E_{in} - E_{out}$的histogram可以發現，大部分的$E_{in} - E_{out}$落在-0.01～0之間，代表$E_{in} > E_{out}$，但E_{in}和E_{out}的差距大部分只有0.01，而整張圖看起來接近Gaussian distribution。

根據上課投影片中的公式：$\mathbb{P}_D[\,|E_{in}(g) - E_{out}(g)| > \epsilon] \leq 4(2N)^{d_{vc}} exp\left(-\frac{1}{8}\epsilon^2 N\right)$，$N$越大，發生壞事的機率會變低。因此第8題的datasize遠比第7題的datasize來得大的情況下，第8題的$E_{in} - E_{out}$也比第7題的小很多。