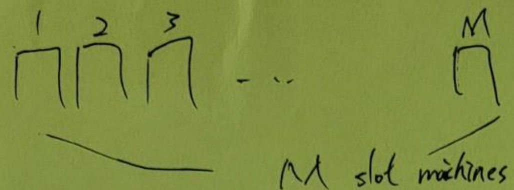
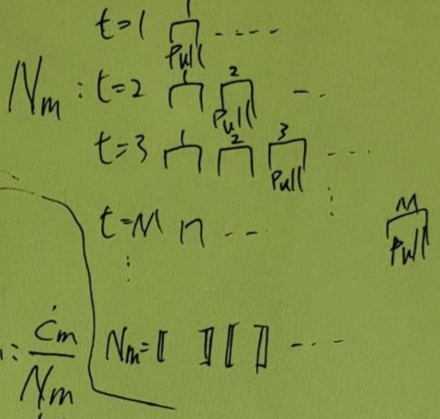
ML2023 HW2 R12631055 林東甫

1





$$N_m = \sum_{1}^{t} [\![m = ((t-1) \bmod M) + 1]\!]$$

$\therefore$ sample prob. of m for returing coin: $\dfrac{c_m}{N_m}$

by one-sided Hoeffding's inequality :

$$P\left(\mu_m > \frac{c_m}{N_m} + \epsilon\right) \le e^{(-2\epsilon^2 N_m)} \quad \text{let } \epsilon = \sqrt{\frac{\ln t - \frac{1}{2}\ln\delta}{N_m}}$$

where $0<\delta<1$: $$P\left(\mu_m > \frac{c_m}{N_m} + \sqrt{\frac{\ln t - \frac{1}{2}\ln\delta}{N_m}}\right) \le e^{(-2\epsilon^2 N_m)}$$

$$= e^{\left(-2\sqrt{\frac{\ln t - \frac{1}{2}\ln\delta}{N_m}}^{\,2} N_m\right)} = e^{-2(\ln t - \frac{1}{2}\ln\delta)} = e^{(-2\ln t + \ln\delta)}$$

$$= t^{-2}\cdot\delta \Rightarrow P\left(\mu_m > \frac{c_m}{N_m} + \sqrt{\frac{\ln t - \frac{1}{2}\ln\delta}{N_m}}\right) \le \delta t^{-2}$$

2. Since there are M slot machines and $t = M+1, M+2, \cdots$

Let $\epsilon = \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln\delta}{N_m}}$, $\sum_{t=M+1}^{\infty}\sum_{M=1}^{M} e^{\left(-2\cdot\sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln\delta}{N_m}}^{2} N_m\right)}$

$$= \sum_{t=M+1}^{\infty}\sum_{M=1}^{M} e^{-2(\ln t + \ln M - \frac{1}{2}\ln\delta)} = \sum_{t=M+1}^{\infty}\sum_{M=1}^{M} t^{-2}\cdot M^{-2}\cdot\delta$$

$$\leq \delta\cdot M^{-2}\cdot\frac{\pi^2}{6}$$ by the magical fact.

$\therefore P\left(\mu_m \geq \frac{C_m}{N_m} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln\delta}{N_m}}\right) \leq \delta M^{-2}\frac{\pi^2}{6} < \delta$

$\forall M \geq 2, M \in N, t = M+1, M+2, \cdots$

$$1 - P\left(\mu_m > \frac{C_m}{N_m} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln\delta}{N_m}}\right) = P\left(\mu_m \leq \frac{C_m}{N_m} + \sqrt{\frac{\ln t + \ln M - \frac{1}{2}\ln\delta}{N_m}}\right) \geq (1-\delta)$$

that is: with prob. at least $1-\delta$:

$$\mu_m \leq \frac{C_m}{N_m} + \sqrt{\frac{\ln t + \ln M + \frac{1}{2}\ln\delta}{N_m}}$$

#

3. P(five tickets / ~~All~~ some number is purely green)

$$= \left(\cancel{\frac{1}{2}}\right)^5 \left(\frac{1}{2^5}+\frac{1}{2^5}+\frac{1}{2^5}+\frac{1}{2^5}\right) - \left(\frac{1}{4^5}+\frac{1}{4^5}+\frac{1}{4^5}+\frac{1}{4^5}\right)$$

$$= \frac{31}{256}$$ #

4. Since five tickets are all green 2's, then B and D will be only two would possibly tickets came from : $\frac{1}{32}$

5. ~~If~~ 4 input vectors be $x_1, x_2, x_3, x_4 \in \mathbb{R}^2$

by brute force

hypo.	x_1	x_2	x_3	x_4
1	-1	-1	-1	-1
2	-1	-1	-1	+1
3	-1	-1	+1	-1
4	-1	+1	-1	-1
5	+1	-1	-1	-1
6	-1	-1	+1	+1
7	-1	+1	+1	-1
8	+1	+1	-1	-1
9	-1	+1	-1	+1
10	+1	-1	-1	+1
11	+1	-1	+1	-1
12	-1	+1	+1	+1
13	+1	+1	+1	-1
14	+1	+1	-1	+1
15	+1	-1	+1	+1
16	+1	+1	+1	+1

-1 = within rectangle

+1 : outside of □

$\forall\, x_1, x_2, x_3, x_4 \in \mathbb{R}^2$

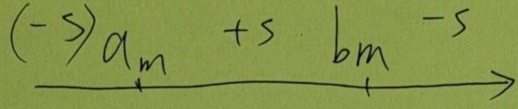
$\exists$ hypothesis sets can shatter it.

$\therefore$ the VC dimension of hypothesis set is no less than 4.

#

6. Known $2M+1$ parameters : $\begin{cases} s \in \{+1, -1\} \\ a_1 < b_1 < a_2 < b_2 < \cdots < a_m < b_m \end{cases}$ ~~$\forall m \in M$~~ $\forall m \le M, m \in \mathbb{N}$

$\forall m \le M$



Since $h_{s,a,b}(x)$ only has 2 outcome

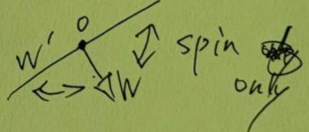
if there is x inputs, to shatter the hypothesis

$\therefore M \ge \frac{x}{2}$ ($M \ge \frac{x-1}{1}$ when x is odd), where the break points of hypothesis : $2M+2$, VC dimension is $2M+1$ #

7. Known $\mathcal{H}_0 = \{h : h(x) = \text{sign}(w_1 x_1 + w_2 x_2)$

Suppose 2 vector $W = (W_1, W_2)$, $x = (x_1, x_2)$

the normal vector of W : W'

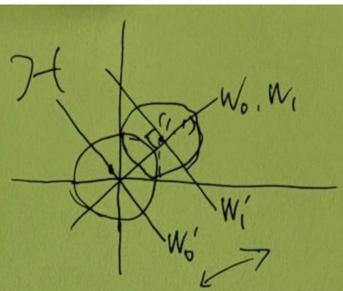


Assume there is n data input, and every data will be denote a sign of + or -, to every point, the perceptron's adjustment will create at least one whole new outcome. for N data, there will be 2N outcome at most.

(Since $w_1 x_1 + w_2 x_2$ pass the origin, means it only "spin", doesn't "shift")

$\therefore$ the growth function of origin-passing perceptrons: $g(x) = 2x$ #

8. known $\mathcal{H} = \mathcal{H}_0 \cup \mathcal{H}_1$



Consider $\mathcal{H}_0 = \mathcal{H}_0 \cup \mathcal{H}_1$ has

two normal vector ~~acr~~crossing $w_0' \times w_1'$ W_1 is like W_0 "shift" from 0 to (1,1)

Lif $N = 3$, $\exists\, x_1, x_2\, x_3$ s.t. $\mathcal{H}$ can shatter

x_1	x_2	x_3
+	+	+
+	+	−
+	−	+
−	+	+
+	−	−
−	+	−
−	−	+
−	−	−

$\therefore$ VC Dimensin is at least 3

Next, let $N = 4$ $\because$ perceptron can't let it be shattered.

Since ~~B~~ the normal vector of $\mathcal{H}$ is including in~~to~~ the sets of perceptron,

Thus the maximum possible $m_H(N)$ is 3

that is, the VC dimension is $\underline{3}$ #

9. known $h_{s,\theta}(x) = s \cdot \text{sign}(x-\theta)$, where $x \sim$ uniform$[-1,1]$,

$y = \text{sign}(x) + \text{noise}$, where the noise flips sign with 10%.

And $s \in \{-1, 1\}$, $\theta \in [-1, 1]$, noise $\updownarrow$ 10% of sign

Let $s = 1$, $E_{out} = P(h_{s,\theta}(x) \neq y)$:

$= 0.9 \cdot P(\text{sign}(x-\theta) \neq \text{sign}(x)) + 0.1 \cdot P(\text{sign}(x-\theta) = \text{sign}(x))$

$$P(\text{sign}(x-\theta) \neq \text{sign}(x)) = \begin{cases} \frac{\theta}{2} & \text{if } 1 \geq \theta \geq 0 \\ -\frac{\theta}{2} & \text{if } -1 \leq \theta < 0 \end{cases}$$

$$P(\text{sign}(x-\theta) = \text{sign}(x)) = \begin{cases} 1 - 0.5\theta & \text{if } 1 \geq \theta > 0 \\ 1 + 0.5\theta & \text{if } 1 \leq \theta < 0 \end{cases}$$

$E_{out} = 0.9 \cdot \frac{|\theta|}{2} + 0.1 \cdot (1 - 0.5|\theta|)$

$= 0.1 + 0.4|\theta|$ where $s = 1$

(if $s = -1$, $E_{out} = 0.9 \cdot (1 - 0.5\theta) + 0.1 \cdot \frac{|\theta|}{2}$

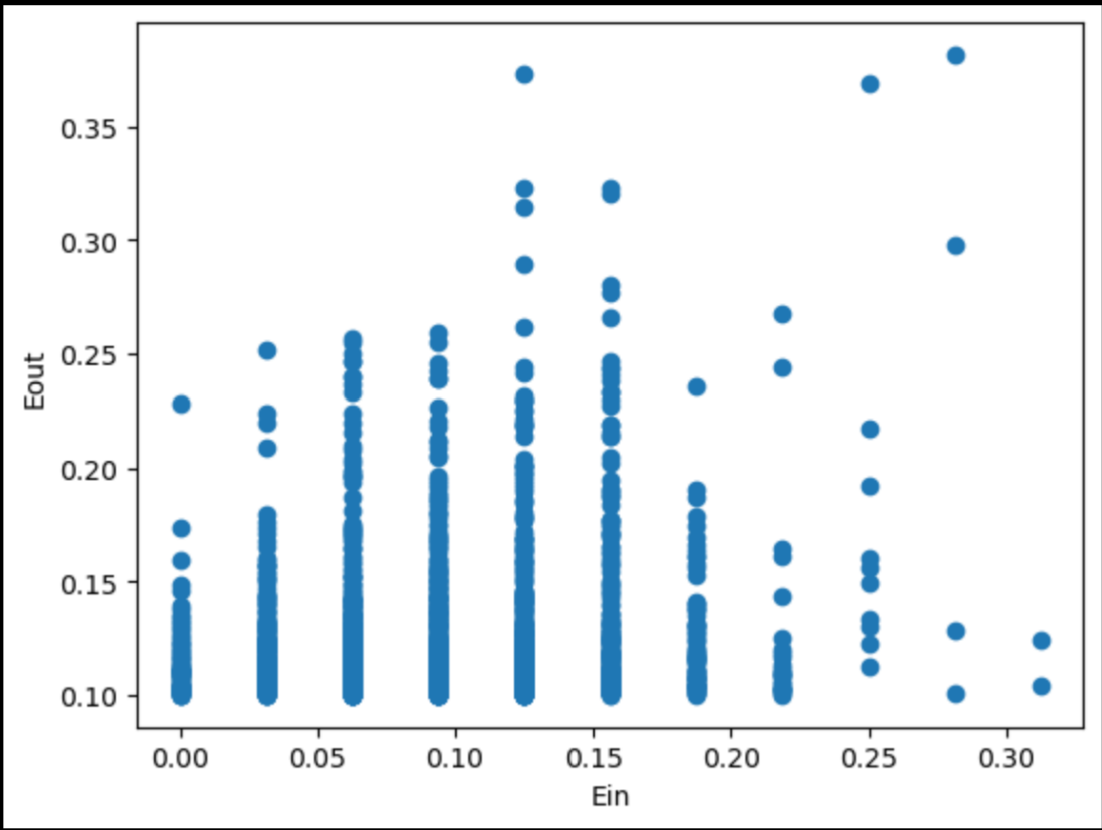
$= 0.9 - 0.4|\theta|$ where $s = -1$)

$\therefore E_{out}(h_{s,\theta}) = 0.5 - 0.4_s + 0.4 s \cdot \theta$

#

10

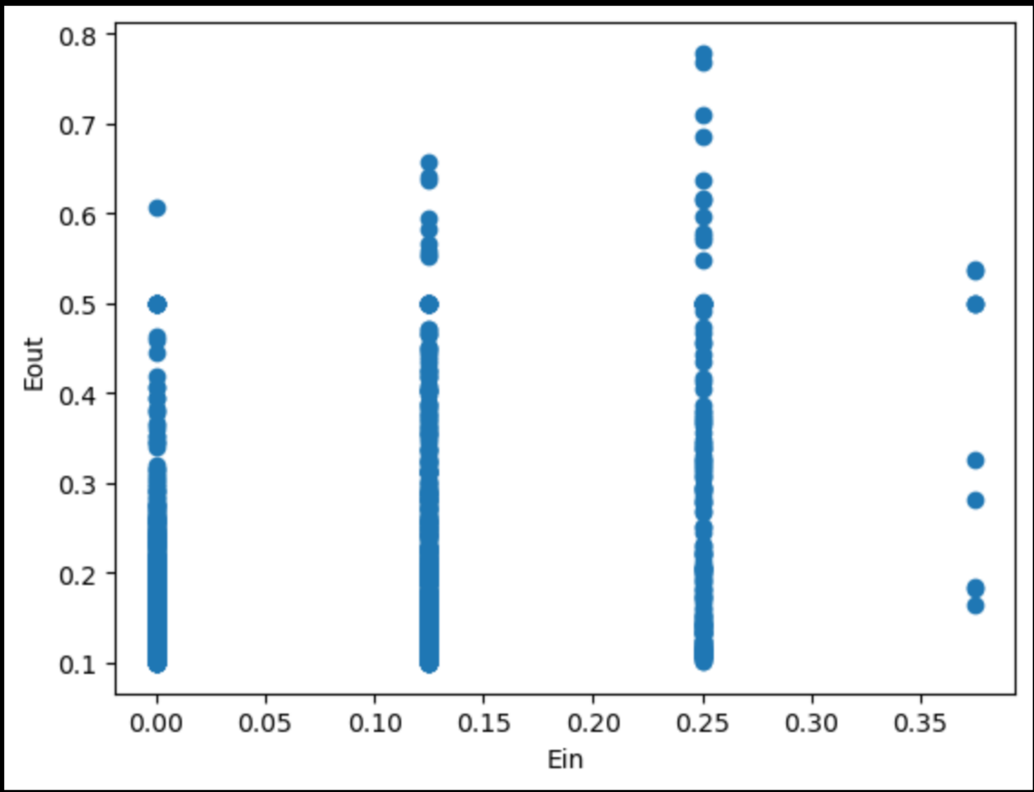
```
the median of Eout(g)-Ein(g): 0.038335814794216126
```



可以觀察到算是有點類似於常態分佈。

11

```
the median of Eout(g)-Ein(g): 0.12032149011795905
```

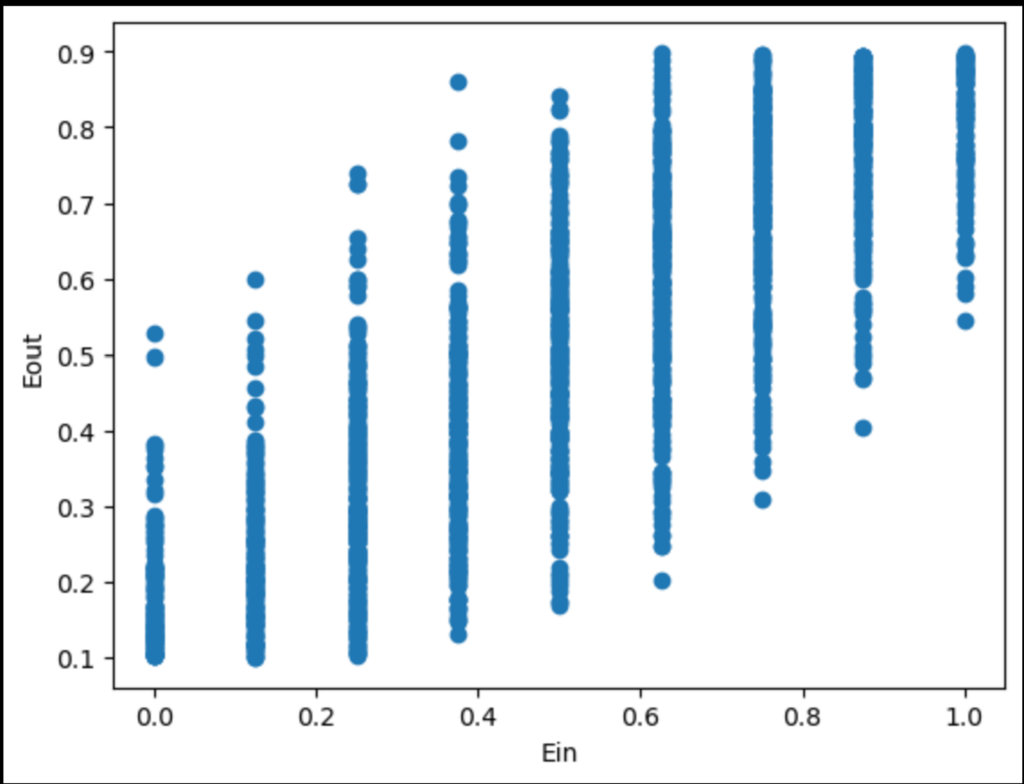


可以觀察到 size 大小從 32 減小到 8 後，Ein 的數量減少了，而 Eout 的範圍擴大，而且 Eout - Ein 的中位數更大了，這可能表示著在較小的 dataset 中，Eout 與 Ein 更不相似。

12

```
the median of Eout(g)-Ein(g): 0.00266144192234391
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...



可以觀察到透過 random uniform 抽樣， Ein 跟 Eout 變化幅度很大，而且相較於 10 和 11，Eout - Ein 的中位數變得非常小，這可能表示，Eout 非常接近 Ein。並且 Ein 和 Eout 有同等的升降幅度，有可能可以推測 Eout 和 Ein 具有相同的趨勢。

13

13. Let the total # of dichotomics be $C\binom{N}{d}$

~~for~~ $\forall$ ~~N~~ $N \in \mathbb{N}$ on $\mathbb{R}^d$

Known
1 1
1 2 1
1 3 3 1
1 4 6 4 1
⋮
pascal prymrid : $C\binom{N}{d} = C\binom{N-1}{d} + C\binom{N-1}{d-1}$

Thus $\forall$ $k = 1$, $M_{\mathcal{H}}(N) = \sum_{i=0}^{d-1} C\binom{N-1}{i}$

$= 2 \cdot \left(C\binom{N-1}{0} + C\binom{N-1}{1} \right) = 2N$

$\therefore$ $\forall$ ~~0~~ $k < d$, $\underline{M_{\mathcal{H}}(N) = 2 \cdot \sum_{i=0}^{d-k} C\binom{N-1}{i}}$