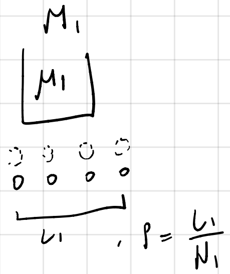
1. IM2 B11705022

If we only look at one machine,

$\mu = \mu_m$ which is the real probability for machine m to return a coin.

$\nu = \frac{c_m}{N_m}$ which is the outcome of the pulling process for machine m



let $\epsilon = \sqrt{\frac{\log t - \frac{1}{2}\log\delta}{N_m}}$

then $e^{-2\epsilon^2 N_m} = e^{-2\cdot\frac{\log t - \frac{1}{2}\log\delta}{\cancel{N_m}}\cdot\cancel{N_m}} = \frac{\delta}{t^2}$

therefore the Hoeffding's inequality becomes

$$P\left(\mu_m > \frac{c_m}{N_m} + \sqrt{\frac{\log t - \frac{1}{2}\log\delta}{N_m}}\right) \le \frac{\delta}{t^2}$$ ※ QED

2. First, we look at one machine only, say m

set $\epsilon = \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_m}}$ then $e^{(-2\epsilon^2 N)} = \frac{\delta}{t^2 M^2}$

$$P\left(\mu_m \le \frac{c_m}{N_m} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_m}}\right) \le \frac{\delta}{t^2 M^2}$$

For all machines, $m = 1, 2, 3 \ldots M$

$$P\left(\mu_1 \le \frac{c_1}{N_1} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_1}} \cup \mu_2 \le \frac{c_2}{N_2} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_2}} \cup \ldots \cup \mu_M \le \frac{c_M}{N_M} + \sqrt{\frac{\log t + \log N - \frac{1}{2}\log\delta}{N_M}}\right)$$

$$\le \sum_{i=1}^{M} P\left(\mu_i \le \frac{c_i}{N_i} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_i}}\right) \le \frac{\delta}{t^2 M^2}\cdot M = \frac{\delta}{t^2 M}$$

For all $t = M+1, M+2 \ldots \infty$ and $m = 1, 2 \ldots M$

$$\sum_{i=1}^{M} P\left(\mu_i \le \frac{c_i}{N_i} + \sqrt{\frac{\log(M+1) + \log M - \frac{1}{2}\log\delta}{N_i}} \cup \mu_i \le \frac{c_i}{N_i} + \sqrt{\frac{\log(M+2) + \log M - \frac{1}{2}\log\delta}{N_i}} \cdots \cup \mu_i \le \frac{c_i}{N_i} + \sqrt{\frac{\log\infty + \log M - \frac{1}{2}\log\delta}{N_i}}\right)$$

$$\le \sum_{i=1}^{M}\sum_{t=M+1}^{\infty} P\left(\mu_i \le \frac{c_i}{N_i} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_i}}\right) \le \sum_{t=M+1}^{\infty} \frac{\delta}{t^2 M} = \frac{\delta}{M}\cdot\frac{\pi^2}{6} < \delta$$

$\left(\because M \ge 2 \quad \therefore \frac{\pi^2}{6M}\cdot\delta < \delta\right)$

therefore, for all $m = 1, 2, 3 \ldots M$, $t = N+1, N+2 \ldots \infty$, $P\left(\mu_m > \frac{c_m}{N_m} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_m}}\right) \le \delta$

which is to say with probability at least $1-\delta$

$$\mu_m \le \frac{c_m}{N_m} + \sqrt{\frac{\log t + \log M - \frac{1}{2}\log\delta}{N_m}}$$ ※ QED

3.

A	B	C	D
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8

$A, B, C, D = \left(\frac{1}{4}\right)^5 = \frac{1}{1024}$

A×4 C×4 ~~A×4~~ D×4

AC AD

BC BD

~~B×4 C×4 D×4~~ D×4

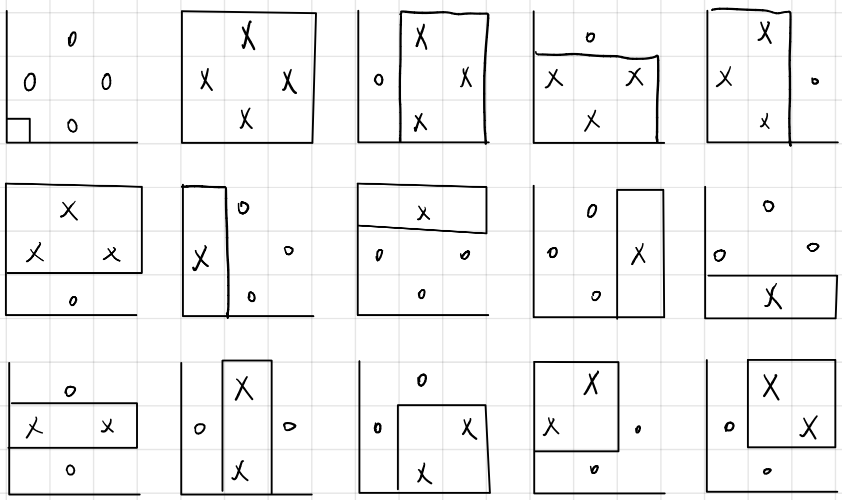
$\left(\frac{1}{2}\right)^5 \times 4 = \frac{1}{8}$

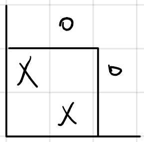
$\frac{1}{8} - \frac{1}{1024}\times 4 = \frac{1}{8} - \frac{1}{256} = \frac{31}{256}$ ※

4. BD

$\left(\frac{1}{2}\right)^5 = \frac{1}{32}$

5. let o be 1 , X be -1





For all of the 16 possibilities, the 4 input vectors are all shattered by the hypothesis set.

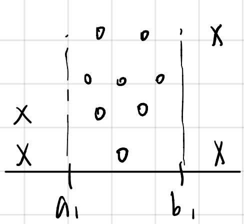
6. Idea:

We can see an $h(x)$ with $2M+1$ parameters can separate the real number line into $2M+1$ sections, when each section is already filled with one element, we add another data which isn't supposed to be inside, the dichotomies can't shatter $2M+2$ points.
Therefore, the break point is $2M+2$. $\Rightarrow$ VC Dimension $= 2M+1$

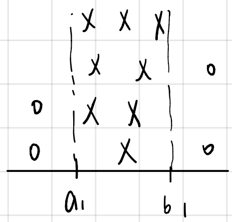
Proof:

Base case $(M=1)$:

when $N=3$, $s=1$

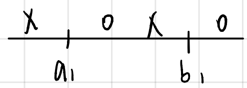


$s=-1$



as we can see, all 8 cases are shattered.

$N=4$, $s=1$



$\Rightarrow$ we can see it is not shattered

therefore when $M=1$, $d_{vc} = 2M+1 = 3$

Inductive case $(M>1)$:

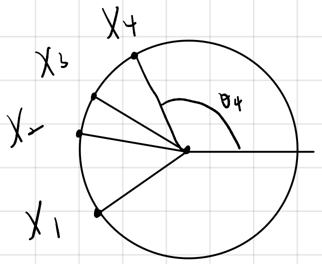
when $N = 2M+1$, all data points will have one interval to stay in.

when we increase M by 1, the interval increases by 2.

Therefore $d_{vc}(H_{M+1}) = d_{vc}(H_M) + 2 = (2M+1) + 2 = 2(M+1)+1$

From induction, $d_{vc}(H_M) = 2M+1$

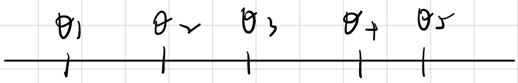
1.



X_1	X_2	X_3	X_4
o	o	o	o
x	x	x	x
o	x	x	x
o	o	x	x
o	o	o	x
x	x	x	o
x	x	o	o
x	o	o	o

Let the angle for each $X_1, X_2 \ldots$ to the origin be $\theta_1, \theta_2 \ldots$
Maximum lines appear when all data appear on the same semi-circle and doesn't appear on the same line that pass the origin. Because only by putting all X on a semi-circle, we can separte each of them 1 by 1.

We can see on the unit circle, when we stretch the circle into a line



Two dichotomy for each line passes between (θ_n, θ_{n+1})
And two for the line putting $\theta_1 \sim \theta_N$ on the same side.
Therefore $m_H(N) = 2N$

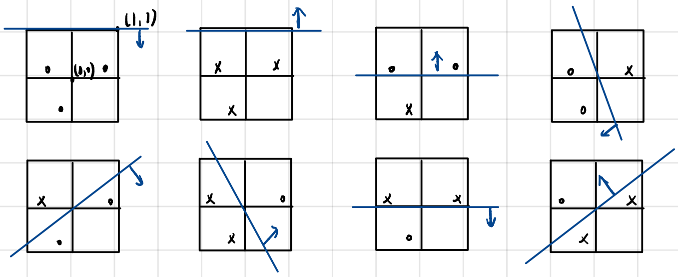
8. The growth function for H is $M_H(N) = M_{H_0 \cup H_1}(N) = M_{H_0}(N) + M_{H_1}(N) - M_{H_0 \cap H_1}(N)$

$H_0 \cup H_1(N)$ concludes lines that passes $(0,0)$ and $(1,1)$, $M_{H_0 \cap H_1}(N) = 2$ for any N

Then $M_H(N) = 2N + 2N - 2 = 4N - 2$, take $N=3$ $H(3) = 10 > 2^3$, $N=4$ $H(4) = 14 < 2^4$

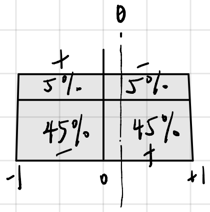
We can see the VC dimension is less than 4

Guess $N=3$:



Since above inputs are shattered by M_H, we conclude $d_{vc} = 3$ #

9. $E_{out}(h_{s,\theta}) = \underset{x\sim P}{\mathcal{E}}\; err(g(x), f(x))$



$$= \underbrace{(0.5 - 0.4s)} + \underbrace{(0.45s - 0.05s)\cdot|\theta|}$$

when $s=1$, 90% data is right

$s=-1$, 90% data is wrong

when θ moves,

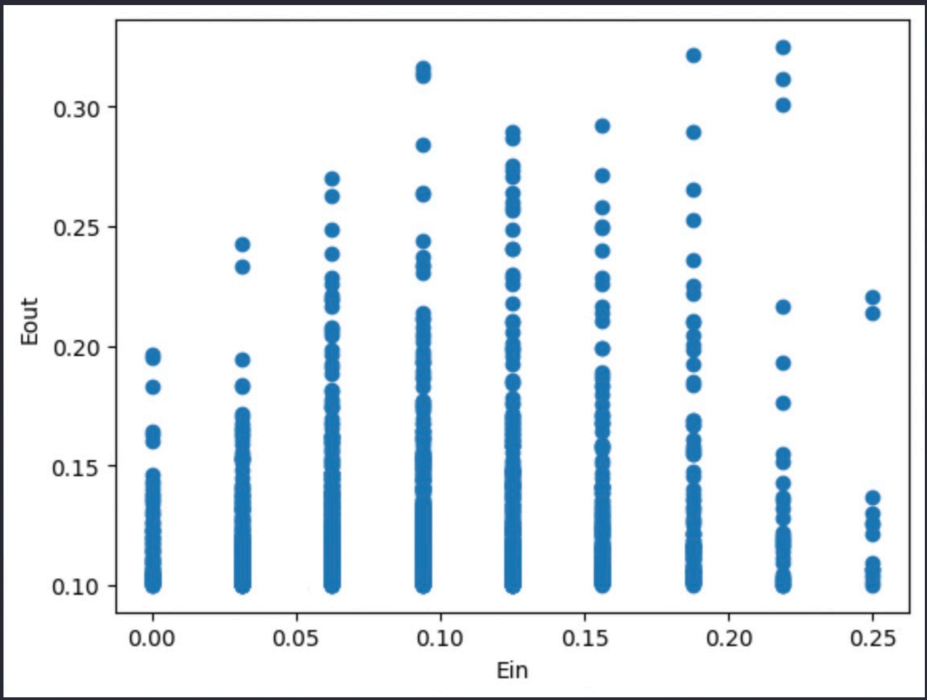
there will be $\begin{cases} 45\%\cdot\theta \text{ data become right/wrong} \\ 5\%\cdot\theta \text{ data become wrong/right.} \end{cases}$

$\Rightarrow$ whether right or wrong depend on s.

$$= -0.4s + 0.5 + 0.4s\cdot|\theta| \quad ※$$

10

Median: 0.03748215433346161



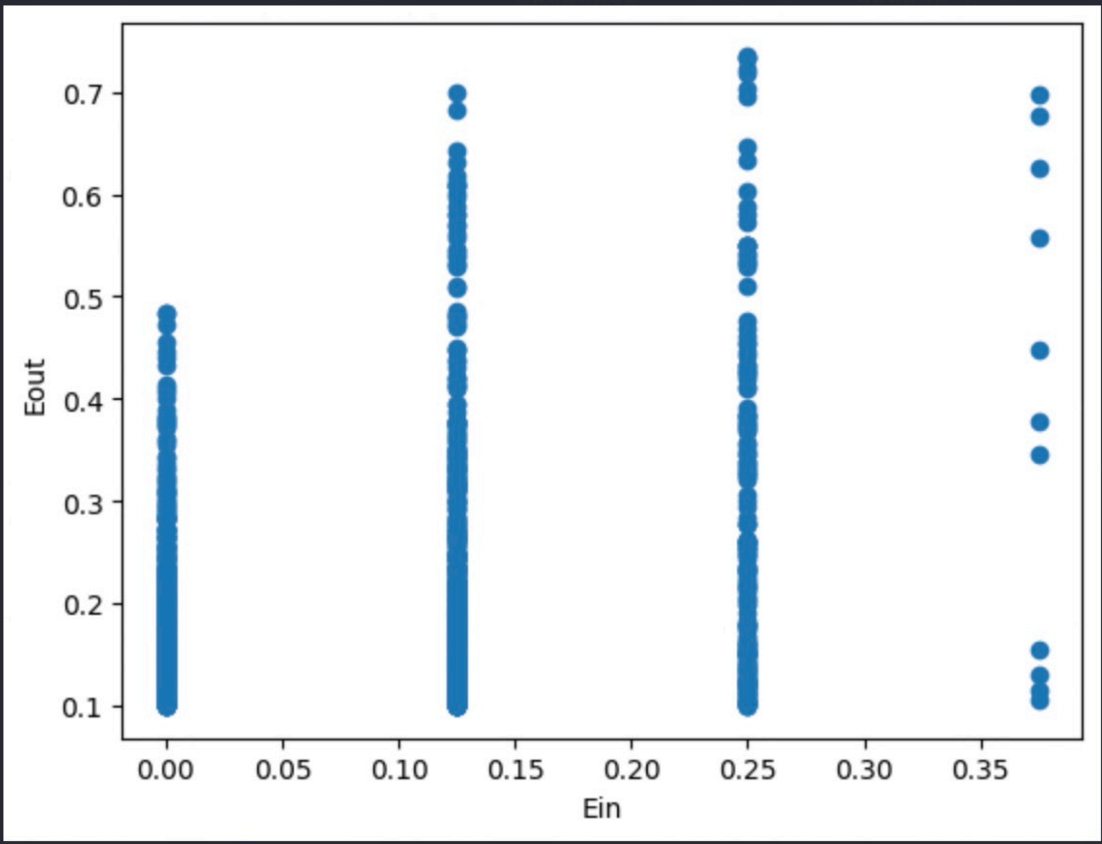
We can see that there are 9 bars on Ein, where the two bars near Ein = 0.1 have the highest ratio, and decreases outwards.

Those nine bars are n/32, n = [0, 8]

Moreover, Eout value appears the most at Ein = 0.1 as well.

The median value is pretty low, meaning that | Eout - Ein | is low

```
Median: 0.11936397811390756
```



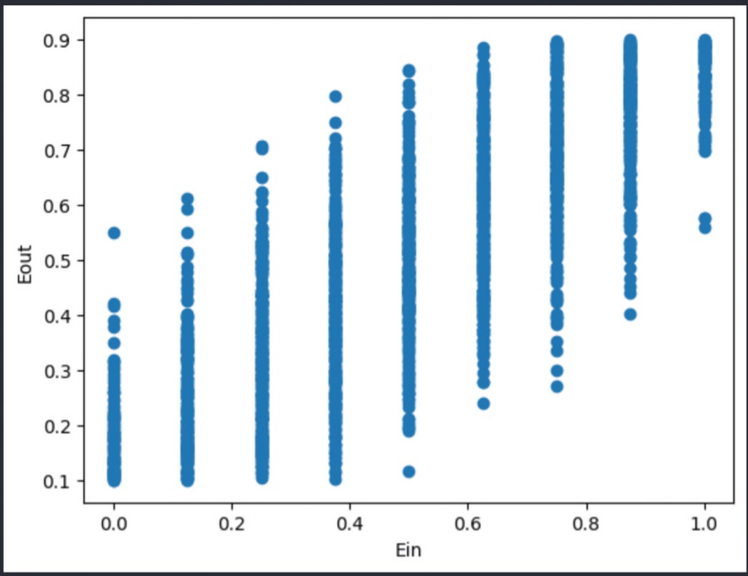
The same as previous, however only 4 bars left this time, which are n/8 n = [0,3]

Eout apppears the most at 0.1 as well.

The bar near Ein = 0.1 should be the most, but in this case it's hard to tell. The bar at Ein = 0.25 is pretty tall.

Moreover, Median grew a bit, with less data size, | Eout - Ein | becomes bigger.

```
Median -0.0016902767255976836
```



In this case, we can see that Ein grows with Eout.

The median became extremely low.

Since we are not looking for the best h, we can get really bad θ and s. from Eout = 0.5-0.4s+0.4s| θ | we can see that it grows with | θ | and s.

Therefore, Eout max value reaches 0.9.

Execution Order

Moreover, since s and θ is totally random, we can be totally right or totally wrong. Therefore, the value of Ein is from 0~1, since data set has size of 8, there are 9 possibility.