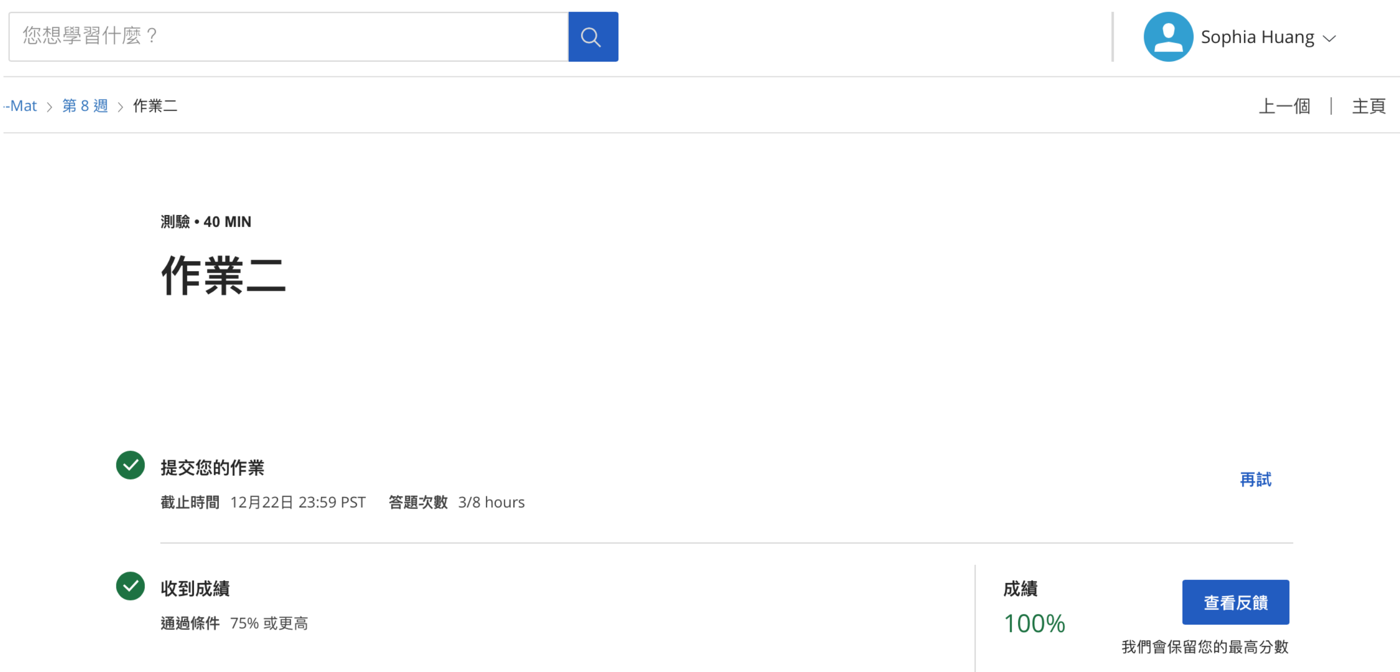
# MLF homework2

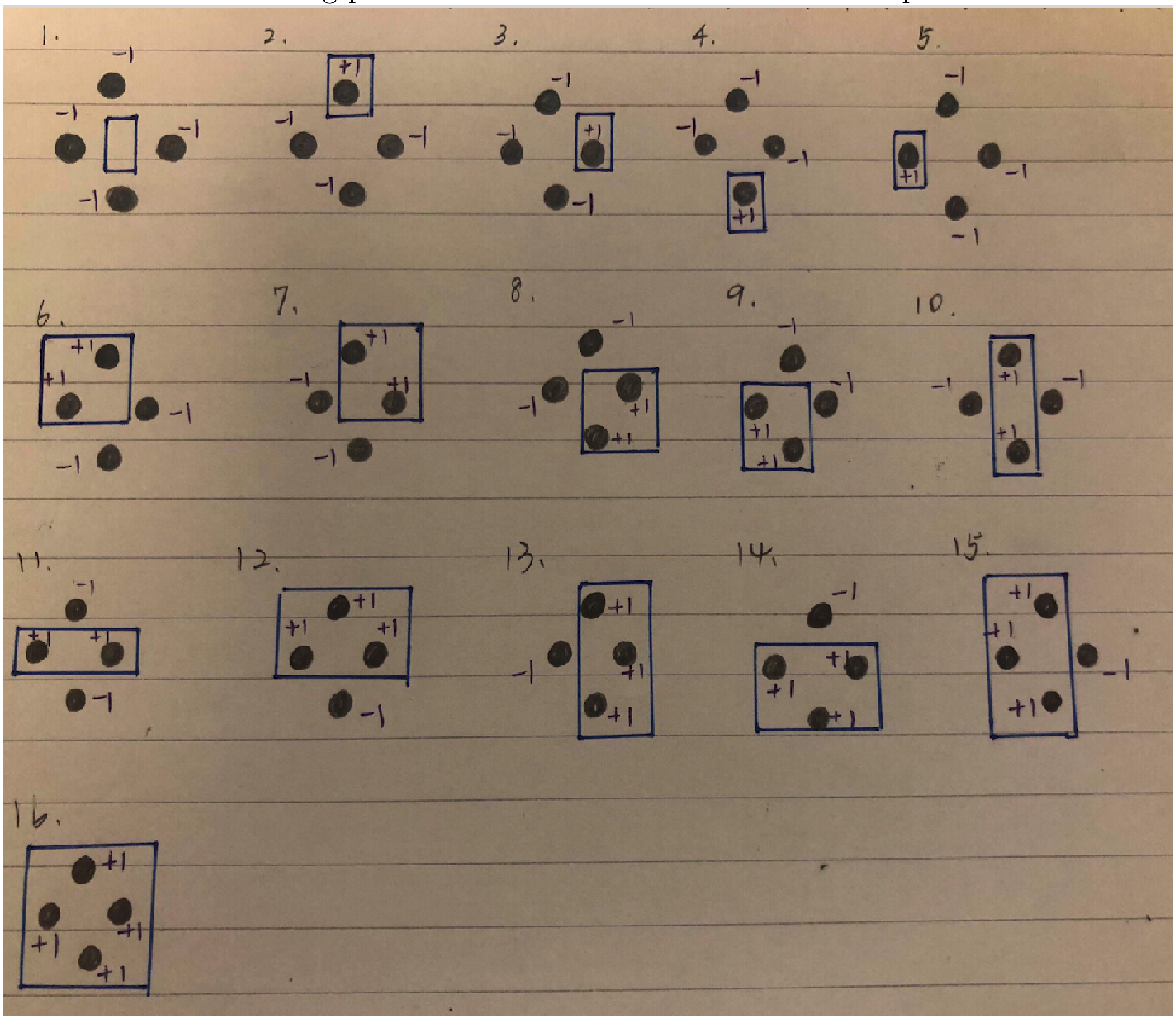
b07902031 資工二 黃永雯

## Problem 1



**Problem 2**

In order to show that the VC-Dimension of the hypothesis set is no less than 4. We can change the statement into: Prove that there exists a set of 4 points that can be shattered. The following picture shows the shatter of the set of 4 points.




The definition of VC dimension is the largest N for which $m_H(\text{N}) = 2^N$. Since there exists a set of 4 points that can be shattered, which means that $m_H(4) = 2^4$, the VC-Dimension of the hypothesis set is no less than 4.

**Problem 3**

Discussion: b07902076 許世儒

The goal is to find the VC-dimension of the hypothesis set, where $h_\alpha$ means choosing hypothesis $\alpha$ .

First, assume that $x_i$ is the i-th point and $x_i = 4^i$. Next, if we use quaternary system to represent $\alpha$, which means that we can change a decimal number $\alpha$ into a $\times$ $4^k$ + b $\times$ $4^{k-1}$ + ... + c $\times$ $4^0$ + d $\times$ $4^{-1}$ + ...

From the assumption above can know that $\alpha$ $x_i$ will be left-shift i places for $\alpha$ under quaternary system. Moreover, $\alpha$x mod 4 will be related to the coefficient of $4^0$ of $\alpha$x.

Now, there are two cases to consider, if the coefficient of term $4^0$ of $\alpha$x = 0, then $h_a$(x) = sign(| $\alpha x \quad mod \quad 4 - 2$ | - 1) = 1. On the other hand, if the coefficient of term $4^0$ of $\alpha$x = 1, then $h_a$(x) = sign(| $\alpha x \quad mod \quad 4 - 2$ | - 1) = -1. (assume that sign(0) = -1)

Therefore, assume that we have N inputs (for j from 1 to N), where N is an arbitrary number. $\alpha$ will be the sum of $4^{-j}$ $\times$ s (where s is determine by $y_j$, for $y_j$ = 1, s will be 0; for $y_j$ = -1, s will be 1). The mathematic representation will be as follows: $\alpha$ = $\sum_{j=1}^{N} 4^{-j} \times s$ (where the definition of s is above) By doing so, we can show which term exists to represent $\alpha$.

From the above can know that the triangle waves hypothesis set can shatter for any arbitrary N on $\mathbb{R}$. Therefore, the VC-dimension is $\infty$.

## Problem 4

Assume for contradiction that $d_{vc}(H_1 \cap H2) > d_{vc}(H_1)$. Using vc1 to represent $d_{vc}(H_1 \cap H2$ and vc2 to represent $d_{vc}(H_1)$. Then we can get vc1 $>$ vc2, and can know that with the hypothesis set $H_1$, there exists a set of vc2 points that can be shattered, and all of the set of vc1 points can not be shattered. However, since $H_1 \cap H2 \subset H_1$, which means that vc1 can be shattered by $H_1$. Contradicted our assumption. Therefore, from contradiction can know that the assumption $d_{vc}(H_1 \cap H2) > d_{vc}(H_1)$ is not true, and get the conclusion that $d_{vc}(H_1 \cap H2) \leq d_{vc}(H_1)$.

**Problem 5**

From mathematic definition can know that $m_{H_1 \cup H_2}$(N) $=$ $m_{H_1}$(N) $+$ $m_{H_2}$(N) - $m_{H1 \cap H_2}$(N). In class can know that $m_{H_1}$(N) = N + 1. In the same way can know that $m_{H_2}$(N) = N + 1. Moreover, $m_{H1 \cap H_2}$(N) is 2 since it contains all "x" and all "o". From the above can know that $m_{H_1 \cup H_2}$(N) = $m_{H_1}$(N) + $m_{H_2}$(N) - $m_{H1 \cap H_2}$(N) = 2 $\times$ (N + 1) - 2 = 2N. Now need to calculate the VC-Dimension. For N = 1, 2N = 2 = $2^1$. For N = 2, 2N = 4 = $2^2$. For N = 3, 2N = 6 $\neq$ $2^3$. The definition of VC-Dimension is the maximal non-break point; therefore, the VC-Dimension $d_{vc}(H_1 \cup H_2)$ is 2 in this case.

**Problem 6**

It can be discussed in four cases:

(1) s = 1, $\theta > 0$: $\mu = \frac{\theta}{2}$

(2) s = -1, $\theta > 0$: $\mu = 1$ - $\frac{\theta}{2}$

(3) s = 1, $\theta < 0$: $\mu = \frac{|\theta|}{2}$

(4) s = -1, $\theta < 0$, $\mu = 1$ - $\frac{|\theta|}{2}$

From the problem 1 of Coursera can know that the error rate is the error rate on the hypothesis + the error rate on noise. Moreover, since the error on noise means that the origin correct data on the hypothesis will be wrong and vice versa, the probability of error will be $\lambda\mu + (1-\lambda)(1-\mu)$. In this case, $\lambda = 0.8$.

Consider the four cases discussed above:

If s = 1, $E_{out}(h_{s,\theta}) = 0.8 \times \frac{|\theta|}{2} + 0.2 \times (1 - \frac{|\theta|}{2}) = 0.2 + 0.3 \times |\, \theta \,|$.

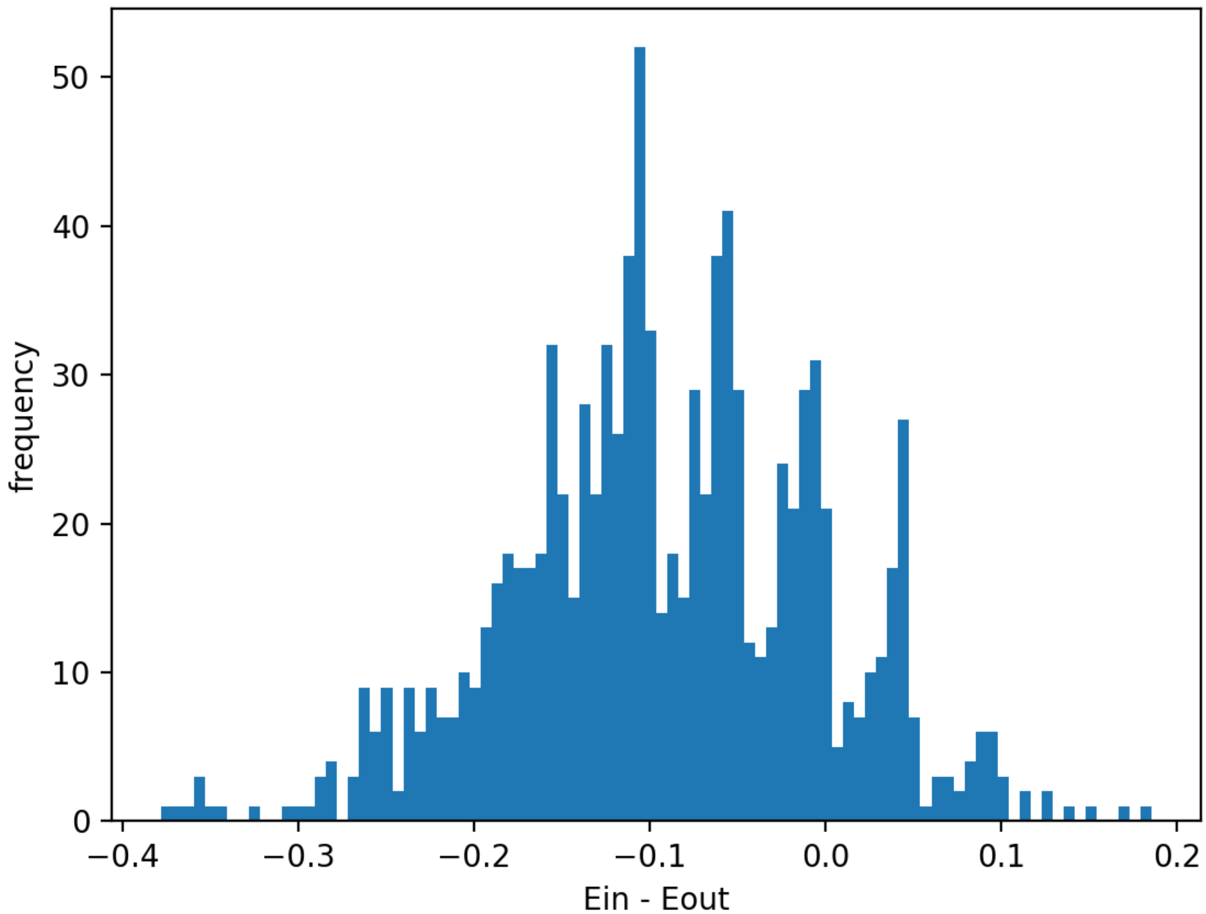
Is s = -1, $E_{out}(h_{s,\theta}) = 0.8 \times (1 - \frac{|\theta|}{2}) + 0.2 \times \frac{|\theta|}{2} = 0.8 - 0.3 \times |\, \theta \,|$.

Considering the sign of s, the above two can be combined into 0.5 + 0.3s(| $\theta$ | - 1).

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

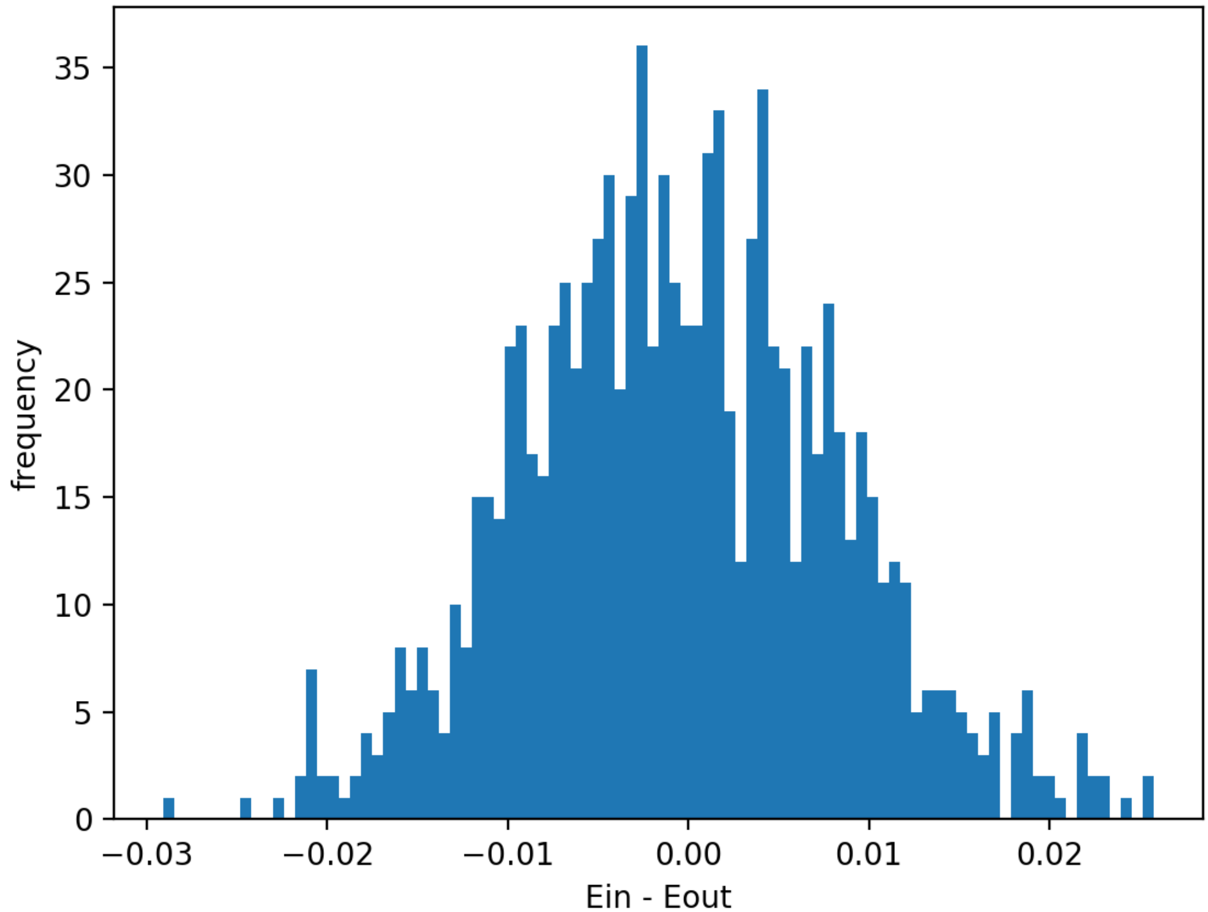
**Problem 7**

The following picture is the histogram of $E_{in}$ - $E_{out}$.




I find that the average of $E_{in}$ approximates from 0.16 to 0.17, and the average of $E_{out}$ approximates from 0.25 to 0.26. Moreover, $E_{in}$ will have more times smaller than $E_{out}$, and the average of $E_{in}$ - $E_{out}$ approximates -0.08. Since it will use 2N $= 2 \times 20 = 40$ dichotomies to find the minimum $E_{in}$ among them, we will find an $E_{in}$ that is small enough by one of the set of s and $\theta$. Then we can use this s and $\theta$ to find $E_{out}$. Since the calculation of $E_{out}$ is realted to the minimum $E_{in}$, $E_{out}$ will not be too large.

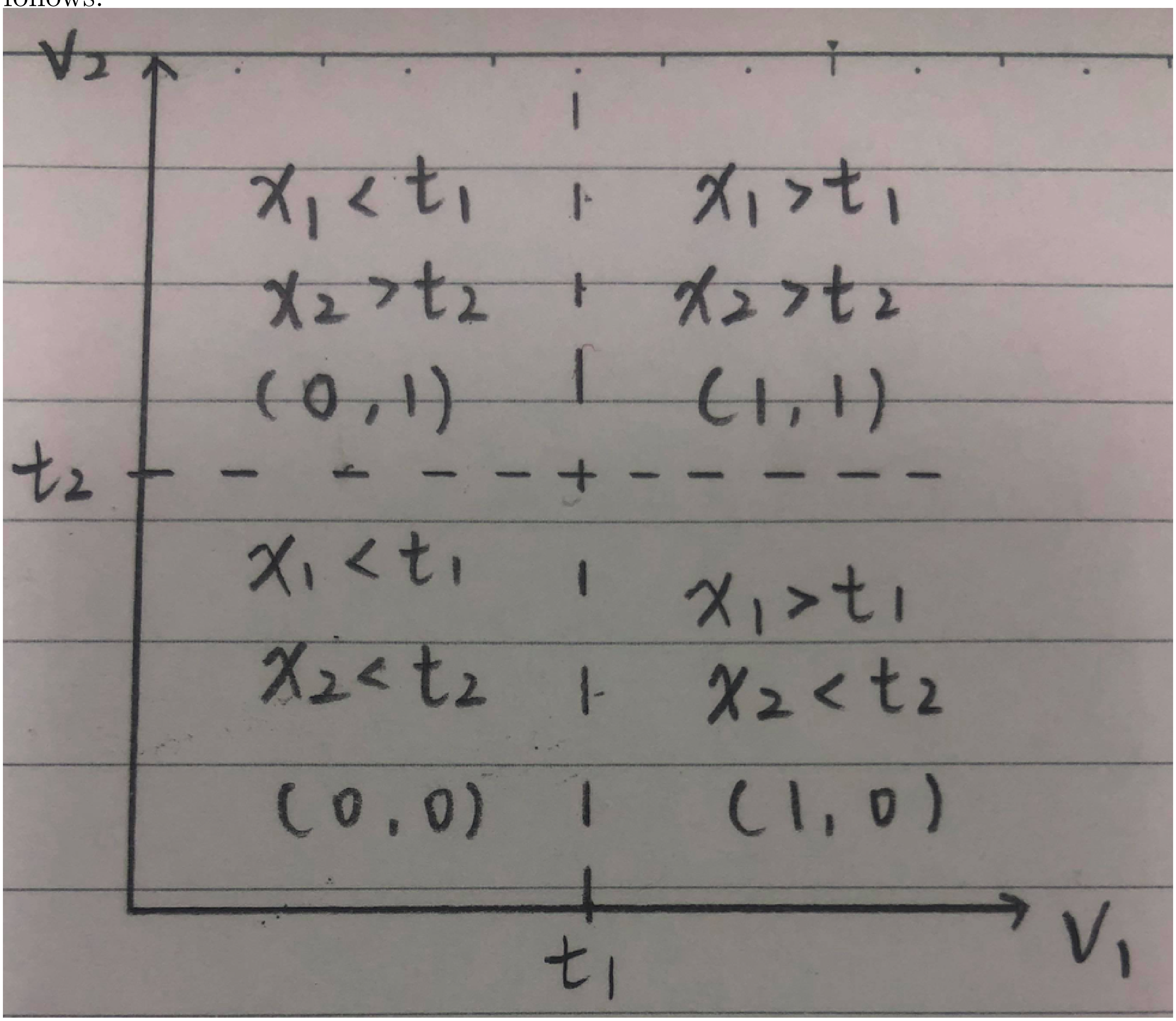
**Problem 8**




I find that the average of $E_{in}$ approximates 0.19, and the average of $E_{out}$ approximates 0.20. Moreover, $E_{in}$ will have more times smaller than $E_{out}$, and the average of $E_{in}$ - $E_{out}$ approximates -0.001. In comparison with problem 7 can find that $E_{in}$ - $E_{out}$ in this problem is much more smaller than that in problem 7. Since the size of data set in this problem is much more larger than that in problem 7, there are more dichotomies. This is what we expected since we want to make sure that the error rate for data we test ($E_{out}$) is close to that of the data we train ($E_{in}$) by larger size of data set. Moreover, it can be found that $E_{in}$ is close to the noise (0.2 in this case). It can be described as follows: Since we will choose the smallest $E_{in}$ where the $\theta$ will be close to zero, and the original data should be approximately correct. Therefore, the error rate $E_{in}$ is almost caused by the error of noise.

**Problem 9**

For d = 2, it can be stated that every vector x in $\mathbb{R}^2$ can be changed into the $2^2$ hyper-rectangular regions. For example, if $x_1 = 3$, $x_2 = 7$, $t_1 = 5$, $t_2 = 4$, then since $x_1 < t_1$ and $x_2 > t_2$, it can be changed into (0, 1) of the hyper-rectangular regions. Therefore, if d = 2, $\mathbb{R}^2$ can be represented by $2^2 = 4$ hyper-rectangular regions. The picture is as follows:




These four hyper-rectangular regions can shatter 4 points since (sign1, sign2, sign3, sign4) can be (o, o, o, o), (x, o, o, o) and so on. Every position have "o" or "x" two possibilities, so there are $2^{2^2}$ kinds of possibilities and can shatter 4 points.

By the same token, for higher dimension, since every vector x in $\mathbb{R}^d$ can be changed into the $2^d$ hyper-rectangular regions. And from the above can know that if changing into hyper-rectangular regions, it can be proved that it can shatter $2^d$ points and there are $2^{2^d}$ possibilities. Therefore, the VC-dimension is at least $2^d$.

Moreover, since there are at most $2^d$ kinds of $\mathbf{V}$, if there are $(2^d + 1)$ points, there will definitely at least a pair of two points that is in the same region of the hyper-rectangular region (there is at least two points choosing the same vector), which means that any set of $(2^d + 1)$ points can not be shattered.

From the above can know that the VC-dimension is $2^d$.