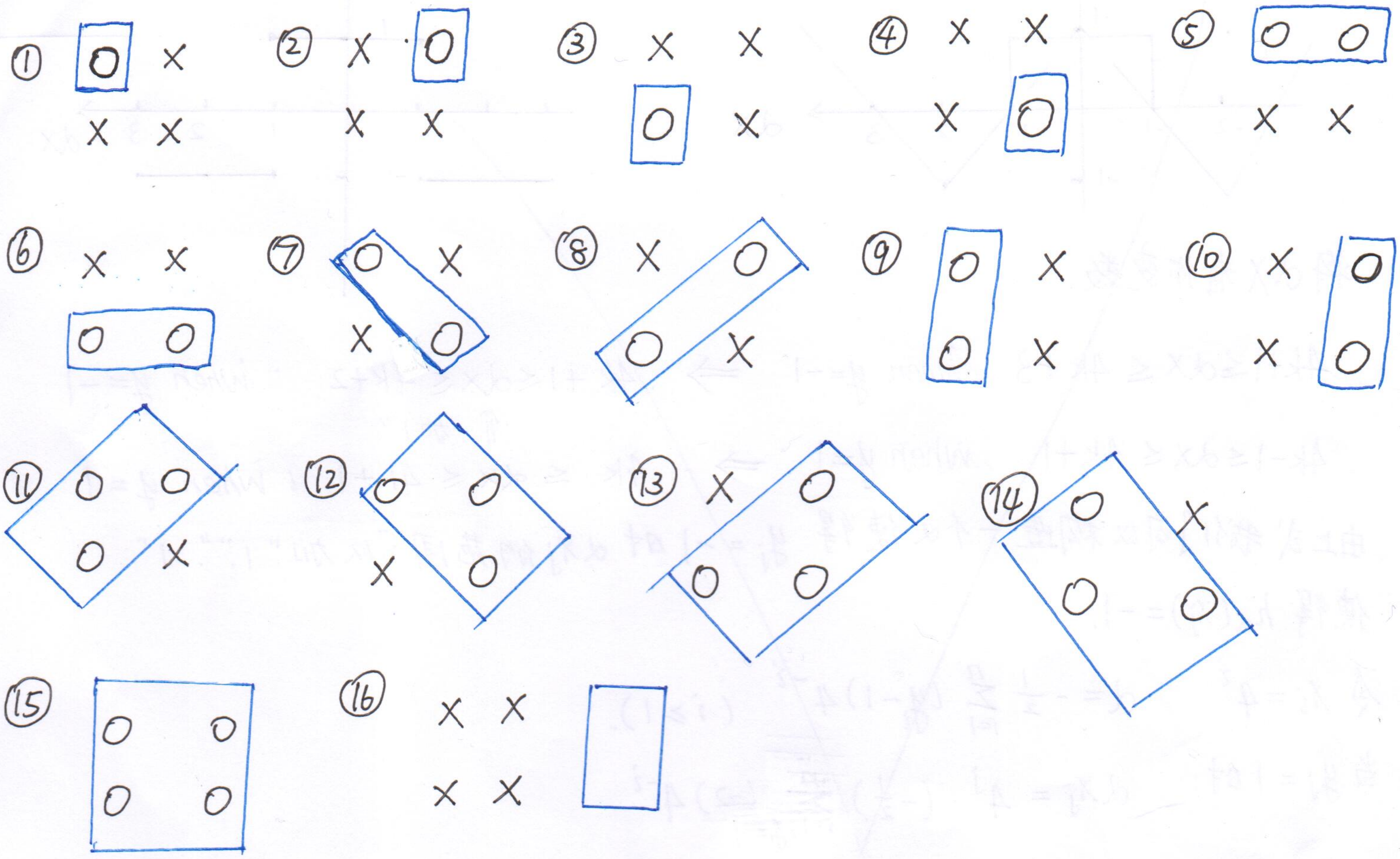
1.


coursera
探索
您想学习什么?
斯曉宇
機器學習基石上 (Machine Learning Foundations)---Mat
第 8 周
作業二
上一个
主页
Noise and Error
视频: Noise and Probabilistic Target
17 min
视频: Error Measure
15 min
视频: Algorithmic Error Measure
13 min
视频: Weighted Classification
16 min
测验: 作業二
20 个问题
测验 • 40 MIN
作業二
向您的目标更进一步
如果您完成此作业，则完成本课程的可能性增加了 54%
提交您的作业
截止时间 Dec 30, 3:59 PM CST
答题次数 3/8 hours
再试
收到成绩
通过条件 75% 或更高
成绩
100%
查看反馈
我们会保留您的最高分数

R08944052 斯晓宇　　ML Foundations HW2.

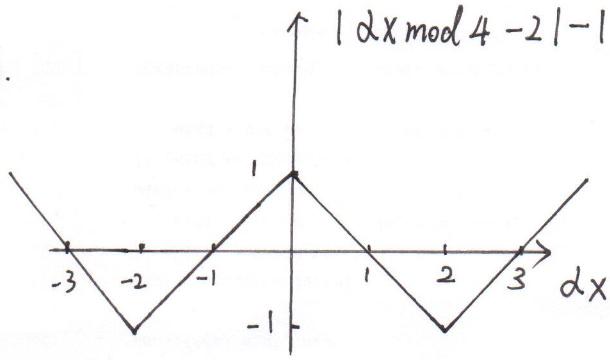
2. Let "O" = +1, "X" = −1

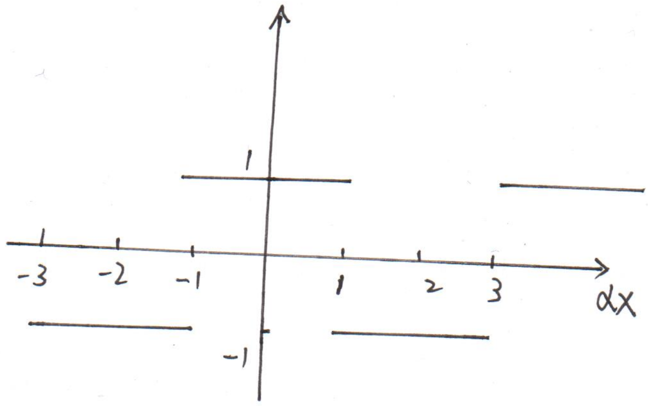


∵ "positive rectangle" hypothesis set 可以 shatter 4 个 inputs ($2^4 = 16$)

∴ VC-Dimension of the hypothesis set ≥ 4

3.





证 $d_{VC}(H) = \infty$

将 αx 看作变数.

$4k+1 \le \alpha x \le 4k+3$ 此时 $y=-1$ $\Longrightarrow$ $4k+1 \le \alpha x \le 4k+2$ 此时 $y=-1$

$\Uparrow$ "加1"

$4k-1 \le \alpha x \le 4k+1$ 此时 $y=1$ $\Longrightarrow$ $4k \le \alpha x \le 4k+1$ 此时 $y=1$

由上式我们可以构造一个 α，当 $y_j=-1$ 时 αx_j 的范围可以通过"加1"使得 $h_\alpha(x_j)=-1$

令 $x_i = 4^i$ $\quad \alpha = \sum\limits_{i=1, y_i=-1}^{n} 4^{-i}$ $\quad (i \ge 1)$.

当 $y_j = 1$ 时： $\alpha x_j = 4^j \cdot \sum\limits_{i=1, y_i=-1}^{n} 4^{-i}$

$$= \sum_{y_i=-1,\, i<j}^{n} 4^{j-i} + 0 + \sum_{y_i=-1,\, i>j}^{n} 4^{j-i}$$

$$= 4k + 0 + \sum_{y_i=-1,\, i>j}^{n} 4^{j-i}$$

$\because 0 \le \sum\limits_{y_i=-1,\, i>j}^{n} 4^{j-i} < \sum\limits_{i=1}^{+\infty} 4^{-i} < \dfrac{1}{1-4^{-1}} - 1 = \dfrac{1}{3}$

$\therefore 4k \le \alpha x_j < 4k + \frac{1}{3}$

$\because$ $y_i = 1$ when $4k \le \alpha x_i \le 4k+1$

$\therefore$ $h_\alpha(x_j) = 1$ when $y_j = 1$.

当 $y_j=-1$ 将在下一页讨论.

接续 problem 3

当 $y_j=-1$ 时:

$$\alpha x_j = 4^j \cdot \sum_{i=1,\, y_i=-1}^{n} 4^{-i}$$

$$= \sum_{y_i=-1,\, i<j}^{n} 4^{j-i} + 1 + \sum_{y_i=-1,\, i>j}^{n} 4^{j-i}$$

$$= 4k + 1 + \sum_{y_i=-1,\, i>j}^{n} 4^{j-i}$$

$\therefore \quad 4k+1 \le \alpha x_j < 4k + \frac{4}{3}$

$\because \quad y_i = -1$ when $4k+1 \le \alpha x_i \le 4k+2$

$\therefore \quad h_\alpha(x_j) = -1$ when $y_j = -1$

综上所述 $y=1$ 或 $y=-1$ 时，都能找到一个 h_α shatter 所有 X

$\therefore \quad d_{vc}(H) = \infty$

4. 设$S_1=\{dichotomies\ of\ H_1\}$ ，$S_2=\{dichotomies\ of\ H_2\}$.

集合中元素数量为 $|S_1|$ ，$|S_2|$.

$\because$ $|S_1 \cap S_2| \leq |S_1|$

$m_{H_1 \cap H_2}(N) = |S_1 \cap S_2|$ ，$m_{H_1}(N) = |S_1|$

如果 $N = d_{vc}(H_1)$

$m_{H_1 \cap H_2}(N) \leq m_{H_1}(N) = 2^N$

$\therefore$ 等号(=)成立，$m_{H_1 \cap H_2}(N)$有可能shatter数量为$d_{vc}(H_1)$的inputs.

如果 $N = d_{vc}(H_1)+1$

$m_{H_1 \cap H_2}(N) \leq m_{H_1}(N) < 2^N$

$\therefore$ $m_{H_1 \cap H_2}(N)$不能shatter数量为$d_{vc}(H_1)+1$的inputs，小于号(<)成立.

$\therefore$ $d_{vc}(H_1 \cap H_2) \leq d_{vc}(H_1)$.

法2. 反证法.

假设 $d_{vc}(H_1 \cap H_2) > d_{vc}(H_1)$

$d_{vc}(H_1 \cap H_2)$代表$H_1 \cap H_2$能shatter最多 $|H_1 \cap H_2|$个 inputs. ,$d_{vc}(H_1)$代表H_2能shatter最多 $|H_1|$个inputs.

而根据集合关系 $H_1 \cap H_2 \subseteq H_1 \Rightarrow |H_1 \cap H_2| \leq |H_1|$

$\therefore$ $d_{vc}(H_1 \cap H_2)$无法shatter 多于 $|H_1|$个inputs.

$\therefore$ $d_{vc}(H_1 \cap H_2) > d_{vc}(H_1)$ 不成立

$\therefore$ $d_{vc}(H_1 \cap H_2) \leq d_{vc}(H_1)$.

5. $m_{H_1}(N) = m_{H_2}(N) = N-1$

∵当 $H_1 \cup H_2$ 在 $N-1$ 个区间内时

会产生 2 种不同的 dichotomies

另外在 2 端会产生全为"O"或全为"X"

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

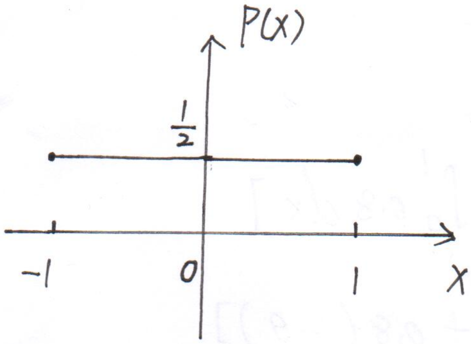
当 $N=1$ 时 $m_{H_1 \cup H_2} = 2N = 2 = 2^N$ 可以 shatter 1个 input

当 $N=2$ 时 $m_{H_1 \cup H_2} = 2N = 4 = 2^N$ 可以 shatter 2个 inputs

当 $N=3$ 时 $m_{H_1 \cup H_2} = 2N = 6 < 2^N = 8$ 不可以 shatter 3个 inputs

∴ $d_{vc}(H_1 \cup H_2) = 2$.

6.



$$P(y|x)=\begin{cases}0.8, & y=f(x)\\ 0.2, & y\neq f(x)\end{cases}$$

$f(x)=\tilde{s}(x)+noise$ where $\tilde{s}(x)=sign(x)$

$-1\le x\le 0\begin{cases}0.8 & y=-1\\ 0.2 & y=1\end{cases}$ $\quad 0<x\le 1\begin{cases}0.8 & y=1\\ 0.2 & y=-1\end{cases}$

$$E_{out}(h_{s,\theta})=\int_{-1}^{1}P(x)\cdot E_{y\sim p(y|x)}[h_{s,\theta}(x)\neq y]\,dx$$

$$=\frac{1}{2}\Big[\int_{-1}^{0}E_{y\sim p(y|x)}[h_{s,\theta}(x)\neq y]dx+\int_{0}^{1}E_{y\sim p(y|x)}[h_{s,\theta}(x)\neq y]dx\Big]$$

$$=\frac{1}{2}\Big[\int_{-1}^{0}0.8[h_{s,\theta}(x)\neq -1]+0.2[h_{s,\theta}(x)\neq 1]\,dx+$$

$$\int_{0}^{1}0.8[h_{s,\theta}(x)\neq 1]+0.2[h_{s,\theta}(x)\neq -1]\,dx\Big]$$

假设 $s=+1$ $\quad h_{s,\theta}(x)=sign(x-\theta)$

$\theta<0$: $\quad E_{out}(h_{s,\theta})=\frac{1}{2}\Big[\int_{-1}^{\theta}0.2dx+\int_{\theta}^{0}0.8dx+\int_{0}^{1}0.2dx\Big]$

$$=\frac{1}{2}[0.2(\theta-(-1))+0.8(0-\theta)+0.2(1-0)]$$

$$=\frac{1}{2}[0.2\theta+0.2-0.8\theta+0.2]$$

$$=-0.3\theta+0.2$$

$\theta\ge 0$: $\quad E_{out}(h_{s,\theta})=\frac{1}{2}\Big[\int_{-1}^{0}0.2dx+\int_{0}^{\theta}0.8dx+\int_{\theta}^{1}0.2dx\Big]$

$$=\frac{1}{2}[0.2(0-(-1))+0.8(\theta-0)+0.2(1-\theta)]$$

$$=\frac{1}{2}[0.2+0.8\theta+0.2-0.2\theta]$$

$$=0.3\theta+0.2$$

假设 $s=-1$ $\quad h_{s,\theta}(x)=-sign(x-\theta)$

$\theta<0$: $\quad E_{out}(h_{s,\theta})=\frac{1}{2}\Big[\int_{-1}^{\theta}0.8dx+\int_{\theta}^{0}0.2dx+\int_{0}^{1}0.8dx$

$$=\frac{1}{2}[0.8(\theta-(-1))+0.2(0-\theta)+0.8(1-0)]$$

$$=\frac{1}{2}[0.8\theta+0.8-0.2\theta+0.8]$$

$$=0.3\theta+0.8.$$

下一页

接续

$S=-1 \quad h_{s,\theta}(x) = -\text{sign}(x-\theta)$

$$\theta \geq 0 \quad E_{out}(h_{s,\theta}) = \frac{1}{2}\left[\int_{-1}^{0} 0.8\,dx + \int_{0}^{\theta} 0.2\,dx + \int_{\theta}^{1} 0.8\,dx\right]$$

$$= \frac{1}{2}\left[0.8(0-(-1)) + 0.2(\theta-0) + 0.8(1-\theta)\right]$$

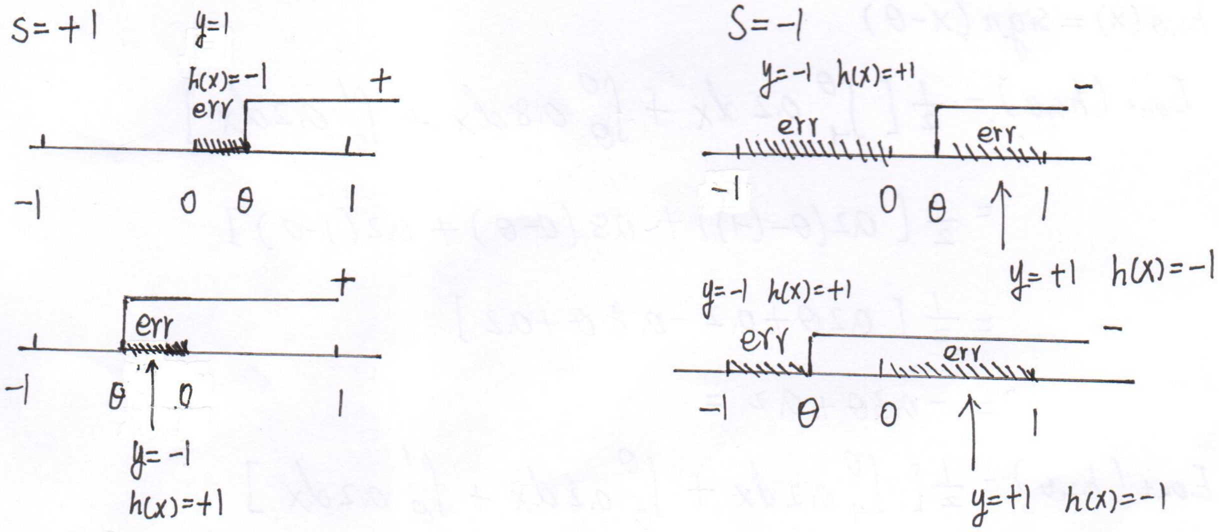
$$= \frac{1}{2}\left[0.8 + 0.2\theta + 0.8 - 0.8\theta\right]$$

$$= -0.3\theta + 0.8$$

综上可得 $E_{out}(h_{s,\theta}) = 0.5 + 0.3S(|\theta|-1)$

方法二.

由coursera第1题可知 $E_{out} = \lambda\mu + (1-\lambda)(1-\mu)$, where $\lambda = 0.8$ ，μ为没有noise的错误率



$$\mu_{S=+1} = \frac{|\theta|}{2} \qquad \mu_{S=-1} = 1 - \frac{|\theta|}{2}$$

$$\mu_{s,\theta} = 0.5(S+1) \times 0.5|\theta| - 0.5(S-1)(1-0.5|\theta|)$$

$$= 0.5S|\theta| - 0.5S + 0.5.$$

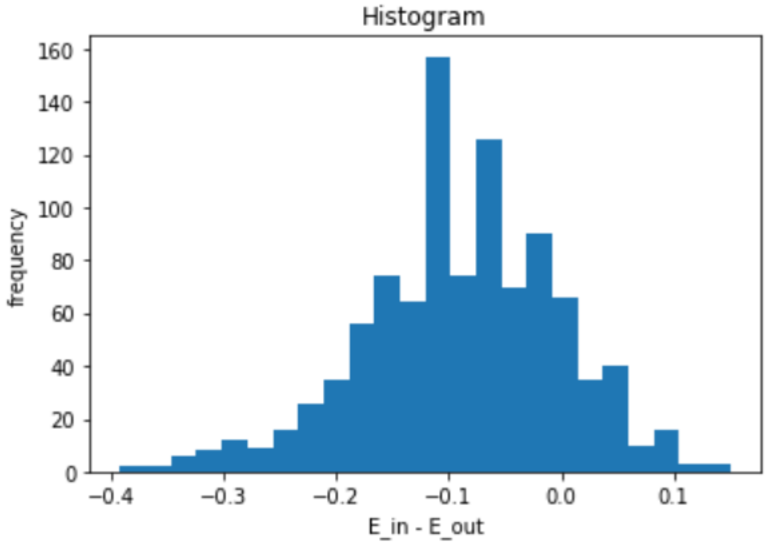
$$E_{out}(h_{s,\theta}) = \lambda\mu_{s,\theta} + (1-\lambda)(1-\mu_{s,\theta})$$

$$= 0.5 + 0.3S(|\theta|-1)$$

7.

通過觀察直方圖，E_{in}-E_{out} 近似高斯分佈，其數值基本位於區間[-0.2, 0.1]之間。當 training data 為 20 的時候，E_{in}-E_{out} 的差值多為負值，說明 E_{in} 通常小於 E_{out}。我們知道在 20% noise 的情況下，真實的 avg E_{in} 和 E_{out} 都約為 0.2，而當前結果顯示 E_{in} 最小的 hypothesis 不一定能保證 E_{out} 最小，且 E_{in} 和 E_{out} 仍不夠接近。

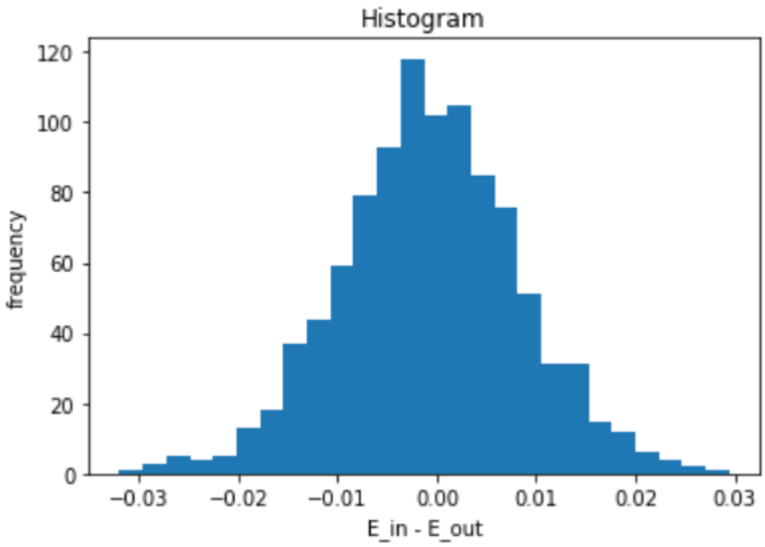
```
Average_E_in = 0.165900
Average_E_out = 0.254686
Ein-Eout= -0.08878643295061384
```



8.

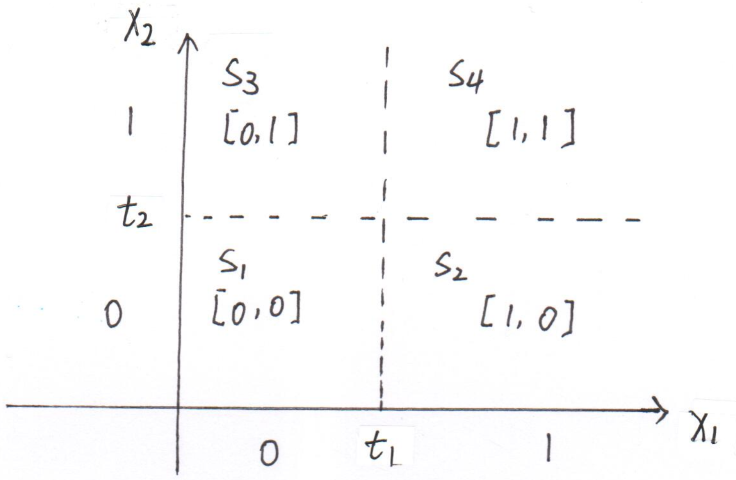
在 training data 擴增到 2000 的情況下，我們可以很明顯觀察到 E_{in}-E_{out} 呈現高斯分佈，其差值基本位於[-0.01, 0.01]之間。在訓練資料增加的情況下，VC bound 限制了 E_{in} 與 E_{out} 的差距，同時 avg E_{in} 與 E_{out} 在此情況下更接近真實值。滿足了機器可以學習的條件，即 $E_{in}\approx 0$，且 $E_{in}\approx E_{out}$。

```
Average_E_in = 0.199878
Average_E_out = 0.200634
Ein-Eout= -0.0007558015278006092
```



9.

假设在二维空间中 $d=2$, $S \subseteq \{[0,0]^T, [0,1]^T, [1,0]^T, [1,1]^T\}$



$t=[t_1, t_2]^T$ 把二维平面划分为4个区域

对任意点 $x\,[x_1, x_2]^T$ 根据 x_1, x_2 与 t_1, t_2 的大小关系都可以

被划分到4个区域之一

∴ 对于 $d=2$ (二维平面), H 可以shatter 最多 4 (2^d, where $d=2$) 个点,

或许可以借此推论, H on R^d 可以shatter 最多 2^d 个点,