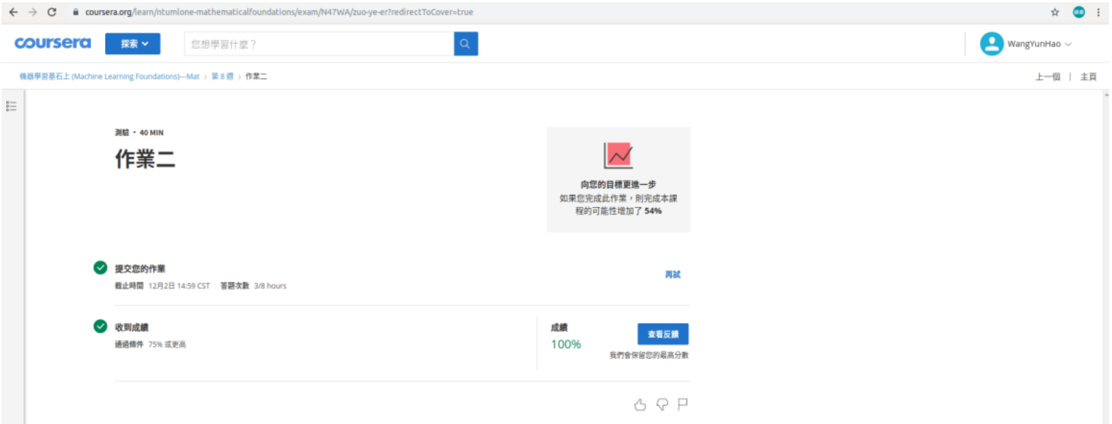
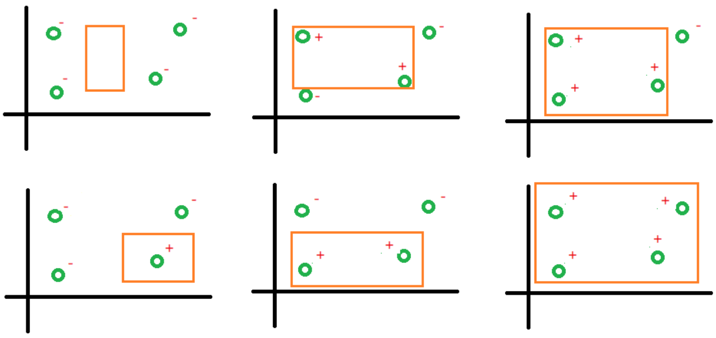
1.



2. An example 4-point set is shown below with all typical labels and the corresponding realization.
   There exists a 4-point set shattered by the concept set, so we have VC-dim≥ 4.



3. $H = \{h_\alpha \mid h_\alpha(x) = \mathrm{sign}(|\alpha x) \bmod 4 - 2| - 1), \ \alpha \in R\}$
   We can easily find that the output range of the hypothesis H set is bounded to $\{-1,1\}$
   Let $Y \in \{+1,-1\}^N$ be the output set and $X \in R$ be the input set
   For each $y \in Y$, we can always find a corresponding input set and an adjusting the α of $h \in H$ that $h_\alpha(x) = y$
   $\because$ $\{+1,-1\}^N$ combinations can be constructed
   $\therefore$ H can shatter any N inputs. That means the VC-dimension of H is ∞.

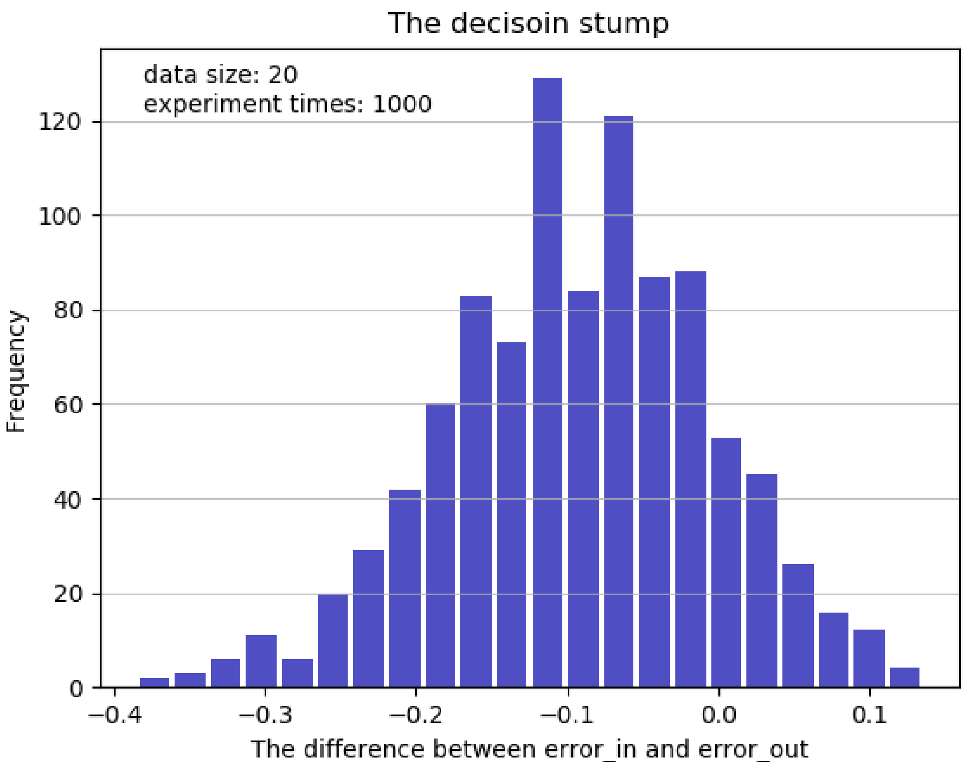
4. Prove that $d_{vc}(H1 \cap H2) \leqq d_{vc}(H1)$ where H1 and H2 that come with non-empty intersection
   First, $\forall n < d_{vc}(H1 \cap H2)$ that means any n inputs can be shattered by $H1 \cap H2$
   $\because H1 \cap H2 \subseteq H1$ and $H1 \cap H2 \notin \{\varnothing\}$
   $\therefore$ Any n inputs can be shattered by H1. That means $d_{vc}(H1 \cap H2) \leqq d_{vc}(H1)$

5. H1 as the positive-ray hypothesis set
   H2 as the negative-ray hypothesis set
   $m_{H1}(N) = N + 1 = m_{H2}(N)$，$m_{H1 \cup H2}(N) = 2(N+1)-2 = 2N$
   When N=3 , $m_{H1 \cup H2}(N) = 2N \neq 2^N$. So the $d_{vc}(H1 \cup H2)$ is 2.

6. $h_{s,\theta}(x) = s \cdot \mathrm{sign}(x-\theta)$，with $\theta \in [-1,1]$
   Given a target function f that $P(x|y) = 0.8$ where $f(x)=y$ and $P(x|y) =0.2$ where $f(x) \neq y$
   Assume that h error rate is x
   $\because$ When s=1, $x = \mathrm{abs}(\theta/2)$ otherwise $x = 1-\mathrm{abs}(\theta/2)$
   $\therefore$ s=1， $E_{out}=0.8* \mathrm{abs}(\theta/2)+0.2(1- \mathrm{abs}(\theta/2)) = 0.2 + 0.3\,\mathrm{abs}(\theta)$
   s=-1， $E_{out}=0.2* \mathrm{abs}(\theta/2)+0.8(1- \mathrm{abs}(\theta/2)) = 0.8-0.3\,\mathrm{abs}(\theta)$
   $E_{out}(h_{s,\theta}) = 0.5+0.3s(\mathrm{abs}(\theta)-1)$

7.

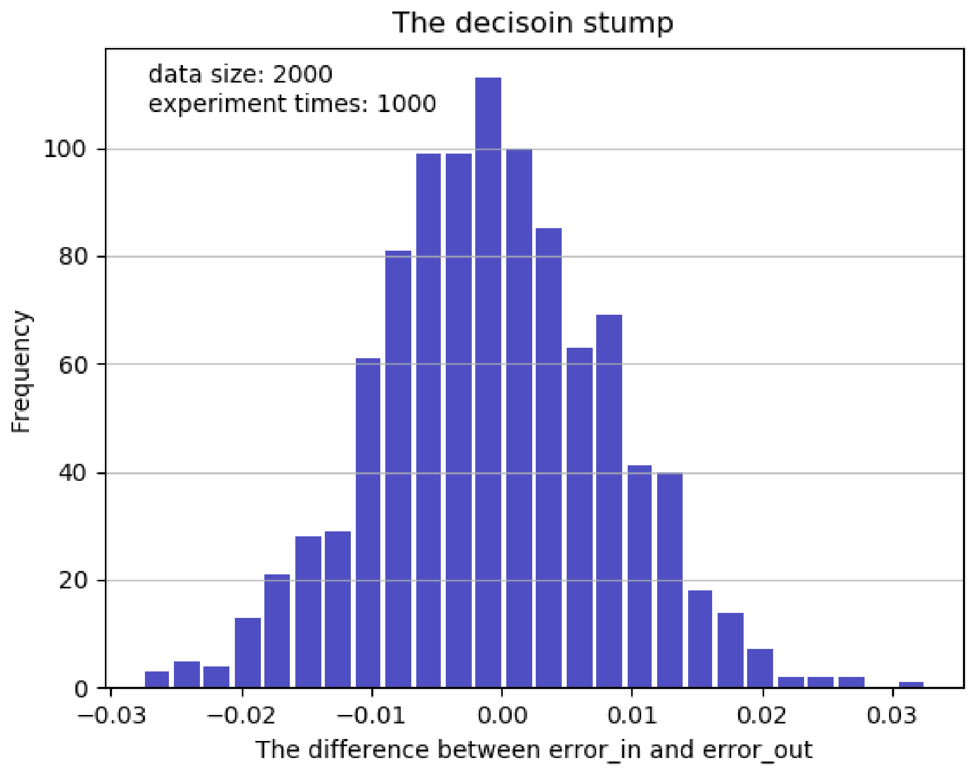
```
PS C:\Users\Wangyunhao\Desktop\MachineLearning\hw2> python .\Q7_8.py 20 1000
 average error_in:  0.17075000000000001
 average error_out:  0.26418036467368383
```



I find that the smaller data size and the larger difference between average error in and average error out. The data size affects the hoeffding inequality and that means the bad sample happened with higher probability.

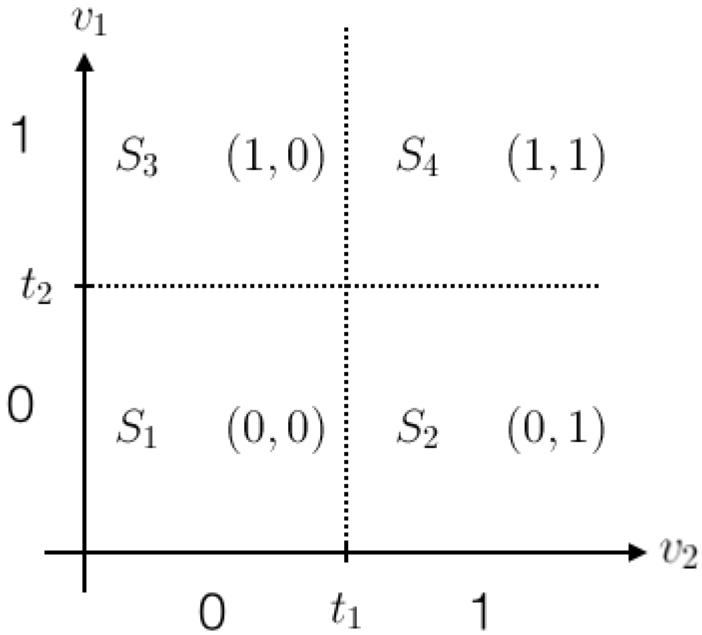
8.

```
PS C:\Users\Wangyunhao\Desktop\MachineLearning\hw2> python .\Q7_8.py 2000 1000
 average error_in:  0.19997299999999932
 average error_out:  0.20070398444370932
```



I find that the difference between average error in and error out gets closer with the increasing of the data size. That proves the correctness of the hoeffding inequality.

9. H = {ht,s | ht,s(x) = 2[[v ∈ s]] -1, where vi = [xi>ti], S a collection of vectors in {0,1}^d, t ∈ R^d}
   What is the VC-dimension of the "simplified decision trees" hypothesis set?



我們將二維度上的 threshold 值 ti 與各向量轉換到二維平面上，上面二维的例子中，simplified decision trees 的 Dvc 與 hyper-rectangular regions 的數量相等。D 維向量可以用 D 條直線最多分割出 2^D 個 hyper-rectangular regions，代表說可以 shatter 掉 2^d 個點。

∴VC-dimension of the "simplified decision trees" is 2^d.