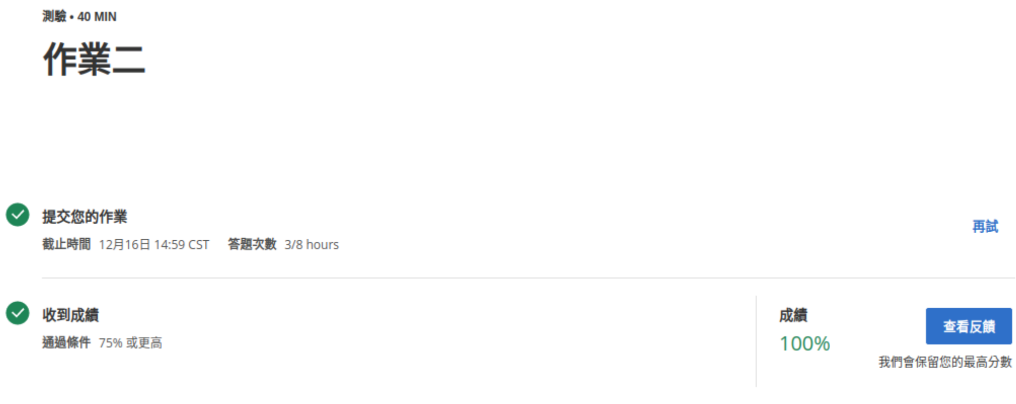
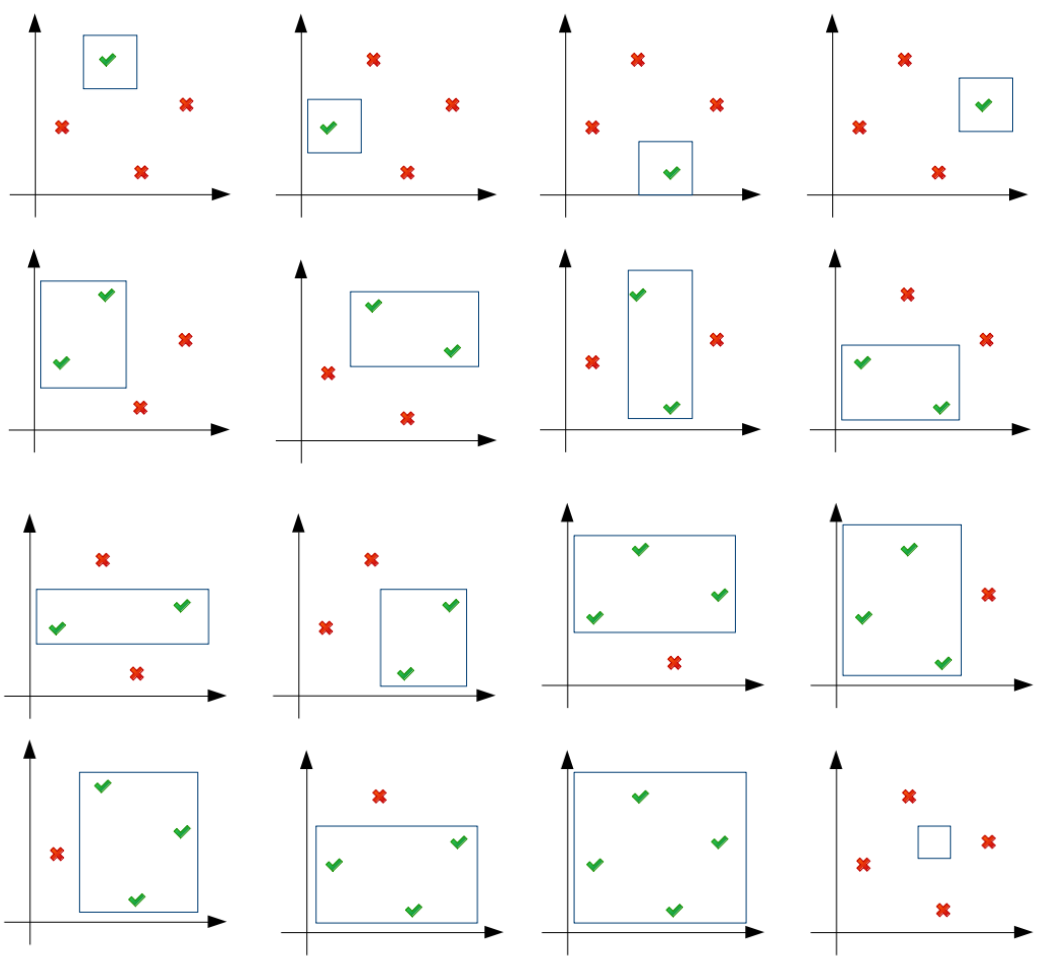
1



2

we need to proof that $d_v c \geq 4$, which means we must find a set of 4 inputs that can be shattered.

Consider 4 points are in the first quadrant in 2D space, which arranged in four corners of a rotated square.



3

Step1

First, we analyze the formula $h_\alpha(x) = sign(|\alpha x \bmod 4 - 2| - 1)$ in the question.

For simplicity, we define $sign(0) = -1$. And we can get that: $h_\alpha(x) = \begin{cases} -1, & \text{when } 1 \le \alpha x \bmod 4 \le 3 \\ 1, & \text{otherwise} \end{cases}$

Step2

Now we need to come up with the configurations of α and x. Suppose there are N points, $x_i = 4^i$ $(1 \le i \le N)$. So we need to construct 2^N kinds of α to satisfiy all $\{+1, -1\}^N$ combinations.

We construct one of the α as follows:

$$\alpha_k \text{ is the kth alpha, } 1 \le k \le 2^N$$
$$\alpha_k = C_1 4^{-1} + C_2 4^{-2} + \cdots + C_N 4^{-N}$$
$$C_i = \{0, 1\}, 1 \le i \le N$$

Let $T = (x_i \alpha_k) \bmod 4$.

$$\begin{aligned} T &= 4^i (C_1 4^{-1} + C_2 4^{-2} + \cdots + C_N 4^{-N}) \bmod 4 \\ &= (C_1 4^{i-1} + C_2 4^{i-2} + \cdots + C_N 4^{i-N}) \bmod 4 \\ &= C_i + C_{i+1} 4^{-1} + \cdots + C_N 4^{i-N} \end{aligned}$$

then we can reduce the formula about $h_\alpha(x)$ as follows,

$$\begin{aligned} h_{\alpha_k}(x_i) &= sign(|\alpha_k x_i \bmod 4 - 2| - 1) \\ &= sign(|T - 2| - 1) \end{aligned}$$

and we will get $h_{\alpha_k}(x_i) = \begin{cases} -1, & \text{when } 1 \le T \le 3 \\ 1, & \text{otherwise} \end{cases}$

By the properties of the sum of geometric sequence, we know $\begin{cases} \text{if } C_i = 1 \text{ then } 1 \le T \le \frac{4}{3} & \Rightarrow h_{\alpha_k}(x_i) = -1 \\ \text{if } C_i = 0 \text{ then } 0 \le T \le \frac{1}{3} & \Rightarrow h_{\alpha_k}(x_i) = 1 \end{cases}$

Step3

For any finite N, we must find N inputs that we can shatter.

Let $X = \{x_i = 4^i\}$ and $Y \in \{0, 1\}^N$ for $1 \le i \le N$, then $\alpha = \Sigma_{i=1}^N C_i 4^{-i}$, $\begin{cases} C_i = 1, \text{ if } y_i = h_\alpha(x_i) = -1 \\ C_i = 0, \text{ if } y_i = h_\alpha(x_i) = 1 \end{cases}$.

Thus $d_{VC} = \infty$

4

First, we assume that $d_{vc}(H_1 \cap H_2) = n$ and $d_{vc}(H_1) = m$, which means n inputs can be shattered by $H_1 \cap H_2$ and m inputs can be shattered by H_1. Proof that $n \leq m$.

Prove by Contradiction.

Suppose that $n > m$.

Since $n > m$, by the definition of VC-Dimemsion, we know that n inputs can be shattered by $H_1 \cap H_2$ but cannot be shattered by H_1.

However, $H_1 \cap H_2 \subseteq H_1$, the inputs shattered by $H_1 \cap H_2$ must also be shattered by H_1, contradiction to the assumption.

Therefore, we have proven that $n \leq m$, $d_{vc}(H_1 \cap H_2) \leq d_{vc}(H_1)$.

5

Since the intersection of H_1 and H_2 is all positive or all negetive. And we know $m_{H_1}(N) = m_{H_2}(N) = N + 1$.

$$
\begin{aligned}
m_{H_1 \cup H_2}(N) &= m_{H_1}(N) + m_{H_2}(N) - m_{H_1 \cap H_2}(N) \\
&= 2(N+1) - 2 \\
&= 2N.
\end{aligned}
$$

when $N = 2$, $m_{H_1 \cup H_2}(N) = 2N = 4 = 2^2$.

when $N = 3$, $m_{H_1 \cup H_2}(N) = 2N = 6 < 2^3$. Thus, $d_{vc}(H_1 \cup H_2) = 2$

6

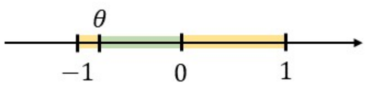
we know $h_{s,\theta}(x) = s \cdot sign(x - \theta)$ and $f(x) = sign(x) + noise$.

μ = average error rate $h(x) \neq f(x)$ = average false accept and false reject .

$$\mu = \frac{green\ part}{yellow\ part + green\ part}$$

when s=1

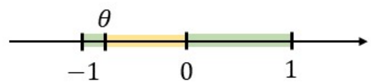
$$h_{s,\theta}(x) = sign(x - \theta)$$
$$f'(x) = sign(x)$$



$$\mu = \frac{|\theta|}{2}$$

when s=-1

$$h_{s,\theta}(x) = -sign(x - \theta)$$
$$f'(x) = sign(x)$$



$$\mu = \frac{(2 - |\theta|)}{2} = 1 - \frac{|\theta|}{2}$$

we can make a small conclusion by eliminating the noise factor (20 percent flip) first.

$$\mu = \begin{cases} \frac{|\theta|}{2}, & \text{when } s = 1 \\ 1 - \frac{|\theta|}{2}, & \text{when } s = -1 \end{cases}$$

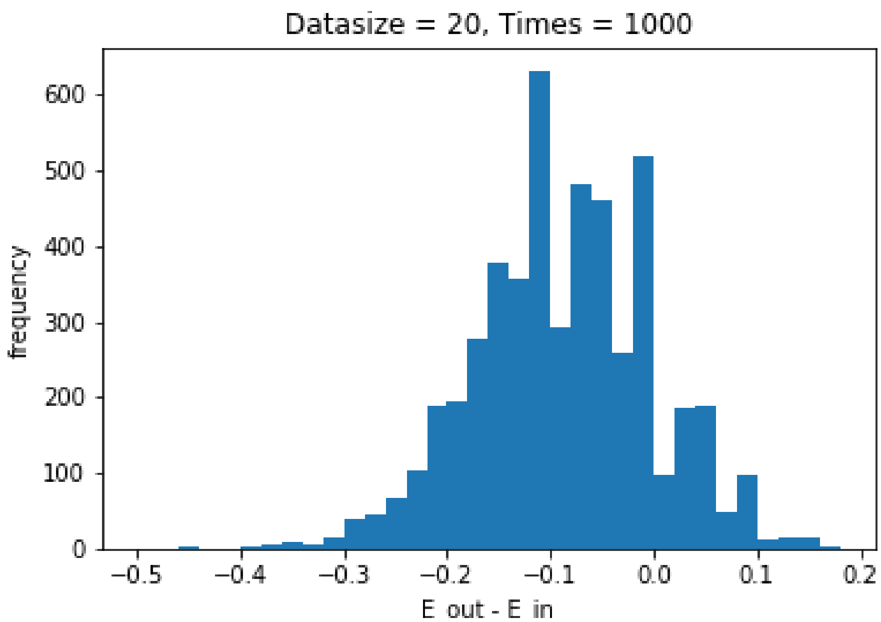
Combine the two cases above, we have $\mu = \frac{s(|\theta|-1)+1}{2}$.

By Problme 1 in coursera, we know $E_{out}(h_s; \theta) = \lambda\mu + (1 - \lambda)(1 - \mu)$. λ is rate of case with no noise.

$$\begin{aligned} E_{out}(h_s; \theta) &= P[\text{no flip}]P[h(x) \neq f(x)] + P[\text{ flip }]P[h(x) = f(x)] \\ &= \lambda\mu + (1 - \lambda)(1 - \mu) \\ &= 0.8\mu + 0.2(1 - \mu) \\ &= 0.2 + 0.6\mu \\ &= 0.2 + 0.6(\frac{s(|\theta| - 1) + 1}{2}) \\ &= 0.5 + 0.3s(|\theta| - 1) \end{aligned}$$

7

The average of $E_{in} - E_{out}$ falls around -0.089, and the range of $E_{in} - E_{out}$ is actually -0.5 to 0.2 .

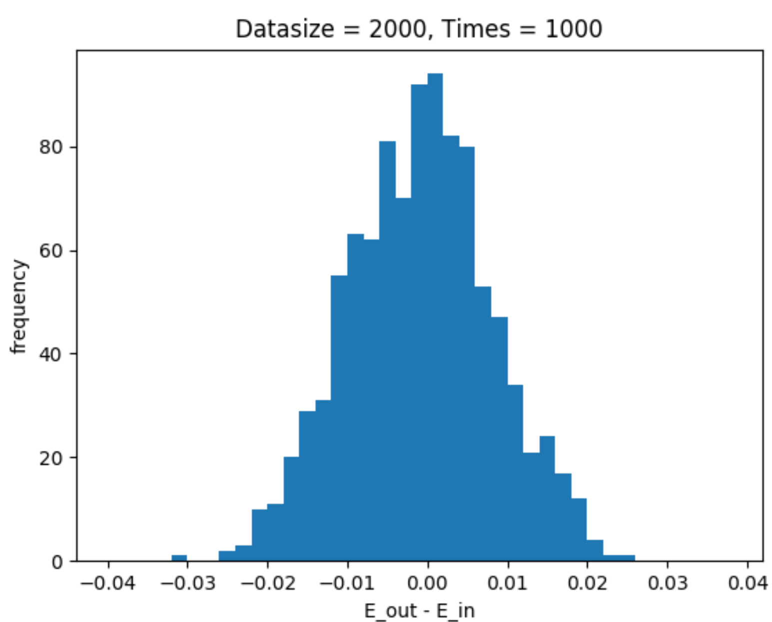


8

The average of $E_{in} - E_{out}$ falls around -0.0008, and the range of $E_{in} - E_{out}$ is actually -0.03 to 0.03 .

Compared with the previous case, $|E_{in} - E_{out}|$ becomes smaller with bigger datasize.

And the distribution of 1000 times experiment with 2000 data is more like a normal distribution whose mean is closer to 0.



9

Consider the "simplified decision trees" hypothesis set on R^d, which is given by

$$H = \{h_{t,s} | h_{t,s}(x) = 2[[v \in S]] - 1, \text{ where } v_i = [[x_i > t_i]],$$

$$\mathbf{S} \text{ a collection of vectors in } \{0,1\}^d, \mathbf{t} \in R^d\}$$

By the definition of H, we know that $v_i = \begin{cases} 1, & \text{when } x_i > t_i \\ 0, & \text{when } x_i \leq t_i \end{cases}$ Each t_i can divide the space into two part.

Therefore if there are d dimension in the space, we can divide the space into 2^d different regoins.

If there are more than 2^d points in the space, it must exist two points in the space belong to the same seperated region, we can't assign these two points to different regions.

For example when $d = 2$. There are 5 points in the space. Always more than 1 points in one of the regions. Suppose p_4 and p_5 in the same region, then the following cases can't appear at the same time.

$$(p_1, p_2, p_3, p_4, p_5)$$

$$(+, -, -, +, +)$$

$$(+, -, -, +, -)$$

Thus, we can know that the VC-dimension of the "simplified decision trees" hypothesis set is 2^d.