Machine Learning HW2

b03902089 林良翰

1.

- The error of h is making an error but predicting f correctly + making no error but predicting f wrongly.
- The result is $\mu\lambda + (1-\mu)(1-\lambda)$

2.

- $\mu\lambda + (1-\mu)(1-\lambda) = \lambda$
  $(1-\mu)\lambda = (1-\mu)(1-\lambda)$
  $\Rightarrow \lambda = \frac{1}{2}$

3.

- $\mathrm{P}[\exists h \in \mathcal{H}\ s.t.\ |\mathrm{E}_{in} - \mathrm{E}_{out}| > \varepsilon] \leq 4m_{\mathcal{H}}(2N)e^{-\frac{1}{8}\varepsilon^2 N}$
  $\varepsilon = 0.1,\ d_{\mathrm{VC}} = 10 \Rightarrow N \geq 98964$
  $\Rightarrow N \geq 99000$

4.

- a. $\varepsilon \leq 0.463$
- b. $\varepsilon \leq 0.631$
- c. $\varepsilon \leq 0.243$
- d. $\varepsilon \leq 0.164$
- e. $\varepsilon \leq 0.158 \Rightarrow$ tightest bound

5.

- a. $\varepsilon \leq 11.106$
- b. $\varepsilon \leq 13.701$
- c. $\varepsilon \leq 5.736$
- d. $\varepsilon \leq 4.033$
- e. $\varepsilon \leq 3.957 \Rightarrow$ tightest bound

6.

- $N-1$ inner intervals
  $\Rightarrow \binom{N-1}{2} = \frac{1}{2}(N^2 - 3N + 2)$
- Considering positive and negative
  $\Rightarrow 2\binom{N-1}{2} = (N^2 - 3N + 2)$
- Dividing into two parts
  $\Rightarrow 2\binom{N}{1} = 2N$
- Final result
  $\Rightarrow N^2 - 3N + 2 + 2N = N^2 - N + 2$

7.

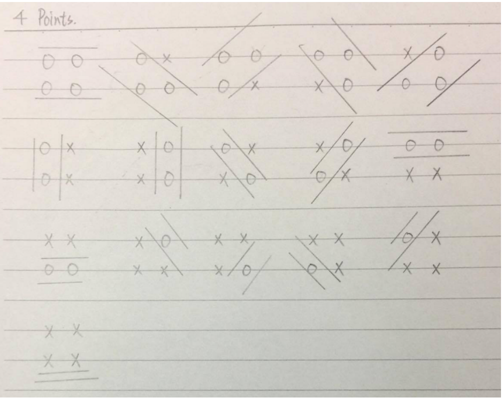
- $k = 1 \Rightarrow k^2 - k + 2 = 2 = 2^1$
- $k = 2 \Rightarrow k^2 - k + 2 = 4 = 2^2$
- $k = 3 \Rightarrow k^2 - k + 2 = 8 = 2^3$
- $k = 4 \Rightarrow k^2 - k + 2 = 14 < 2^4 \Rightarrow$ break point
- VC-dimension $d_{VC} = 3$

8.

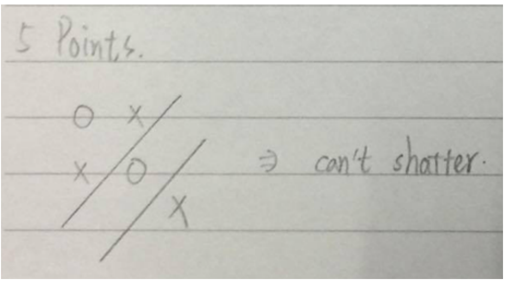
- By using polar coordinate (θ, r), we can observe that positive circles in $\mathbb{R}^2$ is associative with only r, so we can convert this porblem into positive ray in $\mathbb{R}^+$.
- $\Rightarrow m_{\mathcal{H}}(N) = N + 1$

9. picture

- $|\omega_0 + \omega_1 x_1 + \omega_2 x_2| \leq \theta$
  $-\theta \leq \omega_0 + \omega_1 x_1 + \omega_2 x_2 \leq \theta \Rightarrow$ 2D-intervals
- 4 points $\Rightarrow$ can shatter

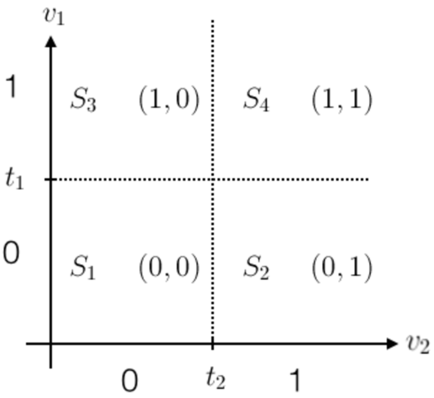


- 5 points $\Rightarrow$ can not shatter



10.

- Consider $d = 2$, we could construct a 2D-graph below:



- Taking $t = [5, 10], x = [6, 8]$ as example, the point x can be converted into region $S_3(1, 0)$ $\because x_1 > t_1, x_2 < t_2$.
- S_1~S_4 is the four regions dividing $\mathbb{R}^2$, we can find that shattering four points *on*\rm I!R$ is available. If there are more than 5 points, we can't shatter them.
- As a result, we can conclude that we can shatter 2^d points on $\mathbb{R}^{\mathrm{d}}$, where d is the depth of the tree.

11.

- First find the range of $h_\alpha(x) = \text{sign}(|(\alpha x) \text{ mod } 4 - 2| - 1), \alpha \in \mathbb{R}$
- $(\alpha x) \text{ mod } 4 \in [0, 4)$
  $(\alpha x) \text{ mod } 4 - 2 \in [-2, 2)$
  $|(\alpha x) \text{ mod } 4 - 2| \in [0, 2]$
  $|(\alpha x) \text{ mod } 4 - 2| - 1 \in [-1, 1]$
  $\Rightarrow h_\alpha(x) \in \{-1, 1\}$
- If there are N inputs $(x_1 \dots x_N)$ where $i \in \mathbb{N}$ and$1 \leq i \leq N$, we can find an x_i
  $s.t.\ d_{min} = \min\limits_i ||(\alpha x_i) \text{ mod } 4 - 2| - 1|$
  where d_{min} is the shortest distance from 0 to $|(\alpha x_i) \text{ mod } 4 - 2| - 1$.
- By adjusting α, we could make x_i change its sign while other x not changed.
  Thus, we can shatter all inputs whatever N is.
- The VC-dimension of $h_\alpha = \infty$

12.

- We have known $m_{\mathcal{H}}(N) \leq 2^N$, implying that $m_{\mathcal{H}}(N - i) \leq 2^{N-i}$, where $i \in \mathbb{N}$ and $0 \leq i \leq N$.
- Therefore,
  $\Rightarrow 2^i m_{\mathcal{H}}(N - i) \leq 2^i \times 2^{N-i} = 2^N$
  $\Rightarrow \min\limits_i 2^i m_{\mathcal{H}}(N - i) \leq 2^N$
- $\min\limits_i 2^i m_{\mathcal{H}}(N - i)$ is an upper bound of growth function $m_{\mathcal{H}}(N)$

13.

- When $N = [1, 9] \Rightarrow 2^{\lfloor 0.1N \rfloor} = 1$, implying that there is always one kind of dichotomy. Thus, we assume that the hypothesis set contains only on hypothesis.
- When $N \geq 10 \Rightarrow 2^{\lfloor 0.1N \rfloor} \neq 1$, showing that the hypothesis set contains not just one hypothesis. $\Rightarrow 2^{\lfloor 0.1N \rfloor}$ is not a valid growth function.

14.

- The lower bound of $d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k)$ is zero, because it is possible that the intersection of all hypothesis set is empty.
- The upper bound of $d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k)$ is the VC-dimension of the smallest hypothesis set, because the maximum intersection of all hypothesis sets is the case that all hypothesis sets contain the smallest hypothesis set.
- $\Rightarrow 0 \leq d_{\text{VC}}(\bigcap_{k=1}^{K} \mathcal{H}_k) \leq \max_{1 \leq k \leq K} d_{\text{VC}}(\mathcal{H}_k)$

15.

- The lower bound of $d_{\text{VC}}(\bigcup_{k=1}^{K} \mathcal{H}_k)$ is the VC-dimension of the biggest hypothesis set, because the minimum union of all hypothesis sets is the case that the biggest hypothesis set contains all the other hypothesis sets.

- The upper bound of $d_{\text{VC}}(\bigcup_{k=1}^{K} \mathcal{H}_k)$ is more complex, thus we first consider the case $K = 2$ with two hypothesis sets $\mathcal{H}_1$ and $\mathcal{H}_2$.

- Consider the growth function while all hypothesis sets are exclusive to each other: $\mathcal{H}_1 \cap \mathcal{H}_2 = \phi$
$m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) \leq m_{\mathcal{H}_1}(N) + m_{\mathcal{H}_2}(N)$
$m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) \leq \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=0}^{d_2} \binom{N}{i}$
$m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) \leq \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=0}^{d_2} \binom{N}{N-i} \leq \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=N-d_2}^{N} \binom{N}{i}$

- Now find the break point on $N \; s.t. \; m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N) < 2^N$
  - $N = d_1$,
  $\sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=N-d_2}^{N} \binom{N}{i} = 2^N + \sum_{i=N-d_2}^{N} \binom{N}{i} > 2^N$
  - $N = d_1 + d_2 + 1$,
  $\sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=N-d_2}^{N} \binom{N}{i} = \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=d_1+1}^{N} \binom{N}{i} = 2^N$

- $N = d_1 + d_2 + 2$,
  $\sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=N-d_2}^{N} \binom{N}{i} = \sum_{i=0}^{d_1} \binom{N}{i} + \sum_{i=d_1+2}^{N} \binom{N}{i} = 2^N - \binom{N}{d_1+1} < 2^N$
  $\Rightarrow$ break point

- The break point of $m_{\mathcal{H}_1 \cup \mathcal{H}_2}(N)$ is $d_1 + d_2 + 2$, therefore the VC-dimension is
  $\Rightarrow d_{\mathrm{VC}}(\mathcal{H}_1 \cup \mathcal{H}_2) = d_1 + d_2 + 1$

- Finally, according to the example when $K = 2$, we can conclude that the upper bound of
  $d_{\mathrm{VC}}(\bigcup_{k=1}^{K} \mathcal{H}_k)$ is
  $\sum_{k=1}^{K} d_{\mathrm{VC}}(\mathcal{H}_k) + (K-1)$

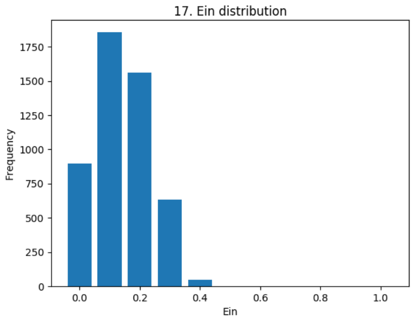
- $\Rightarrow \max_{1 \le k \le K} d_{\mathrm{VC}}(\mathcal{H}_k) \le d_{\mathrm{VC}}(\bigcup_{k=1}^{K} \mathcal{H}_k) \le \sum_{k=1}^{K} d_{\mathrm{VC}}(\mathcal{H}_k) + (K-1)$

16.

- Error = (predict + but real -) + (predict - but real +)
- $s = +1, \mathrm{E}_{out} = 0.8\frac{|\theta|}{2} + 0.2\frac{2-|\theta|}{2} = 0.2 + 0.3|\theta|$
- $s = -1, \mathrm{E}_{out} = 0.2\frac{|\theta|}{2} + 0.8\frac{2-|\theta|}{2} = 0.8 - 0.3|\theta|$
- E_{out} with $s = 0.5 + 0.3s(|\theta| - 1)$

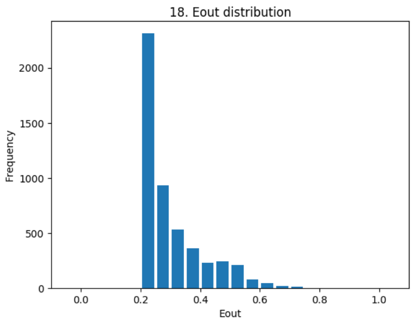
17.

- Average $\mathrm{E}_{in} = 0.142$



18.

- Average $\mathrm{E}_{out} = 0.300$



19.

- Optimal dicision stump: 4-th dimension, $\theta \approx 1.618, s = -1, \mathrm{E}_{in} = 0.25$

```
[Qhan@Qhan-Mac: ~/ml2016/hw2] 257
$ python 19.py
[0.39, 3.883, -1]
[0.39, -3.324, 1]
[0.4, -8.465, 1]
[0.25, 1.6175000000000002, -1]
[0.43, -9.6105, 1]
[0.28, 4.329000000000001, -1]
[0.34, 4.085, -1]
[0.39, -2.8135, -1]
[0.36, -0.147, -1]
best dimension: 3
best Ein: 0.250
best theta: 1.618
best s:  -1
```

20.

- $\mathrm{E}_{out} = 0.355$

```
[Qhan@Qhan-Mac: ~/ml2016/hw2] 258
$ python 20.py
[0.39, 3.883, -1]
[0.39, -3.324, 1]
[0.4, -8.465, 1]
[0.25, 1.6175000000000002, -1]
[0.43, -9.6105, 1]
[0.28, 4.329000000000001, -1]
[0.34, 4.085, -1]
[0.39, -2.8135, -1]
[0.36, -0.147, -1]
best dimension: 3
best Ein: 0.250
best theta: 1.618
best s: -1
Eout: 0.355
```