B06507153
邱律結

Machine Learning Foundations
HOMEWORK 2

1. c	6. d	11. d	16. d
2. d	7. d	12. b	17. b
3. c	8. d	13. b	18. c
4. b	9. [illegible] c	14. d	19. c
5. b	10. [illegible] c	15. b	20. a

1.
c

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 7 & 8 & 9 \\ 1 & 15 & 16 & 17 \\ 1 & 21 & 23 & 25 \end{bmatrix}\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} \to \vec{y} \text{ must be able to construct any 4-D vector}$$

$$\Rightarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} w_0 + \begin{bmatrix} 1 \\ 7 \\ 15 \\ 21 \end{bmatrix} w_1 + \begin{bmatrix} 1 \\ 8 \\ 16 \\ 23 \end{bmatrix} w_2 + \begin{bmatrix} 3 \\ 9 \\ 17 \\ 25 \end{bmatrix} w_3 = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

the above four vectors should be linearly independent

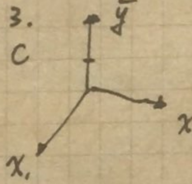
$$\Rightarrow \begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 7 & 8 & 9 \\ 1 & 15 & 16 & 17 \\ 1 & 21 & 23 & 25 \end{bmatrix} \xRightarrow{RREF} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & [illegible] & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \therefore \text{answer is c}$$

2.
d
In 1D perceptrons $\Rightarrow$ growth function $= 2(N-1)+2$

$2 \times$ (1D perceptrons) $-$ 2 overlaps

$$= 2\left[2(N-1)+2\right] - 2 = 4N-2$$

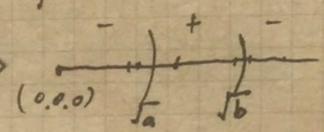
3.
c



$y = w_0 + w_1 x_1 + w_2 x_2$
with one constraint
$\Rightarrow$ VC dimension of 2D perceptrons -1
$= (2+1) - 1 = 2$

4.
b
In polar coordinate, only the distance from points to $(0,0,0)$ is of interest

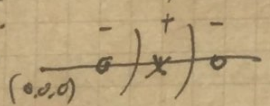
$\Rightarrow$



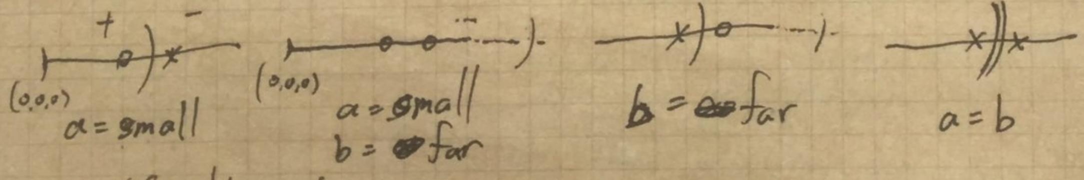
there are $N+1$ choices to cut, two cuts shouldn't be in the same point-interval

$\Rightarrow \binom{N+1}{2} + 1$ ($h(\vec{x}) = -1$ for every points is added).

5.
b



$\Rightarrow$ VC dimension < 3



$\therefore$ VC dimension $= 2$

6. $|E_{in}(h) - E_{out}(h)| \le \sqrt{\frac{8}{N}\left(\ln\frac{4\, m_H(2N)}{\delta}\right)} = \varepsilon$
d

$\Rightarrow E_{out}(g) - \varepsilon \le E_{in}(g) \le E_{out}(g) + \varepsilon$

$E_{out}(g^*) - \varepsilon \le E_{in}(g^*) \le E_{out}(g^*) + \varepsilon$

$E_{in}(g) \le E_{in}(g^*)$

$\Rightarrow E_{out}(g) - \varepsilon \le E_{in}(g) \le E_{in}(g^*) \le E_{out}(g^*) + \varepsilon$

$\Rightarrow E_{out}(g) - E_{out}(g^*) \le 2\varepsilon = 2\sqrt{\frac{8}{N}\ln\frac{4 m_H(2N)}{\delta}}$

7. d

2 hypothesis $\Rightarrow d_{vc} = 1$

4 hypothesis $\Rightarrow d_{vc} = 2$

⋮

M hypothesis's $\Rightarrow d_{vc} = \lfloor \ln M \rfloor$

8. d

For k=0 "+1", need 2 h's to shatter

for k=1 "+1", need 4 h's to shatter

⋮

for k=k "+1" need 2^{k+1} h's to shatter

$\lfloor \log_2 2^{k+1} \rfloor = k+1$

9. c

$d_{vc}(H) \ge d \Rightarrow$ ① some set of d distinct inputs is shattered

~~but ② some set of d distinct inputs is not shattered~~

~~e.g. 2D perceptrons $d_{vc} = 3$, but ... can't be shattered~~

$d_{vc}(H) < d+1 \Rightarrow$ ⑤ any d+1 distinct inputs is not shattered

↳ $d_{vc}(H) = d \Rightarrow$ some set of d distinct inputs is shattered

&

~~any~~ set of d+1 distinct inputs is not shattered

⇓ (implies)

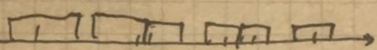
⑥ some set of d+1 distinct inputs is not shattered

③ "some set of d distinct inputs is not shattered by $\mathcal{H}$" is not always true

~~e.g. $d_{vc} = 1$ 2D perceptron $\Rightarrow$~~

e.g. 1D perceptron $d_{vc} = 2$, there is no way one can find 2 distinct inputs which couldn't be shattered

~~10. For each x_i on the number line, a non-overlap positive interval is set on top initially.~~



~~When y_i is chosen to be in the positive region, the left bound of its right positive interval (or the right bound of its left positive interval) is pulled to overlap on x_i. Hence, a target hypotheses is found. Since one can always apply the above algorithm to shatter for any distinct inputs $x \in R$, $d_{vc} = \infty$.~~

10. Let $\alpha = \pi\left(1 + \sum_{i=1}^{m} 2^{i} \frac{1-y_i}{2}\right)$ over a set of points $(x_1, x_2, \cdots, x_m)$
o with arbitrary labels $(y_1, y_2, \cdots, y_m) \in \{-1, +1\}^m$

$$x_j = 2^{-j}$$

$$\Rightarrow \alpha x_j = \pi\left(2^{-j} + \sum_{i=1}^{m} 2^{i-j} \frac{1-y_i}{2}\right)$$

$$= \pi\left(2^{-j} + \sum_{i=1}^{j} 2^{i-j} \frac{1-y_i}{2} + \cancel{\sum_{i=j+1}^{m} 2^{i-j} \frac{1-y_i}{2}}\right)$$

dropped since it only contributes multiples of 2π

$$\Rightarrow \alpha x_j = \pi\left(2^{-j} + \sum_{i=1}^{j-1} 2^{i-j} \frac{1-y_i}{2} + \frac{1-y_j}{2}\right)$$

$$\le \pi\left(\sum_{i=0}^{j-1} 2^{i-j} + \frac{1-y_j}{2}\right)$$

$$< \pi\left(1 + \frac{1-y_j}{2}\right)$$

$$\alpha x_j > \pi \frac{1-y_j}{2}$$

$$\therefore \pi \frac{1-y_j}{2} < \alpha x_j < \pi\left(1 + \frac{1-y_j}{2}\right)$$

for $y_j = -1$

$$\Rightarrow \pi < \alpha x_j < 2\pi$$

$$\Rightarrow h_\alpha(x_j) = -1$$

for $y_j = +1$

$$\Rightarrow 0 < \alpha x_j < \pi$$

$$\Rightarrow h_\alpha(x_j) = +1$$

11. d

$$E_{out}(h, \tau) = (1-\tau) E_{out}(h, 0) + \tau E_{out}(h, 1)$$

$$= (1-\tau) E_{out}(h, 0) + \tau(1 - E_{out}(h, 0))$$

$$= (1-2\tau) E_{out}(h, 0) + \tau$$

$$\Rightarrow E_{out}(h, 0) = \frac{E_{out}(h, \tau) - \tau}{1 - 2\tau}$$

12. Due to the continuity in $[0,1]$, ties happen in negligible probability.
b ~~Since~~ Since x_1, x_2, x_3 are independent and uniformly generated, ~~both~~ all three of them have equal probability being the largest of all three. $p = \frac{1}{3}$

Case 1: $f(\vec{x}) = 1$

$$(1-1)^2 \times 0.7 + (1-2)^2 \times 0.1 + (1-3)^2 \times 0.2 = 0.9$$

Case 2: $f(\vec{x}) = 2$

$$(2-2)^2 \times 0.7 + (2-3)^2 \times 0.1 + (2-1)^2 \times 0.2 = 0.3$$

Case 3: $f(\vec{x}) = 3$

$$(3-3)^2 \times 0.7 + (3-1)^2 \times 0.1 + (3-2)^2 \times 0.2 = 0.6$$

$$E_{out}(f) = \frac{1}{3} \times 0.9 + \frac{1}{3} \times 0.3 + \frac{1}{3} \times 0.6 = 0.6$$

13. $f_*(\vec{x}) = P(y=1|\vec{x}) + 2P(y=2|\vec{x}) + 3P(y=3|\vec{x})$
b

Case 1: $f(\vec{x}) = 1$

$$f_*(\vec{x}) = 0.7 + 2 \times 0.1 + 3 \times 0.2 = 1.5$$

Case 2: $f(\vec{x}) = 2$

$$f_*(\vec{x}) = 0.2 + 2 \times 0.7 + 3 \times 0.1 = 1.9$$

Case 3: $f(\vec{x}) = 3$

$$f_*(\vec{x}) = 0.1 + 2 \times 0.2 + 3 \times 0.7 = 2.6$$

$$\Delta(f, f_*) = \frac{1}{3}(1-1.5)^2 + \frac{1}{3}(2-1.9)^2 + \frac{1}{3}(3-2.6)^2$$

$$= 0.14$$

14. $P[BAD] \le 4 m_H(2N) e^{-\frac{1}{8}\varepsilon^2 N}$, $\varepsilon = 0.1$, $m_H(2N) = 2(2N-1)+2 = 4N$
d ~~when $N = 11543$, $4m_H(2N) e^{-\frac{1}{8}\varepsilon^2 N} > 0.1 = \delta$~~

~~when~~ $4 m_H(2N) e^{-\frac{1}{8}\varepsilon^2 N} = 16 N e^{-\frac{0.01}{8} N}$

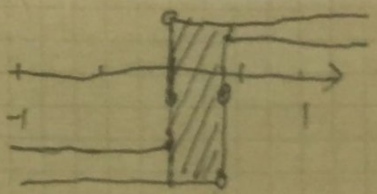
When $N = 11543$, $16 \cdot 11543 \cdot e^{-\frac{0.01}{8} \cdot 11543} > 0.1 = \delta$

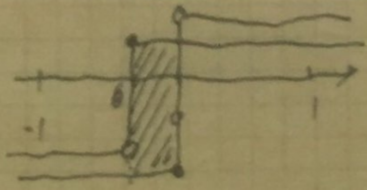
when $N = 11544$, $16 \cdot 11544 \cdot e^{-\frac{0.01}{8} \cdot 11544} < 0.1 = \delta$

$\therefore$ among the five choices, 12000 is the smallest N ~~choice~~

Scanned with CamScanner

15.
b





① $E_{out}(h_{+1,\theta}, g) = P(\text{sign}(x-\theta) \neq \text{sign}(x))$

$= \frac{1}{2}|\theta|$