MACHINE LEARING HW1

3. (d)

$$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$$

$$\downarrow \times \frac{1}{4}$$

$$X = \{x_1, x_2\}$$

$$R^2 = \max_n \|x_n\|^2$$

$$\rho = \min_n y_n \frac{w_f^T}{\|w_f\|} x_n.$$

$$\frac{w_f}{\|w_f\|} \cdot \frac{w_T}{\|w_T\|} \geq \sqrt{T} \cdot \frac{\rho}{R}$$

$$T \leq \frac{\rho^2}{R^2} = \frac{\min_n |w_f \cdot x_n|^2}{\max_n \|x_n\|^2 \cdot \|w_f\|^2}.$$

the answer is (d) unchanged.

explanation:

the time-spent on ML is limited by the upper bound T and T is smaller than $\frac{\rho^2}{R^2}$, the size of $\frac{\rho^2}{R^2}$ is determined by the max length of x_n. In this case, we cannot assure that Dr. Short eliminate the "max x_n", which is the only factor changes the size of T. In the worst case, the time-spent will be unchanged.

4. (c)

$$\rho = \min_n \frac{|w_f^T x_n|}{\|w_f\|\|x_n\|}$$

original : $T \le \frac{\rho^2}{R^2} = \frac{\min_n |w_f \cdot x_n|^2}{\max_n \|x_n\|^2 \|w_f\|^2}$

$$\|w_{t+1}\|^2 = \|w_t + \eta_t y_{n(t)} x_{n(t)}\|^2$$

$$= \|w_t + \frac{x_{n(t)}}{\|x_{n(t)}\|} y_{n(t)}\|^2.$$

$$= \|w_t\|^2 + 2 y_{n(t)} w_t^T \cdot \frac{x_{n(t)}}{\|x_{n(t)}\|} + \|y_{n(t)} \frac{x_{n(t)}}{\|x_{n(t)}\|}\|^2$$

$$\le \|w_t\|^2 + \|y_{n(t)} \frac{x_{n(t)}}{\|x_{n(t)}\|}\|^2.$$

$$\le \|w_t\|^2 + \underbrace{\max_n \|\frac{x_n}{\|x_n\|}\|^2}_{1} = \|w_t\|^2 + 1 \ge$$

$$\le \|w_{t-1}\|^2 + 2 \le \|w_0\|^2 + t + 1 \Rightarrow t + \cancel{1} \ge \|w_t\|^2 \cancel{t}$$

$$t \ge \|w_t\|^2$$

$$w_f^T w_{t+1} \ge w_f^T w_t + \|w_f\| \rho.$$

$$\downarrow$$

$$w_f^T w_{t+1} \ge (t+1)\|w_f\| \cdot \rho$$

$$w_f^T w_t \ge t \|w_f\| \cdot \rho.$$

$$\frac{w_f^T w_t}{\|w_t\|\|w_f\|} \ge \cancel{\|w_f\|} \cdot \rho \cdot \sqrt{t}$$

$$1 \ge \rho \cdot \sqrt{t}$$

$$\sqrt{t} \le \frac{1}{\rho}, \quad t \le \rho^{-2}. \qquad p = 2$$

5.

(C) $y_{n(t)} \cdot w_t^T x_{n(t)} < 0$

$$y_{n(t)} \cdot w_{t+1}^T x_{n(t)} > 0$$

$$w_{t+1}^T = w_t + \eta_t \cdot y_{n(t)} x_{n(t)}$$

$$y_{n(t)} \cdot (w_t + \eta_t \cdot y_{n(t)} x_{n(t)}) x_{n(t)} > 0$$

$$[y_{n(t)} w_t + \eta_t \underbrace{(y_{n(t)})^2}_{=1} x_{n(t)}] x_{n(t)} > 0$$

$$[y_{n(t)} w_t + \eta_t x_{n(t)}] x_{n(t)} > 0.$$

$$y_{n(t)} w_t^T x_{n(t)} + \eta_t \|x_{n(t)}\|^2 > 0$$

$$\eta_t > \frac{-y_{n(t)} w_t^T x_{n(t)}}{\|x_{n(t)}\|^2}.$$

the answer is C

6. (C)

Ensuring that ① w_t gets more aligned with w_f &

② w_t doesn't grow too fast

can give a promise of finding a perfect line

7 (e) There is no any "correct" input and output in reinforcement learing, the result of which will be judged by the enviroment, ex. win or lose.

8. (b)

16. 11 (b)

17. -7 (b)

18. 15 (c)

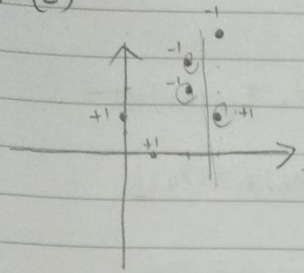
19. 17 (d)

20. 17 (d)

1. (d). create the data set X = mang image
Y = quality of mang.

2. (e).
(a) no training data
(b) no training.
(c) no training
(d)

9. (e)



the smallest $E_{ots}(g)$ is 0 because the data $\subseteq \mathcal{U}$ is linear saparable.

$$E_{ots} = \frac{1}{|\mathcal{U}/D|} \sum_{(x,y)\in \mathcal{U}/D} [\![h(x) \neq y]\!] \quad , \underbrace{|\mathcal{U}/D|}_{差} = 6-3=3$$

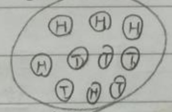
$$E_{ots\,max} = \frac{1}{3} \cdot 3 = 1 \quad \Rightarrow (E_{ots\,min}, E_{ots\,max}) = (0, 1)$$

10. (b)

Hoeffding's Inequality : $P[|\nu-\mu| > \epsilon] \leq 2e^{(-2\epsilon^2 N)}$

a big plate with two types of coin : head & tail

$\mu = \frac{1}{2} + \epsilon \quad \nu = N.(1-\delta)$



N = times

$1-\delta$ = 成功辨別出哪一面機率較高的機率

δ = 失敗的機率

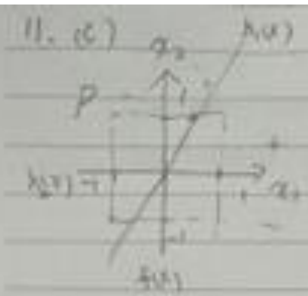
$$\underbrace{P[|\nu-\mu| > \epsilon]}_{失敗的機率} \leq \delta = 2e^{(-2\epsilon^2 N)}$$

$$\ln \frac{\delta}{2} = -2\epsilon^2 N$$

$$N = \frac{1}{-2\epsilon^2} \ln \frac{\delta}{2}$$

$$= \frac{1}{2\epsilon^2} \ln \frac{2}{\delta}$$

11. (c)



$5^*(x_n, f(x_n))$ s.t. $E_{in}(h_*) = 0$

$E_{in}(h_*) = \frac{1}{N}\sum_{n=1}^{N}[\![h(x_n) \neq f(x_n)]\!]$

$f(x) = sign(x_1)$

$h_1(x) = sign(2x_1 - x_2) \rightarrow 2x_1 - x_2 = 0,\quad 2x_1 = x_2$

$h_2(x) = sign(x_2)$

the probability of one sample s.t. $f(x_n) = h(x_n)$ is $\frac{1}{2}$

In the case of five sample : $(\frac{1}{2})^5 = \frac{1}{32}$

12.

$E_{in}(h_2) = E_{in}(h_1)$

$E_{in}(h_2) = \frac{1}{N}\sum_{n=1}^{N}[\![h_2(x_n) \neq f(x_n)]\!]$

$E_{in}(h_1) = \frac{1}{N}\sum_{n=1}^{N}[\![h_1(x_n) \neq f(x_n)]\!]$

Sample(n)	1	2	3	4	5
f(x)	+				
$h_1(x)$	+				
$h_2(x)$	+				
	2^1	2^1	2^3	2^3	2^3

$f(x_n) = h_1(x_n) =$

when $E_{in}(h_2) = E_{in}(h_1)$

$\Rightarrow (\# \, h_2(x) \neq f(x)) = (\# \, h_1(x) \neq f(x))$

$\Rightarrow 0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1$

13. (C)

$$P[\text{BAD } \mathcal{D} \text{ for } \mathcal{H}] \le C \cdot 2e^{(-2\epsilon^2 N)}$$

$$P[\text{BAD } \mathcal{D} \text{ for } \mathcal{H}] = P[\text{BAD } \mathcal{D} \text{ for } h_1] \cup P[\text{BAD for } h_2 \ldots$$

$$\le P[\text{BAD } \mathcal{D} \text{ for } h_1] + P[\text{BAD for } h_2] \ldots$$

$$\le \underbrace{2d}_{2d\text{個 hypothesis}} \cdot 2e^{(-2\epsilon^2 N)}$$

14. (d)

1	2	3	4	5	6		
	O		O		O	→ A	✓
	O	O	O		✓	→ B	✓
					O	→ C	
	O	O		O	✓	— d	✓

6 → green

five green 3: $\left(\frac{3}{4} \times \frac{1}{6}\right)^5 = \left(\frac{1}{12}\right)^5$

five geen 1 : 0

orange 2: $\left(\frac{1}{4} \times \frac{1}{6}\right)^5 = \left(\frac{1}{24}\right)^5$

green 2: $\left(\frac{3}{4} \times \frac{1}{6}\right)^5 = \left(\frac{1}{8}\right)^5$

green 4: $\left(\frac{2}{4} \times \frac{1}{6}\right)^5 = \left(\frac{1}{12}\right)^5$ ✓

green 5: $\left(\frac{1}{4} \times \frac{1}{6}\right)^5 = \left(\frac{1}{24}\right)^5$

15.

2: $\frac{3^5}{4^5}$ 6: $\frac{2^5}{4^5}$

3: $\frac{2^5}{4^5}$

4: $\frac{2^5}{4^5}$

$$\text{sum}(2,3,4,5) = \left(\frac{1}{2}\right)^5 \times 3 + \left(\frac{1}{4}\right)^5 + \left(\frac{3}{4}\right)^5$$

$$= \frac{3}{32} + \frac{1}{1024} + \frac{243}{1024}$$

$$= \frac{342}{\quad}$$