ML foundation HW 1

Electrical Engineering
B06504016 林家宏

Answer: 1~5 dedcd
6~10 cebeb
11~15 Cdbdc
16~20 bbcdd

Solution: 1 (d) satisfy the three condition of ML.
① The appearance of mango has something to do with its quality, so the images of mango exist some underlying pattern to be learned.
② No programmable definition since the size or color of mango cannot be a absolute factor of quality.
③ Without doubt, there are lots of image and quality of mango (data) for machinge to learn.

2 (a) Flipping coins is independent to spam emails.
(b) The email are not classified by machine.
(c) There is no learning process since we use programmable definition.
(d) The rule is determined by human.
(e) The maching is "learning" since it updates the function throughout the whole process. => (e) is the answer.

3 $R^2 = \max_n \|x_n\|^2 \Rightarrow R_{new}^2 = \max_n \|\frac{x_n}{4}\|^2 = \frac{1}{16}R^2$

$\rho^2 = \left(\min_n y_n \frac{w_f^T}{\|w_f\|} x_n\right)^2 = \min_n \left(\frac{w_f^T}{\|w_f\|} x_n\right)^2 \Rightarrow \rho_{new} = \min_n \left(\frac{w_f^T}{\|w_f\|} \frac{x_n}{4}\right)^2 = \frac{1}{16} \min_n \left(\frac{w_f^T}{\|w_f\|} x_n\right)$

$= \frac{1}{16}\rho^2$

$$T_{new} \le \frac{R_{new}^2}{\rho_{new}^2} = \frac{\frac{1}{16}R^2}{\frac{1}{16}\rho^2} = \frac{R^2}{\rho^2} = T$$

4 $w_f^T w_t \ge w_f^T w_{t-1} + \min_n \frac{1}{\|x_{n(t-1)}\|} w_f^T y_{n(t-1)} x_{n(t-1)} = w_f^T w_{t-1} + \hat{\rho} \cdot \|w_f\|$

$$\begin{aligned} w_f^T w_t &\ge w_f^T w_{t-1} + \hat{\rho}\cdot\|w_f\| \\ w_f^T w_{t-1} &\ge w_f^T w_{t-2} + \hat{\rho}\cdot\|w_f\| \\ &\vdots \\ +)\ w_f^T w_0 &\ge 0 \\ \hline w_f^T w_t &\ge \hat{\rho}\|w_f\|\cdot T \quad ① \end{aligned}$$

$$\begin{aligned} \|w_t\|^2 &\ge \|w_{t-1}\|^2 + 1 \\ \|w_{t-1}\|^2 &\ge \|w_{t-2}\|^2 + 1 \\ &\vdots \\ +)\ \|w_0\|^2 &\ge 0 \\ \hline \|w_t\|^2 &\ge T \quad ② \end{aligned}$$

$w_t = w_{t-1} + \frac{1}{\|x_{n(t-1)}\|} y_{n(t-1)} x_{n(t-1)}$

$\Rightarrow \|w_t\|^2 \ge \|w_{t-1}\|^2 + \frac{\|x_{n(t-1)}\|^2}{\|x_{n(t-1)}\|^2} = \|w_{t-1}\|^2 + 1$

Combine ① and ②

$\Rightarrow 1 \ge \frac{w_f^T w_t}{\|w_f\|\|w_t\|} \ge \sqrt{T}\hat{\rho} \Rightarrow \sqrt{T}\hat{\rho} \le 1 \Rightarrow T \le \frac{1}{\hat{\rho}^2} = \hat{\rho}^{-2} \Rightarrow (c)$

5 We want to find the update rule such that $y_{n(t)} w_{t+1}^T x_{n(t)} > 0$ is always true.

(d) $w_{t+1} = w_t + y_{n(t)} x_{n(t)} \cdot \left[\left(\frac{-y_{n(t)} w_t^T x_{n(t)}}{\|x_{n(t)}\|^2}\right) + 1\right]$

$$y_{n(t)} w_{t+1}^T x_{n(t)} = y_{n(t)} \left[w_t + y_{n(t)} x_{n(t)} \left[\frac{-y_{n(t)} w_t^T x_{n(t)}}{\|x_{n(t)}\|^2} + 1 \right] \right]^T x_{n(t)}$$

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

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

$$= \|x_{n(t)}\|^2 \geq 0$$

Therefore, (d) is correct answer.

6 (a) $w_f^T w_t = w_f^T (w_{t-1} + 2 y_{n(t-1)} x_{n(t-1)}) \geq w_f^T w_{t-1} + \boxed{2 \min_n w_f^T x_n y_n} \rightarrow \rho \cdot \|w_f\|$

$\Rightarrow w_f^T w_t \geq t \cdot \rho \cdot \|w_f\|$

$$\|w_t\|^2 = \|w_{t-1} + 2 y_{n(t-1)} x_{n(t-1)}\|^2 = \|w_{t-1}\|^2 + \overbrace{4 y_{n(t-1)} w_{t-1}^T x_{n(t-1)}}^{<0} + \|2 y_{n(t-1)} x_{n(t-1)}\|^2$$

$\leq \|w_{t-1}\|^2 + \boxed{\max_n \|4 x_n\|^2} \rightarrow R^2$ $\quad 1 \geq \frac{w_f^T w_t}{\|w_f\| \|w_t\|} \geq \sqrt{t} \frac{\rho}{R}$

(b) Similar to (a)

(d) $w_f^T w_t = w_f^T \left(w_{t-1} + y_{n(t-1)} x_{n(t-1)} \left[\frac{-y_{n(t-1)} w_{t-1}^T x_{n(t-1)}}{\|x_{n(t-1)}\|^2} + 1 \right] \right)$

$\geq w_f^T w_{t-1} + \max_{(n,t)} w_f^T y_n x_n \left(\frac{-y_n w_t^T x_n}{\|x_n\|^2} + 1 \right) \rightarrow \rho \cdot \|w_f\|$

$\Rightarrow w_f^T w_t \geq t \cdot \rho \cdot \|w_f\|$

$$\|w_t\|^2 = \left\| w_{t-1} + y_{n(t-1)} x_{n(t-1)} \left[\frac{-y_{n(t-1)} w_{t-1}^T x_{n(t-1)}}{\|x_{n(t-1)}\|^2} + 1 \right] \right\|^2 = \|w_{t-1}\|^2 - \frac{\|w_t^T x_t\|}{\|x_t\|^2} + \|x_n$$

$\leq \|w_{t-1}\|^2 + \max_n \|x_n\|^2 \rightarrow R^2$

$$1 \geq \frac{w_f^T w_t}{\|w_f\| \|w_t\|} \geq \sqrt{t} \frac{\rho}{R}$$

(a)(b)(d) Choose (c)

7 They are getting judgement on their performance in the learning process. It's reinforcement learning. $\Rightarrow$ (e) is correct.

8 structured: Records of human driving can be divided into several part for machine to learn, just like steering, braking.

semi-supervised: Half of the data contain the human record, the other half do not.

Batch learning: The model is trained by batch of data before application.

Raw feature: Video view are simple feature. $\Rightarrow$ The answer is (b)

9

g_1

Choose $(\vec{x}, y) = \{((3,2),1), ((0,2),1), ((2,3),-1)\}$

Insert $x_0 = 1$, and start with $\vec{w} = (0,0,0)$

We can obtain $\vec{w} = (10,-1,-3)$, and check the remaining data

$((1,1,0),1), ((0,2,4),-1), ((0,3,5),-1)$. We find that all are correct

$\Rightarrow E_{ots}(g_1) = 0$ → Choose (e)

g_2

Choose $(\vec{x}, y) = \{((1,0),1), ((2,4),-1), ((3,5),-1)\}$

Insert $x_0 = 1$, and start with $\vec{w} = (0,0,0)$

We can obtain $\vec{w} = (1,1,0)$, and check the remaining data

$((2,3),-1), ((2,4),-1) ((3,5),-1)$. We find that all are wrong. $\Rightarrow E_{ots}(g_2) = 1$

10 X_{ob}: The coin coming up probability we observe.

X_r: The real coin coming up probability. $(\frac{1}{2} + \varepsilon)$

$P(|X_{ob} - X_r| > \varepsilon) \leq 2\exp(-2\varepsilon^2 N)$ — ①

$1 - P(|X_{ob} - X_r| > \varepsilon) \geq 1 - \delta \Rightarrow P(|X_{ob} - X_r| > \varepsilon) \leq \delta$ — ②

Combine ① and ②

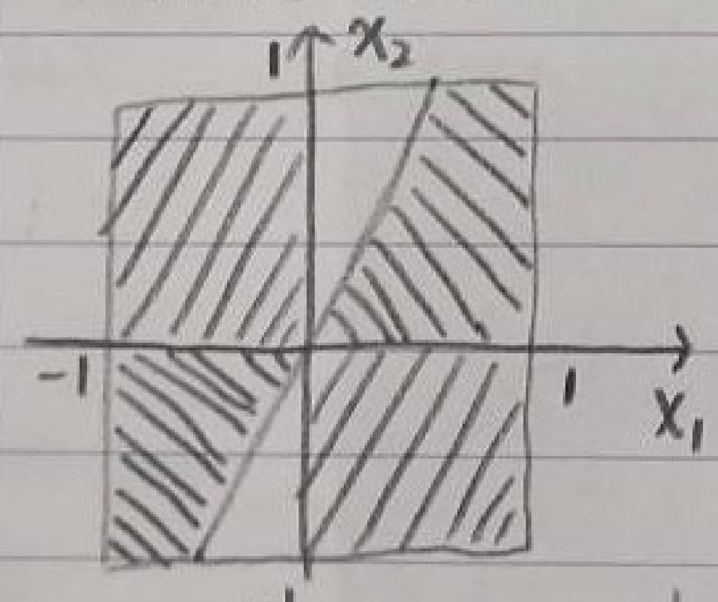
$\delta \leq 2\exp(-2\varepsilon^2 N) \Rightarrow N \geq \frac{1}{2\varepsilon^2} \log \frac{2}{\delta}$ $\Rightarrow$ Choose (b)

11 $E_{in}(h_2) = 0$ implys $\text{sign}(x_1) = \text{sign}(x_2)$ for all five data

For each data, $P(\text{sign}(x_1) = \text{sign}(x_2)) = \frac{1}{2}$. Therefore, $P(E_{in}(h_2) = 0) = (\frac{1}{2})^5 = \frac{1}{32}$

$\Rightarrow$ Choose (c)

12



$h_1: O$ $h_2: X$ $P = \frac{1}{2}$

$h_1: X$ $h_2: O$ $P = \frac{1}{8}$

$h_1: O$ $h_2: O$ $P = \frac{3}{8}$

$$E_{in}(h_1) = E_{in}(h_2) = \begin{cases} 0 : (\frac{3}{8})^5 & \text{— ①} \\ 1 : C^5_3 C^2_1 C^1_1 (\frac{3}{8})^3 (\frac{1}{2})(\frac{1}{8}) & \text{— ②} \\ 2 : C^5_1 C^4_2 C^2_2 (\frac{3}{8})(\frac{1}{2})^2(\frac{1}{8})^2 & \text{— ③} \end{cases}$$

① + ② + ③ $= \frac{3843}{32768}$ $\Rightarrow$ Choose (d)

13 $P(\text{BAD } D \text{ for } H) = P(\text{BAD } D \text{ for } h_1) \text{ or } P(\text{BAD } D \text{ for } h_2) \cdots \text{ or } P(\text{BAD } D \text{ for } h_d)$

$= P(\text{BAD } D \text{ for } h_1) \text{ or } P(\text{BAD } D \text{ for } h_2) \text{ or } \cdots P(\text{BAD } D \text{ for } h_d)$

$\leq P(\text{BAD } D \text{ for } h_1) + P(\text{BAD } D \text{ for } h_2) + \cdots + P(\text{BAD } D \text{ for } h_d)$

$\leq d \cdot 2\exp(-2\varepsilon^2 N)$

$\Rightarrow C = d \Rightarrow$ Choose (b)

14

	Green	Orange
A	2,4,6	1,3,5
B	2,3,4	1,5,6
C	6	1~5
D	2,3,5	1,4,6

3 shows up two times(B,D) in the green column, and so does 4.

$\Rightarrow$ Choose (d)

15

Types of dice among 5 dices	Combination	number of each combination	Total number
1	A,B,C,D	(5)	1X4 = 4
2	AB,AC,AD,BD	(4,1)(3,2)(2,3)(1,4)	30X4 = 120
3	ABD	(3,1,1)(2,2,1)(2,1,2) (1,3,1)(1,1,3)(1,2,2)	150X1 = 150

$\frac{4+120+150}{4^5} = \frac{274}{1024} \Rightarrow$ Choose (c)