Machine Learning Foundations HW_1 R09946023 吳偉樂

SOLUTIONS :

1. (d) 2. (e) 3. (d) 4. (c) 5. (d)
6. (d) 7. (e) 8. (b) 9. (e) 10. (b)
11. (c) 12. (d) 13. (b) 14. (d) 15. (c)
16. (b) 17. (b) 18. (c) 19. (d) 20. (d)

1. The answer is **(d)**

We can simply use CNN to rank it. For classical machine learning method, we can find features from mango images and classify it by multiclass classification.

For example, we have many images that contain mangoes and its rank is given (A,B,C), these are the data (x,y) where x is image , y is the rank .

The mango A contains more red color (more ripe , or beautiful) than mango B (unripe, so contain green color or have some black part) than mango C (mostly unripe or contain big broken black part，即芒果上的黑色斑點), so the mean of Red color (pick R in RGB channel of images) of mango A > Red color of mango B > Red color of mango C. Thus mean of Red color of image can be a feature .

Also, we can find another features (variance, texture feature , …) from image. Finally, we can use some method like SVM , RandomForest, Boosting to classify it. (I have participated in Mango Competition in AI HUB, and I use these methods so I can say so much XD)

2. The answer is **(e)**

[a] It is not machine learning because it is just classify it randomly ,
flip 3 fair coins = {HHH, HHT, HTH, HTT, TTT, TTH, THT, THH}
2 heads = 4/8 = 0.5 , so it just depends on 1/2 probability, it does not contain any data to let machine learn !

[b] This is not machine learning because It doesn't have any data set ,
it depends on 'human feeling' , for human 1 and 2 , this email may considered as spam because they don't like the content , so it is classified by whether have 2 humans dislike same thing.

[c] Not machine learning because it just produce the list of words for spams by 3 humans. Machine Learning needs the data answer to analyse, but no any answer (spam, not spam) given here so machine can't classify it , this method just a simple algorithm

[d] Although it has data set , but it does not define how the spam words found, produce a list of words that appear more than 5 times seems not reasonable because when the non-spam word like preposition 'is, of ' appear more than 5 times , then it consider as spam word .

3) (d) unchanged

According to the 2 conditions:

① $y_{n(t)} w_f^T x_{n(t)} \cdot \frac{1}{4} \geq \min_n y_n w_f^T x_n \frac{1}{4} > 0$

$\Rightarrow w_f^T w_{t+1} = w_f^T \left(w_t + y_{n(t)} \frac{x_{n(t)}}{4} \right) \geq w_f^T w_t + \min_n y_n w_f^T \frac{x_n}{4}$

$\Rightarrow w_f^T w_{t+1} \geq w_f^T w_t + \frac{1}{4} \|w_f\| \rho$ where, $\rho = \min_n y_n \frac{w_f^T}{\|w_f\|} x_n$

$w_f^T w_0 = 0$

$w_f^T w_1 \geq \cancel{w_f^T w_0} + \frac{1}{4} \|w_f\| \rho$

$w_f^T w_2 \geq w_f^T w_1 + \frac{1}{4} \|w_f\| \rho \geq \frac{1}{4} \cdot 2 \|w_f\| \rho$

$\vdots$

$w_f^T w_T \geq \frac{1}{4} \cdot T \|w_f\| \rho \Leftrightarrow \underline{\frac{w_f^T w_T}{\|w_f\|} \geq \frac{1}{4} T \rho}$ — $(*)$.

② w_t changed when mistake

$\|w_{t+1}\|^2 = \|w_t + y_{n(t)} \frac{x_{n(t)}}{4}\|^2 \leq \|w_t\|^2 + \|y_{n(t)} \frac{x_{n(t)}}{4}\|^2 \leq \|w_t\|^2 + \frac{1}{16} \underbrace{\max_n \|x_n\|^2}_{=R^2}$

$= \|w_t\|^2 + \frac{1}{16} R^2$

$\|w_0\|^2 = 0$

$\|w_1\|^2 \leq \cancel{\|w_0\|^2} + \frac{1}{16} R^2$

$\vdots$

$\|w_T\|^2 \leq \frac{T}{16} R^2 \Leftrightarrow \underline{\|w_T\| \leq \frac{\sqrt{T}}{4} R}$ — $(**)$

$\frac{(*)}{(*)(*)} : \frac{\frac{1}{4} T \rho}{\frac{1}{4} \sqrt{T} R} \leq \frac{w_f^T w_T}{\|w_f^T\| \|w_T\|} \leq 1 \Leftrightarrow \boxed{T \leq \frac{R^2}{\rho^2}}$

---

4) (c) 2

According to 2 conditions:

① $y_{n(t)} w_f^T x_{n(t)} \cdot \frac{1}{\|x_{n(t)}\|} \geq \min_n y_n w_f^T \frac{x_n}{\|x_n\|} > 0$

$\Rightarrow w_f^T w_{t+1} = w_f^T \left(w_t + y_{n(t)} \frac{x_{n(t)}}{\|x_{n(t)}\|} \right) \geq w_f^T w_t + \min_n y_n w_f^T \frac{x_n}{\|x_n\|}$

$\Rightarrow w_f^T w_{t+1} \geq w_f^T w_t + \|w_f\| \hat{\rho}$ where $\hat{\rho} = \min_n \frac{|w_f^T x_n|}{\|w_f\| \|x_n\|}$

$\Rightarrow w_f^T w_0 = 0$, $w_f^T w_1 \geq \|w_f\| \hat{\rho}$

$\vdots$

$w_f^T w_T \geq T \|w_f\| \hat{\rho}$ — $(*)$

② w_t changed under mistake:

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

$\Rightarrow \|w_0\|^2 = 0$

$\|w_1\|^2 \leq \|w_0\|^2 + 1$

$\|w_T\|^2 \leq T \Leftrightarrow \|w_T\| \leq \sqrt{T}$ — $(**)$

$\frac{(*)}{(**)} = \frac{T \|w_f\| \hat{\rho}}{\sqrt{T}} \leq \frac{w_f^T w_T}{\|w_T\|} \Leftrightarrow \sqrt{T} \hat{\rho} \leq \frac{w_f^T w_T}{\|w_f\| \|w_T\|} \leq 1 \Leftrightarrow \boxed{T \leq \hat{\rho}^{-2}}$

5) (d)

$y_{n(t)} W_t^T x_{n(t)} \leq 0$

$W_{t+1} = W_t + \eta_{n(t)} y_{n(t)} x_{n(t)}$

$\Rightarrow$. find $\eta_{n(t)}$ such that $y_{n(t)} W_{t+1}^T x_{n(t)} > 0$

$\Rightarrow$. $y_{n(t)} W_{t+1} = y_{n(t)} W_t + \eta_{n(t)} \underbrace{y_{n(t)} y_{n(t)}}_{=+1} x_{n(t)} = y_{n(t)} W_t + \eta_{n(t)} x_{n(t)}$

$y_{n(t)} W_{t+1}^T x_{n(t)} = y_{n(t)} W_t^T x_{n(t)} + \eta_{n(t)} x_{n(t)} \cdot x_{n(t)} > 0$

$\eta_{n(t)} \|x_{n(t)}\|^2 > - y_{n(t)} W_t^T x_{n(t)}$

$\eta_{n(t)} > \frac{- y_{n(t)} W_t^T x_{n(t)}}{\|x_{n(t)}\|^2} \Rightarrow$ take $\eta_{n(t)} = \frac{- y_{n(t)} W_t^T x_{n(t)}}{\|x_{n(t)}\|^2} + 1$

6) (d) 4

By the fact, w_t changed only when mistake

$\Leftrightarrow \text{sign}(W_t^T x_{n(t)}) \neq y_{n(t)} \Leftrightarrow y_{n(t)} W_t^T x_{n(t)} \leq 0$

$y_{n(t)} \cdot W_t^T \cdot \frac{x_{n(t)}}{\eta_{n(t)}} \leq 0$ ($x_{n(t)}$ is scale by some factors)

ⓐ $\eta_{n(t)} = \eta = 2$

ⓑ $\eta_{n(t)} = \eta = 0.1126$

ⓒ $\eta_{n(t)} = \left(\frac{- y_{n(t)} W_t^T x_{n(t)}}{\|x_{n(t)}\|^2} \right) > 0$

ⓓ $\eta_{n(t)} = \left\lfloor \frac{- y_{n(t)} W_t^T x_{n(t)}}{\|x_{n(t)}\|^2} \right\rfloor > 0$

$\Rightarrow$. $\therefore y_{n(t)} W_t^T \frac{x_{n(t)}}{\eta_{n(t)}} \leq 0$

$\Leftrightarrow y_{n(t)} W_t^T x_{n(t)} \leq 0$

ⓔ $\eta_{n(t)} = - \left\lfloor \frac{- y_{n(t)} W_t^T x_{n(t)}}{\|x_{n(t)}\|^2} + 1 \right\rfloor < 0 \Rightarrow$. $y_{n(t)} W_t^T \frac{x_{n(t)}}{\eta_{n(t)}} \leq 0$

$\Leftrightarrow$. $y_{n(t)} W_t^T x_{n(t)} > 0$ (reject!)

$\therefore$ ⓔ is not the choice!

$\Rightarrow$. only ⓐ, ⓑ, ⓒ, ⓓ

7. The answer is **(e)**

{ Learning to play the game by practicing with itself and getting the feedback from the ‘judge’ environment } is equivalent to

{ Do not show the correct answer but can give reward or punishment as feedback }, which is the idea of Reinforcement Learning

8. The answer is **(b)**

Raw features : since the input is a video (dynamic pixels), it often need human or machines to convert to concrete ones

Semi-Supervised Learning : Given records of a car, but no given record for 1126 cars

Structured Learning : input = video, output = multiclass classification, it is classification problrm without ‘explicit’ class definition

Batch Learning : it gives all output of a car

9) (e) $(0,1)$

	X	y	g	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
D	$(1,0)$	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
	$(3,2)$	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
	$(0,2)$	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
	$(2,3)$	-1	?	1	1	1	-1	-1	1	-1	-1
	$(2,4)$	-1	?	1	1	-1	1	-1	-1	1	-1
	$(3,5)$	-1	?	1	-1	1	1	1	-1	-1	-1
	$E_{ots}(g)$			1	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0

$E_{in}(g) = 0$

smallest largest

$\Rightarrow E_{ots}(g) = (0, 1)$

10). (b) $\frac{1}{2\varepsilon^2}\log\frac{2}{\delta}$

$\nu = \frac{1}{2} + \varepsilon$, $\mu = \frac{H}{T+H} \geq 1-\delta$

$P[|\nu-\mu| > \varepsilon] \leq \delta = 2e^{-2\varepsilon^2 N}$

$\Rightarrow \delta = 2e^{-2\varepsilon^2 N} \Rightarrow \ln\frac{\delta}{2} = -2\varepsilon^2 N \Rightarrow N = \frac{1}{2\varepsilon^2}\ln\frac{2}{\varepsilon}$

11). (c) $\frac{1}{32}$

$X = (x_1, x_2)^T \in \mathbb{R}^2$ $\quad E_{in}(h_2) = \frac{1}{N}\sum_{i=1}^{N}[\![h(y_n) \neq y_n]\!]$

$f(X) = \text{sign}(x_1)$, $h_2(\mathbf{X}) = \text{sign}(x_2)$

when $X = (+,+)$, $f(X) = +1$, $h_2(X) = +1$
$X = (+,-)$, $f(X) = +1$, $h_2(X) = -1$
$X = (-,+)$, $f(X) = -1$, $h_2(X) = +1$
$X = (-,-)$, $f(X) = -1$, $h_2(X) = -1$

if $\text{sign}(x_1) = \text{sign}(x_2)$, then $f(X) = h_2(X)$
$\Leftrightarrow [\![h(X) \neq y]\!] = 0$

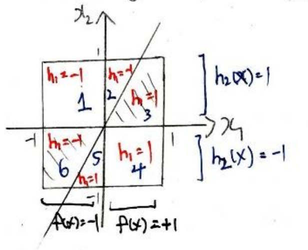
$\Rightarrow P(E_{in}(h_2) = 0) = P\left[\sum_{i=1}^{5}[\![h(X_i) \neq y_i]\!] = 0\right] = \underbrace{\frac{2}{4}}_{X_1} \times \underbrace{\frac{2}{4}}_{X_2} \times \underbrace{\frac{2}{4}}_{X_3} \times \underbrace{\frac{2}{4}}_{X_4} \times \underbrace{\frac{2}{4}}_{X_5} = \frac{1}{32}$

(x_1, x_2) — $\frac{1}{2}$ chance get + (or −); $\frac{1}{2}$ chance get + (or −) $\Rightarrow \frac{2}{4}$ for $++$, $--$

12.) (d) $\frac{3843}{32768}$

$h_1(x) = \text{sign}(2x_1 - x_2) = \begin{cases} 1, & 2x_1 > x_2 \\ 0, & 2x_1 < x_2 \end{cases}$, $h_2(x) = \text{sign}(x_2) = \begin{cases} 1, & x_2 > 0 \\ 0, & x_2 < 0 \end{cases}$

$f(x) = \text{sign}(x_1)$



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

$\Leftrightarrow \underbrace{\frac{1}{N}\sum_{i=1}^{N} [\![h_2(x) \neq f(x)]\!]}_{} = \frac{1}{N}\sum_{i=1}^{N} [\![h_1(x) \neq f(x)]\!]$

$\hookrightarrow = \begin{cases} 1 & \text{if } h_2(x) \neq f(x) \\ 0 & \text{if } h_2(x) = f(x) \end{cases}$

Case 1: all 5 points are either at region 3 or region 6

Case 2: 3 points are either at region 3 or 6,
1 point either at region 1 or 4, 1 point either at region 2 or 5

Case 3: 1 point either at region 3 or 6,
2 points either at region 1 or 4
2 points either at region 2 or 5

$\Rightarrow P[E_{in}(h_2) = E_{in}(h_1)]$

$= P(\text{case}1) + P(\text{case }2) + P(\text{case }3)$

$= (2\times\frac{1}{4}\times\frac{3}{4})^5 + (2\times\frac{1}{4}\times\frac{3}{4})^3 (2\times\frac{1}{4})(2\times\frac{1}{4}\times\frac{1}{4}) \frac{5!}{3!1!1!}$

$+ (2\times\frac{1}{4}\times\frac{3}{4})(2\times\frac{1}{4})^2(2\times\frac{1}{4}\times\frac{1}{4})^2 \frac{5!}{1!2!2!}$

$= \frac{3843}{32768}$

13. (b) $\therefore d$.

- For $i = 1, 2, .., d$, $h_i(x) = \text{sign}(x_i)$ • For $i = d+1, .., 2d$, $h_i(x) = -\text{sign}(x_{i-d})$

$$h_1(x) = \text{sign}(x_1) \xleftrightarrow{\text{same BAD data}} h_{d+1}(x) = -\text{sign}(x_1)$$

$$h_2(x) = \text{sign}(x_2) \xleftrightarrow{\text{same BAD data}} h_{d+2}(x) = -\text{sign}(x_2)$$

$$\vdots \qquad\qquad \vdots$$

$$h_d(x) = \text{sign}(x_d) \xleftrightarrow{\text{same BAD data}} h_{2d}(x) = -\text{sign}(x_d)$$

$\Rightarrow$ P[BAD D for H] $\leftarrow$ d ($\because$ union bound)

$\le$ P[BAD D for h_1] + .. + P[BAD D for h_{2d}]

$$\le 2e^{-2\varepsilon^2 N} + \ldots + 2e^{-2\varepsilon^2 N} = d \cdot 2e^{-2\varepsilon^2 N}$$

14. (d) 5 green 4's

$$P(\text{Getting five green 3's}) = \Big[\underbrace{\tfrac{1}{4} \times \tfrac{1}{6} + \tfrac{1}{4} \times \tfrac{1}{6}}_{\text{Each Draw}}\Big]^{5 \leftarrow \text{5 Draws}}$$

In each draw, $\frac{1}{4}$ chance get dice A. In dice A, $\frac{1}{6}$ chance get green 3

$\frac{1}{4}$ chance get dice D. In dice D, $\frac{1}{6}$ chance get green 3

$$P(\text{Five green 4's}) = \left[\tfrac{1}{4} \times \tfrac{1}{6} + \tfrac{1}{4} \times \tfrac{1}{6}\right]^5$$

In each draw, $\frac{1}{4}$ chance get dice A, $\frac{1}{6}$ chance get green 4 in dice A

$\frac{1}{4}$ chance get dice B, $\frac{1}{6}$ chance get green 4 in dice B

15. (C) $\frac{274}{1024}$

A : green : 2, 4, 6

B green : 2, 3, 4

C green : 6

D green : 2, 3, 5

$\Rightarrow$ 4 dices, draw 5 times (with replacement),

the combinations get some number that is purely green

$\rightarrow \left[\binom{4}{1}\right]^5 = 4^5 = 1024$

2 : A, B, D

3 : B, D

4 : A, B

5 : D

$\}$ included in

6 : A, C

$\longrightarrow \left[\binom{3}{1}\right]^5 + \left[\binom{2}{1}\right]^5 - 1 = 3^5 + 2^5 - 1 = 274$

chose 1 from ABD

chose 1 from AC

chose A (repeated)

$\Rightarrow \frac{274}{1024}$