Machine Learning Foundations HW1

B06502152, 許書銓

March 2, 2021

1

The answer is (d)

(a) A lottery is basicly a random process, no logic or pattern. (b) No need ML. (c) No need ML.

2

The answer is (e)

(a) It is basicly a random process. (b) It's a "human-determining" process. (c) No "machine-learning" process include, human-determining. (d) No "machine-learning" process include, human-determining.

3

The answer is unchanged, (d).

From the lecture note, we realize that will be bounded by $\frac{R^2}{\rho^2}$, and the fraction of R^2 and ρ^2 will cancel the X_n^2 term. Hence the scale of X_n^2 term is no function.

4

The answer should be (c). From lecture note of the coverage of PLA, we realize that,

$$\mathbf{w_f^T w_{t+1}} \geq \mathbf{w_f^T w_t} + min\frac{y_n \mathbf{w_f^T x_n}}{||\mathbf{x_n}||}$$

After T times iteration, we get

$$\mathbf{w_f^T w_{t+1}} \geq T min\frac{y_n \mathbf{w_f^T x_n}}{||\mathbf{x_n}||}$$

From the other bound equation

$$||\mathbf{w_{T+1}}||^2 \leq max\frac{||\mathbf{x_n}||^2}{||\mathbf{x_n}||^2}$$

After T times iteration, we get

$$||\mathbf{w_{T+1}}||^2 \leq T$$

From the above equations, we get

$$cos\theta \leq \frac{\mathbf{w_f^T w_t}}{||\mathbf{w_f}|| * ||\mathbf{x_n}||} \leq T * \frac{y_n \mathbf{w_f^T x_n}}{||\mathbf{w_f}|| * ||\mathbf{x_n}||} * \frac{1}{T^{1/2}}$$

Thus, after change the variable into ρ

$$T \leq \frac{1}{\rho^2}$$

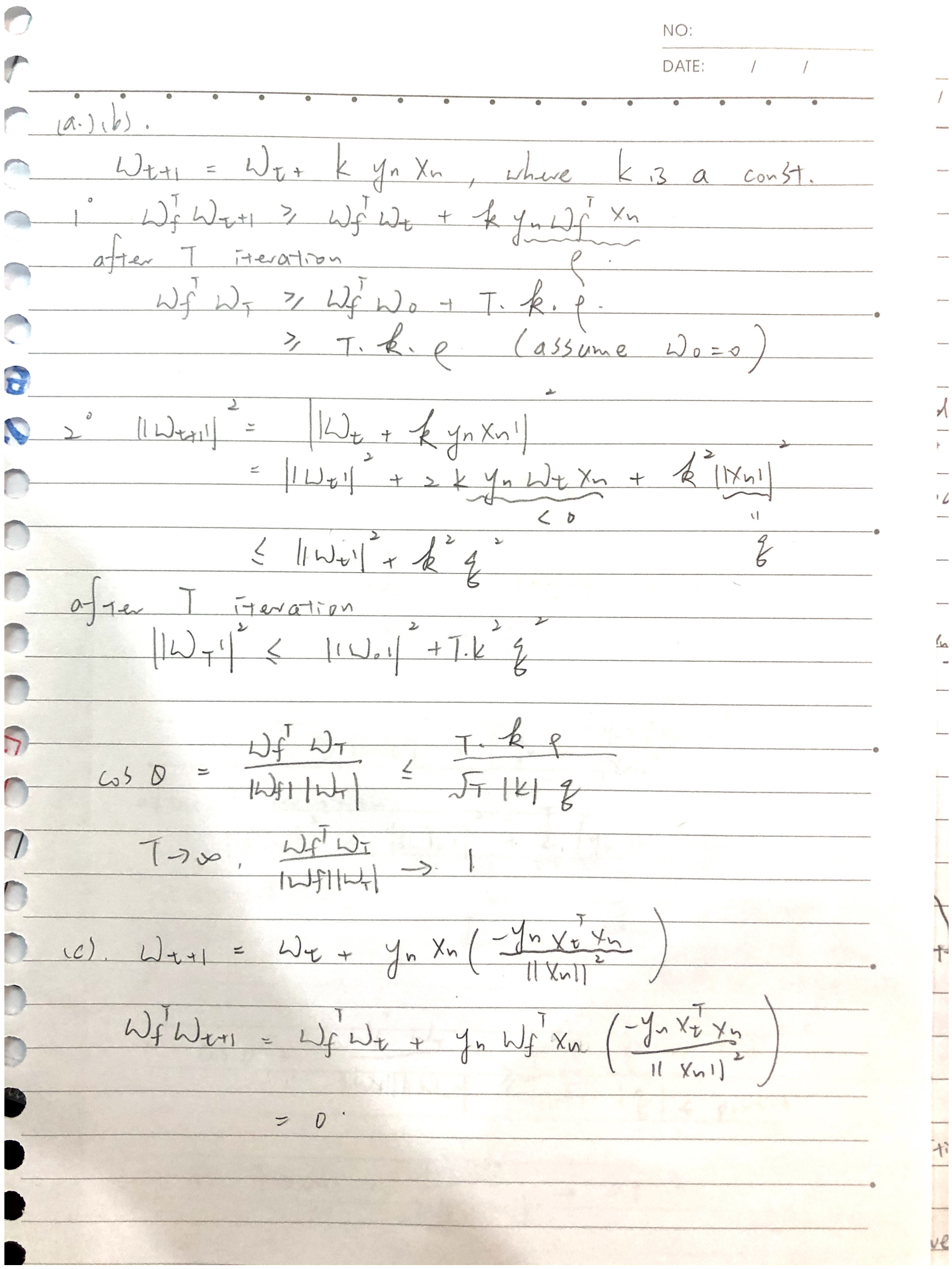
5

The answer should be (d).

(a) If we multiply each side with y_n and x_n, the right hand side will be $y_n w_t x_n + 2||y_n||^2||x_n||^2$. The first term is negative, and the second term is positive; however, we can't identify which one is larger, which means we are not sure that the entire tight hand side will be a positive number. (b) Same reason as (a). (c) The second term in right hand side will be $-y_n w_t x_n$ after multiply with y_n and x_n, which means the entire right hand side will be 0. (d) This condition share some value will (c); however, after plus 1 in the second term, which will leave the right hand side with $||y_n||^2||x_n||^2$. (e) This condition share some value will (c); however, the sign before the second term is "minus", shich means after work with the first term, the first term will not be canceled.

6

The answer should be (c).



NO:

DATE: / /

(a.),(b).

$w_{t+1} = w_t + k\, y_n x_n$, where k is a const.

1° $w_f^T w_{t+1} \geq w_f^T w_t + k \underbrace{y_n w_f^T x_n}_{\rho}$

after T iteration

$w_f^T w_T \geq w_f^T w_0 + T \cdot k \cdot \rho$

$\geq T \cdot k \cdot \rho$ (assume $w_0 = 0$)

2° $\|w_{t+1}\|^2 = \|w_t + k\, y_n x_n\|^2$

$= \|w_t\|^2 + 2k \underbrace{y_n w_t x_n}_{< 0} + k^2 \underbrace{\|x_n\|^2}_{\xi^2}$

$\leq \|w_t\|^2 + k^2 \xi^2$

after T iteration

$\|w_T\|^2 \leq \|w_0\|^2 + T \cdot k^2 \xi^2$

$\cos\theta = \frac{w_f^T w_T}{\|w_f\| \|w_T\|} \leq \frac{T \cdot k\, \rho}{\sqrt{T}\, |k|\, \xi}$

$T \to \infty$, $\frac{w_f^T w_T}{\|w_f\| \|w_T\|} \to 1$

(c). $w_{t+1} = w_t + y_n x_n \left(\frac{-y_n x_t^T x_n}{\|x_n\|^2} \right)$

$w_f^T w_{t+1} = w_f^T w_t + y_n w_f^T x_n \left(\frac{-y_n x_t^T x_n}{\|x_n\|^2} \right)$

$= 0$.

Figure 1: question 6. hand-writing

DATE: / /

(d)

1°. $w_f^T \cdot w_{t+1} = w_f^T w_t + \underbrace{y_n w_f^T \cdot x_n}_{>0} \left[\underbrace{\frac{-y_n w_t^T x_n}{\|x_n\|^2}+1}_{\theta\cdot\theta>0.} \right]$

$$\geq w_f^T w_t + y_n w_f^T x_n \left(\frac{-y_n x_t^T x_n}{\|x_n\|^2}\right)$$

after T iteration

$$w_f^T w_T \geq w_f^T w_0 + \sum_{t=0}^{T} y_n w_f^T x_n \left(\frac{-y_n x_t^T x_n}{\|x_n\|}\right)$$

$\boxed{\text{assume } w_0=0}$, where ρ is the lower bound of the set

$$\geq w_f^T w_0 + T\cdot\rho \geq T\rho.$$

2°. $\|w_{t+1}\|^2 = \left| w_t + y_n x_n \cdot \left[\frac{-y_n w_t x_n}{\|x_n\|^2}+1\right]\right|^2$

$$= \|w_t\|^2 + 2 \underbrace{y_n w_t x_n \left[\frac{-y_n w_t x_n}{\|x_n\|^2}+1\right]}_{<0.} + \left|y_n x_n\left[\frac{-y_n w_t x_n}{\|x_n\|^2}+1\right]\right|^2$$

$$\leq \|w_t\|^2 + \left|y_n x_n \left(\frac{-y_n w_t x_n}{\|x_n\|^2}+1\right)\right|^2$$

after T iteration

$$\|w_T\|^2 \leq \|w_0\|^2 + \sum_{t=0}^{T}\left|y_n x_n\left(\frac{-y_n w_t x_n}{\|x_n\|^2}+1\right)\right|^2$$

where ξ is the upper bound of the set

$$\leq \|w_0\|^2 + T\cdot \xi^2.$$

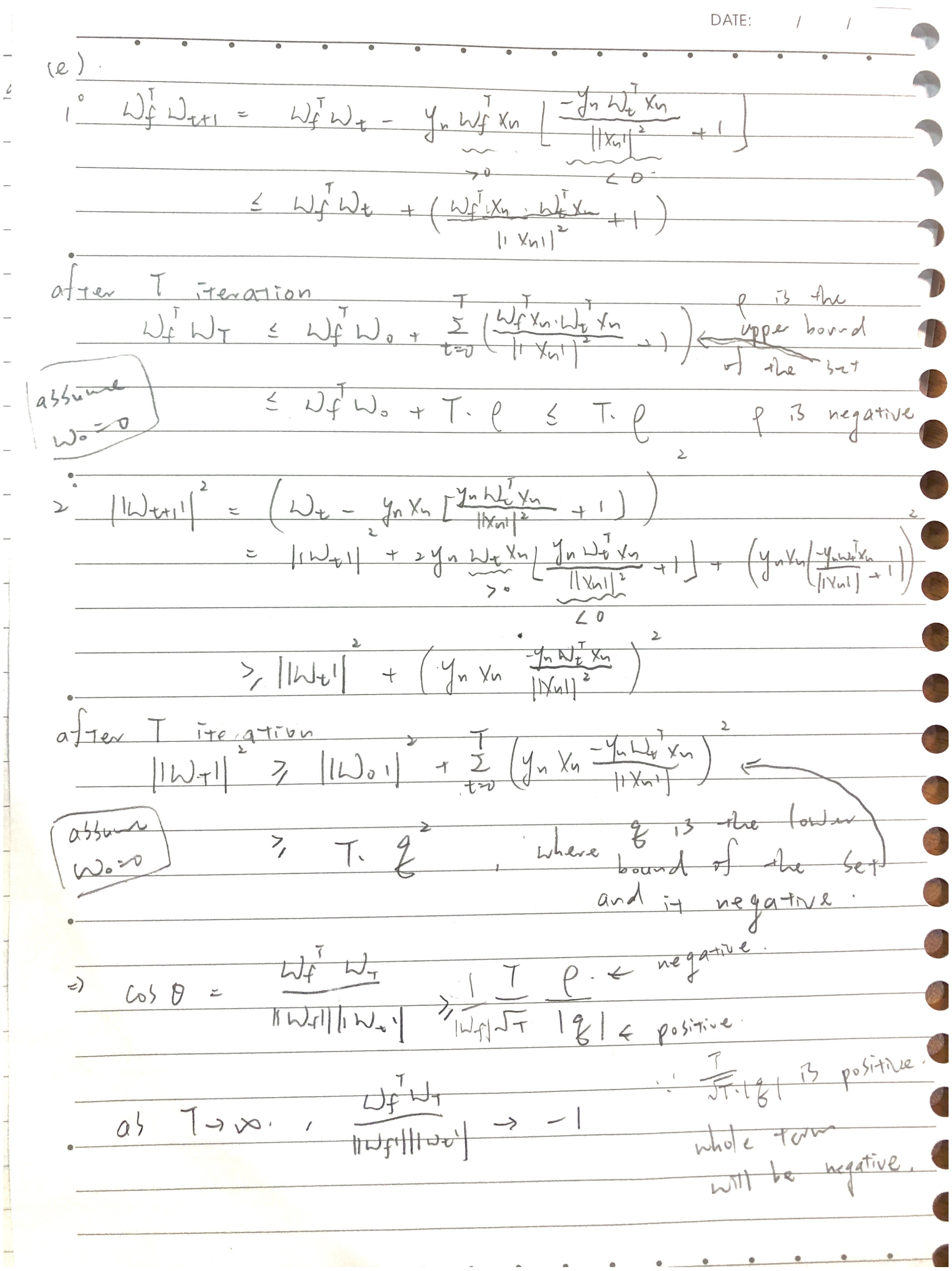
$\boxed{\text{assume } w_0=0}$ $\leq T\cdot\xi^2$

⇒

$$\cos\theta = \frac{w_f^T\cdot w_T}{\|w_f\|\|w_T\|} \leq \frac{T}{\sqrt{T}}\cdot\frac{\rho}{\xi}\cdot\frac{1}{\|w_f\|}$$ ρ and ξ is positive.

as $T\to\infty$, $\frac{w_f^T\cdot w_T}{\|w_f\|\|w_T\|}\to 1.$

Figure 2: question 6. hand-writing



(e)

1° $w_f^T w_{t+1} = w_f^T w_t - \underbrace{y_n w_f^T x_n}_{>0} \left[\underbrace{\frac{-y_n w_t^T x_n}{\|x_n\|^2}}_{<0} + 1 \right]$

$\le w_f^T w_t + \left(\frac{w_f^T x_n \cdot w_t^T x_n}{\|x_n\|^2} + 1 \right)$

after T iteration

$w_f^T w_T \le w_f^T w_0 + \sum_{t=0}^{T} \left(\frac{w_f^T x_n \cdot w_t^T x_n}{\|x_n\|^2} + 1 \right)$ ← ρ is the upper bound of the set

assume $w_0 = 0$

$\le w_f^T w_0 + T \cdot \rho \le T \cdot \rho$ ρ is negative

2° $\|w_{t+1}\|^2 = \left(w_t - y_n x_n \left[\frac{y_n w_t^T x_n}{\|x_n\|^2} + 1 \right] \right)^2$

$= \|w_t\|^2 + 2 \underbrace{y_n w_t x_n}_{>0} \left[\underbrace{\frac{y_n w_t^T x_n}{\|x_n\|^2}}_{<0} + 1 \right] + \left(y_n x_n \left(\frac{-y_n w_t^T x_n}{\|x_n\|} + 1 \right) \right)^2$

$\ge \|w_t\|^2 + \left(y_n x_n \frac{-y_n w_t^T x_n}{\|x_n\|^2} \right)^2$

after T iteration

$\|w_T\|^2 \ge \|w_0\|^2 + \sum_{t=0}^{T} \left(y_n x_n \frac{-y_n w_t^T x_n}{\|x_n\|} \right)^2$ ←

assume $w_0 = 0$

$\ge T \cdot \xi^2$, where ξ is the lower bound of the set and it negative.

$\Rightarrow \cos\theta = \frac{w_f^T w_T}{\|w_f\| \|w_T\|} \ge \frac{1}{\|w_f\|} \frac{T}{\sqrt{T}} \frac{\rho}{|\xi|}$ ← negative ← positive

as $T \to \infty$, $\frac{w_f^T w_T}{\|w_f\| \|w_T\|} \to -1$ $\because \frac{T}{\sqrt{T} \cdot |\xi|}$ is positive. whole term will be negative.

Figure 3: question 6. hand-writing

7

The answer should be (e) since the "judge" environment denotes that the learning process includes some kind of feedbacks.

8

The answer should be (b). Human actes with constrained choices implies multi-class. Another video without human record implies semi-supervised. Learning from all video to obtain a hypothesis implies batch learning. No concrete feature implies raw feature.

9

The answer should be (e) since whenever your $E_{in}(g)$ is small as 0, the posibility that the data outside D is still incorrect at all.

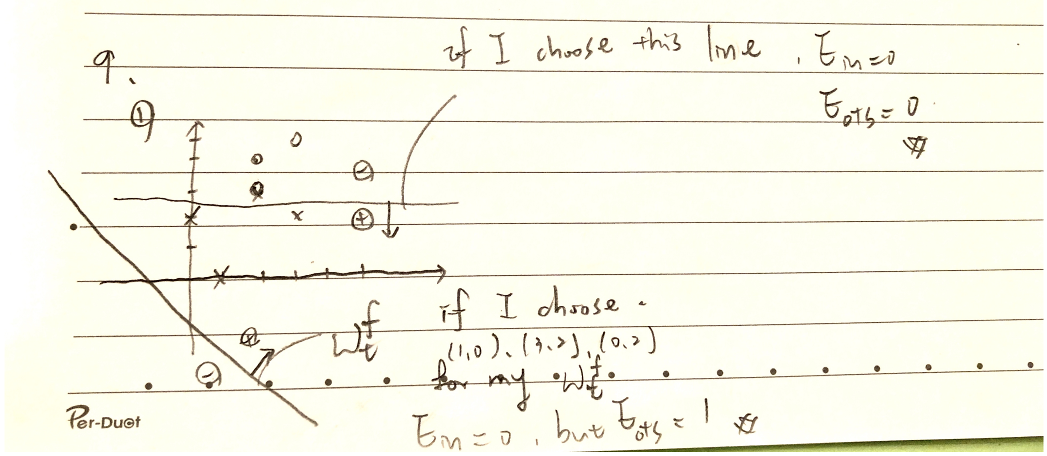


Figure 4: question 9. hand-writing

10

The answer should be (b). Since, the possibility of coming up with the unwanted side is $1-(\frac{1}{2}-\epsilon)$, and the related possibility is $1-(1-\delta)=\delta$. Hence,

$$P[|1-(\frac{1}{2}-\epsilon)|] \leq \delta = 2 * exp(-2*(\epsilon^2)N)$$

$$\ln(\delta/2) \geq -2*(\epsilon^2)N$$

$$N \geq \frac{\ln(2/\delta)}{2\epsilon^2}$$

11

The answer should be (c). Since for each example in sample data, the possibility of $E_{in}(x)=0$ is $\frac{1}{2}$ ($P[sign(x_1) = sign(x_2)] = 1/2$). Thus, the possibility of the whole process is $(\frac{1}{2})^5$

12

The answer should be (d).

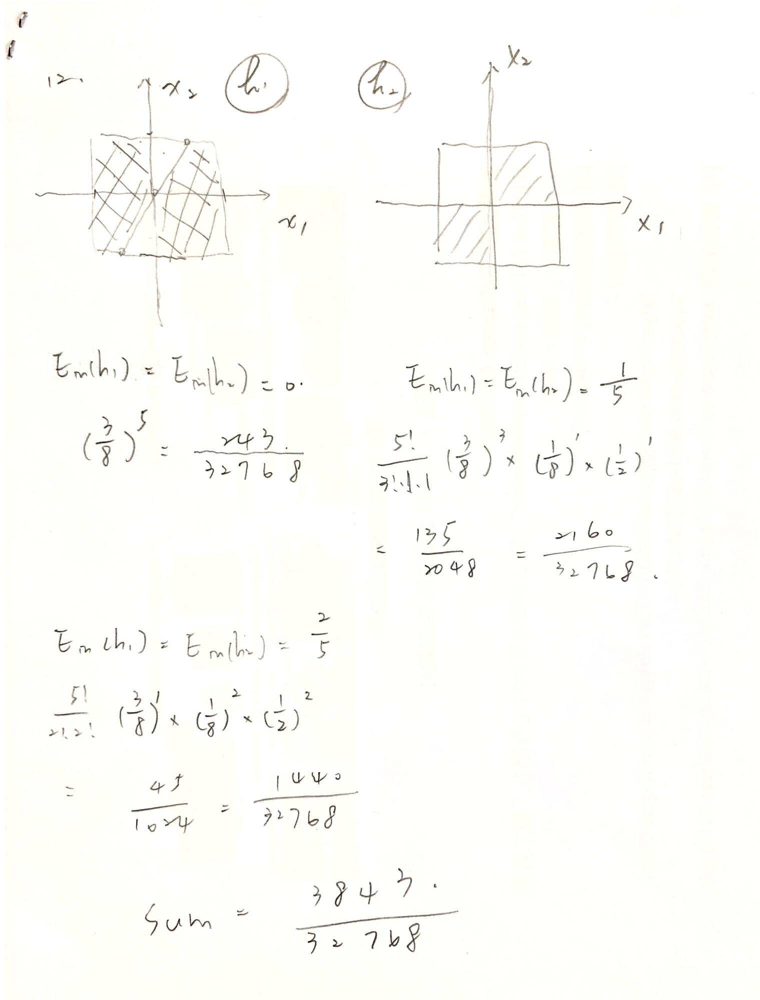


Figure 5: question 12. hand-writing

13

The answer should be (b). Since the bad data for h_1 is exactly the same as h_{d+1}, and so as the relationship between h_2 and h_{d+2} ... to h_d and h_{2d}

14

The answer should be (d). For picking five green 3's, the possibility in each term is $\frac{2}{4}$. The possiblity for each condition is (a) $\frac{1}{4}$ (b) $\frac{1}{4}$ (c) $\frac{3}{4}$ (d) $\frac{2}{4}$ (e) $\frac{1}{4}$

15

The answer should be (d). The possible combination of all green 1's is none. All green 2's is (A,B,D). All green 3's is (B,D). All green 4's is (A,B). All green 5's is (D). All green 6's is (A,C). However, (B,D) combination, (A,B) combination and (D) will be cover in the combination of (A,B,D). All we have to count is all green 2's and all green 6's case, and minus the combination of all (A). Hence, the possiblity is

$$\frac{3^5 + 2^5 - 1}{4^5} = \frac{274}{1024}$$

16

The following my code in python3.6.10.

```
import numpy as np
import random
import py_compile
import matplotlib.pyplot as plt

#main
data = np.genfromtxt("hw1_train.dat.txt")
x_0 = input('input x_0: ')
scale = input('input the value of scaling down the vector: ')

x_0 = int(x_0)
scale = int(scale)

result = []
result_w = []

X = data[:, : -1]
y = data[:, data.shape[1]-1]
X = np.c_[(x_0 * np.ones(X.shape[0]), X)]
X = X / scale

for i in range(1000):
        w = np.zeros((1, 11), np.float)
        t = 0 #count the number of iteration
        correct = 0

        while correct < 5 * X.shape[0]:
                k = random.randint(0, X.shape[0]-1)
                sign = np.sign(np.inner(X[k, :], w))

                if sign*y[k] <= 0:
                        correct = 0
                        w += y[k] * X[k, :]
                        t += 1
                else:
                        correct += 1

        result.append(t)
        result_w.append(w[0][0])

#plt.hist(result)
#plt.show()
print("median of updates: ", np.median(result))
print("median of w_0: ", np.median(result_w))
py_compile.compile("/Users/leo/Desktop/hello/ML/hw1/ml_hw1.py")
```

```
        input x_0: 1
        input the value of scaling down the vector: 1
        median of updates:  11.0
        median of w_0:  -7.0
```

17

Same code in question (16).

18

most close to 14.

```
        input x_0: 10
        input the value of scaling down the vector: 1
        median of updates:  15.0
        median of w_0:  0.0
```

19

```
        input x_0: 0
        input the value of scaling down the vector: 1
        median of updates:  17.0
        median of w_0:  0.0
```

20

```
        input x_0: 0
        input the value of scaling down the vector: 4
        median of updates:  17.0
        median of w_0:  0.0
```