ML HW2

B06303126 Lo Yun Chien

October 2020

Question 1

Correct answer:(c)
With writing it as a matrix A, we want each row is linear independent, by doing row operations. We have

$$A = \begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 7 & 8 & 9 \\ 1 & 15 & 16 & 17 \\ 1 & 21 & 23 & 25 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 6 & 7 & 6 \\ 0 & 14 & 15 & 14 \\ 0 & 20 & 22 & 22 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 6 & 7 & 6 \\ 0 & 0 & \frac{-8}{6} & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

So we know that A is full rank, then it's rows are linear independent. So given all points, $\boldsymbol{w}$ exists to shatter.

Question 2

Correct answer:(d)
Start with 4 points, WLOG, assume $x_1 \leq x_2 \leq x_3 \leq x_4$. The vertical perceptrons can produce 8 dichotomy. For their y values, to maximize the possible dichotomy, we ask them to be arranged differently, or they form the same dichotomy. So, we also form 8 dichotomy, but (OOOO) and (XXXX) are repeated. So the total dichotomy is $8 + 8 - 2 = 14$. Later, for any new point, in x-axis, we have 2n dichotomy. In y-axis, we have 2n different dichotomy, and -2 for repeating. So the growth rate is $4N - 2$.

Question 3

Correct answer:(c)
We want to show that given any three points, the hypothesis set can't shatter it. Assume exists a full rank 3×3 matrix A, such that $A\boldsymbol{w} = \boldsymbol{b}$. Then we have $\boldsymbol{w} = A^{-1}\boldsymbol{b}$. Pick any $\boldsymbol{b} \in R^3$ that no entry is 0. Consider $A^{-1}\boldsymbol{b} = \boldsymbol{w}_1$ and $A^{-1}(-\boldsymbol{b}) = \boldsymbol{w}_2$, then we have $\boldsymbol{w}_2 = -\boldsymbol{w}_1$, but $w_0 > 0$ in both vector, which occurs contradiction. Then A is not full rank.
For any 3×3 matrix A is not full rank, by lecture 7, can't be shattered. But for two points like$(1, 1), (2, 1)$, H_0 can shatter it. So VC bound is 2.

Question 4

Correct answer: (b)
Notice that $x_1^2 + x_2^2 + x_3^2$ is Euclidean norm, it's the same question as positive interval. by lecture 4, the growth function is $\binom{n+1}{2} + 1$.

Question 5

Correct answer: (b)
For k=2, consider point c and d, define $f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2$. We let $f(c) > f(d)$. So $b < f(d)$, $b < f(d) < a$ and $b < f(c)$, $b < f(d) < a$ and $b < f(c) < a$, $f(d) < a$ and $b < f(c) < a$. It construct exactly four dichotomy. So $d_{vc} \geq 2$.
For k = 3, given any three point c, d, e. WLOG, suppose $f(c) \leq f(d) \leq f(e)$. If we want to construct a dichotomy which is $h(c) = h(e) = 1$, $h(d) = -1$. We'll have $a \leq f(c), f(e) \leq b$ but $f(d) < a$ or $f(d) > b$. Which occurs contradiction. So this dichotomy is impossible for all three arbitrary points, $d_{vc} \leq 2$. Then $d_{vc} = 2$.

Question 6

Correct answer: (d)
Consider $E_{in}(g)$, we can control $E_{out}(g) < E_{in}(g) + \epsilon$ by Hoeffding. Since g^* can minimize E_{out}, and combine with $E_{in}(g^*) - E_{out}(g^*) < \epsilon$. So the worst case is $E_{in}(g^*)$ closed to $E_{in}(g)$ but E_{out} fall in total different direction. Which create $E_{out}(g) - E_{out}(g^*) < 2\epsilon$.

So we have

$$\begin{aligned}
\delta &= 4m_H(2N)e^{-\frac{1}{8}\epsilon^2 N} \\
&\Longrightarrow \frac{1}{4m_H(2N)}\delta = e^{-\frac{1}{8}\epsilon^2 N} \\
&\Longrightarrow \ln \frac{\delta}{4m_H(2N)} = -\frac{1}{8}\epsilon^2 N \\
&\Longrightarrow \epsilon = \sqrt{\frac{8}{N} \ln \frac{4m_H(2N)}{\delta}}
\end{aligned}$$

2ϵ means $2\sqrt{\frac{8}{N} \ln \frac{4m_H(2N)}{\delta}}$

Question 7

Correct answer: (d)
For point k, there exists 2^k dichotomy. If the number of hypothesis in H is less this number. Then it definitely can't shatter k points.

Question 8

Correct answer: (d)
For symmetric boolean function, we only need to count how many 1 or -1 in the input. For k

dimensional input, the number of 1 can be chosen from 0 to k. create $k+1$ possibility. So any $k+2$ points can't be shattered.

Question 9

Correct answer: (c)
Recall that d_{vc} means any $k > d_{vc}$ points can't be shattered. And some $k \leq d_{vc}$ points can be shattered. So we have 1, 6, 8 are right.

Question 10

Correct answer: (c)
For any sequence $\{x_n\}$ and each $x_n \in [0, 2\pi]$. If $\frac{\pi}{\alpha} > x_n > 0$, then $\sin \alpha x > 0$. If $\frac{2\pi}{\alpha} > x_n > \frac{\pi}{\alpha}$, then $\sin \alpha x < 0$. Furthermore, it's periodic by $\frac{2\pi}{\alpha}$. Thus, $[0, 2\pi]$ is partitioned into 2α intervals with each width $\frac{\pi}{\alpha}$. With α large enough, there's $+1$ and -1 almost everywhere. So each dichotomy must can be found by increasing α to shorten the size of width. Since it holds $\forall n$, $d_{vc} \to \infty$.

Question 11

Correct answer: (d)
By definition we have

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

For any wrong case $E_{out}(h, 0)$, we have $(1-\tau)$ chance to remain wrong. For any correct case $1 - E_{out}(h, 0)$, we have τ chance to flip it to wrong. Compute the above equation, we get

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

Question 12

Correct answer: (c)
By intuition, three indexes have exactly the same probability to be sampled. so each has probability $\frac{1}{3}$. By the definition of $E_{out}(f)$.

$$\frac{1}{3}[0.1(2-1)^2 + 0.2(3-1)^2 + 0.1(3-2)^2 + 0.2(1-2)^2 + 0.1(1-3)^2 + 0.2(1-3)^2]$$

After computing, the answer is 0.6.

Question 13

Correct answer: (b)
First we derive $f^*(x)$
For x s.t f(x) = 1, $f^*(x) = 0.7 * 1 + 0.1 * 2 + 0.2 * 3 = 1.5$
For x s.t f(x) = 2, $f^*(x) = 0.7 * 2 + 0.1 * 3 + 0.2 * 1 = 1.9$

For x s.t f(x) = 3, $f^*(x) = 0.7 * 3 + 0.1 * 1 + 0.2 * 2 = 2.6$
Now we compute $\Delta(f, f^*)$,

$$\Delta(f, f^*) = \frac{1}{3}(1 - 1.5)^2 + \frac{1}{3}(2 - 1.9)^2 + \frac{1}{3}(3 - 2.6)^2 = 0.14$$

Question 14

Correct answer: (d)
Using the inequality, we have

$$\begin{aligned}
&\delta \geq 4(4N)e^{-\frac{1}{8}\epsilon^2 N} \\
&\Longrightarrow 0.1 \geq 4(4N)e^{-\frac{1}{8}\frac{1}{100}N} \\
&\Longrightarrow \frac{1}{160N} \geq e^{-\frac{1}{8}\frac{1}{100}N} \\
&\Longrightarrow \ln \frac{1}{160N} \geq -\frac{1}{8}\frac{1}{100}N \\
&\Longrightarrow \ln 160N \leq \frac{N}{800}
\end{aligned}$$

To use numeric solution, from option (a) to (e), we find (d) is the minimum among the options.

Question 15

Correct answer: (b)
$h_{1,\theta}(x) = sign(x - \theta)$, $y = f(x) = sign(x)$. If $\theta > 0$, $h_{1,\theta}(x) = f(x)$ iff $x < 0$, or $x > \theta$. If $\theta < 0$, $h_{1,\theta}(x) = f(x)$ iff $x > 0$, or $x < \theta$. So the probability of $h_{1,\theta}(x) \neq f(x)$ is the proportion of $[0, |\theta|]$ and $[-1, 1]$, which is $\frac{1}{2}|\theta|$.

HTML HW2: Conduct decision_stump

Name: 羅允謙
Department: 經濟四
ID: B06303126

Correct answer:
16: (d)
17: (b)
18: (d)
19: (c)
20: (a)

In [75]:

```
import pandas as pd
import numpy as np
import random
```

Define sign function:

sign(x) = 1 iff x > 0

Define h function:

h(s, t, x) = s * sign(x - t)

In [76]:

```
def sign(x: float) -> int:
    if x > 0:
        return 1
    else:
        return -1
```

In [77]:

```
def h(s: int, theta: float, x:float) -> int:
    return s * sign(x - theta)
```

Define decision_stump:

x: 一維向量
n: 樣本量
tau: noise(噪音)

decision_stump(n, tau):

1. Create n samples from uniform[-1, 1], named x
2. With probability tau, create (x, -sign(x)). Otherwise, (x, sign(x))
3. Define Theta as a list contain {-1} and (x*{i}* + x{i+1})/2 for i from 0 to n-2
4. For each element in Theta, compute E_in
5. By question 11, we have E_out(h, tau) = E_out(h, 0) (1 - 2 * tau) + tau
6. By question 15, we have E_out(h, 0) = |theta| / 2
7. Return E_in, E_out as final result.

In [78]:

```
def decision_stump(n: int, tau: float):
    data = []
    for i in range(n):
        x = random.uniform(-1,1)
        t = random.uniform(0,1)
        if t <= tau:
            data.append((x, (-1) * sign(x)))
        else:
            data.append((x, sign(x)))
    Theta = [-1]
    for i in range(n - 1):
        Theta.append((data[i][0] + data[i+1][0]) / 2)

    E_in = {}
    for theta in Theta:
        total_error = 0
        for i in range(n):
            if h(1, theta, data[i][0]) == data[i][1]:
                pass
            else:
                total_error += 1
        avg_error = total_error / n
        E_in[avg_error] = (1, theta)
        total_error = 0
        for i in range(n):
            if h(-1, theta, data[i][0]) == data[i][1]:
                pass
            else:
                total_error += 1
        avg_error = total_error / n
        E_in[avg_error] = (-1, theta)
    E_out = 1 / 2 * abs(E_in[min(E_in)][1])
    E_out = E_out * (1 - 2 * tau) + tau
    return min(E_in), E_out
```

Problem 16: Calculate mean of (E_out - E_in) with n = 2, tau = 0

answer: 0.3

In [79]:

```
df = []
for i in range(10000):
    df.append(decision_stump(2, 0))

answer_df = pd.DataFrame(df)
answer_df.rename(columns = {0: "E_in", 1: "E_out"}, inplace =True)
answer_df["answer"] = answer_df["E_out"] - answer_df["E_in"]
answer_df.mean()
```

Out[79]:

```
E_in      0.000000
E_out     0.293283
answer    0.293283
dtype: float64
```

Problem 17: Calculate mean of (E_out - E_in) with n = 20, tau = 0

answer: 0.03

In [80]:

```
df = []
for i in range(10000):
    df.append(decision_stump(20, 0))

answer_df = pd.DataFrame(df)
answer_df.rename(columns = {0: "E_in", 1: "E_out"}, inplace =True)
answer_df["answer"] = answer_df["E_out"] - answer_df["E_in"]
answer_df.mean()
```

Out[80]:

```
E_in      0.00628
E_out     0.03441
answer    0.02813
dtype: float64
```

Problem 18: Calculate mean of (E_out - E_in) with n = 2, tau = 0.1

answer: 0.3

In [81]:

```
df = []
for i in range(10000):
    df.append(decision_stump(2, 0.1))

answer_df = pd.DataFrame(df)
answer_df.rename(columns = {0: "E_in", 1: "E_out"}, inplace =True)
answer_df["answer"] = answer_df["E_out"] - answer_df["E_in"]
answer_df.mean()
```

Out[81]:

```
E_in      0.000000
E_out     0.345077
answer    0.345077
dtype: float64
```

Problem 19: Calculate mean of (E_out - E_in) with n = 20, tau = 0.1

answer: 0.05

In [82]:

```
df = []
for i in range(10000):
    df.append(decision_stump(20, 0.1))

answer_df = pd.DataFrame(df)
answer_df.rename(columns = {0: "E_in", 1: "E_out"}, inplace =True)
answer_df["answer"] = answer_df["E_out"] - answer_df["E_in"]
answer_df.mean()
```

Out[82]:

```
E_in      0.096105
E_out     0.139563
answer    0.043458
dtype: float64
```

Problem 20: Calculate mean of (E_out - E_in) with n = 200, tau = 0.1

answer: 0.005

In [83]:

```
df = []
for i in range(10000):
    df.append(decision_stump(200, 0.1))

answer_df = pd.DataFrame(df)
answer_df.rename(columns = {0: "E_in", 1: "E_out"}, inplace =True)
answer_df["answer"] = answer_df["E_out"] - answer_df["E_in"]
answer_df.mean()
```

Out[83]:

```
E_in      0.099394
E_out     0.104552
answer    0.005158
dtype: float64
```