

Homework 2 of Machine Learning

B01902032 江東峻

1. Answer: d

真正的錯誤率是（對的可是判斷錯誤 + 錯誤的判斷正確）
 $= \lambda\mu + (1-\lambda)(1-\mu)$

2. Answer: d

λ 不會受到 μ 的影響
 $1-\lambda = \lambda\mu + (1-\lambda)(1-\mu)$

3. Answer: c

$$\sqrt{(8/N * \ln(4 * (2N)^{10}/0.05))} \leq 0.05$$

代入：

$$[b]: 0.0506$$

$$[c]: 0.0496$$

4. Answer: e

$$[a]: \epsilon \leq 0.6322$$

$$[b]: \epsilon \leq 0.8921$$

$$[c]: \epsilon \leq 0.3313$$

$$[d]: \epsilon \leq 0.2237$$

$$[e]: \epsilon \leq 0.2152$$

5. Answer: d

$$[a]: \epsilon \leq 4.25$$

$$[b]: \epsilon \leq 4.25$$

$$[c]: \epsilon \leq 2.81$$

$$[d]: \epsilon \leq 1.74$$

$$[e]: \epsilon \leq 2.27$$

6. Answer: a

先選一個當起點 s ，再選一個當終點 f ，若 f 在 s 的右邊，則 $s \sim f$ 這段是正，否則就是負。
這可以選到所有可能， $s \sim f$ 這段是負的可能在 s, f 順序顛倒的情況會發生，再加上兩個特例：
 s, f 是同一個點的時候，可能是全負或全正。

$$P(N, 2) + 2 = n(n-1) + 2 = [a]$$

7. Answer: b

$$mH(2) = 4 = 2^2$$

$$mH(3) = 8 = 2^3$$

$$mH(4) = 14 < 2^4$$

$$d^{vc} = 3$$

8. Answer: c

平面上有 N 個點，若將 N 個點離原點的距離投影到數線上，再依照大小分成 $a_1 \sim a_n$ ，就變成 positive interval 的問題，是 $C(N+1, 2) + 1$ (a, b 在同一個區間的情形)

9. Answer: b

$$d_{vc} \geq D+1:$$

$\sum c_i x_i$ 是一個 D 次方的多項式，若值 >0 則 h_c 為正，否則為負。那取任意 D 個 x ，每個 x_i, x_{i+1} 的sign不一樣的話，就代表這之中有一個根，則最多有 D 個根（包含全部都一樣），則用根可以構造出方程式 $\pm(x-a_1)(x-a_2)\dots=0$ 滿足任意的sign組合。

$$d_{vc} \leq D+1:$$

若 $d_{vc} = D+2$ ，如果 $x_1, x_2, x_3 \dots$ 是正負交錯的，則有 $D+1$ 個根，但是 D 次方的方程式只能最多有 D 個根，所以 $D+2$ 不能shatter

10. Answer: a

$$d_{VC} \geq 2^d:$$

假設在各個象限都有一個點 x ，共有 2^d 個點(d 個象限)，則對於任何一組 h ，設定 $t=(0,0,0,\dots,0)$
 S =所有可能組合($|S|=2^d$)，若 $h(x_i)$ 為-1，就把 S 裡與 x_i 同一個的象限的組合拔掉
 \Rightarrow 可以shatter

$$d_{VC} \leq 2^d:$$

若多一個點，則一定有某個點 x' 與其中一個 x 同個象限，則無法找到一個 t 把這兩個分開(分開的話就有另一個點與 x' 一起被 t 分在同一區)，則無法作出符合 $h(x) \neq h(x')$ 異號的情況。
 \Rightarrow 不能shatter

11. Answer: d

N 個 x 時可以被shatter：取 $x_1=2^{2-\epsilon}$, $x_2=2^{3-\epsilon}$, $x_3=2^{4-\epsilon}$, ..., $x_n=2^{(n+1)-\epsilon}$ 。 ϵ 為極小正數。
 α 從 $0/2^N \sim (2^N-1)/2^N$ 裡取出來，若要求的是 $y_1, \dots, y_n \in \{0,1\}$ (0為負,1為正)，則將 y 照順序組
成一個 N -bits的2進位數字，再將每個bit 0,1互換，再轉成10進位數字 A ，則把 $A/2^N$ 當作 α 。
Ex: 若 $N=4$ ，則 $x_1 \sim x_4 = 4, 8, 16, 32$ ， α 從 $0/16 \sim 15/16$ 中取，若要求的是0, 1, 0, 1，
那先把bit顛倒成1010，10進位表示是10，那取 $\alpha = 10/16$ ，

for $x_1 = 4-\epsilon$, $h = -1$ 為負(0)

for $x_2 = 8-\epsilon$, $h = 1$ 為正(1)

for $x_3 = 16-\epsilon$, $h = -1$ 為負(0)

for $x_4 = 32-\epsilon$, $h = 1$ 為正(1)

符合所要求的0, 1, 0, 1。

而對所有的 N 都可以用同樣的方法取值， N 有無限多個。

12. Answer: a c d

[a]: [a]為將 N 個變數分成兩半，分別的可能組合，而相乘表示將兩邊的組合一一結合，但是
有可能經過組合後，產生不行的情況。例如positive rays，在兩個變數的情況下，有3種組合
： xx , $x0$, 00 ，在一個變數的情況下，有0, x ，但是結合後 $x0x$ 就變成不行的情況需要扣掉，
所以分段相乘的組合數會大於實際上組合後長度的組合數(需要扣掉組合後不行的情況)。

[b]: 為常數，若 N 夠大就bound不了。

[c]: $2^m mH(N-1) \geq mH(N)$ ，因為多了一個變數可以放 $0x$ (2倍)，再扣掉不行的情況(ex: positive
rays 可以是 $x0$ ，但是 $x0x$ 就不行。) 推回 $N-1, N-2, \dots$ 可以知道 $2^m mH(N-1)$ 是[c]裡最小的，可
還是可以bound住 $mH(N)$ 。

$$[d] mH(N) \leq B(N, K) \leq \sum_{(i=0 \sim k-1)} C(N-1, i) \leq N^{k-1} = N^{d_{VC}} \leq N^{d_{VC}+1}$$

[e] 若[e]是upper bound，則有 $mH(N-1) + N \cdot d_{VC} \geq mH(N)$ ，用連加方式消去，可以得到：

$$mH(2) + \sum_{(i=3 \sim N)} d_{VC} \geq mH(N) \Rightarrow \sum_{(i=3 \sim N)} d_{VC} \text{ 為2次式，則在某些情況會不是upper bound}$$

ex: 4D-perceptron, $mH(N) = O(N^4)$ ，當 N 夠大時會bound不了。

13. Answer: a

[a]: ex: convex sets

$$[b]: N=1, mH(N) = 2 = 2^1$$

$$N=2, mH(N) = 2 < 2^2 \text{ break point!!}$$

因為有break point \Rightarrow 被 $\text{poly}(N)=N^k$ bound

但是當 N 夠大時， N^k 就bound不了 $\Rightarrow mH(N)$ 不存在。

[c]: 同[b]，break point = 1 ($2^0 < 2^1$)，但無法被 $\text{poly}(N)$ bound $\Rightarrow mH(N)$ 不存在。

[d]: 同[b]，break point = 2 ($2^1 < 2^2$)，但無法被 $\text{poly}(N)$ bound $\Rightarrow mH(N)$ 不存在。

[e]: break point = 2 ($mH(2) = 3 < 2^2$)，若存在 \Rightarrow 可被 $\text{poly}(N)=N^2$ bound住，但是

$[e]=O(N^3)$ ，無法被 N^2 bound $\Rightarrow mH(N)$ 不存在。

14. Answer: b

[b]交集一定被最小的集合包含住，也就是說，交集可以shatter的，最小的集合一定也可以shatter，但最小的集合可以shatter的，交集不一定可以shatter。最小的集合的VC dimension會大於交集的VC dimension。

[a][c]的bound都沒有[b]緊，而[d][e]的大小關係是錯誤的。

15. Answer: e

同14題，聯集一定包含住最大的集合，所以聯集的VC dimension會大於最大的集合的VC dimension。

對於 $dvc(\bigcup_{k=1}^K H_k) \leq \sum_{k=1}^K dvc(H_k)$ ，取反例：

若A有 $\{(0, 0, 0), (0, 0, x), (0, x, 0), (x, 0, 0)\}$ $dvc(A) = 1$

若B有 $\{(x, x, x), (x, 0, x), (0, x, x), (x, x, 0)\}$ $dvc(B) = 1$

但是 $dvc(A \cup B) = 3 \neq 1 + 1$ 。

對於 $dvc(\bigcup_{k=1}^K H_k) \leq K - 1 + \sum_{k=1}^K dvc(H_k)$ ，

右側可以看成是 $k=1 \sim K$ 的最小break point相加再-1，

因為 $k=1 \sim K$ 都是break point都不能shatter，那聯集也不能shatter，

所以聯集的break point $\leq \sum_{k=1}^K \text{break point}(H_k)$ ，

而聯集的dvc一定小於等於聯集的break point-1。（“=”:break point是最小的break point）

16. Answer: c

當 $s=1$ ，將 $[-1, 1]$ 分成三段，沒有 θ 的那側(以原點分)、原點 $\sim \theta$ 、 θ 到另一端 $[-1 \text{ or } 1]$ ，其中除了原點 $\sim \theta$ 這段noise不將結果改變的話算錯以外，其他都是有noise改變結果時算是錯。

可得： $E_{out} = 0.5 * [(1-|\theta|) * 20\% + |\theta| * 80\% + 1 * 20\%] = 0.2 + 0.3|\theta| = 0.5 - 0.3 + 0.3|\theta|$

同理， $s=-1$ 時，分成三段，但是算錯的方式顛倒，只有原點 $\sim \theta$ 這段需要noise改變。

可得： $E_{out} = 0.5 * [(1-|\theta|) * 80\% + |\theta| * 20\% + 1 * 80\%] = 0.8 - 0.3|\theta| = 0.5 + 0.3 - 0.3|\theta|$

綜合以上，加入 s 可得 $E_{out} = 0.5 + 0.3s(|\theta| - 1)$

17. Answer: b

18. Answer: c

19. Answer: c

20. Answer: d

21.

22.