Homework 1 of Machine Learning

B01902032 江東峻

1. answer: a
   (i) 和 (iii) 有明確的定義。
   (v) 只要搜集足夠多的資料，也許可以找出最適合的年紀，或根本沒有關係。 沒有關係。

2. answer: b
   銷售數字上升或下降不一定代表推薦的好壞，但是銷售得好對於賣書來說算是一件好事，所以算是一種獎勵。

3. answer: c
   用對老鼠實驗時，有很多變因被控制，對於某個要素去找出結果，因為老鼠數量不多，所以不能使用大量藥物來獲取大量的資料，所以是策略性的學習。

4. answer: d
   沒有給定主題，只有分成幾堆。

5. answer: a
   用以往的資料判斷有某些特質的人在借貸超過多少錢的狀況會不還錢，再把這些資料餵給演算法。

6. answer: d
   假設4種情況帶入$E_{OTS}(g,f)$：
   N, L all even, $E_{OTS}=1/2$
   N even, L odd, $E_{OTS}=(L-1)/2L$
   N odd, L even, $E_{OTS}=1/2$
   N, L all odd, $E_{OTS}=(L+1)/2L$
   將4種情況帶入(a)(b)(c)(d)式，選擇結果與上述相符的。

7. answer: c
   存在另一個f’ 可以”generate” D, 這個f 在X_1~X_N裡都符合$f'(X)=y_n= f(X)$, 但在X_{N+1}~X_{N+L}裡就沒有限制，所以有2^L種可能。

8. answer:a c e

(a)(b):E_{OTS}=k/L 意思是在X_{N+1}~X_{N+L}裡，有k個X使得f(X)≠g(X)，有C(L,k)種組合，在X_1~X_N裡，f(X)為noiseless，所以只有1種組合，f共有C(L,k)種組合。

(c)(d):E_{OTS} (g, f) 的期望值，X_1~X_N都不影響E_{OTS} (期望值=0),在X_{N+1}~X_{N+L}裡，每個X都有1/2的機率使得f(X)≠g(X),有L個X,f(X)≠g(X)的期望值 = L/2, EOTS的期望值 = 1/2。

(e): A_1產生g_1, A_2產生g_2, 對於每個 g 來說，因為 g 有任何可能，所以f(X)≠g_1(X)和f(X)≠g_2(X)的機率是一樣的，所以E_{OTS} (g, f)的期望值都一樣。

9. answer:a

因為u = 0.5, v = u的狀況發生在取樣取到剛好5個的狀況，
C(10, 5)*(1/2)^10 = 0.246

10. answer: b

因為u = 0.9, v = u的狀況發生在取樣取到剛好9個的狀況，
C(10, 9)*(0.9)^9 * (0.1)^1 = 0.387

11. answer: d

因為u = 0.9, v ≤ 0.1的狀況發生在取樣取到1個或0個的狀況，
C(10, 1)*(0.9)^1 * (0.1)^9 + C(10, 0)*(0.9)^0 * (0.1)^10 = 9.1E-9

12. answer: b

以ε = 0.8, N = 10代入 Hoeffding's Inequality
得 P(v ≤ 0.1) ≤ 5.521E-6, (0.8+0.9 ≤ v 的狀況不存在)

13. answer: b

4種骰子中，只有2種有橘色的1，所以每次拿到有橘色1的骰子的機率為1/2, 而骰子數量很多，所以每次拿到橘色1的機率都一樣，
為(1/2)^5=1/32

14. answer: c

其中要有一個數字是同色的狀況只有4種組合，
(A,C) (B,C) (A,D) (B,D)
(4 (4種只有單一種骰子的狀況) + C(5,1)+C(5,2)+C(5,3)+C(5,4) (同時有兩種骰子的狀況))
* (1/4 (每個骰子有4種))^5 = 124/1024 = 31/256

15. answer: c
16. answer: c
17. answer: c
18. answer: a
19. answer: b
20. answer: a

21. answer: No

根據lecture 2的第15, 16頁可知道：

A. linear separable D => $\exists$ w_f such that $\forall$ $x_n \in X$, $y_n=\text{sign}(w_f^T x_n)$

=> $y_n(t) w_f^T x_n(t) \geq \min y_n w_f^T x_n > 0$

=> $w_f^T w_{t+1} \geq w_f^T w_t + \min y_n w_f^T x_n > w_f^T w_t + 0$

B. w_t changed only when mistake

=>sign $(w_t^T x_n(t)) \neq y_n(t) \Leftrightarrow y_n(t) w_t^T x_n(t) \leq 0$

=>$\|w_{t+1}\|^2 \leq \|w_t\|^2 + \max \|x_n\|^2$

用連加消去以及$w_0 = 0$ 的條件可以得出：

C. $w_f^T w_T \geq T * \min y_n w_f^T x_n > 0$ (linear separable D)

D. 0(絕對值) $\leq \|w_T\|^2 \leq T * \max \|x_n\|^2$ => $0 \leq \|w_T\| \leq \sqrt{T} * \max \|x_n\|$

C/D 可得(當D不為0的狀況)：

$w_f^T w_T / \|w_T\| \geq \sqrt{T} * \min y_n w_f^T x_n / \max \|x_n\|$

(根據C, $w_T = 0$ => $\min y_n w_f^T x_n = 0$ 且$w_f^T w_T = 0$, 符合上式 “=“ 發生的情況)

同除 $\|w_f\|$ (D不為0):

1(向量內積) $\geq w_f^T w_T / (\|w_T\| * \|w_f\|) \geq \sqrt{T} * \min y_n w_f^T x_n / (\max \|x_n\| * \|w_f\|)$

又 $R = \max \|x_n\|$, $\rho = \min y_n w_f^T x_n / \|w_f\|$

$T \leq R^2 / \rho^2$

當x_n變為1/10時，R, ρ都變為1/10

=> T 的upper bound不變

實驗：

拿Problem 15的測資來跑，輸入測資後再將測資/10，結果總共更新了777次，是變慢。（閾值的影響力變大）