MLFoundation HW1

r07922100 資工碩一 楊力權

1.

2.

因為E_{OTS}是在x_{N+1}到x_{N+L}中g(x)與f(x)的相同函數值的數量再除上L，而在任何x下f(x)=1，在k是奇數條件下$g(x_k)=1$，因此E_{OTS}是x_{N+1}到x_{N+L}中的奇數個數除以L，而x_{N+1}到x_{N+L}中的奇數個數也等於的x_1到x_{N+L}中的奇數個數減上x_1到x_N中的奇數個數。

x_1到x_{N+L}中的奇數個數 $= \lfloor \frac{N+L}{2} \rfloor$

x_1到x_N中的奇數個數 $= \lfloor \frac{N}{2} \rfloor$

因此$E_{OTS} = \frac{1}{L}(\lfloor \frac{N+L}{2} \rfloor - \lfloor \frac{N}{2} \rfloor)$

3.

對於N+1~N+L的test input而言， f 有2^L種不同形式，因此每種 f 出現的機率都是$\frac{1}{2^L}$，每個 f 對$A_1(D)$生出的g_1來說 ，有$\binom{L}{i}$個 f 與 g_1 產生 i 個error。

$E_f\{E_{OTS}(A_1(D), f)\} = \frac{1}{2^L}\sum_{\forall f} E_{OTS}(A_1(D), f) = \frac{1}{2^L}(\frac{1}{L}\sum_{i=1}^{L} i\binom{L}{i})$，且

$(1+x)^L = \sum_{i=0}^{L}\binom{L}{i}x^i => \frac{d}{dx}(1+x)^L = L(1+x)^{L-1} = \sum_{i=0}^{L}\binom{L}{i}ix^{i-1} = \sum_{i=1}^{L}\binom{L}{i}ix^{i-1}$

把x帶1，得$L(2)^{L-1} = \sum_{i=1}^{L} i\binom{L}{i}$

把$\sum_{i=1}^{L} i\binom{L}{i} = L(2)^{L-1}$帶回$E_f\{E_{OTS}(A_1(D), f)\} = \frac{1}{2^L}(\frac{1}{L}\sum_{i=1}^{L} i\binom{L}{i}) = \frac{1}{2^L}(\frac{1}{L}(L2^{L-1})) = \frac{1}{2}$

而對A2(D)，也有$\binom{L}{i}$個 f 與 g_2 產生 i 個error

因此$E_f\{E_{OTS}(A_2(D), f)\} = \frac{1}{2^L}(\frac{1}{L}\sum_{i=1}^{L} i\binom{L}{i}) = \frac{1}{2}$

故可得$E_f\{E_{OTS}(A_1(D), f)\} = \frac{1}{2} = E_f\{E_{OTS}(A_2(D), f)\}$

4.

$$P(v \le 0.1) = P(v = 0.1) + P(v = 0) = \binom{10}{1}(0.8)(0.2)^9 + \binom{10}{0}(0.2)^{10} = 4.1984 \times 10^{-6}$$

$$P(v \ge 0.9) = P(v = 0.9) + P(v = 1) = \binom{10}{9}(0.8)^9(0.2) + \binom{10}{10}(0.8)^{10} = 0.3758$$

5.

從ABCD中只要全挑AD就可以得到全部綠色的1，因此$P(five\ green\ 1s) = (\frac{2}{4})^5 = \frac{1}{32}$

6.

討論各點數要全綠的條件如下

全綠的1：AD 全綠的2：BD 全綠的3：AD 全綠的4：BC 全綠的5：AC 全綠的6：BC

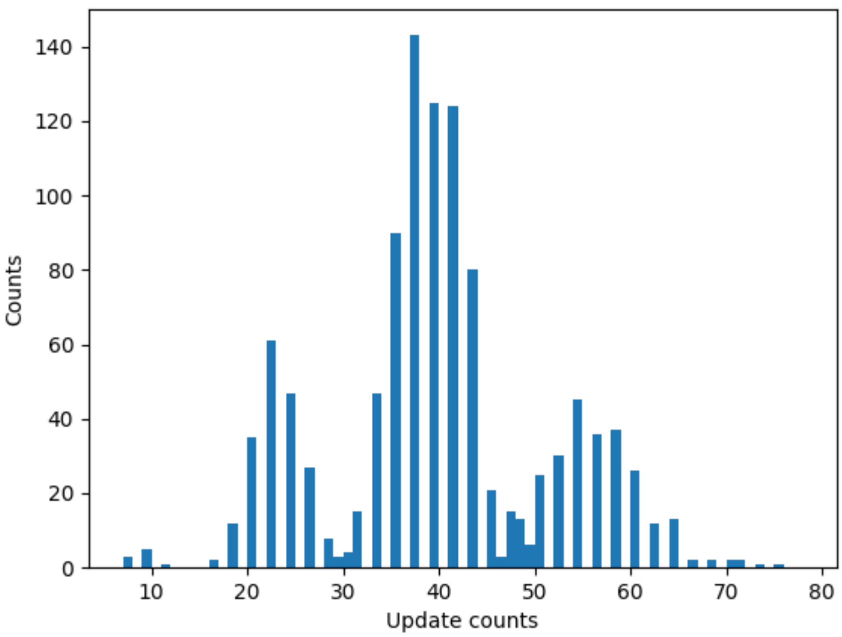
因此若要至少一個數字全綠，則要滿足只有AD or AC or BD or BC，但還要扣去重複計算的部分，如AC與AD中重複計算了四個A的機率。

$P(some\ number\ is\ purely\ green)$

$= P(only\ AD) + P(only\ AC) + P(only\ BC) + P(only\ BD) - P(only\ A) - P(only\ B) - P(only\ C) - P(only\ D)$

$= (\frac{1}{2})^5 \times 4 - (\frac{1}{4})^5 \times 4 = \frac{31}{256}$，即使骰子有6個面，但滿足條件僅只四種可能AD BD AD BC，且有重複的情況全A全B全C全D。

7.



8.

pf. M exists

如果data是linear separable，且在$y_{n(t)}w_tx_{n(t)} < 0$執行一次$w_{t+1} \leftarrow w_t + y_{n(t)}x_{n(t)}$未必能夠使$y_{n(t)}w_{t+1}^Tx_{n(t)} > 0$，可能要執行M次的更新，才能保證$y_{n(t)}w_{t+1}^Tx_{n(t)} > 0$，因此應該存在某M能夠保證$w_{t+1} \leftarrow w_t + y_{n(t)}x_{n(t)} \times M$能使$y_{n(t)}w_{t+1}^Tx_{n(t)} > 0$。

$y_{n(t)}w_{t+1}^Tx_{n(t)} > 0$且$w_{t+1} \leftarrow w_t + y_{n(t)}x_{n(t)} \times M$ => $y_{n(t)}w_{t+1}^Tx_{n(t)} = y_{n(t)}(w_t + y_{n(t)}x_{n(t)} \times M)^Tx_{n(t)}$

$= y_{n(t)}w_t^Tx_{n(t)} + My_{n(t)}^2x_{n(t)}^Tx_{n(t)} > 0 \Rightarrow M > \dfrac{-y_{n(t)}w_t^Tx_{n(t)}}{y_{n(t)}^2x_{n(t)}^Tx_{n(t)}} = \dfrac{-w_t^Tx_{n(t)}}{y_{n(t)}x_{n(t)}^Tx_{n(t)}}$

$M > \dfrac{-w_t^Tx_{n(t)}}{y_{n(t)}x_{n(t)}^Tx_{n(t)}}$能保證存在最小M使 $w_{t+1} \leftarrow w_t + y_{n(t)}x_{n(t)} \times M$能滿足$y_{n(t)}w_{t+1}^Tx_{n(t)} > 0$

pf. the new update rule will still ensure halting with a perfect line

$\|w_{t+1}\|^2 = \|w_t + y_{n(t)}x_{n(t)}M\|^2 = \|w_t\|^2 + 2y_{n(t)}w_t^Tx_{n(t)}M + \|y_{n(t)}x_{n(t)}M\|^2 \le \|w_t\|^2 + 0 + \|y_{n(t)}x_{n(t)}M\|^2$

$\le \|w_t\|^2 + \max_n\|x_nM\|^2$ (y_n是1, -1因此不影響長度可拿掉)

$\because \|w_{t+1}\|^2 \le \|w_t\|^2 + M^2\max_n\|y_nx_n\|^2 \quad \therefore \|w_T\|^2 \le \|w_0\|^2 + TM^2\max_n\|y_nx_n\|^2 \le TM^2\max_n\|y_nx_n\|^2$

得$\|w_T\| \le \sqrt{T}M\max_n\|x_n\|$且對真實函數 f 的單位內積最大為1：$\dfrac{w_f^T}{\|w_f\|}\dfrac{w_T}{\|w_T\|} \le 1$

$w_f^Tw_T = w_f^T(w_{T-1} + y_{n(T-1)}x_{n(T-1)}M) = w_f^T(w_0 + \sum_{t=0}^{T-1}y_{n(t)}x_{n(t)}M) \ge w_f^Tw_0 + TMw_f^T\min_n y_{n(t)}x_{n(t)} \simeq TMw_f^T\min_n y_nx_n$

$1 \ge \dfrac{w_f^T}{\|w_f\|}\dfrac{w_T}{\|w_T\|} \ge \dfrac{w_f^T}{\|w_f\|}\dfrac{w_T}{\sqrt{T}M\max_n\|x_n\|} \ge \dfrac{TMw_f^T\min_n y_nx_n}{\|w_f\|\sqrt{T}M\max_n\|x_n\|} = \sqrt{T}\dfrac{\min_n(y_nw_f^Tx_n)}{\|w_f\|max_n\|x_n\|}$

$\because y_nw_f^Tx_n \ge 0 \quad \therefore T \le (\dfrac{\|w_f\|max_n\|x_n\|}{\min_n(y_nw_f^Tx_n)})^2$

所以在data是線性可分的條件下，T有上限，因此終究會停在一個線上，並能將data完美分割。