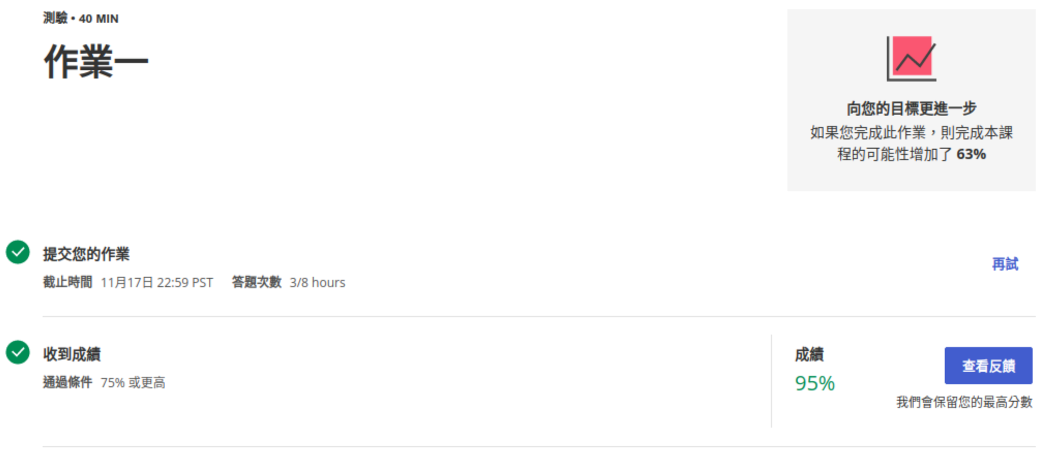
Student ID: R08942062
name: 林志皓

Problem 1



Problem 2

We can labeled some of the news from different media companies with labels such as “left or right on political spectrum”, “for or against particular issues/politicians” …, of course we are not able to label all of them. With these labeled data, we can classified other news without labels, furthermore discover the political position of each media, and report to the publics. (Taiwan definitely need this.)

Problem 3

③ $\because$ All f that can generate D are equally likely,

$$P\left(E_{OTS}(A(D), f) = \frac{l}{L}\right) = \frac{C^L_l}{2^L}$$

$$\therefore E_f\left\{E_{OTS}(A(D), f)\right\}$$

$$= \sum_{l=0}^{L} P\left(E_{OTS}(A(D), f) = \frac{l}{L}\right) \cdot \frac{l}{L}$$

$$= \sum_{l=1}^{L} \left(\frac{C^L_l}{2^L} \cdot \frac{l}{L}\right) = \frac{1}{2^L} \sum_{l=1}^{L} \left(C^L_l \cdot \frac{l}{L}\right)$$

$$= \frac{1}{2^L} \sum_{l=1}^{L} \left(\frac{L(L-1)\cdots(L-l+1)}{l!} \cdot \frac{l}{L}\right)$$

$$= \frac{1}{2^L} \sum_{l=1}^{L} \frac{(L-1)(L-2)\cdots(L-l+1)}{(l-1)!}$$

$$= \frac{1}{2^L} \sum_{l=1}^{L} C^{L-1}_{l-1} = \frac{1}{2^L} \cdot 2^{L-1} = \frac{1}{2}$$

$\therefore E_f\left\{E_{OTS}(A(D), f)\right\} = \frac{1}{2}$ (costant) regardless of A

#

Problem 4, Problem 5

④

green 1 → dice A or dice C

$P(\text{green } 1 \text{ for each dice}) = \frac{2}{4} = \frac{1}{2}$.

$P(\text{green } 1 \text{ for all dices}) = (\frac{1}{2})^5 = \frac{1}{32}$ #

⑤

one green number:

for number 1~6. prob. are same $= (\frac{1}{2})^5 = \frac{1}{32}$. $\Rightarrow$ sum $= \frac{6}{32}$

two green number:

let the prob of green x & y = $p(x, y)$

$p(1,2) = (\frac{1}{4})^5$ $\quad p(1,6) = 0$ $\quad p(2,6) = (\frac{1}{4})^5$ $\quad p(4,5) = (\frac{1}{4})^5$

$p(1,3) = (\frac{1}{2})^5$ $\quad p(2,3) = (\frac{1}{4})^5$ $\quad p(3,4) = 0$ $\quad p(4,6) = (\frac{1}{2})^5$

$p(1,4) = 0$ $\quad p(2,4) = (\frac{1}{4})^5$ $\quad p(3,5) = (\frac{1}{4})^5$ $\quad p(5,6) = (\frac{1}{4})^5$

$p(1,5) = (\frac{1}{4})^5$ $\quad p(2,5) = 0$ $\quad p(3,6) = 0$

sum $= (\frac{1}{4})^5 \cdot 8 + (\frac{1}{2})^5 \cdot 2 = \frac{8}{1024} + \frac{1}{16} = \frac{9}{128}$

three green number

Let prob of green x & y & z = $p(x, y, z)$

$p(1,2,3) = (\frac{1}{4})^5$ $\quad p(2,4,6) = (\frac{1}{4})^5$

$p(1,3,5) = (\frac{1}{4})^5$ $\quad p(4,5,6) = (\frac{1}{4})^5$

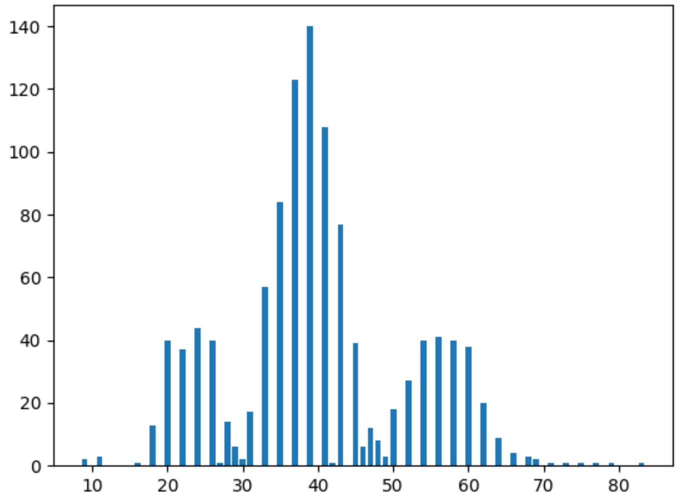
sum $= (\frac{1}{4})^5 \cdot 4 = \frac{1}{256}$

$\therefore$ P(all green for "some number")

$= \frac{6}{32} - \frac{9}{128} + \frac{1}{256} = \frac{48-18+1}{256} = \frac{31}{256}$ #

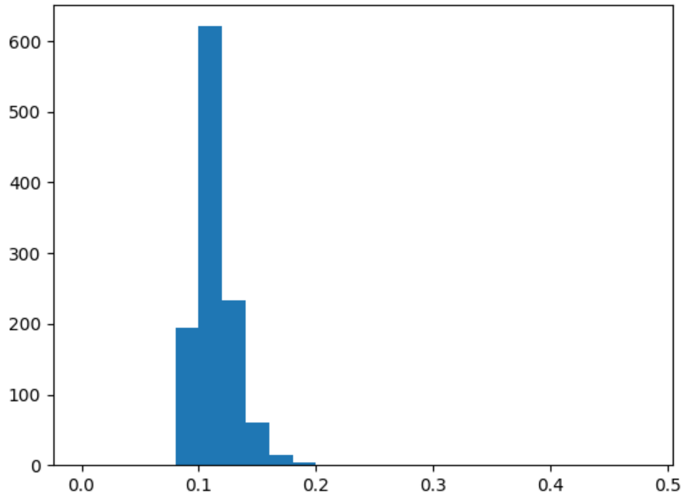
Problem 6

Average number of updates: 40.12
histogram:



Problem 7

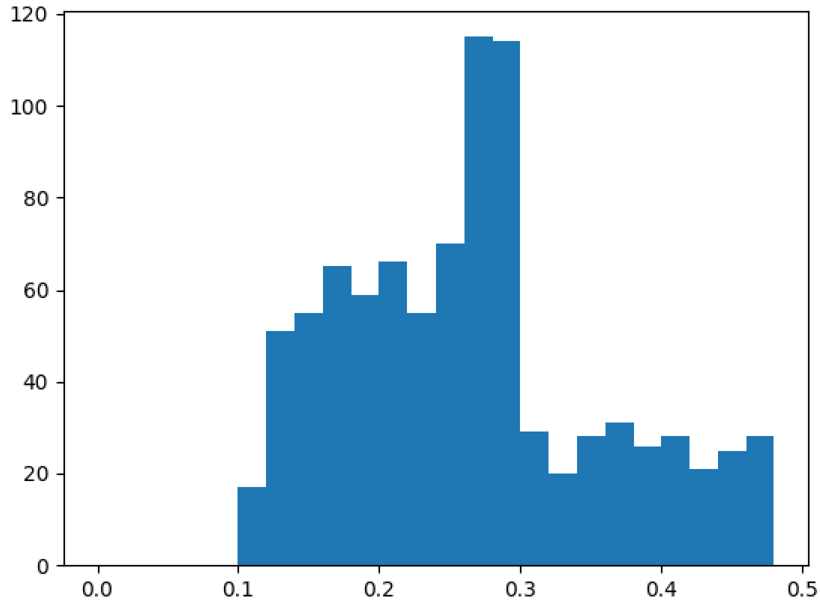
Average error rate: 0.1138
histogram:



Problem 8

Average error rate: 0.328

histogram:



Compared with the result of problem 7, the average error rate is larger, because w_pocket holds the "best" w during training, while w_t is not guaranteed to be better than previous ones, so the error rate on test set are larger.

Problem 9

9

$$\frac{w_f}{\|w_f\|} w_T = \frac{w_f}{\|w_f\|}\left(w_{T-1} + y_{n(T-1)} x_{n(T-1)}\right)$$

$$= \frac{w_f}{\|w_f\|} w_{T-1} + \frac{w_f}{\|w_f\|} y_{n(T-1)} x_{n(T-1)}$$

$$\ge \frac{w_f}{\|w_f\|} w_{T-1} + \min_n \frac{w_f}{\|w_f\|} y_n x_n \ge T \cdot \min_n \frac{w_f}{\|w_f\|} y_n x_n$$

$$\|w_T\|^2 = \|w_{T-1} + y_{n(T-1)} x_{n(T-1)}\|^2 = \|w_{T-1}\|^2 + 2 w_{T-1} y_{n(T-1)} x_{n(T-1)} + \|y_{n(T-1)} x_{n(T-1)}\|^2$$

$$\le \|w_{T-1}\|^2 + \max_n \|x_n\|^2 \le T \cdot \max_n \|x_n\|^2$$

$$\|w_T\| \le \sqrt{T} \cdot \max_n \|x_n\| \qquad \frac{1}{\|w_T\|} \ge \frac{1}{\sqrt{T}} \cdot \frac{1}{\max_n \|x_n\|}$$

$$\frac{w_f}{\|w_f\|} \frac{w_T}{\|w_T\|} \ge \sqrt{T} \cdot \frac{\min_n \frac{w_f}{\|w_f\|} y_n x_n}{\max_n \|x_n\|} \qquad T \le \left(\frac{w_f}{\|w_f\|} \cdot \frac{w_T}{\|w_T\|} \cdot \frac{\max_n \|x_n\|}{\min_n \frac{w_f}{\|w_f\|} y_n x_n}\right)^2$$

$$\le \left(\frac{\max \|x_n\|}{\min_n \frac{w_f}{\|w_f\|} y_n x_n}\right)^2 = UB$$

scale down all x_n linearly won't change the value of UB ⇒ plan won't work!