Problem 1

1. Assume $P[+1|x] \sim U[0,1]$

Consider $y \neq f_{CIA}(x)$ 的 events

I. false negtive, $y=+1$、$f_{CIA}(x)=-1$

II. false positive, $y=-1$、$f_{CIA}(x)=+1$

∵ 題目說明「false positive is 1000 times more important than a false negtive」

相當於 false negtive 發生的次數是 false positive 發生的次數 1000 倍以上

∴ 令抽樣的次數為 N

$$\cancel{N} * P[\text{I}] \geq 1000 * \cancel{N} * P[\text{II}]$$

又 event I 發生在 $P(H|X) \le \alpha$

且 event II 發生在 $P(H|X) > \alpha$ 時

$\Rightarrow$ $P[P(H|X) \le \alpha] \geq 1000 * P[P(H|X) > \alpha]$

$\because$ $P(H|X)$ 符合 uniform distribution

$$\alpha \geq 1000(1-\alpha)$$

$$1001\alpha \geq 1000$$

$$\alpha \geq \frac{1000}{1001}$$

$$\alpha = \frac{1000}{1001}$$ #

Problem 2

2\. 原本 environment

令 O correct 機 : $(1-E_{out})$

X incorrect 機 : E_{out}

新 environment

令 O' correct 機 : $1-E_{(x,y)\sim p(x,y)}[[g(x)\neq y]]$

X' incorrect 機 : $E_{(x,y)\sim p(x,y)}[[g(x)\neq y]]$

∵ $P(y=+f(x)|x)=1-\varepsilon,\ P(y=-f(x)|x)=\varepsilon$

A $P(y=+f(x)|x)$ 發生在 $O\to O'$, $X\to X'$; $P(y=-f(x)|x)$ 發生在 $O\to X'$, $X\to O'$

$E_{(x,y)\sim p(x,y)}[[g(x)\neq y]]$

$= \underbrace{\varepsilon(1-E_{out})}_{O\to X'} + \underbrace{(1-\varepsilon)E_{out}}_{X\to X'}$

$= \varepsilon - 2\varepsilon E_{out} + E_{out}$ #

Problem 3

3.

$$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}(wx_n - y_n)^2$$

$$= \frac{1}{N}\sum_{n=1}^{N}(w^2x_n^2 - 2wx_ny_n + y_n^2)$$

∵ square error function 是一凸函数

∴ 最小值在 $\nabla E_{in}(w) = 0$

$$\nabla E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}(2wx_n^2 - 2x_ny_n)$$

$$= \frac{2}{N}\sum_{n=1}^{N} x_n(wx_n - y_n)$$

assume $\forall x_i \neq 0,\ 1 \leq i \leq N$

$$\Rightarrow \sum_{n=1}^{N} wx_n - y_n = 0 \quad \Rightarrow \quad w_{lin} = \frac{y_1 + \cdots + y_n}{x_1 + \cdots + x_n}$$

Problem 4

4. $\min_{w} E_{in}$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(h(x)-f(x)\right)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left((w_0+w_1x)-ax^2-b\right)^2$$

令 $c = ax^2-b$, $X = [1\ x]$, $W = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(2W^TX^TX \cdot W + c^2 - 2X^TWc\right)$$

$$\nabla E_{in} = \frac{2}{N}\sum_{i=1}^{N}\left(x^Tx\,W - x^Tc\right)$$

as $W = (X^TX)^{-1}X^Tc$, $\nabla E_{in} = 0$

$W = (X^TX)^{-1}X^Tc$

$$= \frac{ax^2-b}{1+x^2}\begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$= \begin{bmatrix} \frac{ax^2-b}{1+x^2} \\ \frac{ax^3-bx}{1+x^2} \end{bmatrix} = \begin{bmatrix} W_0^* \\ W_1^* \end{bmatrix}$$ ✗

Problem 5

5. $Y' = aY + b$

$$= aXW_{Lin} + b$$

$$= a\begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & x_{nn} \end{bmatrix}\begin{bmatrix} w_0 \\ \vdots \\ w_n \end{bmatrix} + a\begin{bmatrix} \frac{b}{a} \\ \vdots \\ \frac{b}{a} \end{bmatrix}$$

$$= a\begin{bmatrix} w_0 & w_1x_{11} & \cdots & w_nx_{1n} \\ w_0 & & & \vdots \\ \vdots & & & \\ w_0 & w_1x_{n1} & \cdots & w_nx_{nn} \end{bmatrix} + a\begin{bmatrix} \frac{b}{a} \\ \vdots \\ \frac{b}{a} \end{bmatrix}$$

$$= a\begin{bmatrix} w_0+\frac{b}{a} & w_1x_{11} & \cdots & w_nx_{1n} \\ \vdots & & & \vdots \\ w_0+\frac{b}{a} & w_1x_{n1} & \cdots & w_nx_{nn} \end{bmatrix}$$

$$= a \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{nn} \end{bmatrix} \begin{bmatrix} w_0 + \frac{b}{a} \\ w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ \vdots & & & \\ 1 & x_{n1} & \cdots & x_{n1} \end{bmatrix} \begin{bmatrix} a(w_0 + \frac{b}{a}) \\ aw_1 \\ \vdots \\ aw_n \end{bmatrix}$$

$$= X \, W'_{Lin}$$

$$W'_{Lin} = \begin{bmatrix} a(w_0 + \frac{b}{a}) \\ aw_1 \\ \vdots \\ aw_n \end{bmatrix}$$

$$= a W_{Lin} + \begin{bmatrix} b \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Problem 6

Problem 7

7.

initialize w_0

For $t = 0, 1, \cdots$

① compute $\nabla E_{in}(w_t)$

② update by $w_{t+1} = w_t + \eta \nabla E_{in}(w_t)$

... until $\nabla E_{in}(w_t) = 0$ or enough iterations

return last w_{t+1} as g

derive $\nabla E_{in}(w_t)$ ...

$$E_{in}(w_t) = \frac{1}{N}\sum_{n=1}^{N}\left(\max(0, 1 - y_n w_t^T x_n)\right)^2$$

$$= \begin{cases} \frac{1}{N}\sum_{n=1}^{N}(1 - y_n w_t^T x_n)^2 & \text{if } y_n w_t^T x_n \le 1 \\ 0 & \text{else} \end{cases}$$

$$= \begin{cases} \frac{1}{N}\sum_{n=1}^{N}(1 + y_n^2 w_t^{T^2} x_n^2 - 2 y_n w_t^T x_n), & \text{if } y_n w_t^T x_n \le 1 \\ 0, & \text{else} \end{cases}$$

$$\nabla E_{in}(w_t) = \begin{cases} \frac{1}{N}\sum_{n=1}^{N}\left(2y_n^2 w_t x_n^2 - 2y_n x_n\right), & \text{if } y_n w_t^T x_n \le 1 \\ 0 & \text{, else} \end{cases}$$

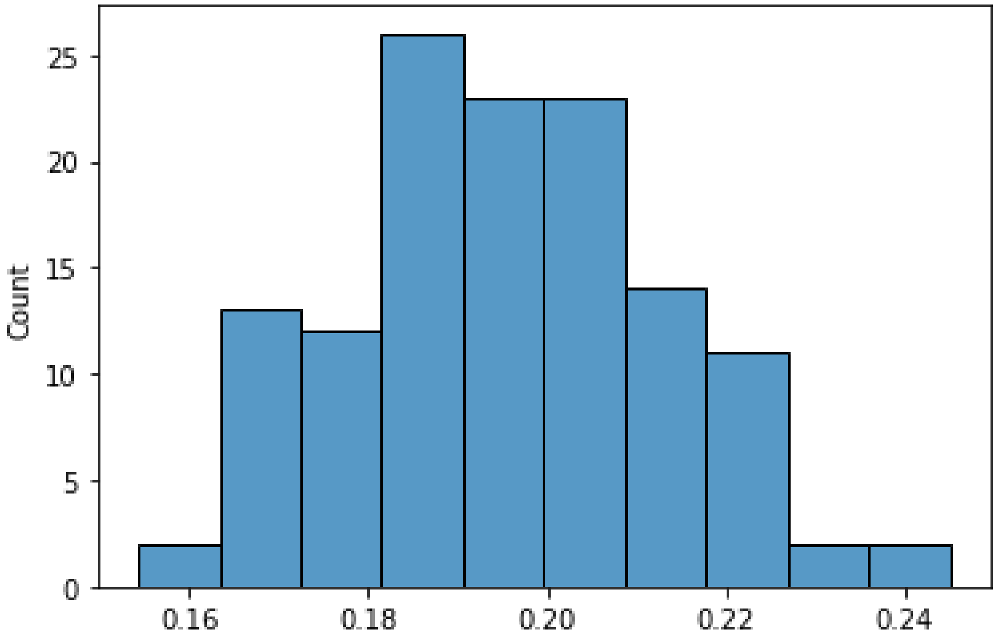
$$= \begin{cases} \frac{1}{N}\sum_{n=1}^{N}\left(2yx(w_t x - 1)\right), & \text{if } y_n w_t^T x_n \le 1 \\ 0 & \text{, else} \end{cases}$$

相同之處在於當 $ys \geq 1$, error 皆為0，而不同之處在於，此种方法能允許資料無 linear seperable，並且在 $0 \le y_n w_t^T x_n \le 1$ 時會進行更行 反觀 PLA 不會。

Problem 8

Problem 9

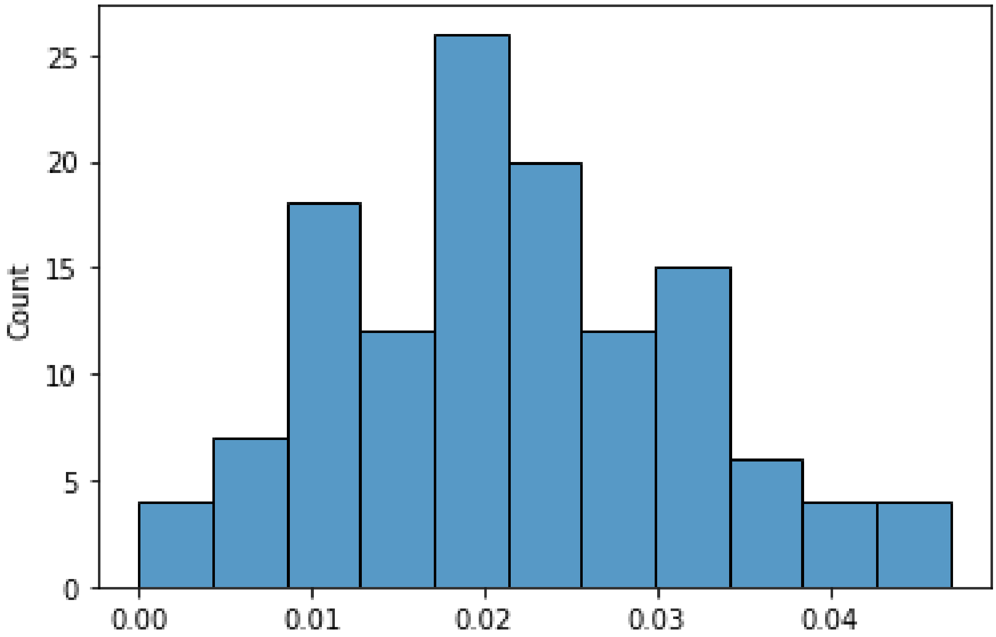
The distribution of $E_{in}^{sqr}(W_{LIN})$:



Median $E_{in}^{sqr}(W_{LIN})$ over 128 experiments : 0.1960

Problem 10

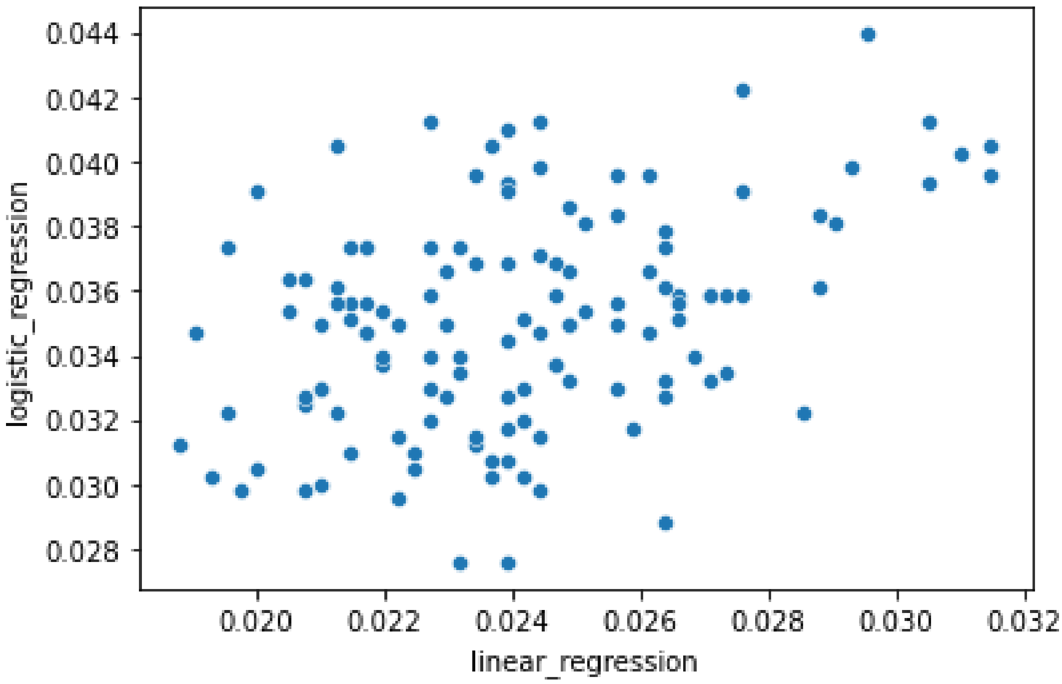
The distribution of $E_{in}^{0/1}(W_{LIN})$:



Median $E_{in}^{0/1}(W_{LIN})$ over 128 experiments : 0.0195

Problem 11

Scatter plot for $[E_{out}^{0/1}(A(D)), E_{out}^{0/1}(B(D))]$:

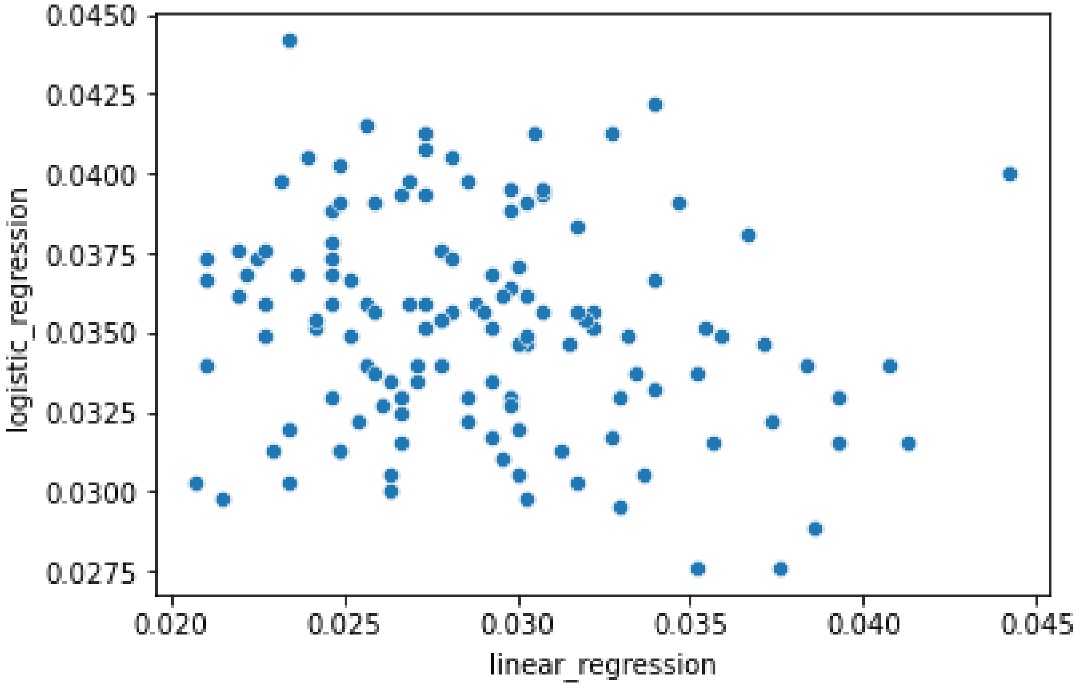


Median $E_{out}^{0/1}(A(D))$ over 128 experiments : 0.0239

Median $E_{out}^{0/1}(B(D))$ over 128 experiments : 0.0351

Problem 12

Scatter plot for $[E_{out}^{0/1}(A(D)), E_{out}^{0/1}(B(D))]$:



Median $E_{out}^{0/1}(A(D'))$ over 128 experiments : 0.0283

Median $E_{out}^{0/1}(B(D'))$ over 128 experiments : 0.0351

My finding : 結果可以看出加入 noise 前後，linear regression 的 $E_{out}^{0/1}$ 差距較大，推測是因為 linear regression 會盡量擬和這些 noise 導致$E_{out}^{0/1}$ 加大，反觀 logistic regression 會往 $E_{out}^{0/1}$ 最小的地方更新，較不受資料影響。