Machine Learning 2023 HW3 R12631055 林東甫

1

2

1.

$$f_{0/1}(x) = \arg\max_{y\in\{-1,+1\}} P(y|x) = \text{sign}\left(P(+1|x) - \tfrac{1}{2}\right)$$

Since +1, -1 are Binary : $\text{sign}\left(P(+1|x) - \frac{1}{2}\right) = \begin{cases} +1 & \text{if } P(+1|x) > \frac{1}{2} \\ -1 & \text{if } P(+1|x) < \frac{1}{2} \end{cases}$

that is $f(x) = \arg\max\limits_{y\in\{-1,+1\}} P(y|x)$, where $err(\tilde{y}, y) = [\![\tilde{y} \neq y]\!]$

Known CIA's false positive, false negative :

$$\begin{array}{c|c|cc} & & g & \\ & & +1 & -1 \\ \hline f & +1 & 0 & 1 \\ & -1 & 1000 & 0 \end{array}$$

$E_{in}(g)$ for err_{CIA} : $\frac{1}{N}\left(\sum_{y_n=+1}[\![y_n \neq g(x_n)]\!] + 1000\cdot\sum_{y_n=-1}[\![y_n \neq g(x_n)]\!]\right)$

$\because f_{\text{origin}}(x) = \begin{cases} +1 & \text{if } P(+1|x) > \frac{1}{2} \\ -1 & \text{if } P(-1|x) < \frac{1}{2} \end{cases}$ where $\begin{array}{c|c|cc} & & g & \\ & & +1 & -1 \\ \hline f & +1 & 0 & 1 \\ & -1 & 1 & 0 \end{array}$

$\therefore f_{CIA}(x) = \begin{cases} +1 & \text{if } P(+1|x) > \frac{1000}{1+1000} \\ -1 & \text{if } P(-1|x) < \frac{1000}{1+1000} \end{cases}$ which is

$$f_{CIA}(x) = \text{sign}\left(P(+1|x) - \frac{1000}{1001}\right) \text{ where } \alpha = \frac{1000}{1001}$$

\#

2.

Known $E_{out}(g) = \mathcal{E}_{x\sim P(x)}[\![g(x) \neq f(x)]\!]$, $P(y|x) : \begin{cases} 1-\epsilon & \text{if } y = +f(x) \\ \epsilon & \text{if } y = -f(x) \end{cases}$

$err(\tilde{y}, y) = [\![\tilde{y} \neq y]\!]$

$\tilde{y}\begin{cases} f(x), & \text{avg. err } \epsilon \\ -f(x), & \text{avg. err } (1-\epsilon) \end{cases}$ $\quad \mathop{E}_{(x,y)\sim P}[\![g(x) \neq y]\!] = \begin{cases} \epsilon & \text{if } y = +f(x) \\ 1-\epsilon & \text{if } y = -f(x) \end{cases}$

3. Known $E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}(h(x_n)-y_n)^2 \quad \forall x_n, y_n, \; n=1,2,3,\cdots,N$

Since $h(x) = wx \;\forall x \in R$ where $h \in \mathcal{H}$

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

$$\nabla E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}(2x_n^2 w - 2y_n x_n) \quad \text{(Assume all denominators} \neq 0\text{)}$$

$$= \frac{2}{N}\left(w\sum_{n=1}^{N}x_n^2 - \sum_{n=1}^{N}y_n x_n\right) \qquad \min_w E_{in}(w) \Leftrightarrow \nabla E_{in}(w) = 0$$

$$\therefore w_{LIN} = \frac{\sum_{n=1}^{N} y_n x_n}{\sum_{n=1}^{N} x_n^2} = \sum_{n=1}^{N}\frac{y_n}{x_n}$$

#

4. Known $f(x) = ax^2 + b$, $h(x) = w_0 + w_1 x \;\; h \in \mathcal{H}$

$$E_{out}(w) = \mathcal{E}_{(x,y)\sim p}(h(x)-f(x))^2 = \int_0^1 (w_0 + w_1 x - ax^2 - b)^2 dx$$

$$= \int_0^1 (w_0^2 + w_0 w_1 x - w_0 a x^2 - w_0 b + w_0 w_1 x + w_1^2 x^2 - w_1 a x^3 - b w_1 x$$

$$- w_0 a x^2 - w_1 a x^3 + a^2 x^4 + bax^2 - w_0 b - w_1 b x + bax^2 + b^2) dx$$

$$= \int_0^1 (a^2 x^4 - 2w_1 a x^3 + (w_1^2 - 2w_0 a + 2ba)x^2 + (2w_0 w_1 - 2bw_1)x + (w_0^2 - 2w_0 b + b^2)) dx$$

$$= \left(\frac{a^2}{5}x^5 - \frac{2w_1 a}{4}x^4 + \frac{(w_1^2 - 2w_0 a + 2ba)}{3}x^3 + (w_0 w_1 - bw_1)x^2 + (w_0^2 - 2w_0 b + b^2)x\right)\Big|_{x=0}^{1}$$

$$= \frac{1}{5}a^2 - \frac{a}{2}w_1 + \frac{1}{3}w_1^2 - \frac{2a}{3}w_0 + \frac{2}{3}ba + w_0w_1 - bw_1 + (w_0 - b)^2$$

$$\cancel{\partial w_0 : \frac{2}{3}w_0 - \frac{2a}{3} + w_1 + 2w_0 - 2b = \frac{8}{3}w_0 + w_1 - 2(b + \frac{a}{3})}$$

$$\therefore \partial w_1 : \frac{a}{2} + w_0 - b + \frac{2}{3}w_1 = \frac{2}{3}w_1 + w_0 + \frac{a}{2} - b = 0 \text{ — ①}$$

$$\partial w_0 : \frac{-2a}{3} + w_1 + 2w_0 - 2b = w_1 + 2w_0 - \frac{2a}{3} - 2b = 0 \text{ — ②}$$

$$\Rightarrow ①\times 2 - ② : \frac{1}{3}w_1 + \frac{5a}{3} = 0, \quad w_1 = -5a$$

$$② - ①\times\frac{3}{2} : \frac{1}{2}w_0 - \frac{3a}{4} - \frac{2a}{3} - \frac{1}{2}b = 0, \quad w_0 = \frac{17}{6}a + b$$

$$(w_0^*, w_1^*) = \left((\tfrac{17}{6}a + b), (-5a + 0b)\right) \text{ or } \left((\tfrac{17}{6}, 1), (-5, 0)\right)$$

#

5. Assume X^TX is invertible, known $y_n' = ay_n + b$

$$E_{in}(w') = \frac{1}{N}\sum_{n=1}^{N}(w'^Tx_n - y_n')^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n^Tw' - (ay_n + b))^2$$

$$= \frac{1}{N}\left\|\begin{bmatrix} \cdots x_1^T \cdots \\ \vdots \\ \cdots x_N^T \cdots \end{bmatrix} w' - \begin{bmatrix} ay_1 + b \\ \vdots \\ ay_N + b \end{bmatrix}\right\|^2 = \frac{1}{N}\| Xw' - (ay + b)\|^2$$

$$= \frac{1}{N}\left(w'^TX^TXw' - 2w'^TX^T(ay+b) + (ay+b)^T(ay+b)\right) \therefore \nabla E_{in}(w') = \frac{2}{N}(X^TXw - X^T(ay+$$

$$w_{LIN} = (X^TX)^{-1}X^T(ay+b) = X^\dagger(ay+b)$$

#

6. Known $h_t(x) = \dfrac{1}{1+\exp(W_t^T x)}$. $E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \ln(1+\exp(-y_n W^T x_n))$

Since $\nabla E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} \left(\frac{\partial \ln(\square)}{\partial \square}\right)\left(\frac{\partial(1+\exp(\circ))}{\partial \circ}\right)\left(\frac{\partial -y_n W^T x_n}{\partial w_i}\right)$

$= \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{\square}\right)(\exp(\circ))(-y_n x_{n,i}) = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\exp(\circ)}{1+\exp(\circ)}\right)(-y_n x_{n,i})$

where $\square = 1+\exp(-y_n W^T x_n)$ $\circ = -y_n W^T x_n$ $\therefore \frac{1}{N}\sum_{n=1}^{N}\theta(-y_n W^T x_n)(-y_n x_n)$

$A_E(W_t) = \nabla(\nabla E_{in}(w)) = \nabla\left(\frac{1}{N}\sum_{n=1}^{N} h_t(-y_n x_n)(y_n x_n)\right)$

$= \nabla\left(\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{1+e^{-y_n W_t^T x_n}}\right)(-y_n x_n)\right) = \frac{1}{N}\sum_{n=1}^{N} -\left(\frac{-e^{-y_n W_t^T x_n}}{(1+e^{-y_n W_t^T x_n})^2}\right)\cdot x_n x_n^T$

$= \frac{1}{N}\sum_{n=1}^{N} \frac{1}{\frac{1}{e^{-y_n W_t^T x_n}}+1}\cdot\frac{1}{1+e^{-y_n W_t^T x_n}}\cdot x_n x_n^T$

$= \frac{1}{N}\sum_{n=1}^{N} \frac{1}{e^{y_n W_t^T x_n}+1}\cdot h_t(-y_n x_n)\cdot x_n x_n^T$

$= \frac{1}{N}\sum_{n=1}^{N} h_t(y_n x_n)\cdot h_t(-y_n x_n)\cdot x_n x_n^T$

#

7

7. Origin PLA : $W_{t+1} \leftarrow W_t + 1 \cdot [\![y_n \neq \text{sign}(W_t^T x_n)]\!] (y_n x_n)$

PLA applying SGD on the truncated squared loss $err(s,y) = (\max(0, 1-ys))^2$

$\nabla err(s,y) = 2 * (\max(0, 1-ys)) \nabla (\max(0, 1-ys))$, $s = W^T x$

$$= \begin{cases} 2(\max(0, 1-\cancel{ys} \; y_n W_t^T x_n))(-y_n x_n) & \text{if } 1-y > 0 \\ 0 & \text{if } 1-ys < 0 \end{cases}$$

(same ~~with~~ as $2(\max(0, 1-y_n W_t^T x_n))(-y_n x_n)$ when $1-ys < 0$)

$W_{t+1} \longleftarrow W_t + \eta \cdot 2(\max(0, 1-y_n W_t^T x_n)) \cdot (-y_n x_n)$

$\Rightarrow W_{t+1} \longleftarrow W_t + [\![y_n \neq \text{sign}(\max(0, 1-W_t^T x_n))]\!] (-y_n x_n)$

這兩者都同樣由正負號來判斷,並都使用 0/1 ~~loss~~ error.
但一個部份可微,因此可以用 SGD,而另一個則不行,在更新 W_t 時也很相似,

8. $E_{in}(W) = \frac{1}{N}\sum_{n=1}^{N} err(W, x_n, y_n)$, where $W = \begin{bmatrix} | & & | \\ w_1 & \cdots & w_k \\ | & & | \end{bmatrix}_{(d+1)\times K}$

$$\cancel{\frac{\partial E_{in}(W)}{\partial W_{ik}} = \frac{1}{N}\sum_{n=1}^{N} \theta(-y_n W^T x_n)(-y_n x_n) \frac{\partial}{\partial}}$$

$$\frac{\partial err(W, x, y)}{\partial W_{ik}} = \frac{\partial(-\ln h_y(x))}{\partial W_{ik}} = \frac{\partial(-\sum_{k=1}^{K}[y=k]\ln h_k(x))}{\partial W_{ik}}$$

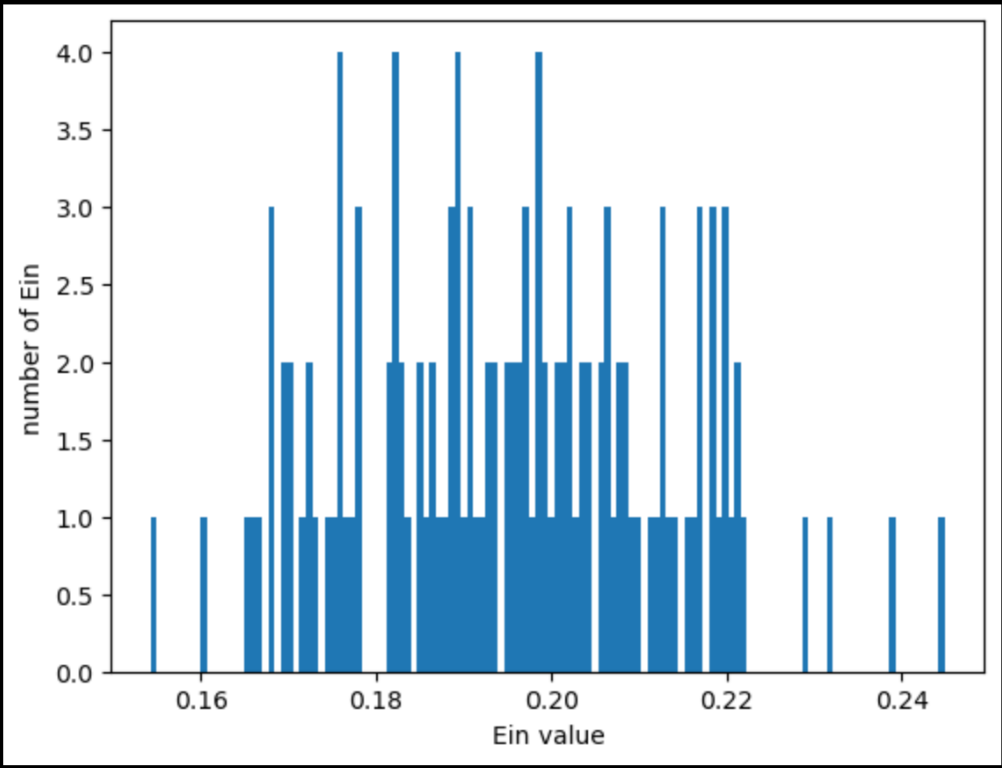
$$= -\left([y=k] - \left(\frac{\exp(W_k^T x)}{\sum_{i=1}^{K}\exp(W_i^T x)}\right)\right)x_i = -([y=k] - h_k(x))\, x_i$$

$$= (h_k(x) - [y=k])\, x_i$$

$$\therefore \nabla E_{in}(W) = \frac{1}{N}\sum_{n=1}^{N} \begin{bmatrix} | & & | \\ ([y=1] - h_1(x))\, x_i & \cdots & -([y=k] - h_k(x))\, x_i \\ | & & | \end{bmatrix}$$

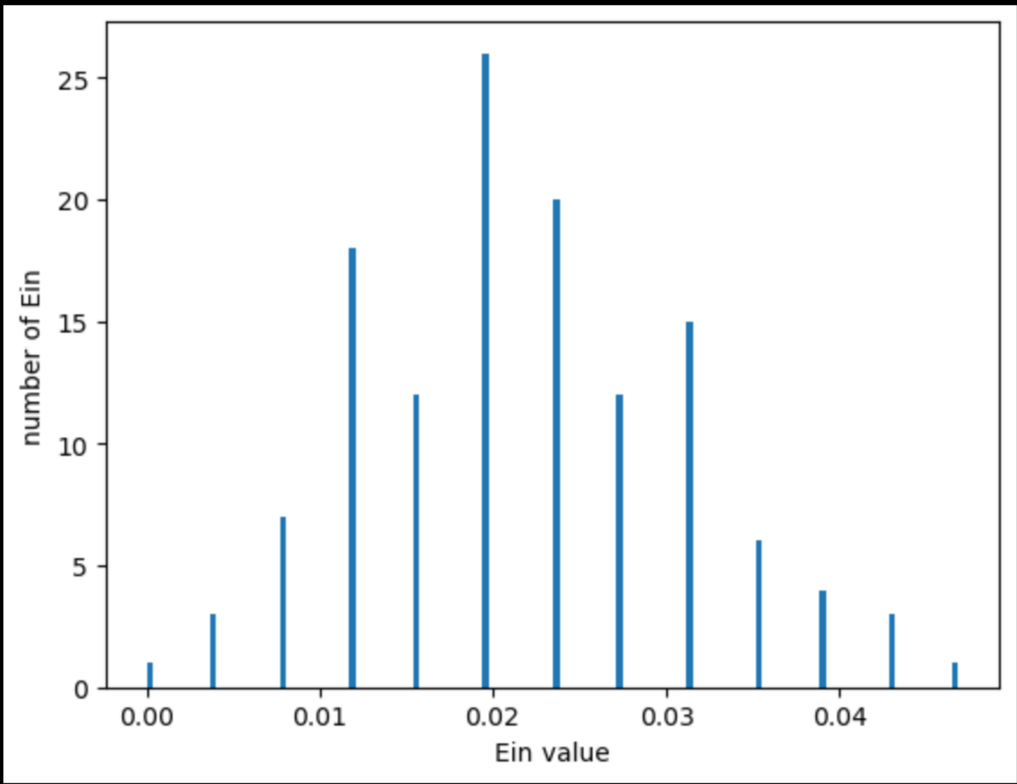
9

```
the median of Ein: 0.1960551435603019

Text(0, 0.5, 'number of Ein')
```



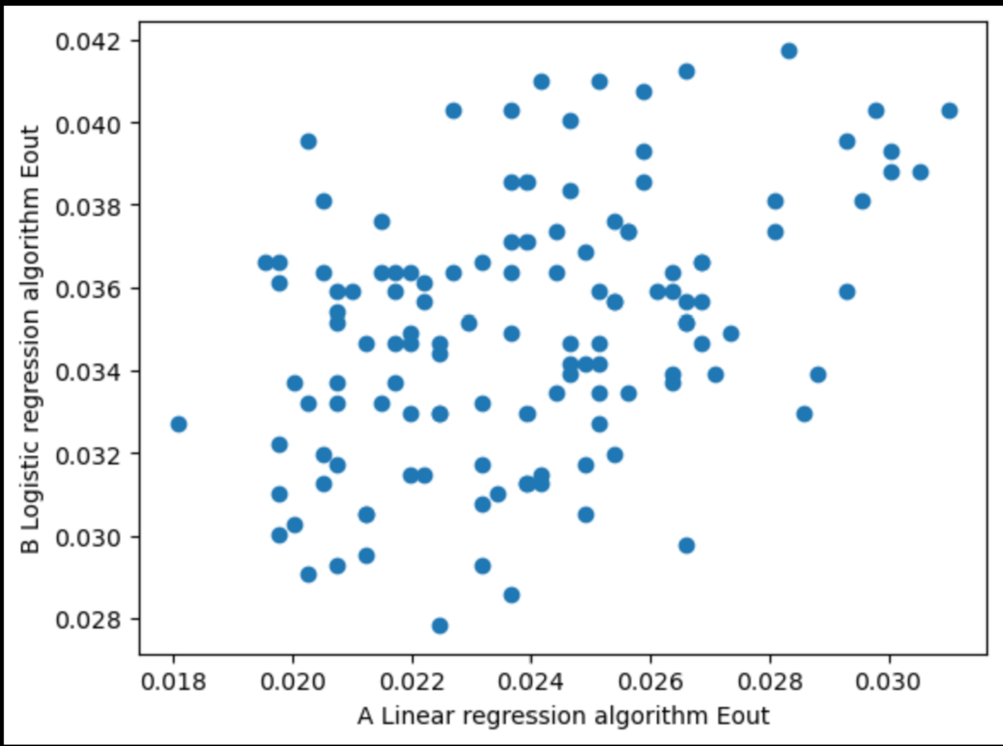
10

```
the median of Ein: 0.01953125

Text(0, 0.5, 'number of Ein')
```



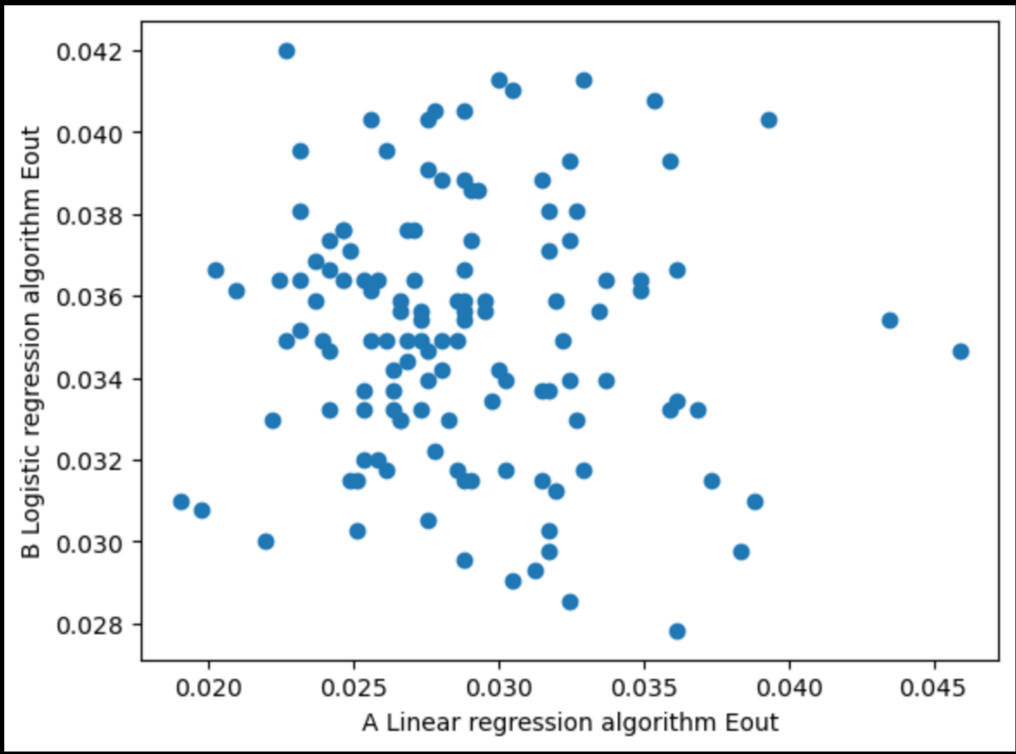
11

```
the median of Eout of A Linear regression algorithm: 0.02392578125
the median of Eout of B Logistic regression algorithm 0.0350341796875

Text(0, 0.5, 'B Logistic regression algorithm Eout')
```



12

```
the median of Eout of A Linear regression algorithm: 0.028076171875
the median of Eout of B Logistic regression algorithm 0.034912109375

Text(0, 0.5, 'B Logistic regression algorithm Eout')
```



與第 11 題相比,Linear regression 的 Eout 有顯著的上升,而 Logistic regression 則沒有顯著的變化,相較之下,第 11 題的散步圖有呈線性的趨勢,而第 12 題的分布趨勢則較觀察不出來,估計是 Logistic regression 在對付 outliner 的效果較好,而一般線性回歸則較容易受 outliner 影響.

13

13.

Known $V = -(A_E(u))^{-1} b_E(u) \rightarrow v = -(X^T D X)^{-1} \nabla E_{in}(w_t)$

Linear regression's $E_{in}(w) = \frac{1}{N} \|Xw - y\|^2$

$= \frac{1}{N}(w^T X^T X w - 2w^T X^T y + y^T y)$

Since $b_E(w) = \nabla E_{in}(w) = \frac{2}{N}(X^T X w - X^T y) = (X^T(Xw - y))$

$A_E(w) = \nabla(\nabla E_{in}(w)) = \frac{2}{N}(X^T X) \quad \forall w = w_t$

$\tilde{X} = \sqrt{\frac{2}{N}} X, \quad \tilde{y} = \sqrt{\frac{2}{N}}(Xw - y)$

#