Machine Learning (NTU, Fall 2023)

HW6 R12631055 林東甫

1.

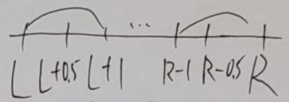
1. known $K_{ds}(x, x') = (\phi_{ds}(x))^T(\phi_{ds}(x'))$

$$= \left(g_{+1,1,\theta_1}(x), \cdots, g_{-1,d,\theta_k}(x)\right) \cdot \left(g_{+1,1,\theta_1}(x'), \cdots, g_{-1,d,\theta_k}(x')\right)$$

$g_{s,i,\theta}(x) = s \cdot \text{sign}(x_i - \theta)$, where $i \in \{1, 2, \cdots, d\}$

$s \in \{-1, +1\}$

$$g_{\underbrace{s}_{\in\, -1, \cdots, +1},\ \underbrace{i}_{\in\, 1, \cdots, d},\ \underbrace{\theta}_{\in\, L+0.5, \cdots, R-0.5}}(x) = s \cdot \text{sign}(x_i - \theta)$$



$L+0.5 \le \theta \le R-0.5$

$\Rightarrow$ ~~θ has~~ # of θ is $R - L = k$

$\theta \in \{\theta \mid L+0.5 \le \theta \le R-0.5\}$

$$K_{ds}(x, x') = \left(g_{+1,1,\theta_1}(x), \cdots, g_{-1,d,\theta_k}(x)\right) \cdot \left(g_{+1,1,\theta_1}(x'), \cdots, g_{-1,d,\theta_k}(x')\right)$$

$$= \sum_{s\in\{-1,+1\}} \sum_{i=1}^{d} \sum_{j=1}^{k} g_{s,i,\theta_j}(x) \cdot g_{s,i,\theta_j}(x') = \sum_{s\in\{-1,+1\}} \sum_{i=1,j=1}^{d,k} \left(s \cdot \text{sign}(x_i - \theta_j)\right) \cdot \left(s \cdot \text{sign}(x_i' - \theta_j)\right)$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{k} \left[(+1 \cdot \text{sign}(x_i - \theta_j)) \cdot (+1 \cdot \text{sign}(x_i' - \theta_j)) + (-1 \cdot \text{sign}(x_i - \theta_j)) \cdot (-1 \cdot \text{sign}(x_i' - \theta_j)) \right]$$

$$= 2 \sum_{i=1}^{d} \sum_{j=1}^{k} \left[\text{sign}(x_i - \theta_j) \cdot \text{sign}(x_i' - \theta_j) \right]$$, (considering all situation)

where $k = R - L$

$$= 2 \sum_{i=1}^{d} \left[k + 2 \cdot \frac{\| x_i - x_i' \|}{2} \right] = 2dk + 2 \sum_{i=1}^{d} \| x_i - x_i' \|$$

$$= \underline{2dk + 2\| x - x' \|_1}$$ #

2.

2. Known $\tilde{K}(x, x') = uK(x, x') + v$ where $u > 0$, $v \in \mathbb{R}$, $\tilde{C} = \frac{C}{u}$

that is $\max W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$

subject to $0 \le \alpha_i \le C \quad \forall i, \quad \sum_{i=1}^{n} \alpha_i y_i = 0$

since $\hat{K}(x, x') = uK(x, x') + v$, $\tilde{C} = \frac{C}{u}$

Let $\max \tilde{W}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \hat{K}(x_i, x_j)$

$= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j (uK(x_i, x_j) + v)$

subject to $0 \le \alpha_i \le \tilde{C} = \frac{C}{u} \quad \forall i. \quad \sum_{i=1}^{n} \alpha_i y_i = 0$

$\therefore f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right)$

$\Rightarrow \tilde{f}(x) = \text{sign}\left\{\sum_{i=1}^{n} \alpha_i y_i [uK(x_i, x) + v] + b\right\}$

which is equivalent gsvm classifier.

3.

3. Known total error $E = \sum_{t=1}^{17} e_t$; 17 classifiers $\{g_t\}_{t=1}^{17}$

$G(x) = \text{sign}\left(\sum_{t=1}^{17} g_t(x)\right)$

$\because E_{out}(g_t) = e_t = \mathcal{E}[\![g_t \cdot f \neq 1]\!]$

$\therefore E_{out}(G) = \mathcal{E}[\![G \cdot f \neq 1]\!] = \mathcal{E}[\![\text{sign}(\sum_{t=1}^{17} g_t \cdot f) \neq 1]\!]$

Since there are 17 classifiers

$[\![\text{sign}(\sum_{t=1}^{17} g_t \cdot f) \neq 1]\!] = 0$ (for 0 to 8) $\quad [\![\text{sign}(\sum_{t=1}^{17} g_t \cdot f) \neq 1]\!] = 1$ (for 9 to 17)

# of g_t classified	0	1	2	…	8	9	…	16	17
$\sum_{t=1}^{17}[\![\text{sign}(g_t f) \neq 1]\!]$:					$\frac{8}{9}$	$\frac{9}{9}$		$\frac{16}{9}$	$\frac{17}{9}$

$$\frac{E_{out}(G)}{E} = \frac{\mathcal{E}[\![\text{sign}(\sum_{t=1}^{17} g_t \cdot f) \neq 1]\!]}{\sum_{t=1}^{17} e_t}$$

$\because [\![\text{sign}(\sum_{t=1}^{17} g_t \cdot f) \neq 1]\!] \leq \frac{1}{k} \sum_{t=1}^{17} [\![\text{sign}(g_t \cdot f) \neq 1]\!]$ where $1 \leq k \leq 9$

$\therefore$ if $k = 9$, $\quad \frac{E_{out}(G)}{E} \leq \frac{1}{9} \frac{\sum_{t=1}^{17} e_t}{\sum_{t=1}^{17} e_t}$ $\quad \left(\text{if } k=1 \ \frac{E_{out}(G)}{E} \leq 1\right)$

4.

4. If N is large enough:

$$\left(1-\frac{1}{N}\right)^{N'} = \frac{1}{\left(\frac{N}{N-1}\right)^{N'}} = \frac{1}{\left(1+\frac{1}{N-1}\right)^{N'}} \simeq \frac{1}{e}$$

where $N' = \frac{3}{4}N$, $N \to \infty$

$$\left(1-\frac{1}{N}\right)^{\frac{3}{4}N} = \frac{1}{\left(1+\frac{1}{N-1}\right)^{\frac{3}{4}N}} \simeq \frac{1}{e^{\frac{3}{4}}} = e^{-\frac{3}{4}} \approx 0.472366\ldots$$

$$\approx 47.2\%$$

5. Known Data of examples are 98% $y_n > 0$, 2% $y_n < 0$

Assume $N = 100$, # of $y_n > 0$ is 98, # of $y_n < 0$ is 2

$$\frac{\sum_{n: y_n > 0} u_n^{(2)}}{\sum_{n: y_n < 0} u_n^{(2)}} = \frac{\sum_{n: y_n>0} u_n^{(1)} \times \overset{\epsilon_t}{\frac{2}{100}}}{\sum_{n: y_n<0} u_n^{(1)} \times \underset{1-\epsilon_t}{\frac{98}{100}}} = \frac{\frac{2}{100} \times 98 u_n^{(1)}}{\frac{98}{100} \times 2 u_n^{(1)}}$$

Since first iteration $u_n^{(1)} = \frac{\cancel{0.98N}\ 1}{N}$, $\cancel{\frac{\frac{2}{100} \times 98 \times \frac{1}{N}}{\frac{98}{100} \times 2 \times \frac{1}{N}}}$

$$\frac{\sum_{n: y_n > 0} u_n^{(2)}}{\sum_{n: y_n < 0} u_n^{(2)}} = 1$$

5.

5. Known $g_1(x) = +1$, $err_{0/1} = 1 - 0.98 = 0.02$

Assume total # of data is N :

$$\epsilon_t = \frac{\sum_{n=1}^{N} u_n^{(t)} [\![y_n \neq g_t(x_n)]\!]}{\sum_{n=1}^{N} u_n^{(t)}} , \quad \diamond_t = \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}$$

If $t=1$, $u_n^{(1)} = \frac{1}{N}$, $\epsilon_1 = \sum_{n=1}^{N} \frac{[\![y_n \neq g_1(x_n)]\!]}{N} = err_{0/1} = 0.02$

Since $\sum_{n: y>0} u_n^{(1)} = \sum_{n: y<0} u_n^{(1)} = \frac{1}{N}$, positive sample $\rightarrow$ correct , negative sample $\rightarrow$ incorrect

$$\therefore \frac{\sum_{n: y>0} u_n^{(2)}}{\sum_{n: y<0} u_n^{(2)}} = \frac{\sum_{n} y>0\, u_n^{(1)} / \diamond_1}{\sum_n y<0\, u^{(1)} \cdot \diamond_1} = \frac{1}{\diamond_1^2} = \frac{\epsilon_1}{1-\epsilon_1} = \frac{0.02}{0.98}$$

#

6.

7.

6. The recursive function of U_{T+1}:

$$U_{T+1} = \sum_{n=1}^{N} u_n^{(T+1)} = \epsilon_T \sum_{n=1}^{N} u_n^{(T)} \sqrt{\frac{1-\epsilon_T}{\epsilon_T}} + (1-\epsilon_T) \sum_{n=1}^{N} u_n^{(T)} \sqrt{\frac{\epsilon_T}{1-\epsilon_T}}$$

$$= 2\sqrt{\epsilon_T(1-\epsilon_T)} \sum_{n=1}^{N} u_n^{(T)} = 2\sqrt{\epsilon_T(1-\epsilon_T)}\, U_T$$

$$\Rightarrow \frac{U_{T+1}}{U_1} = \frac{2^T \prod_{t=1}^{T} \sqrt{\epsilon_t(1-\epsilon_t)}\, U_1}{U_1} = 2^T \prod_{t=1}^{T} \sqrt{\epsilon_t(1-\epsilon_t)}$$ #

$\cancel{\frac{U_{T+1}}{U_I}}$

7. Since minimizing $\frac{1}{N} \sum_{n=1}^{N} ((y_n - s_n) - \eta g_t(x_n))^2$

after taking the derivative of $\eta = 0$;

$$\eta = \frac{\sum_{n=1}^{N} (y_n - s_n) g_t(x_n)}{\sum_{n=1}^{N} g_t^2(x_n)}$$ with $s_n^{(t)} = s_n^{(t-1)} + \alpha_t g_t(x_n)$, $y_n = 0$

$$\sum_{n=1}^{N} s_n^{(t)} g_t(x_n) = \sum_{n=1}^{N} (s_n^{(t-1)} + \alpha_t g_t(x_n)) g_t(x_n)$$

$$= \sum_{n=1}^{N} s_n^{(t-1)} g_t(x_n) + \eta \sum_{n=1}^{N} g_t^2(x_n) = \sum_{n=1}^{N} s_n^{(t-1)} g_t(x_n) - \sum_{n=1}^{N} s_n^{(t-1)} g_t(x_n)$$

$$= 0$$ #

8.

8- Back-propagation : $a_j^{(l)} = \tanh\left(\sum_i W_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)}\right)$ (b_j is bias

weights update : $\frac{\partial L}{\partial W_{ij}^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot \frac{\partial a_j^{(l)}}{\partial W_{ij}(l)}$, since $a_j^{(l)}=0$, $\frac{\partial L}{\partial a_j^{(l)}}=0$

When all inital weights $W_{ij}^{(l)}$ are set to 0, $\tanh(0)=0$

Cause all weights of hidden layers are zero,

13- Input Layer : $x_1, x_2, \cdots, x_d$: d nodes

Hidden Layer : $h_1, h_2, \cdots, h_d$: d nodes

Output Layer : y 1 node

9.

9

```python
cart = DecisionTreeCART()
cart.fit(X_train, y_train)
```

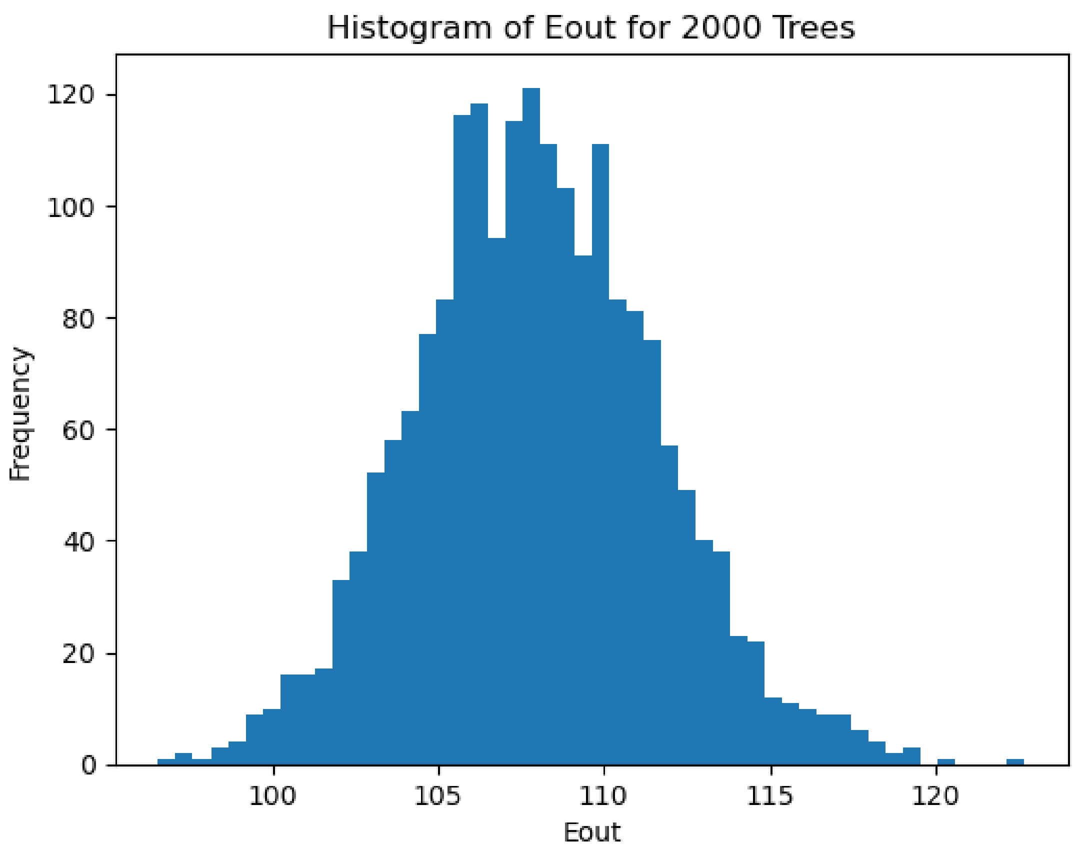
✓ 32.5s

```python
random_forest = RandomForest()
random_forest.fit(X_train, y_train)
eouts = random_forest.calculate_eout(X_test, y_test)
eouts[0]
```

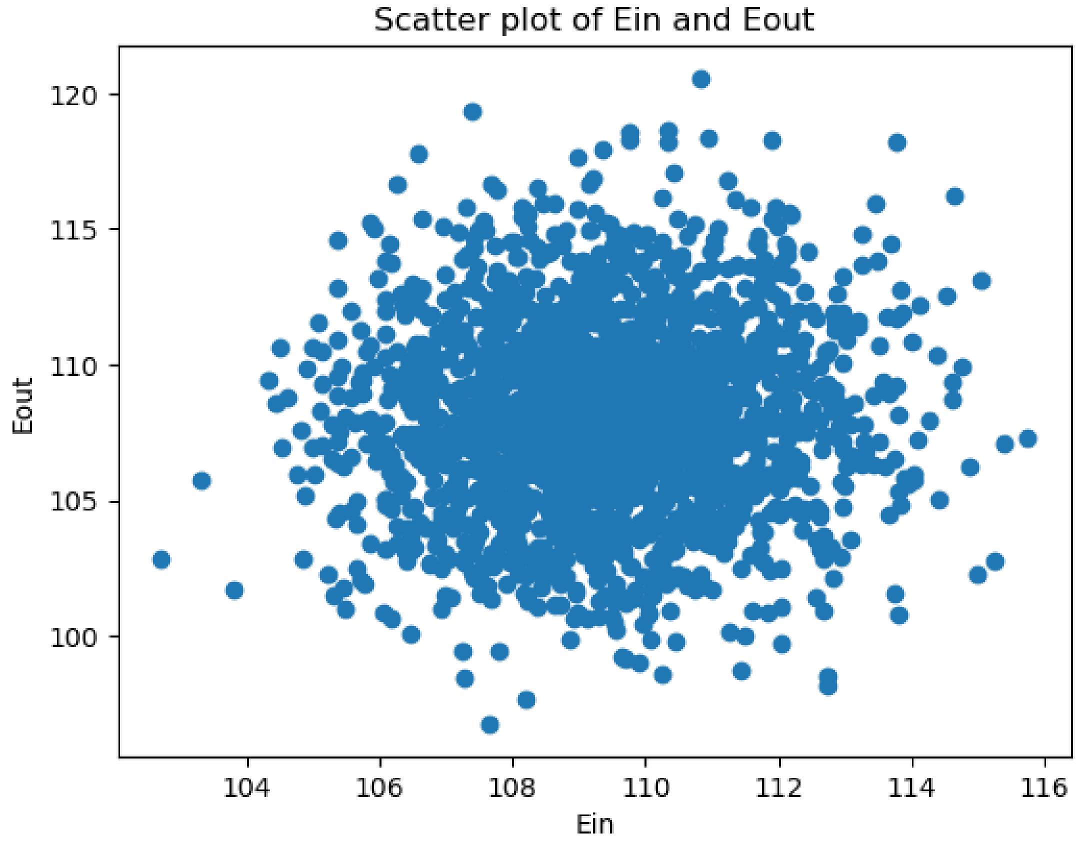
✓ 8.0s

```
106.60245901639344
```

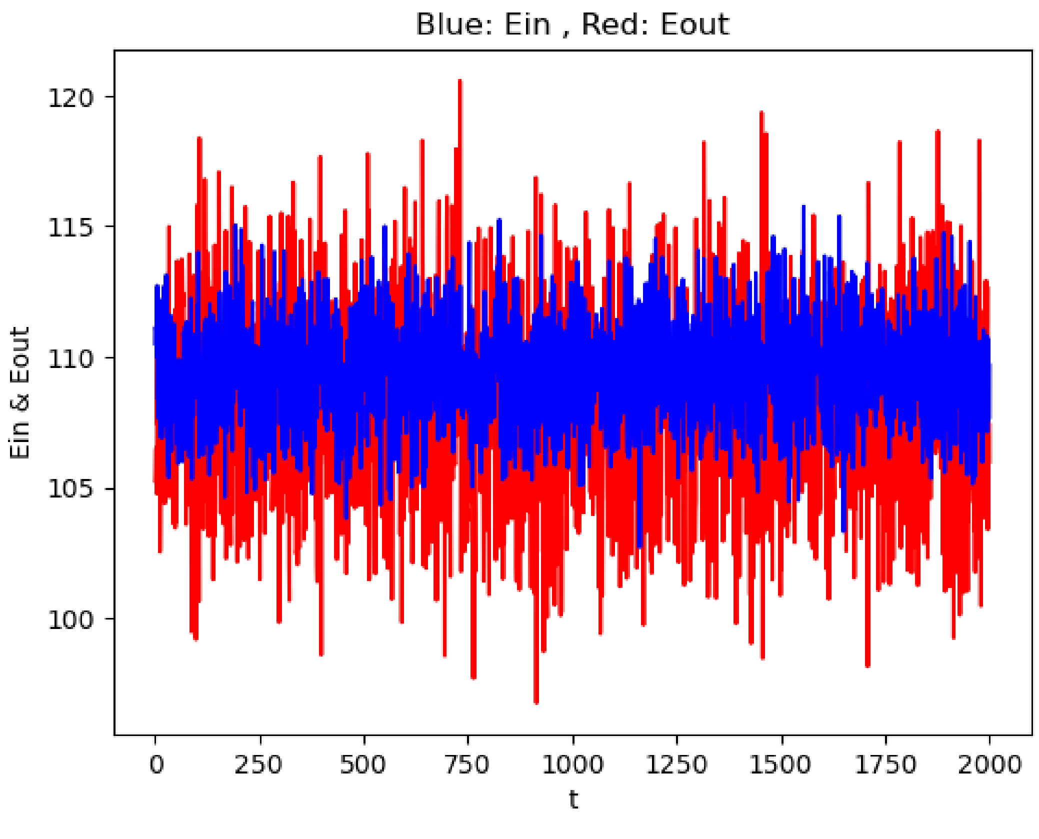
10.



11.


Scatter plot of Ein and Eout
120
115
110
105
100
Eout
104
106
108
110
112
114
116
Ein

12.


Blue: Ein , Red: Eout
120
115
110
105
100
Ein & Eout
0
250
500
750
1000
1250
1500
1750
2000
t