Machine Learning HW7

B03902089 資工三 林良翰

Boosting

1.

- $\min_{w} E_{in}^{u}(w) = \frac{1}{N}\sum_{n=1}^{N} u_n\left(y_n - w^T x_n\right)^2$
- For every $\left(x_n, y_n\right)$
  $u_n\left(y_n - w^T x_n\right)^2 = \left(\sqrt{u_n} y_n - w^T \sqrt{u_n} x_n\right)^2 = \left(\tilde{y}_n - w^T \tilde{x}_n\right)^2$
- $\Rightarrow \left\{\left(\tilde{x}_n, \tilde{y}_n\right)\right\}_{n=1}^{N} = \left\{\left(\sqrt{u_n} x_n, \sqrt{u_n} y_n\right)\right\}_{n=1}^{N}$

2.

- $u^{(1)} = \left[\frac{1}{N}, \frac{1}{N}, \ldots m, \frac{1}{N}\right]$
- 99% of the examples are positive $\Rightarrow \epsilon_1 = 0.99$
- $u_{-}^{(2)} \propto \frac{1}{N} \cdot \epsilon_1 \Rightarrow \frac{k \cdot 0.99}{N}$
  $u_{-}^{(2)} \propto \frac{1}{N} \cdot (1 - \epsilon_1) \Rightarrow \frac{k \cdot 0.01}{N}$
- $\frac{u_{+}^{(2)}}{u_{-}^{(2)}} = \frac{(1-0.99)}{0.99} = \frac{1}{99}$

Kernel for Decision Stumps

3.

- $g_{s,i,\theta}(x) = s \cdot \text{sign}\left(\mathrm{x_i} - \theta\right)$

  $i \in \{1, 2, \ldots, d\}$, d is the finite dimensionality of the input space.

  $s \in \{+1, -1\}$, $\theta \in \mathbb{R}$, and $\text{sign}(0) = +1$
- $L = 1$ and $R = 6 \Rightarrow$ There are 6 kinds of θ.
- $d = 2 \Rightarrow$ Two dimensions.
- $s \in \{+1, -1\} \Rightarrow$ Positive ray or negative ray.
- Thus, there are $6 \times 2 \times 2 = 24$ kinds of $g_{s,i,\theta}(x)$.

4.

Decision Tree

u_+ : Fraction of positive examples.

$u_- = 1 - u_+$: Fraction of Negative examples.

5.

- Find the maximum value of Gini index
  $\max_{u_+} 1 - u_+^2 - u_-^2$, where $u_+ \in [0, 1]$
- By $u_- = 1 - u_+$
  $$\begin{aligned} & 1 - u_+^2 - u_-^2 \\ = \quad & 1 - u_+^2 - (1 - u_+)^2 \\ = \quad & -2u_+^2 + 2u_+ \end{aligned}$$
- $\frac{d\left(-2u_+^2 + 2u_+\right)}{du_+} = -4u_+ + 2 = 0$
  $u_+ = \frac{1}{2}$ can obtain maximum value $1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{1}{2}$

6.

- Normalized Gini index is
  $\frac{1 - u_+^2 - u_-^2}{\frac{1}{2}} = 2\left(1 - u_+^2 - u_-^2\right) = -4u_+^2 + 4u_+ = -4\left(u_+ - \frac{1}{2}\right)^2 + 1$
- [a]
  As stated in the problem, the normalized classification error is $2\min\left(u_+, u_-\right)$
- [b]
  $$\begin{aligned} & u_+\left(1 - \left(u_+ - u_-\right)\right)^2 + u_-\left(-1 - \left(u_+ - u_-\right)\right)^2 \\ = \quad & u_+\left(1 - u_+ + u_-\right)^2 + u_-\left(-1 - u_+ + u_-\right)^2 \\ = \quad & u_+(2 - 2u_+)^2 + (1 - u_+)(-2u_+)^2 \\ = \quad & 4u_+(1 - u_+) = -4u_+^2 + 4u_+ = -4\left(u_+ - \frac{1}{2}\right)^2 + 1 \end{aligned}$$
  $\Rightarrow$ Maximum is 1 when $u_+ = \frac{1}{2}$, $\Rightarrow$ Normalized form is $-4\left(u_+ - \frac{1}{2}\right)^2 + 1$
- [c]
  Find the derivative on u_+
  $$\begin{aligned} & \frac{d\left(-u_+ \ln(u_+) - u_- \ln\left(u_-\right)\right)}{du_+} \\ = \quad & \frac{d(-u_+ \ln(u_+) - (1 - u_+)\ln(1 - u_+))}{du_+} \\ = \quad & -\ln(u_+) - 1 + \ln(1 - u_+) + 1 \\ = \quad & -\ln(u_+) + \ln(1 - u_+) = 0 \end{aligned}$$
  $\Rightarrow$ Maximum entropy is $-\ln(2)$ when $u_+ = \frac{1}{2}$

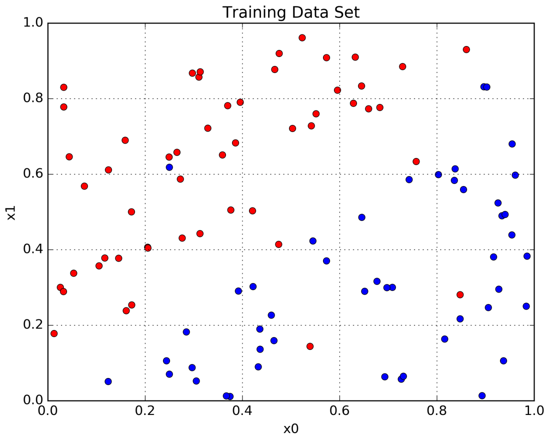
$\Rightarrow$ Normalized form is $\frac{u_+ \ln(u_+) + u_- \ln(u_-)}{\ln(2)}$

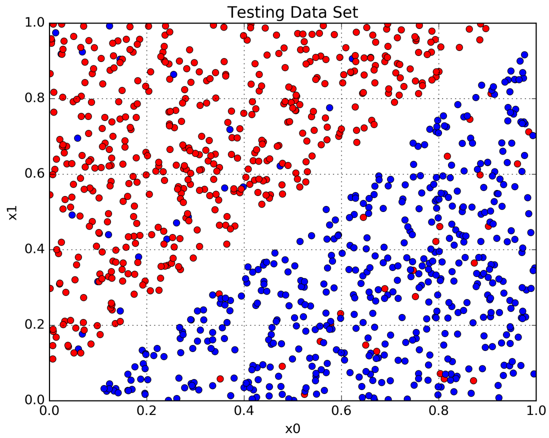
- [d]
  The maximum of $1 - |u_+ - u_-|$ is 1 when $u_+ = u_- = \frac{1}{2}$
  $\Rightarrow$ Normalized form is still $1 - |u_+ - u_-|$

- The answer is [b]

Experiments with Adaptive Boosting

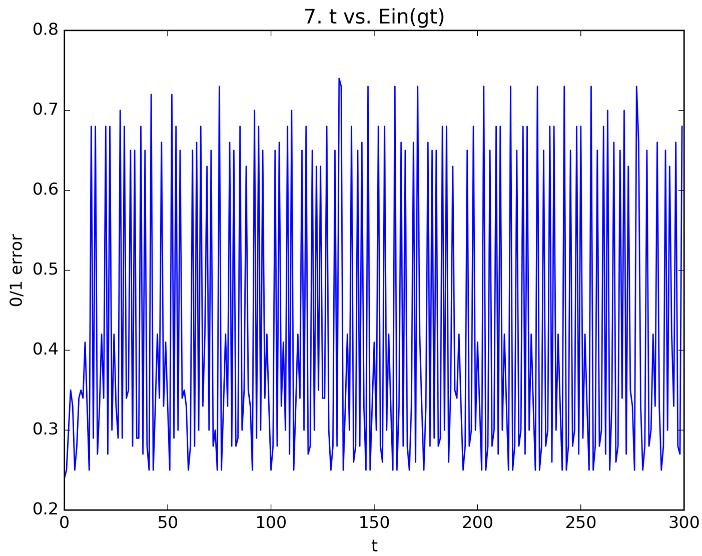
- Training data: 100
- Testing data: 1000
- Iteration $T = 300$, dimension $D = 2$





7.

- $E_{in}(g_1) = 0.24, \alpha_1 = 0.57634$

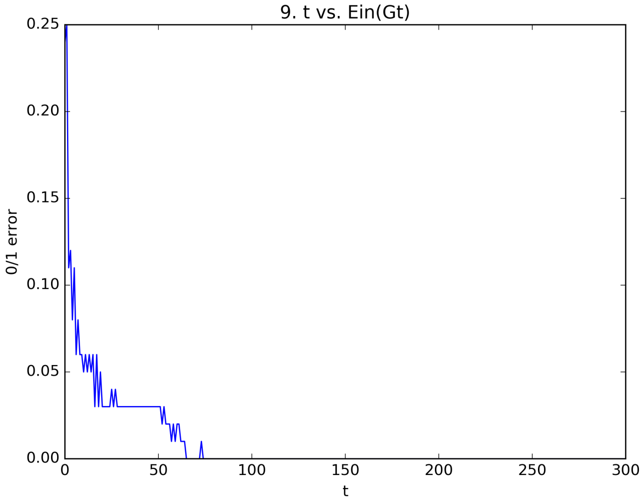


8.

- There is no obvious trend that $E_{in}(g_t)$ is decreasing or increasing, because it goes up and down frequently.

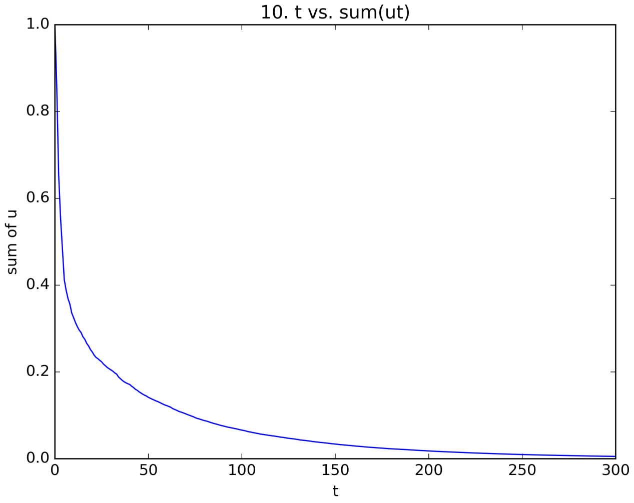
9.

- $E_{in}(G_T) = 0.0$



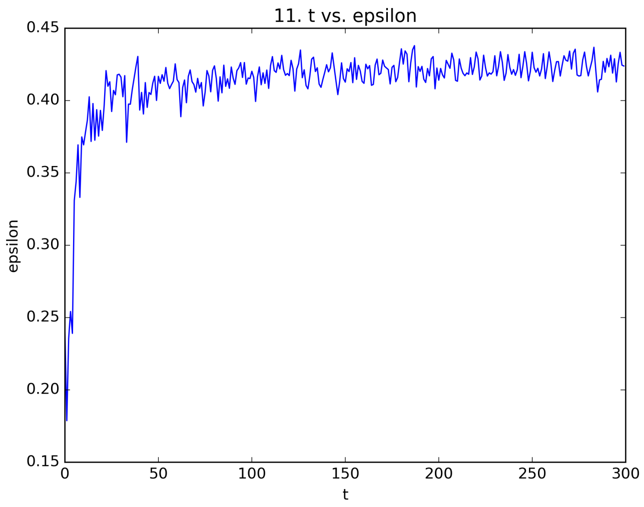
10.

- $U_2 = 0.85417$, $U_T = 0.00547$



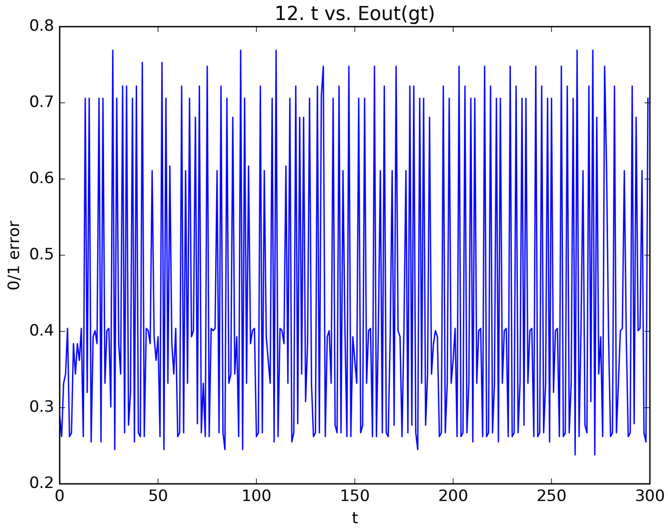
11.

- $\min_t \epsilon_t = 0.17873$, $t = 1$



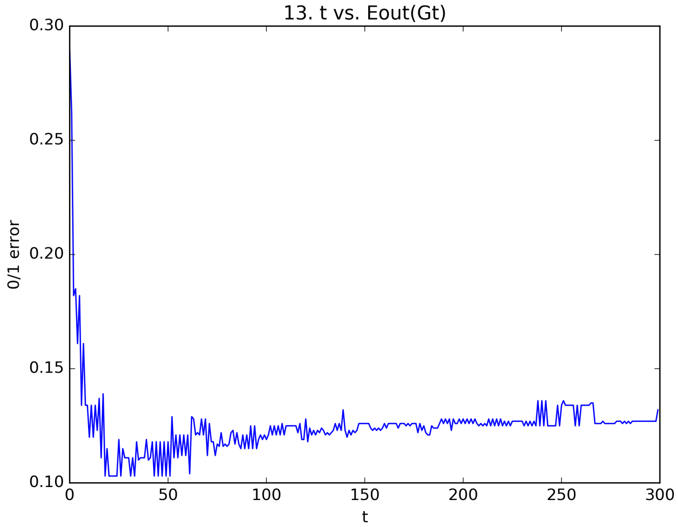
12.

- $E_{out}\left(g_1\right) = 0.29$



13.

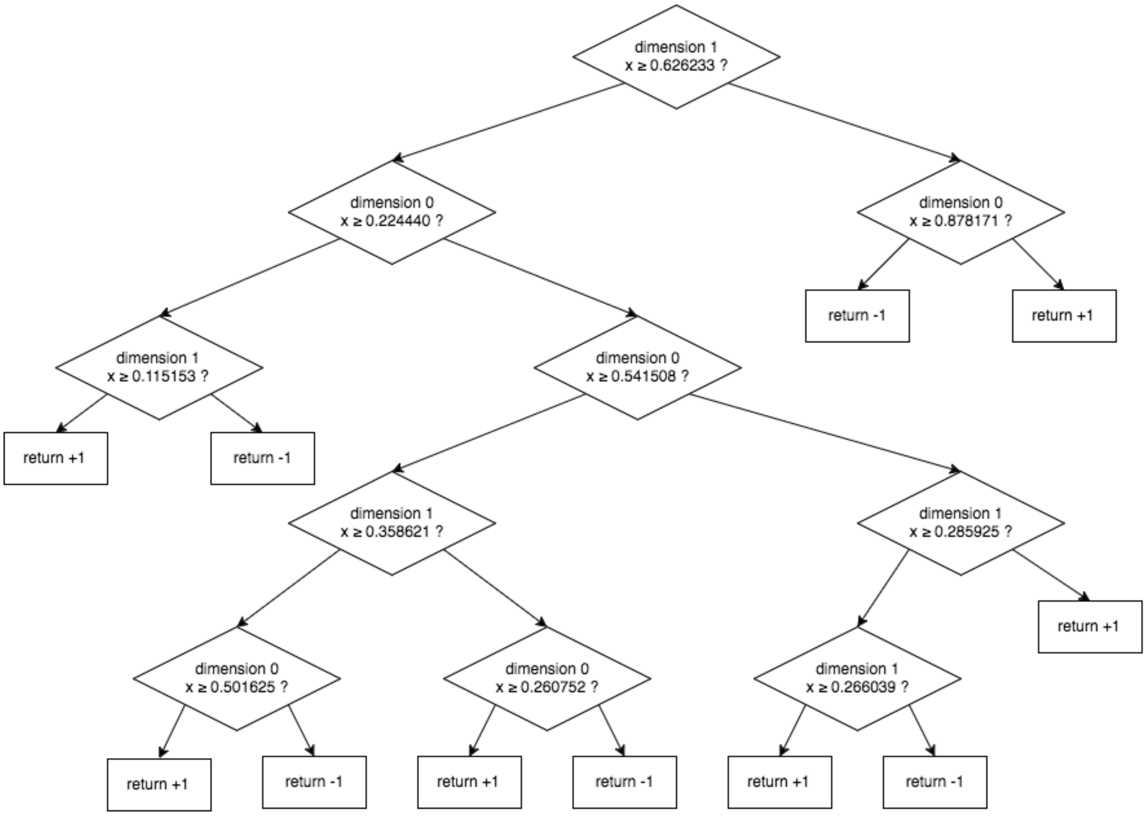
- $E_{out}\left(G_T\right) = 0.132$



Experiments with Unpruned Decision Tree

14.

- C&RT Tree



15.

- $E_{in} = 0.0$, $E_{out} = 0.126$

16.

- $\min E_{in} = 0.01$, corresponding $E_{out} = 0.109$

```
Problem 16.
prune: dim = 1, theta = 0.115153, prune_left, Ein: 0.010000, Eout: 0.144000
prune: dim = 1, theta = 0.115153, prune_right, Ein: 0.140000, Eout: 0.215000
prune: dim = 0, theta = 0.501625, prune_left, Ein: 0.140000, Eout: 0.203000
prune: dim = 0, theta = 0.501625, prune_right, Ein: 0.010000, Eout: 0.109000
prune: dim = 0, theta = 0.260752, prune_left, Ein: 0.010000, Eout: 0.117000
prune: dim = 0, theta = 0.260752, prune_right, Ein: 0.060000, Eout: 0.173000
prune: dim = 1, theta = 0.266039, prune_left, Ein: 0.090000, Eout: 0.242000
prune: dim = 1, theta = 0.266039, prune_right, Ein: 0.010000, Eout: 0.116000
prune: dim = 1, theta = 0.285925, prune_right, Ein: 0.200000, Eout: 0.279000
prune: dim = 0, theta = 0.878171, prune_left, Ein: 0.300000, Eout: 0.383000
prune: dim = 0, theta = 0.878171, prune_right, Ein: 0.030000, Eout: 0.153000
```