Machine Learning Technique Homework 3

1

Let

$$(\tilde{x}_n, \tilde{y}_n) = (\sqrt{u_n} y_n, \sqrt{u_n} x_n)$$

then

$$\begin{aligned}
\min_w E^u_{\in}(w) &= \frac{1}{N} \sum_{n=1}^{N} u_n (y_n - w^T x_n)^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} (\sqrt{u_n} y_n - w^T \sqrt{u_n} x_n)^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} (\tilde{y}_n - w^T \tilde{x}_n)^2
\end{aligned}$$

Thus they are equivalent.

2

All positive examples are classified correctly in the first iteration, while all negative examples are classified incorrectly. And we know that the examples should be reweighted so that the weighted number of correct examples is equal to number of incorrect examples. Therefore,

$$u_+^{(2)} \text{number of positive example} = u_-^{(2)} \text{number of negative example}$$

$$\begin{aligned}
u_+^{(2)} / u_-^{(2)} &= \text{number of negative example/number of positive example} \\
&= \frac{1}{99}
\end{aligned}$$

3

For each dimension, consider θ between $(0,1), (1,2), (2,3), \cdots, (5,6)$ with $s = +1, -1$, totally 12 different decision stumps. And there are two dimension; therefore, there are 24 decision stumps.

4

$$
\begin{aligned}
K_{ds}(x, x') =& (\phi_{ds}(x))^T(\phi_{ds}(x')) \\
=& g_1(x)g_1(x') + g_2(x)g_2(x') + \cdots + g_{|\mathcal{G}|}(x)g_{|\mathcal{G}|}(x')
\end{aligned}
$$

Let $g_n(x) = s_n \cdot \text{sign}(x_{i_n} - \theta_n)$

$$
\begin{aligned}
g_n(x)g_n(x') &= s_n \cdot \text{sign}(x_{i_n} - \theta_n)s_n \cdot \text{sign}(x'_{i_n} - \theta_n) \\
&= s_n^2 \cdot \text{sign}(x_{i_n} - \theta_n)\text{sign}(x'_{i_n} - \theta_n) \\
&= \text{sign}(x_{i_n} - \theta_n)\text{sign}(x'_{i_n} - \theta_n) \\
&= \begin{cases} 1 & \text{sign}(x_{i_n} - \theta_n) = \text{sign}(x'_{i_n} - \theta_n) \\ -1 & \text{sign}(x_{i_n} - \theta_n) \neq \text{sign}(x'_{i_n} - \theta_n) \end{cases}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
g_n(x)g_n(x') =& (+1) \cdot |\{g|g(x) = g(x')\}| + (-1) \cdot |\{g|g(x) \neq g(x')\}| \\
=& (+1) \cdot (2(R - L + 1) - |x_1 - x'_1| - |x_2 - x'_2|) + \\
& (-1) \cdot (|x_1 - x'_1| + |x_2 - x'_2|)
\end{aligned}
$$

where the second equality is because $g_n(x) \neq g_n(x')$ only when θ_n is between $(x_i, x_i + 1), (x_i + 1, x_i + 2), \cdots, (x'_i - 1, x'_i)$, totally $|x'_i - x_i|$ decision stumps, and trivially, $|\{g|g(x) = g(x')\}| = |\text{decision stumps}| - |\{g|g(x) \neq g(x')\}|$.

To work with continuous input vector, we can first sort x_1, x_2 respectively. Then we can replace (x_1, x_2) with (d_1, d_2), where d_1, d_2 is the indice of x_1, x_2 in the sorted x_1, x_2. So

$$
g_n(x)g_n(x') = (-1) \cdot (|d_1 - d'_1| + |d_2 - d'_2|)
$$

5

$$
\begin{aligned}
1 - \mu_+^2 - \mu_-^2 &= 1 - \mu_+^2 - (1 - \mu_+^2) \\
&= 1 - 2\mu_+^2 + 2\mu_+ - 1 \\
&- 2\mu_+^2 + 2\mu_+
\end{aligned}
$$

it gets the maximum value $\frac{1}{2}$ when $\mu_+ = \mu_- = \frac{1}{2}$.

6

The answer is **[e]**.

Since $\max_{\mu_+, \mu_-} 1 - \mu_+^2 - \mu_-^2 = \frac{1}{2}$, it is normalized to

$$
\tilde{f}_0(\mu_+, \mu_-) = 2(1 - \mu_+^2 - \mu_i^2) = 2(1 - \mu_+^2 - (1 - \mu_+)^2) = 4\mu_+ - 4\mu_+^2
$$

[a]

$$\min(\mu_+, \mu_-)|_{\mu+=\frac{1}{4}, \mu_-=\frac{1}{4}} = \frac{1}{4} \neq \frac{3}{4} = \tilde{f}_0(\frac{1}{4}, \frac{3}{4})$$

thus not [a].

[b]

Let

$$f_b(\mu_+, \mu_-) = \mu_+(1 - (\mu_+ - \mu_-))^2 + \mu_-(-1 - (\mu_+ - \mu_-))^2$$

$$\begin{aligned}
&\frac{\partial}{\partial \mu_+}\mu_+(1 - (\mu_+ - \mu_-))^2 + \mu_-(-1 - (\mu_+ - \mu_-))^2 \\
=&\frac{\partial}{\partial \mu_+}\mu_+(1 - (\mu_+ - (1 - \mu_+)))^2 + (1 - \mu_+)(-1 - (\mu_+ - (1 - \mu_+)))^2 \\
=&\frac{\partial}{\partial \mu_+}4\mu_+^2(1 - \mu_+) \\
=&4(2\mu_+(1 - \mu_+) - \mu_+^2) \\
=&4(2\mu_+ - 3\mu_+^2) = 0 \text{ when } \mu_+ = \frac{2}{3} \text{ or } \mu_+ = 0.
\end{aligned}$$

Since $f_b(0, 1) = f_b(1, 0) = 0 < f_b(\frac{2}{3}, \frac{1}{3}) = \frac{16}{27}$,

$$\arg\max f_b(\mu_+, \mu_-) = (\frac{2}{3}, \frac{1}{3})$$

Thus the normalized f_b is

$$\tilde{f}_b(\mu_+, \mu_-) = \frac{27}{16}(\mu_+(1 - (\mu_+ - \mu_-))^2 + \mu_-(-1 - (\mu_+ - \mu_-))^2)$$

But

$$\tilde{f}_b(\frac{1}{2}, \frac{1}{2}) = \frac{27}{32} \neq 1 = \tilde{f}_0(\frac{1}{2}, \frac{1}{2})$$

thus not [b].

[c]

Let

$$\begin{aligned}
f_c(\mu_+, \mu_-) =& - \mu_+ \ln \mu_+ - \mu_- \ln \mu_- \\
=& - \mu_+ \ln \mu_+ - (1 - \mu_+) \ln(1 - \mu_+)
\end{aligned}$$

$$\frac{\partial}{\partial \mu_+} f_c(\mu_+, \mu_-)$$
$$= -\ln \mu_+ - 1 + \ln(1-\mu_+) + 1 = 0 \text{ when } \mu_+ = \frac{1}{2}.$$

Thus

$$\max f_c(\mu_+, \mu_-) = \ln 2$$

and the normalized f_c is

$$\tilde{f}_c(\mu_+, \mu_-) = -\mu_+ \log_2 \mu_+ - \mu_- \log_2 \mu_-$$

But

$$\tilde{f}_c(\frac{3}{4}, \frac{1}{4}) = -\frac{3}{4}\log_2 \frac{3}{4} - \frac{1}{4}\log_2 \frac{1}{4}$$
$$= 2 - \frac{3}{4}\log_2 3 \neq \frac{3}{4} = \tilde{f}_0(\frac{3}{4}, \frac{1}{4})$$

thus not [c].

[d]

Let

$$f_d(\mu_+, \mu_-) = 1 - |\mu_+ - \mu_-|$$

$$\max f_d(\mu_+, \mu_-) = 1$$

thus nromalized f_d is

$$\tilde{f}_d(\mu_+, \mu_-) = f_d(\mu_+, \mu_-) = 1 - |\mu_+ - \mu_-|$$

But

$$\tilde{f}_d(\frac{3}{4}, \frac{1}{4}) = \frac{1}{2} \neq \frac{3}{4} = \tilde{f}_0$$

Thus not [d].

7

$E_{in}(g_t) = 0.24$, $\alpha_1 = 0.5763397549691922$

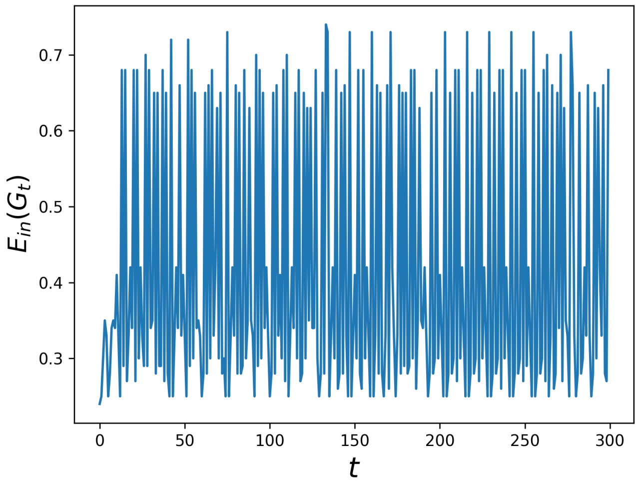

Figure 1: t versus $E_{in}(g_t)$

8

It is neither increasing nor decreasing. In just oscillates severely. The reason may be the changing of sample weights during the training process.

9

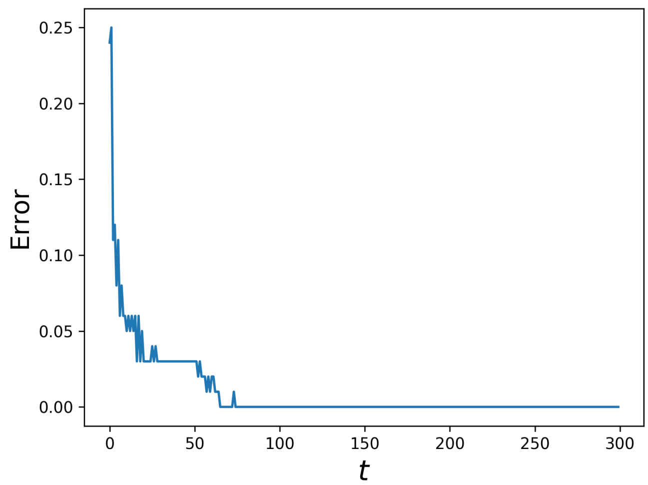

Figure 2: t versus $E_{in}(G_t)$

10

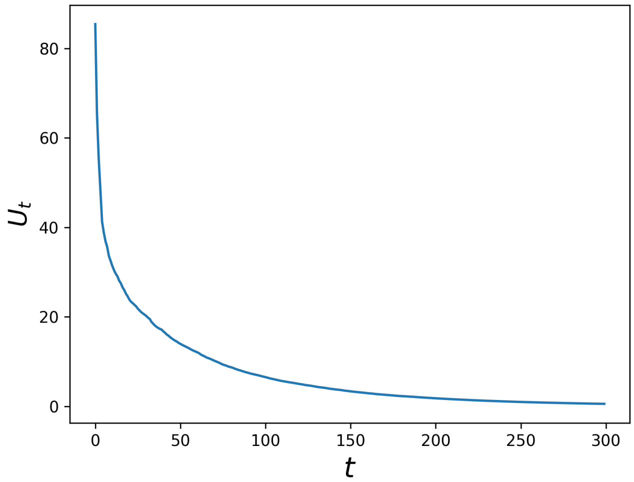

Figure 3: t versus U_t

$U_2 = 65.450396, U_T = 0.540149$.

11

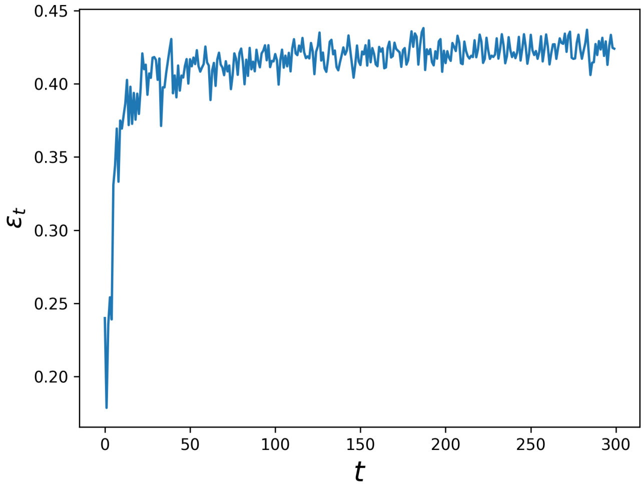

Figure 4: t versus ϵ_t

The minimum $\epsilon = 0.178728070175$

12

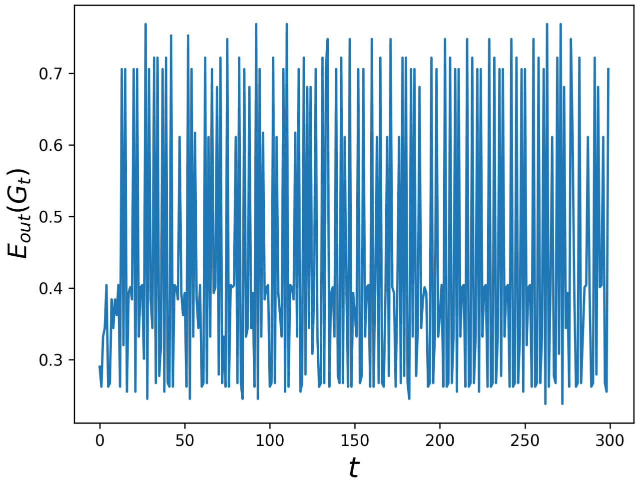

Figure 5: t versus $E_{out}(g_t)$

$E_{out}(G_1) = 0.29$

13

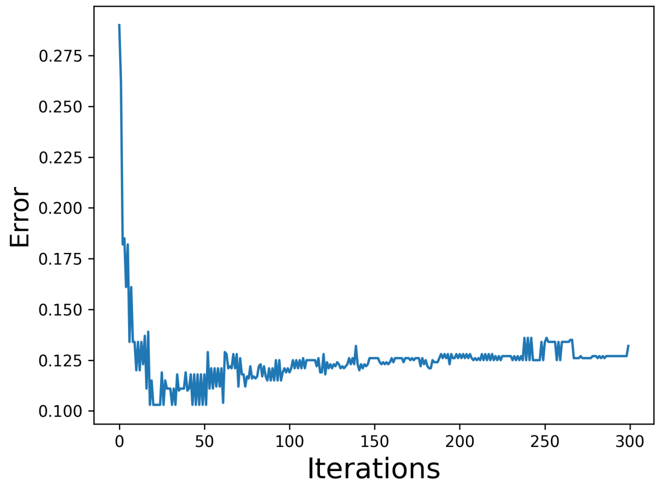

Figure 6: t versus $E_{out}(G_t)$

$E_{out}(G) = 0.132$

14

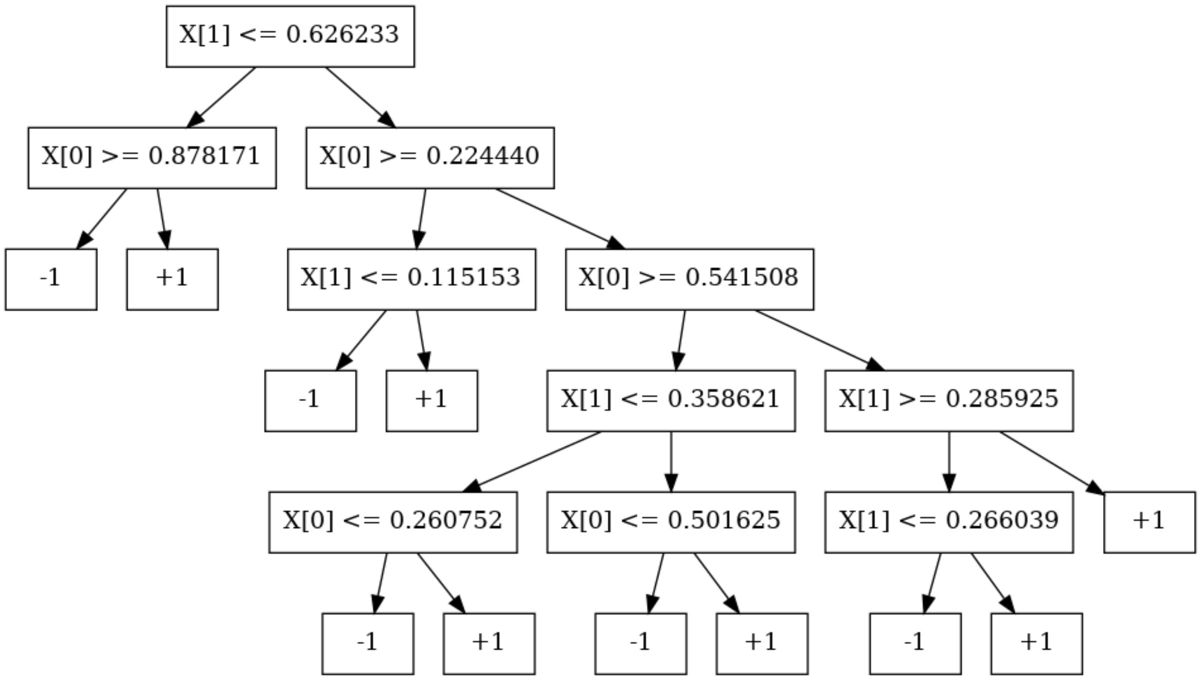

Figure 7: Decision Tree

Note: left path implies if the node predict false, while right implies true.

15

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

16

Purining three leaves results in same lowest $E_{in} = 0.01$, while the corresponding $E_{out} = 0.144, E_{out} = 0.117, E_{out} = 0.109$.

17

$$\begin{aligned}
U_{t+1} = & \sum_{i=1}^{N} u_i^{(t+1)} \\
= & \sum_{i=1}^{N} \begin{cases} u_i^{(t)} \cdot \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} & g_t(x_i) \neq y_i \\ u_i^{(t)} / \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} & g_t(x_i) = y_i \end{cases} \\
= & \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} \sum_{\{i|g_t(x_i)\neq y_i\}} u_i^{(t)} + \sqrt{\frac{\epsilon_t}{1-\epsilon_t}} \sum_{\{i|g_t(x_i)=y_i\}} u_i^{(t)} \\
= & \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} \cdot \epsilon_t U_t + \sqrt{\frac{\epsilon_t}{1-\epsilon_t}} \cdot (1-\epsilon) U_t \\
= & U_t \cdot 2\sqrt{\epsilon_t(1-\epsilon_t)} \\
\leq & U_t \cdot 2\sqrt{\epsilon(1-\epsilon)}
\end{aligned}$$

The last $\leq$ is because

$$\begin{aligned}
& \epsilon(1-\epsilon) - \epsilon_t(1-\epsilon_t) \\
= & \epsilon - \epsilon^2 - \epsilon_t + \epsilon_t^2 \\
= & (\epsilon_t + \epsilon)(\epsilon_t - \epsilon) + \epsilon - \epsilon_t \\
= & (\epsilon_t + \epsilon - 1)(\epsilon_t - \epsilon) \geq 0 & \because \epsilon_t + \epsilon \leq 1; \epsilon_t \leq \epsilon \\
\implies & \epsilon_t(1-\epsilon_t) \leq \epsilon(1-\epsilon)
\end{aligned}$$

18

$$\begin{aligned}
E_{in}(G_{t+2}) \leq U_{t+1} = & U_t \cdot 2\sqrt{\epsilon_t(1-\epsilon_t)} \\
\leq & U_t \cdot 2\sqrt{\epsilon(1-\epsilon)} \\
\leq & U_t \exp(-2(\frac{1}{2}-\epsilon)^2)
\end{aligned}$$

Since $\exp(-2(\frac{1}{2}-\epsilon)^2) > 1$, after $T = \lceil \log 2N / \log \exp(-2(\frac{1}{2}-\epsilon)^2) \rceil$ iterations,

$$\begin{aligned}
E_{in}(G_{T-1}) \leq U_T \leq & \exp(-2(\frac{1}{2}-\epsilon)^2)^{-T} U_1 \\
\leq & \frac{1}{2N} U_1 \\
\leq & \frac{1}{2N} \\
= & 0 & \because \ E_{in} \in \{0, \frac{1}{N}, \frac{2}{N}, \cdots, 1\}
\end{aligned}$$

Therefore, after $T = O(\log(N))$ iterations, $E_{in}(G_T) = 0$.