Machine Learning Techniques HW3

B04902004 王佑安

1. $1 - \mu_+^2 - \mu_-^2$
   $= 1 - \mu_+^2 - 1 + 2\mu_+ - \mu_+^2$
   $= -2(\mu_+ - 0.5)^2 + 0.5$
   $\Rightarrow 1 - \mu_+^2 - \mu_-^2$ has maximum value 0.5

2. Normalized Gini index $= \frac{-2(\mu_+ - 0.5)^2 + 0.5}{0.5} = -4\mu_+^2 + 4\mu_+$
   is equivalent to **[b]**
   $\mu_+(1 - (\mu_+ - \mu_-))^2 + \mu_-(-1 - (\mu_+ - \mu_-))^2$
   $= \mu_+(2 - 2\mu_+)^2 + (1 - \mu_+)4\mu_+^2$
   $= -4\mu_+^2 + 4\mu_+$

3. Each example has $(1 - \frac{1}{N})^{pN}$ probability not be sampled.
   $\lim\limits_{n \to \infty} (1 - \frac{1}{N})^N = e^{-1}$
   $\mathbb{E}[\#$ examples not be sampled$] = N(1 - \frac{1}{N})^{pN} \approx e^{-p} \cdot N$

4. For worst case, every examples are predicted wrong exactly $\frac{K+1}{2}$ times.
   So max $E_{out} = \frac{\sum_{k=1}^{K} e_k}{\frac{K+1}{2}} = \frac{2}{K+1} \sum_{k=1}^{K} e_k$

5. $\frac{\partial(\frac{1}{N}\sum_{n=1}^{N}((0+\eta g_1(\mathbf{x}_n)) - y_n)^2)}{\partial \eta}$
   $= \frac{1}{N} \sum_{n=1}^{N} 4(2\eta - y_n) = 0$
   $\Rightarrow \eta = \frac{\sum_{n=1}^{N} y_n}{2N}$
   $s_n = \eta g_1(x_n) = \frac{\sum_{n=1}^{N} y_n}{N}$

6. $\frac{\partial(\frac{1}{N}\sum_{n=1}^{N}(s_n - y_n)^2)}{\partial \eta}$
   $= \frac{1}{N} \sum_{n=1}^{N} 2g_t(x_n)(s_n - y_n)$
   $= 2(\frac{1}{N} \sum_{n=1}^{N} g_t(x_n)s_n - \frac{1}{N} \sum_{n=1}^{N} g_t(x_n)y_n) = 0$
   $\Rightarrow \frac{1}{N} \sum_{n=1}^{N} s_n g_t(x_n) = \frac{1}{N} \sum_{n=1}^{N} y_n g_t(x_n)$

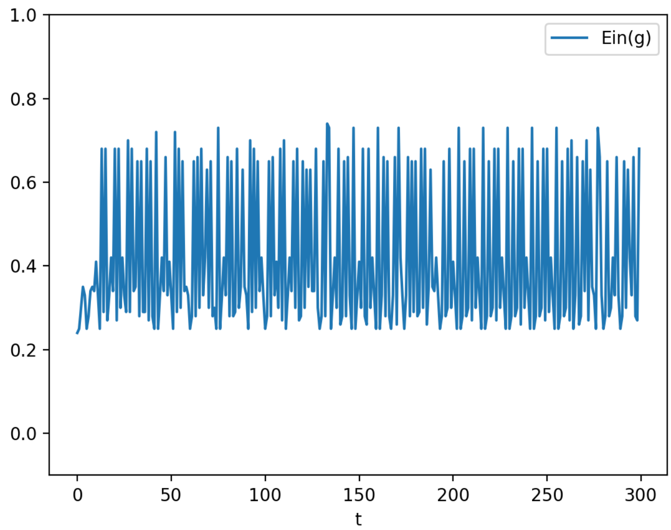
7. Let $g_t(\mathbf{x}) = w_t\mathbf{x} + b_t$, and w_t, b_t is optimal.
   $G_2(\mathbf{x}) = \alpha_1(w_1\mathbf{x} + b_1) + \alpha_2(w_2\mathbf{x} + b_2)$
   $= (\alpha_1 w_1 + \alpha_2 w_2)\mathbf{x} + (\alpha_1 b_1 + \alpha_2 b_2)$
   Suppose G_2 has optimal solution $w' = \alpha_1 w_1 + \alpha_2 w_2$, $b' = \alpha_1 b_1 + \alpha_2 b_2$
   If $g_2(\mathbf{x}) \neq 0$
   $\Rightarrow w_2 \neq 0$ or $b_2 \neq 0$
   $\Rightarrow w' \neq \alpha_1 w_1$ or $b' \neq \alpha_1 b_1$
   $\Rightarrow G_1(\mathbf{x}) = \alpha_1(w_1\mathbf{x} + b_1)$ isn't optimal.
   So $g_2(\mathbf{x})$ must be 0.

8.
$$w_i = \begin{cases} 1 & x_i = +1 \\ 0 & x_i = -1 \end{cases}$$
$g_A(\mathbf{x}) = +1$ if there's any $x_i = +1$, which is equivalent to OR operation.

9. $\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} \cdot (x_i^{(l)})$
$= \delta_j^{(l)} \cdot (\tanh(\sum_i w_{ij}^{(l-1)} x_i^{(l-2)}))$
$= \delta_j^{(l)} \cdot (\tanh(0))$
$= \delta_j^{(l)} \cdot 0$
$= 0$
$\Rightarrow$ for each i, j, l, $\frac{\partial e_n}{\partial w_{ij}^{(l)}} = 0$

10. $\frac{\partial e}{\partial s_k^{(L)}} = \frac{\partial v_k \ln q_k}{\partial q_k} \frac{q_k}{\partial s_k^{(L)}}$
$= -\frac{v_k}{q_k} \frac{\exp(s_k^{(L)})(\sum_{i=1}^K \exp(s_i^{(L)}) - \exp(s_i^{(L)}))}{\sum_{i=1}^K \exp(s_i^{(L)})}$
$= -\frac{v_k}{q_k} q_k(1 - q_k)$
$= q_k - v_k$

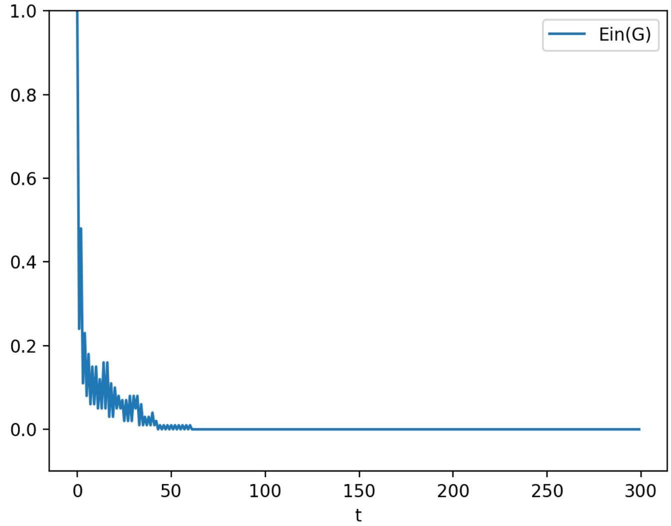
11.



$E_{in}(g_1) = 0.24$, $\alpha_1 = 0.576$

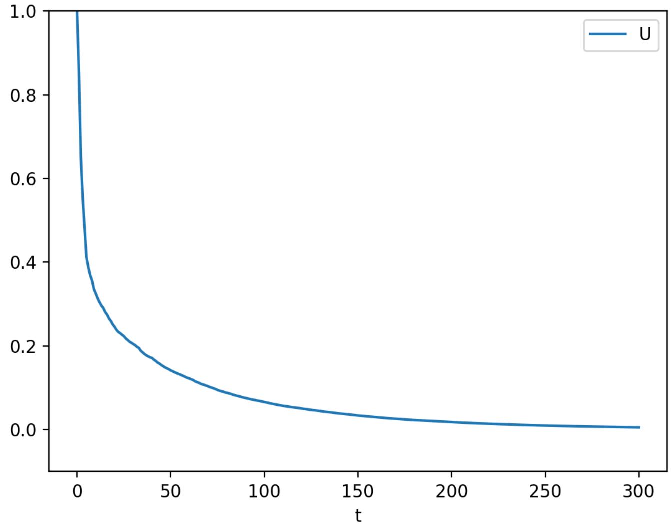
12. $E_{in}(g_t)$ 沒有明顯的增減，因為 g_t 的目的是 minimize $E_{in}(G_{t_1} + \alpha_t g_t)$，跟 $E_{in}(g_t)$ 沒有直接的遞增或遞減關係。

13.



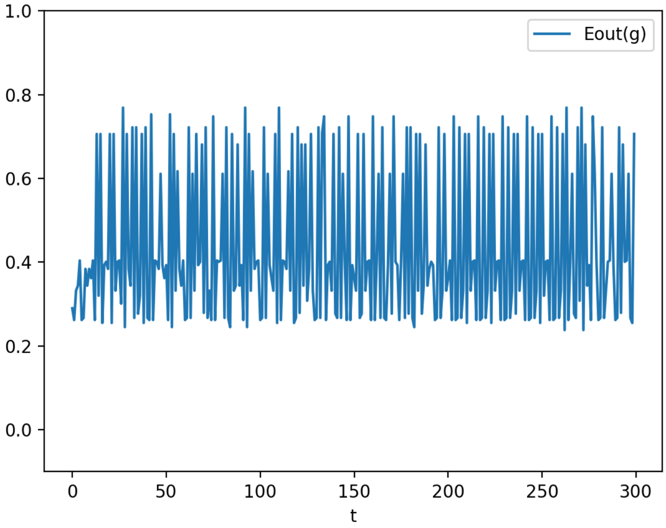
$E_{in}(G) = 0$

14.



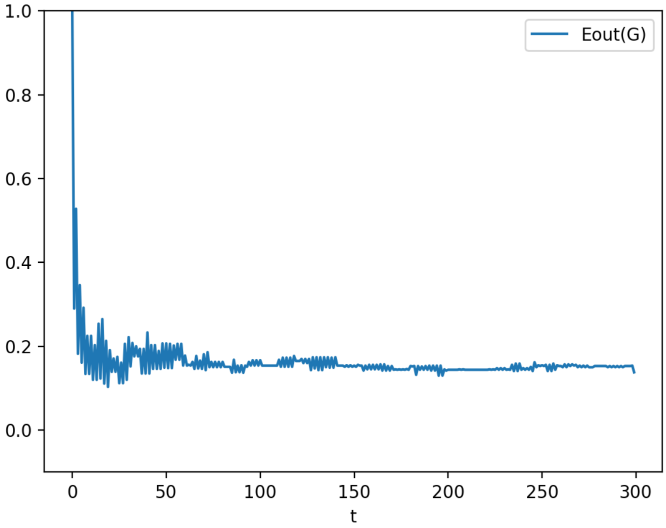
$U_2 = 0.854$, $U_T = 0.005$

15.



$E_{out}(g_1) = 0.29$

16.



$E_{out}(G) = 0.138$

17. $U_{t+1} = \sum_{i=1}^{N} \begin{cases} u_i^t \cdot \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} & y_n \neq g_t(x_n) \\ u_i^t / \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} & y_n = g_t(x_n) \end{cases}$

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

$= \sqrt{\frac{1-\epsilon_t}{\epsilon_t}} \epsilon_t U_t + \sqrt{\frac{\epsilon_t}{1-\epsilon_t}} (1-\epsilon_t) U_t$

$= U_t 2\sqrt{\epsilon_t(1-\epsilon_t)}$

$\leq U_t 2\sqrt{\epsilon(1-\epsilon)}$ (Since $\epsilon_t \leq \epsilon < 1$, $\epsilon_t(1-\epsilon_t) = \epsilon_t - \epsilon_t^2 \leq \epsilon - \epsilon^2 = \epsilon(1-\epsilon)$)

18. $E_{in}(G_T) \leq U_{T+1}$
$\leq U_T 2\sqrt{\epsilon_t(1-\epsilon_t)}$
$\leq U_T \exp(-2(\frac{1}{2}-\epsilon_t)^2)$
$= \exp(-2(\frac{1}{2}-\epsilon_0)^2)^T$
Let $\exp(-2(\frac{1}{2}-\epsilon_0)^2)^T = \lim_{n\to\infty} \frac{1}{n} = 0$
$\Rightarrow T\log(\exp(-2(\frac{1}{2}-\epsilon_0)^2)) = \lim_{n\to\infty} -\log n$
$\Rightarrow T = \lim_{n\to\infty} \frac{\log n}{2(\frac{1}{2}-\epsilon_o)^2} = O(\log n)$
$\Rightarrow$ When n is large, $E_{in}(G_{O(logn)}) \leq \frac{1}{n} \approx 0$