ML Technique Hw3 r06922114 陳政元.

1. Gini Index $= 1-\mu_+^2-\mu_-^2 = 1-\mu_+^2-(1-\mu_+)^2 = -2\mu_+^2+2\mu_+$.

$$0=\frac{\partial(-2\mu_+^2+2\mu_+)}{\partial\mu_+} = -4\mu_++2 \Leftrightarrow \mu_+=\frac{1}{2}$$

and by $\frac{\partial^2(-2\mu_+^2+2\mu_+)}{\partial\mu_+^2}=-4$, we know Gini index has max ($1-(\frac{1}{2})^2-(\frac{1}{2})^2=\frac{1}{2}$) as $\mu_+=\frac{1}{2}$

2. normalized Gini index $= \frac{1-\mu_+^2-\mu_-^2}{1/2} = -4\mu_+^2+4\mu_+$, denoted as $G(\mu_+)$

(a) $2\min(\mu_+,\mu_-)=0.2$ as $\mu_+=0.1$

$G(\mu_+)=0.36$ as $\mu_+=0.1$ (X)

(b) $\mu_+(1-(\mu_+-\mu_-)^2)+\mu_-(-1-(\mu_+-\mu_-)^2) = \mu_+(1-(2\mu_+-1))^2+(1-\mu_+)(-1-(2\mu_+-1))^2$

$=\mu_+(2-2\mu_+)^2+(1-\mu_+)(2\mu_+)^2 = 4\mu_+-8\mu_+^2-4\mu_+^2 = -4\mu_+^2+4\mu_+$

$$0=\frac{\partial(-4\mu_+^2+4\mu_+)}{\partial\mu_+} = -8\mu_++4 \Leftrightarrow \mu_+=\frac{1}{2},\ \max=\frac{1}{2}(1-0)^2+\frac{1}{2}(-1-0)^2=1$$

its normalized form $=\frac{-4\mu_+^2+4\mu_+}{1} = -4\mu_+^2+4\mu_+$ (V)

(c) $-\mu_+\ln\mu_+-(1-\mu_+)\ln(1-\mu_+)$

$$0=\frac{\partial\mu_+\ln\mu_+-(1-\mu_+)\ln(1-\mu_+)}{\partial\mu_+} = \ln\mu_++1+\ln(1-\mu_+)+1$$

$$= 2+\ln\mu_+(1-\mu_+)$$

$\Leftrightarrow \ln\mu_+(1-\mu_+)=\frac{1}{4} \Leftrightarrow \mu_+=\frac{1}{2}$, $\max=\ln(2)$

its normalized form $=\frac{-\mu_+\ln\mu_+-(1-\mu_+)\ln(1-\mu_+)}{\ln(2)}$ (X)

(d). normalized form $=\frac{1-|2\mu_+-1|}{1}$ (X).

Choose (b).

3. $(1-\frac{1}{N})^{PN}$ is the probability of not been sampled after PN times.

$(1-\frac{1}{N})^{PN} \longrightarrow \lim_{N\to\infty}(1-\frac{1}{N})^{PN} = e^{-P}$

So, $e^{-P}\cdot N$ of the samples will not be sampled at all.

---

4. For a fixed x_n, $E_{out}(G(x_n)) = E_{out}(\Sigma g_k(x_n))$

Suppose there are m kinds of g_k s.t. $E_{out}(g_k(x_n)) = 1$

we have $k-m$ kinds of g_k s.t. $E_{out}(g_k(x_n)) = 0$

i) $E_{out}(G(x_n)) = \underbrace{1}_{\text{as } m>k-m} \le \frac{1}{m}\sum_{k=1}^{k} E_{out}(g_k(x_n))$ as $m>k-m$.

ii) $E_{out}(G(x_n)) = \underbrace{0}_{\text{as } m<k-m} \le \frac{1}{m}\sum_{k=1}^{k} E_{out}(g_k(x_n))$ as $m>k-m$

by i) ii), we have $E_{out}(G(x_n)) \le \frac{1}{m}\sum_{k=1}^{k} E_{out}(g_k(x_n))$ as $m>k-m$

Consider all x_n, $E_{out}(G) = \frac{1}{N}\sum_{n=1}^{N} E_{out}(G(x_n)) \le \frac{1}{N}\sum_{n=1}^{N}\frac{1}{m}\sum_{k=1}^{k} E_{out}(g_k(x_n))$

$$= \frac{1}{m}\sum_{k=1}^{k}\frac{1}{N}\sum_{n=1}^{N} E_{out}(g_k(x_n))$$

$$= \frac{1}{m}\sum_{k=1}^{k} e_k \quad \text{as } m>k-m.$$

So, $E_{out}(G) \le \frac{1}{m}\sum_{k=1}^{k} e_k \le \frac{2}{k+1}\sum_{k=1}^{k} e_k$ (since $m>k-m$)

5. $s_n^{(1)} \leftarrow 0 + \alpha_1 g_1(x_n)$.

Obtain α_1 by $0 = \frac{\partial \sum (y_n - 0) - \eta g_1(x_n))^2}{\partial \eta} = -4\sum(y_n - 2\eta)$.

$\Leftrightarrow \eta = \frac{1}{2N} \sum y_n = \alpha_1$.

$s_n^{(1)} = \alpha_1 \cdot g_1(x_n) = \frac{1}{2N} \sum y_n \cdot 2 = \frac{1}{N} \sum y_n$ #

---

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

$= -2 \sum ((y_n - s_n) - \eta g_t(x_n)) g_t(x_n)$

$\eta = \frac{\sum g_t(x_n)(y_n - s_n)}{\sum g_t^2(x_n)} = \alpha_t$.

$\sum s_n^{(t)} g_t(x_n) = \sum (s_n^{(t-1)} + \alpha_t g_t(x_n)) g_t(x_n) = \sum s_n^{(t-1)} g_t(x_n) + \sum g_t(x_n)(y_n - s_n^{(t-1)})$

$= \sum g(x_n) y_n$ #

---

7. $s_n^{(1)} \leftarrow 0 + \alpha_1 g_1(x_n)$

obtain α_1 by $0 = \frac{\partial \sum_{n=1}^{N} ((y_n - 0) - 2\eta)^2}{\partial \eta} = -4 \sum_{n=1}^{N} (y_n - 2\eta)$.

$\eta = \frac{1}{2N} \sum_{n=1}^{N} y_n = \alpha_1$

$s_n^{(1)} = \frac{1}{N} \sum_{n=1}^{N} y_n$

$\min_{g_2} \sum_{n=1}^{N} (g_2(x_n) - (y_n - s_n^{(1)}))^2 = \sum (y_n - (s_n^{(1)} + g_2(x_n))^2$

Now, we find an fixed x minimize $\sum (y_n - x)^2$ by $0 = \frac{\partial \sum (y_n - x)^2}{\partial x}$

$= \sum (y_n - x)$.

$x = \frac{1}{N} \sum y_n = s_n^{(1)}$

so, $g_2(x_n) = 0$ #

$1-d+1$

$-(d-1)+1$ $\quad -(d-1)+1+d-1$

8.

let $w_0 = d-1$

$w_1 = \dots = w_d = 1$

if $(x_1, \dots, x_d) = (-1, \dots, -1)$ $\quad \sum_{i=0}^{d} w_i x_i = -1$

if at least one x_i is 1, $\quad \sum_{i=0}^{d} w_i x_i \geq 1$

This is $OR(x_1, \dots, x_d)$.

---

9. Output layer: $\frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = -2(y_n - s_1^{(L)}) \underbrace{(x_i^{(L-1)})}_{0} = 0$

hidden layer: $\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} x_i^{(l-1)}$, where $\delta_j^{(l)} = \sum \delta_k^{(l+1)} \underbrace{w_{jk}^{(l+1)}}_{0} \tanh(s_j^{(l)}) = 0$

all layer's gradient is 0.

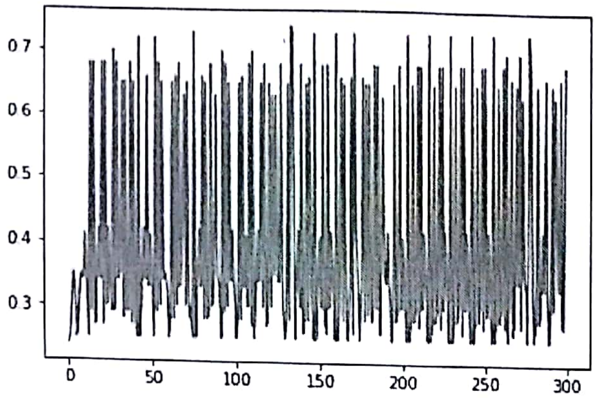
---

10. $\frac{\partial e}{\partial s_k^{(L)}} = \frac{\partial - \sum_{k=1}^{K} v_k (s_k^{(L)} - \ln \sum \exp s_k^{(L)})}{\partial s_k^{(L)}} = -\sum v_k (1 - \underbrace{\frac{\exp s_k^{(L)}}{\sum \exp s_k^{(L)}}}_{q_k}) =$

$= -1(1-q_k)$ when $y=k$

$= q_k - 1 = q_k - v_k$

11.



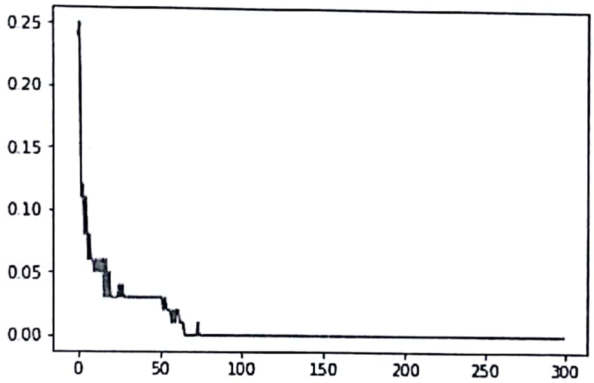
Ein(g1) = 0.24 alpha_1 = 0.576

12.

Ein(g_t) should varies as t increases.

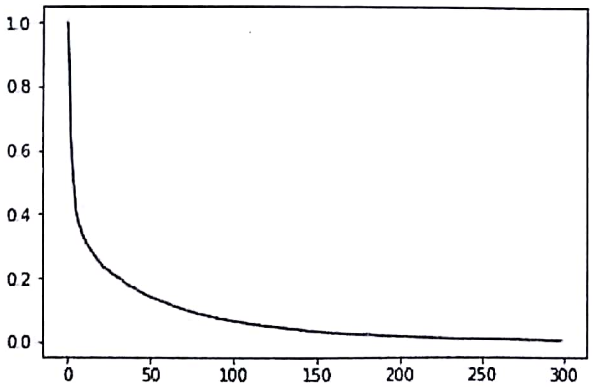
We want g_t be very different, so we can learn a better G by diversity.

13.



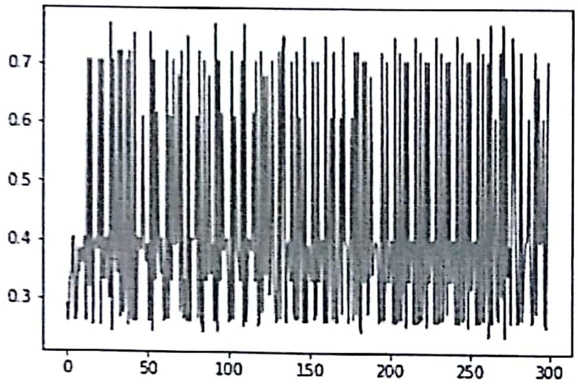
Ein(G) = 0

14.



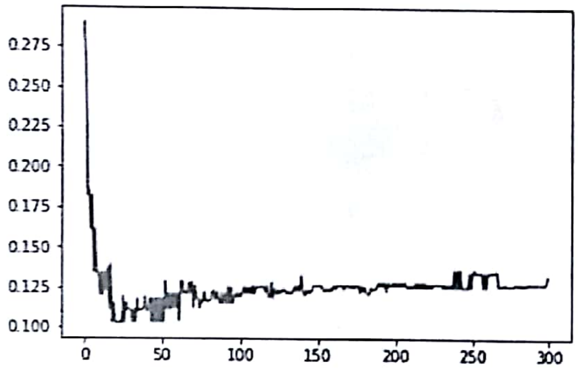
U_2 = 0.854 U_T = 0.0054

15.



Eout(g_1) = 0.29

16.



Eout(G) = 0.132

17. $$\frac{U_{t+1}}{U_t} = \frac{\sum u_n^{(t+1)}}{\sum u_n^{(t)}} = \frac{\sum_{x_n\text{分對}} u_n^{(t)}}{\sum u_n^{(t)}} \Big/ \blacklozenge_t + \frac{\sum_{x_n\text{分錯}} u_n^{(t)}}{\sum u_n^{(t)}} \cdot \blacklozenge_t$$

$$= (1-\varepsilon_t)\sqrt{\frac{\varepsilon_t}{1-\varepsilon_t}} + \varepsilon_t\sqrt{\frac{1-\varepsilon_t}{\varepsilon_t}} = 2\sqrt{(1-\varepsilon_t)\varepsilon_t}$$

18. $E_{in}(G_T) \le U_{T+1} \le U_T \cdot 2\sqrt{\varepsilon(1-\varepsilon)} \le 2^T\left(\sqrt{\varepsilon(1-\varepsilon)}\right)^T$

$\le 2^T\left(\frac{1}{2}\right)^T \exp\left(-2T\left(\frac{1}{2}-\varepsilon\right)^2\right) = \exp\left(-2T\left(\frac{1}{2}-\varepsilon\right)^2\right)$

We say $E_{in}(G_T)=0$ as $E_{in}(G_T) < \frac{1}{N}$

$\exp\left(-2T\left(\frac{1}{2}-\varepsilon\right)^2\right) < \frac{1}{N} \Leftrightarrow 2T\left(\frac{1}{2}-\varepsilon\right)^2 > \log N \Leftrightarrow T > \frac{\log N}{2\left(\frac{1}{2}-\varepsilon\right)^2}$

So, $T = O(\log N)$