ML HW3 r07922003 劉濬慶

1. Gini impurity $= 1-\sum_{k=1}^{K}\mu_k^2$

$[\mu_1, \mu_2, \ldots, \mu_K]$; $\mu_k \geq 0$ and $\sum_{k=1}^{K}\mu_k = 1$

Sol: 套用柯西不等式

$$(\mu_1^2+\mu_2^2+\cdots+\mu_K^2)(1^2+1^2+\cdots+1^2) \geq \underbrace{(\mu_1+\mu_2+\cdots+\mu_K)}_{=1}{}^2$$

$$\Rightarrow \mu_1^2+\mu_2^2+\cdots+\mu_K^2 \geq \frac{1}{K}$$

$$\Rightarrow \underbrace{1-\sum_{k=1}^{K}\mu_k^2 \leq 1-\frac{1}{K}} \quad \#$$

A: $1-\frac{1}{K}$

2.

squared regression error : $\mu_+(1-\mu_++\mu_-)^2+\mu_-(-1-\mu_++\mu_-)^2$

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

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

$$= 4\mu_+-8\mu_+^2+4\mu_+^3+4\mu_+^2-4\mu_+^3$$

$$= 4\mu_+-4\mu_+^2$$

$$= \underline{4(\mu_+-\mu_+^2)}$$

Gini impurity : $1-\mu_+^2-\mu_-^2$

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

$$= 2\mu_+-2\mu_+^2$$

$$= \underline{2(\mu_+-\mu_+^2)}$$

A: squared regression error is a scale version of the Gini impurity

3. 一个 g_t OOB 之机率 $= (1-\frac{1}{N})^{N'} = (1-\frac{1}{N})^{PN} = \frac{1}{((1+\frac{1}{N-1})^N)^P} \approx \frac{1}{e^P}$

N examples OOB 的数量大約 $= N \times \frac{1}{e^P} = \underline{Ne^{-P}}$ #

4. K binary classification trees $\{g_k\}_{k=1}^K$, k = odd ; $E_{out}(g_k) = e_k$

一个 data 被認為錯誤必須滿足 $\frac{k+1}{2}$ binary classification trees 分類為 error

total error $= E_{out}(g_1) + E_{out}(g_2) + \cdots + E_{out}(g_k)$

$= e_1 + e_2 + \cdots + e_k$

$$\frac{\text{total error}}{\text{one data error}} = \frac{e_1+e_2+\cdots+e_k}{\frac{k+1}{2}} = \underline{\frac{2}{k+1} \times \sum_{k=1}^{K} e_k} \text{ 得證}$$

最極端之情形

5. $g_1(x) = 11.26$; $S_n = 0$

$$F = \frac{1}{N}\sum_{n=1}^{N}\left[(y_n - s_n) - \alpha_1 g_1(x_n)\right]^2$$

$$\frac{\delta F}{\delta \alpha_1} = \frac{-2}{N}\sum_{n=1}^{N} g_1(x_n)(y_n - s_n - \alpha_1 g_1(x_n)) = 0$$

$\Rightarrow \sum_{n=1}^{N} g_1(x_n)(y_n - s_n - \alpha_1 g_1(x_n)) = 0$ 代入 $g_1(x)$ 及 S_n

$$\Rightarrow \sum_{n=1}^{N} 11.26\,(y_n - 11.26\alpha_1) = 0$$

$$\Rightarrow (y_1 - 11.26\alpha_1) + (y_2 - 11.26\alpha_1) + \cdots + (y_n - 11.26\alpha_1) = 0$$

$$\Rightarrow \sum_{n=1}^{N} y_n = 11.26N\alpha_1$$

$$\Rightarrow \underline{\alpha_1 = \frac{1}{11.26N}\sum_{n=1}^{N} y_n} \quad \#$$

A. $\underline{\frac{\sum_{n=1}^{N} y_n}{11.26N}}$

6.

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

$$\Rightarrow \alpha_t \sum_{n=1}^{N} g_t^2(x_n) = \sum_{n=1}^{N} g_t(x_n)(y_n - s_n^{(t-1)}) = \sum_{n=1}^{N} g_t(x_n) y_n - \sum_{n=1}^{N} g_t(x_n) s_n^{(t-1)}$$

$$\Rightarrow \sum_{n=1}^{N} g_t(x_n) \cdot s_n^{(t-1)} = \sum_{n=1}^{N} g_t(x_n) y_n - \alpha_t \sum_{n=1}^{N} g_t^2(x_n)$$

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

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

$$= \underline{\sum_{n=1}^{N} g_t(x_n) y_n} \quad \#$$

$$A: \sum_{n=1}^{N} y_n g_t(x_n)$$

8. $g_A(x) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right)$ implement $OR(x_1, x_2, \ldots, x_d)$

最壞的情況為全錯 $(-1, -1, \ldots, -1)$

A: $\underline{w_0 = d-1,\ w_1 = w_2 = \cdots = w_d = 1}$ #

9. $e_n = (y_n - NNet(x_n))^2 = (y_n - s_1^{(L)})^2 = \left(y_n - \sum_{i=0}^{d} w_{i1}^{(L)} x_i^{(L-1)}\right)^2$

$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = -2(y_n - s_1^{(L)})(x_i^{(L-1)})$$

$$\frac{\partial e_n}{\partial w_{ij}^{(\ell)}} = \delta_j^{(\ell)} x_i^{(\ell-1)}$$

$$\delta_j^{(\ell)} = \sum_k (\delta_k^{(\ell+1)})(w_{jk}^{(\ell+1)})(\tanh'(s_j^{(\ell)}))$$

$\because w_{ij}^{(\ell)} = 0 \quad \therefore s_j^{(\ell)} = 0,\ (\ell = 1, \ldots, L) \Rightarrow x_i^{(\ell)} = 0,\ \ell \geq 1,\ i > 0$

$1 \leq \ell \leq L$; $\delta_j^{(\ell)} = 0$, 梯度 $\frac{\partial e_n}{\partial w_{ij}^{(\ell)}} = 0$

A: $\ell = L$; if $i > 0$, 梯度 $\frac{\partial e_n}{\partial w_{i1}^{(L)}} = 0$

if $i = 0$, 梯度 $\frac{\partial e_n}{\partial w_{01}^{(L)}} = -2 y_n x_0^{(L-1)}$ (可能不為 0)

10.

$$e = -\sum_{k=1}^{K} v_k \ln q_k$$

$$= -\sum_{k=1}^{K} v_k \ln\left(\frac{\exp(s_k^{(L)})}{\sum_{k=1}^{K} \exp(s_k^{(L)})} \right)$$

$$= -\sum_{k=1}^{K} v_k \left(s_k^{(L)} - \ln\left(\sum_{k=1}^{K} \exp(s_k^{(L)}) \right) \right)$$

$$= \sum_{i=1, i\neq k}^{K} v_i \left(\ln\left(\sum_{j=1}^{K} \exp(s_j^{(L)}) \right) - s_i \right) + v_k \left(\ln\left(\sum_{j=1}^{K} \exp(s_j^{(L)}) \right) \right) - v_k s_k^{(L)}$$

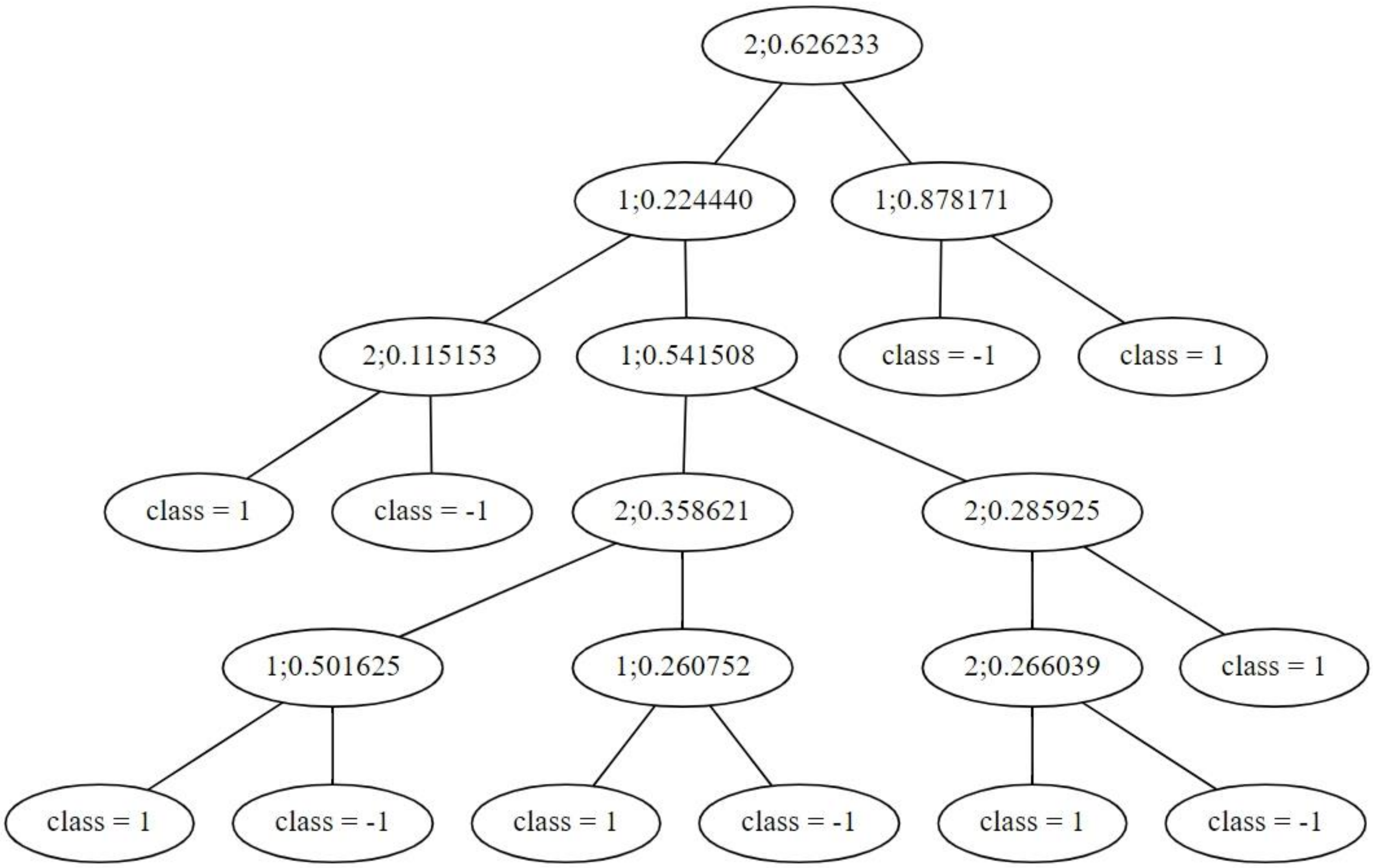
$$\frac{\partial e}{\partial s_k^{(L)}} = \sum_{i=1, i\neq k}^{K} v_i \times \frac{s_k^{(L)}}{\sum_{j=1}^{K} \exp(s_j^{(L)})} + v_k \frac{s_k^{(L)}}{\sum_{j=1}^{K} \exp(s_j^{(L)})} - v_k$$

$$= \sum_{i=1, i\neq k}^{K} v_i q_k + v_k q_k - v_k$$

$$= (v_1 + v_2 + \cdots + v_K) q_k - v_k$$

$$= \underline{q_k - v_k} \quad \#$$

Q11

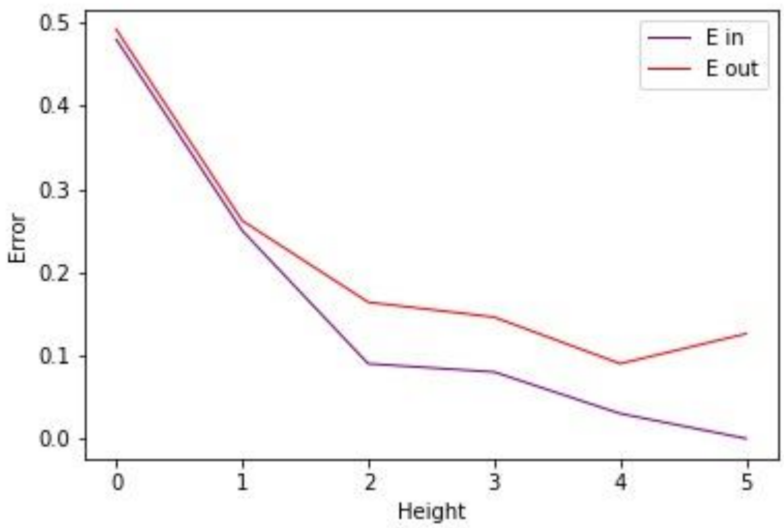


Q12

E in = 0.000

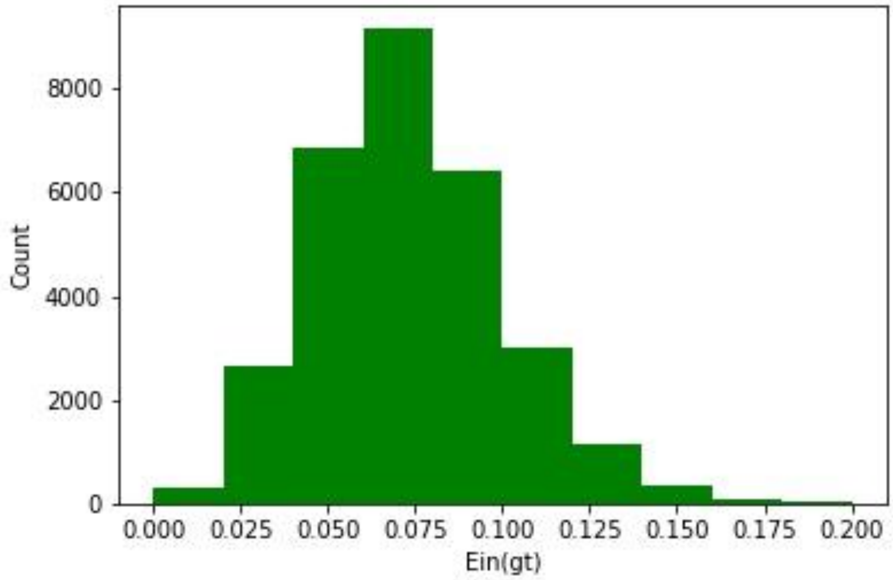
E out = 0.126

Q13



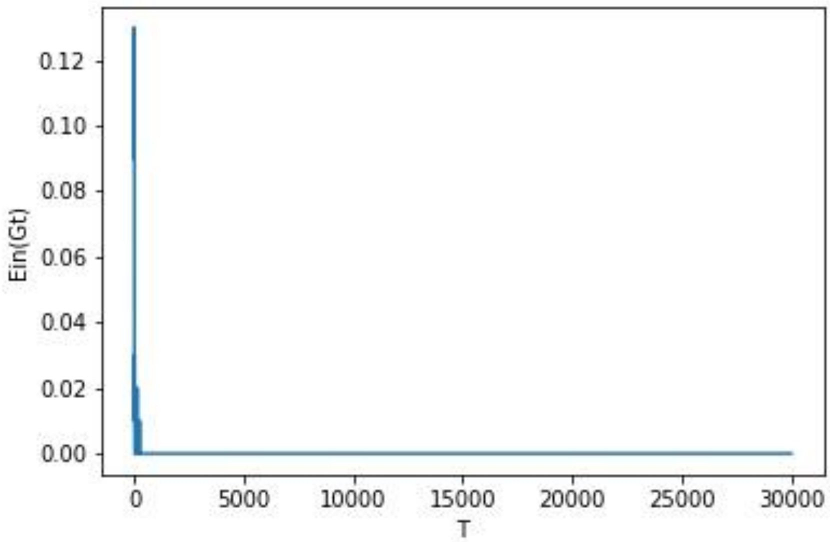
在 Height 為 4 時，Eout 最高，並不是 fully-grown tree 最好

Q14



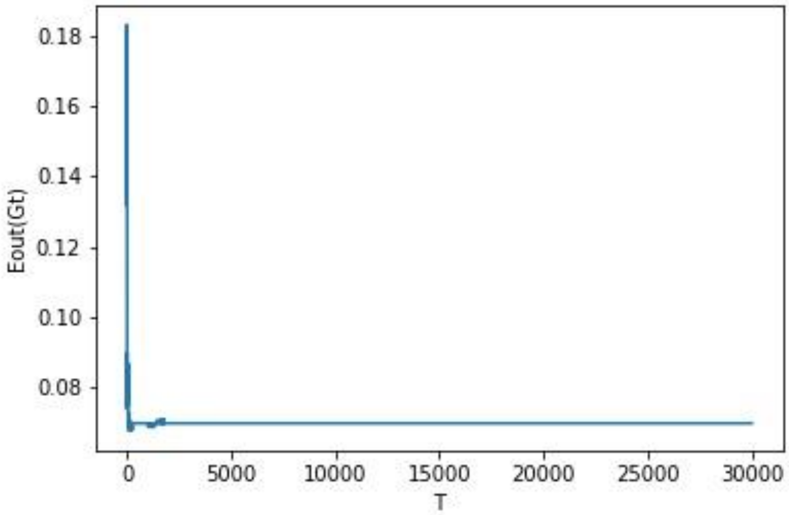
8000
6000
4000
2000
0
Count
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
Ein(gt)

Q15



0.12
0.10
0.08
0.06
0.04
0.02
0.00
Ein(Gt)
0
5000
10000
15000
20000
25000
30000
T

Q16



E_out 跟 E_in 最後都會趨於平坦平均維持在同一個 error