ML hw3

r07922100 楊力權

1.

柯西不等式 $|\overrightarrow{A}\cdot\overrightarrow{B}| \le |\overrightarrow{A}||\overrightarrow{B}| \Rightarrow |\overrightarrow{A}\cdot\overrightarrow{B}|^2 \le |\overrightarrow{A}|^2|\overrightarrow{B}|^2$

設$\overrightarrow{A} = (\mu_1, \mu_2, \cdots), \overrightarrow{B} = (1,1,\cdots)$得 $1^2 = (\sum_{k=1}^{K}\mu_k)^2 \le (\sum_{k=1}^{K}\mu_k^2)(1+1+\cdots) = K\sum_{k=1}^{K}\mu_k^2$

因此$\sum_{k=1}^{K}\mu_k^2 \ge \frac{1}{K}$且等號成立時$\frac{\mu_1}{1} = \frac{\mu_2}{1} = \cdots = \frac{\mu_K}{1}$

$\max_{\mu}$ Gini $= 1 - \sum_{k=1}^{K}\mu_k^2 = \frac{K-1}{K}$且$\mu_1 = \mu_2 = \cdots = \mu_K$

2.

square regression error = $\mu_+(1-(\mu_+-\mu_-))^2 + \mu_-(-1-(\mu_+-\mu_-))^2$

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

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

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

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

給定任一μ_+, μ_-，square regression error是Gini impurity的兩倍

因此為Gini impurity的scaled version。

3.

N很大的情況，example不被抽到的機率=$(1-\frac{1}{N})^{pN} = \frac{1}{(\frac{N}{N-1})^{pN}} = \frac{1}{(1+\frac{1}{N-1})^{Np}} \approx \frac{1}{e^p}$

因此約有Ne^{-p}個example不會被sample到的。

4.

一sample出錯，表示至少有過半數$\frac{K+1}{2}$個樹出錯，要求upper bound所以假設每個Error發生點都是剛好$\frac{K+1}{2}$個樹出錯，又假設總共有N個sample，共有$\sum_{k=1}^{K}Ne_k$棵樹出錯，因此Random Forest中有$\frac{\sum_{k=1}^{K}Ne_k}{\frac{K+1}{2}} = \frac{2N}{K+1}\sum_{k=1}^{K}e_k$個點出錯，則

$E_{out}(G) = \frac{1}{N}\frac{2N}{K+1}\sum_{k=1}^{K}e_k = \frac{2}{K+1}\sum_{k=1}^{K}e_k$

5.

確定$g_1(x) = 11.26$，$\alpha_1 = \arg\min_\eta \frac{1}{N}\sum_{n=1}^{N}((y_n - s_n) - \eta g_1(x))^2 = \arg\min_\eta \frac{1}{N}\sum_{n=1}^{N}(y_n - 11.26\eta)^2$

$\frac{\partial}{\partial\eta}\sum_{n=1}^{N}(y_n - 11.26\eta)^2 = 2N\eta(11.26)^2 - 2(11.26)\sum_{n=1}^{N}y_n = 0 \Rightarrow \alpha_1 = \frac{1}{11.26N}\sum_{n=1}^{N}y_n$

6.

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

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

$\Rightarrow \alpha_t = \eta = \frac{\sum_{n=1}^{N}g_t(x_n)(y_n - s_n^{(t-1)})}{\sum_{n=1}^{N}g_t^2(x_n)}$帶入上第一式可得$\sum_{n=1}^{N}s_n g_t(x_n)$

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

7.

$g_1(x) = w_1x + w_2x^2 + \cdots + b$，令$\sum_{n=1}^{N}(g_1(x_n) - (y_n - s_n))^2$偏微分為零，得optimal解為

$g_1^*(x) = w_1^*x + w_2^*x^2 + \cdots + b^*$，能使$\sum_{n=1}^{N}(g_1^*(x_n) - (y_n - s_n))^2$最小。

此時若存在一$\alpha_1 \neq 1$使$\sum_{n=1}^{N}(\alpha_1 g_1^*(x_n) - (y_n - s_n))^2$更小，表示必定存在一個

$g_1'(x) = \alpha_1(w_1^*x + w_2^*x^2 + \cdots + b^*) = w_1'x + w_2'x^2 + \cdots + b'$

能使$\sum_{n=1}^{N}(g_1'(x_n) - (y_n - s_n))^2 < \sum_{n=1}^{N}(g_1^*(x_n) - (y_n - s_n))^2$

因此$g_1^*(x)$不是optimal solution造成矛盾。因此optimal $\alpha_1 = 1$。

8.

設$x_0 = 1,\ w_0 = d - 1,\ w_i = 1,$ for $1 \leq i \leq d$

當全部$x_i = -1$，$g_A(x) = $ sign($\sum_{i=0}^{d}w_ix_i$) = sign($(d-1) - d$) $= -1$符合OR結果False。

當其中有一$x_n = +1$，$g_A(x) = $ sign($\sum_{i=0}^{d}w_ix_i$) = sign($(d-1) - (d-1) + 1$) $= +1$

符合OR結果True。愈多$x_i = +1$只會讓$\sum_{i=0}^{d}w_ix_i$愈大符合OR結果True。

9.

由output layer來看(第L-1層->output)，$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \frac{\partial err(s_1^{(L)}, y_n)}{\partial s_1^{(L)}} (x_i^{(L-1)})$

因為$w_{i1} = 0$所以$x_i^{(L-1)} = \tanh(s_i^{(L-1)}) = \tanh(\sum_{k=1}^{d^{(L-2)}} w_{ki}^{(L-1)} x_k^{(L-2)}) = \tanh(0) = 0$

由其他layer來看(第 $l-1$ 層->第 l 層)，$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} (x_i^{(l-1)})$

因為$w_{ij} = 0$所以$\delta_j^{(l)} = \frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{(l+1)}} \frac{\partial e_n}{\partial s_k^{(l+1)}} \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{(l+1)}} \delta_k^{(l+1)} w_{jk}^{(l+1)} \frac{\partial \tanh(s_j^{(l)})}{\partial s_j^{(l)}} = 0$

所有的gradient components皆為0。

10.

$$e = -\sum_{i=1}^{K} v_i \ln q_i$$

$$\frac{\partial e}{\partial s_k^{(L)}} = \frac{\partial e}{\partial q_k} \frac{\partial q_k}{\partial s_k^{(L)}} = \frac{\partial}{\partial q_k}(-v_k \ln q_k) \frac{\partial q_k}{\partial s_k^{(L)}} + \sum_{i \neq k} \frac{\partial}{\partial q_i}(-v_i \ln q_i) \frac{\partial q_i}{\partial s_k^{(L)}}$$

$$= (-v_k \frac{1}{q_k}) \frac{\partial q_k}{\partial s_k^{(L)}} + \sum_{i \neq k} (-v_i \frac{1}{q_i}) \frac{\partial q_i}{\partial s_k^{(L)}}$$

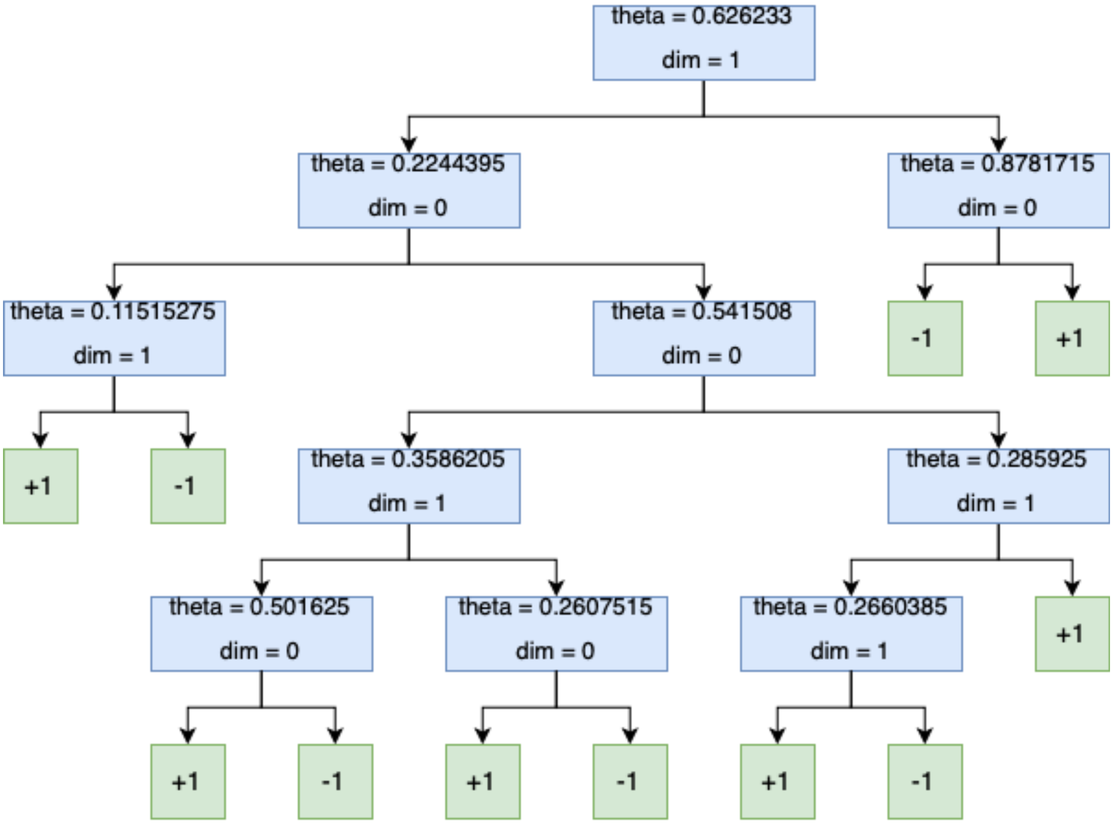
$$= (-v_k \frac{1}{q_k}) \frac{(\sum_{i=1}^{K} \exp(s_i^{(L)})) \exp(s_k^{(L)}) - \exp(s_k^{(L)})^2}{(\sum_{i=1}^{K} \exp(s_i^{(L)}))^2} + \sum_{i \neq k} (-v_i \frac{1}{q_i}) \frac{-\exp(s_i^{(L)}) \exp(s_k^{(L)})}{(\sum_{i=1}^{K} \exp(s_i^{(L)}))^2}$$

$$= (-v_k \frac{1}{q_k})(q_k - q_k^2) + \sum_{i \neq k} (-v_i \frac{1}{q_i})(-q_k q_i) = -v_k(1 - q_k) + \sum_{i \neq k} v_i q_k = -v_k + v_k q_k + \sum_{i \neq k} v_i q_k$$

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

11.

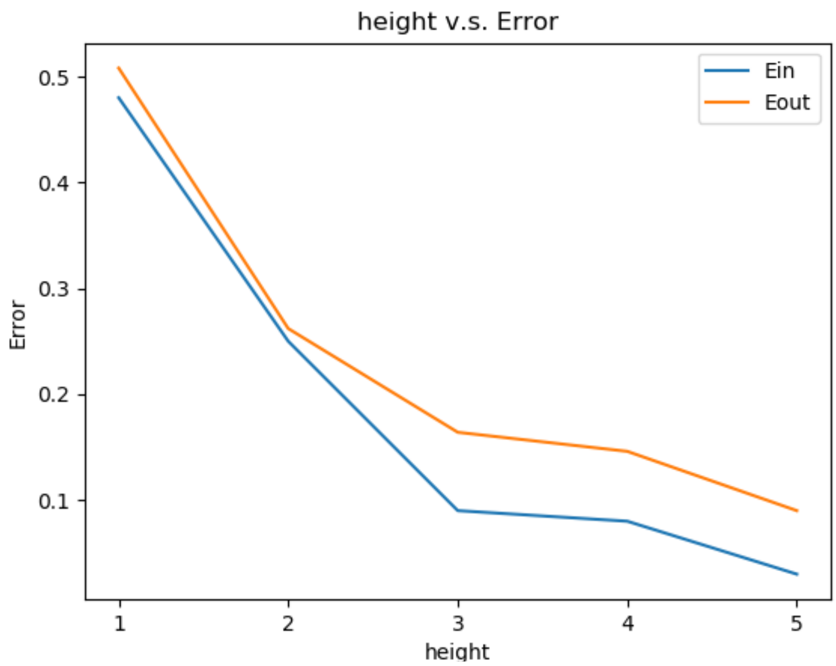
左分支為小於關係，右分支為大於等於關係。



12.

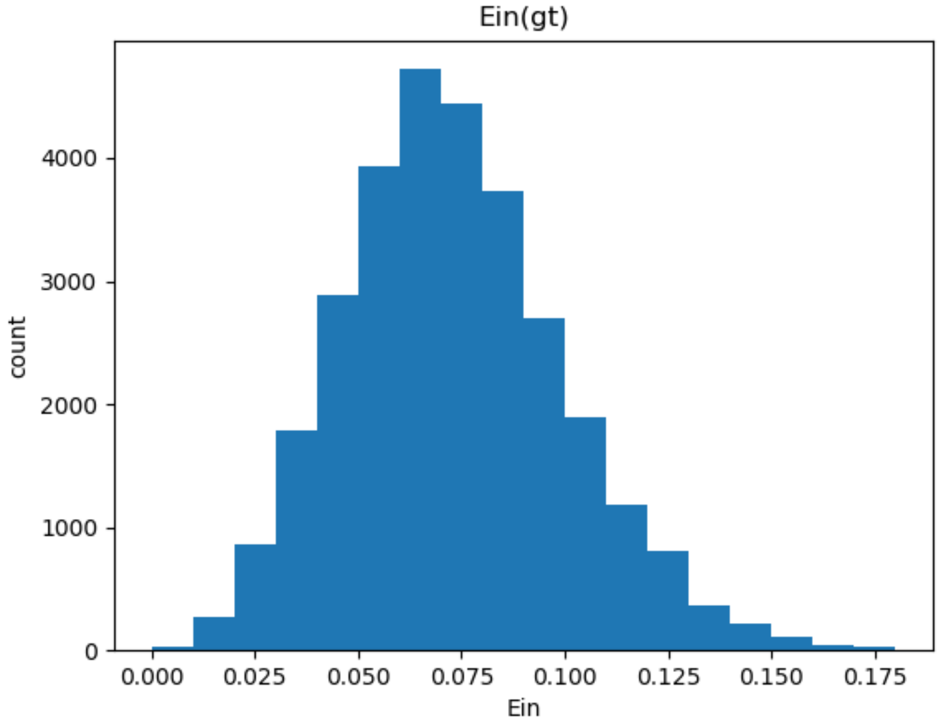
$E_{in} = 0,\ E_{out} = 0.126$

13.

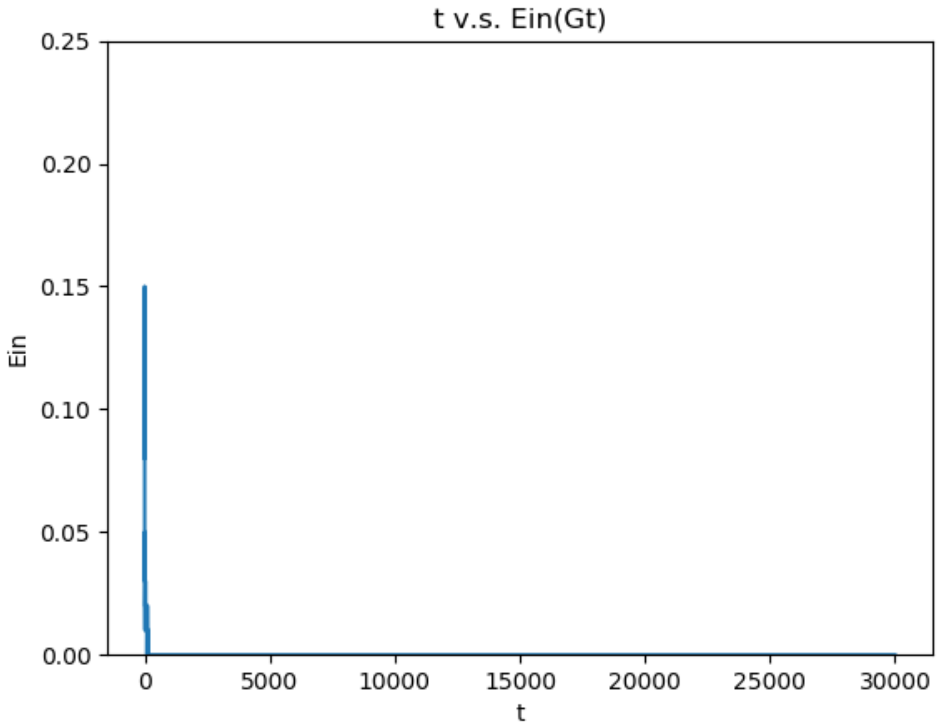


Ein與Eout同步的下降，沒有明顯overfit的情況。

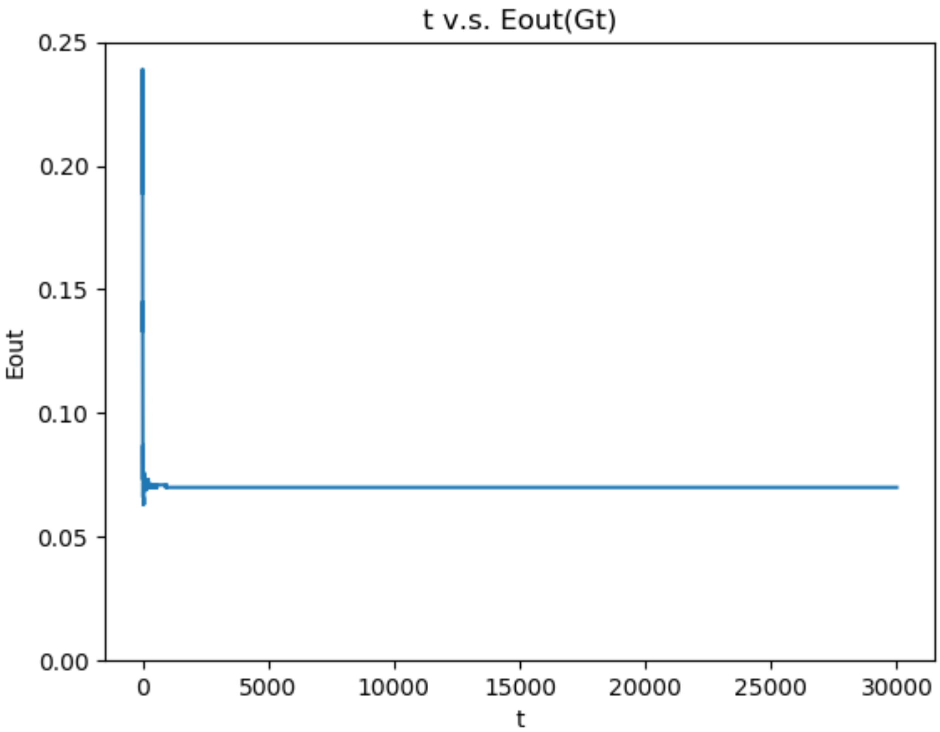
14.


Ein(gt)
4000
3000
2000
1000
0
count
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
Ein

15.


t v.s. Ein(Gt)
0.25
0.20
0.15
0.10
0.05
0.00
Ein
0
5000
10000
15000
20000
25000
30000
t

16.



觀察與15.的圖後，發現Random Forest在Ein非常快就能達到0，而Eout的曲線與Ein幾乎是同步下降並趨近一個值，也是非常快便能得到穩定不錯的預測結果。

17.

XOR$(x_1, x_2, \cdots, x_d)$是x為+1的個數是否為奇數。因此個數直接反映在如下加總上：(-d,-d+2,-d+4,-d+6,...,d-2,d)紅字者為True，(d-2與d的True False要視d為奇數或偶數)，因此以這個方式出發，hidden layer每個neuron判斷是否大於間隔中間值。

透過以下架構結合sign(s)，能得到XOR的效果。**藍色權重部分全為+1**。