Problem1

1、$K_{ds}(x,x') = (\phi_{ds}(x)^T, \phi_{ds}(x'))$

$= \sum_{s,i,\theta} g_{s,i,\theta}(x)\, g_{s,i,\theta}(x')$

而 x 與 x' 有 2 种情況 ① $x = x'$ ② $x \neq x'$

case ① ⇒ $g_{s,i,\theta}(x)\ g_{s,i,\theta}(x') = 1$ 且

每個 decision stump 的結果相同，則總數為

$2 * d * (R-L)$

case ② ⇒ 若值在 x_i 與 x_i' 間範圍則會不同

$g_{+1,i,\theta}(x)\ g_{+1,i,\theta}(x') = -1$ 且

$g_{-1,i,\theta}(x)\ g_{-1,i,\theta}(x') = -1$，因此有

$2\sum_{i=1}^{d} |x_i - x_i'| = 2\|x - x'\|$ 個 -1

∴ total 為 $2 * d * (R-L) - 2 \times 2\|x - x'\|$

⇒ $K_{ds}(x,x') = 2 * d \times (R-L) - 4\|x - x'\|$ #

Problem2

2. 找到(α, b)與$(\tilde{\alpha}, \tilde{b})$的關係式 以證明
新的PC 與原問題等價

assume $\tilde{\alpha}_n = \dfrac{\alpha_n}{u}$

$$\min_{\tilde{\alpha}} \frac{1}{2} \sum_{n}^{N} \sum_{m}^{N} \tilde{\alpha}_n \tilde{\alpha}_m y_n y_m \tilde{K}(x_n, x_m) - \sum_{n}^{N} \tilde{\alpha}_n$$

$$= \min_{\alpha} \frac{1}{2} \sum_{n}^{N} \sum_{m}^{N} \frac{\alpha_n}{u} \frac{\alpha_m}{u} y_n y_m \left(u K(x_n, x_m) + v \right) - \sum_{n}^{N} \frac{\alpha_n}{u}$$

$$= \min_{\alpha} \frac{1}{2} \sum_{n}^{N} \sum_{m}^{N} \alpha_n \alpha_m y_n y_m \left(K(x_n, x_m) + \frac{v}{u} \right) - \sum_{n}^{N} \alpha_n$$

$$= \min_{\alpha} \frac{1}{2} \sum_{n}^{N} \sum_{m}^{N} \alpha_n \alpha_m y_n y_m K(x_n, x_m) - \sum_{n}^{N} \alpha_n$$

$$\tilde{g}_{svm}(x) = \text{sign}\left(\sum_{SV} \tilde{\alpha}_n y_n \tilde{K}(x_n,x) + \tilde{b}\right)$$

$$= \text{sign}\left(\sum_{SV} \frac{\alpha_n}{u} y_n \left(uK(x_n,x)+v\right) + \tilde{b}\right)$$

$$= \text{sign}\left(\sum_{SV} \alpha_n y_n \left(K(x_n,x)+\frac{v}{u}\right) + \tilde{b}\right)$$

$$= \text{sign}\left(\sum_{SV} \alpha_n y_n K(x_n,x) + b\right)$$

$$\therefore \begin{cases} \tilde{\alpha}_n = \dfrac{\alpha_n}{u} \\ b = \tilde{b} + \alpha_n y_n \dfrac{v}{u} \Rightarrow \tilde{b} = b - \alpha_n y_n \dfrac{v}{u} \end{cases}$$

以上關係式 新 $\tilde{K}$、$\tilde{C}$ 與原問題

等價。

Problem3

3.

$$\frac{E_{out}(G)}{E} = \frac{E_{out}(G)}{\sum_{t=1}^{17} E_{out}(g_t)} = \frac{\frac{1}{N} E_{out}(G)}{\frac{1}{N}\sum_{t=1}^{17} E_{out}(g_t)}$$

$$= \frac{P[G(x_n) \neq y_n]}{\sum_{n=1}^{17} P[g(x_n) \neq y_n]} = \frac{P[\sum_{t=1}^{17} [[g_t(x_n) \neq y_n]] \geq 9]}{\sum_{n=1}^{17} P[g(x_n) \neq y_n]}$$

根據 Markov's inequality：$P(X \geq \alpha) \leq \frac{E(X)}{\alpha}$

$$\leq \frac{1}{9} \frac{E[\sum_{t=1}^{17} [[g_t(x_n) \neq y_n]]]}{\sum_{t=1}^{17} P[g(x_n) \neq y_n]}$$

$$= \frac{1}{9} \frac{\frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{17} [[g_t(x_n) \neq y_n]]}{\sum_{t=1}^{17} P[g_t(x_n) \neq y_n]}$$

$$= \frac{1}{9} \cancel{\frac{\sum_{t=1}^{17} P[g_t(x_n) \neq y_n]}{\sum_{t=1}^{17} P[g_t(x_n) \neq y_n]}}$$

$\therefore \frac{E_{out}(G)}{E} \leq \frac{1}{9}$，最大為 $\frac{1}{9}$

Problem4

4.

抽一次 一个sample 没被抽到的机率

$$\left(1-\frac{1}{N}\right)$$

抽 $\frac{3}{4}N$ 次 一个sample没被抽到的机率

$$\left(1-\frac{1}{N}\right)^{\frac{3}{4}N}$$

$$= \frac{1}{\left(\frac{N}{N-1}\right)^{\frac{3}{4}N}} = \left(\frac{1}{\left(1+\frac{1}{N-1}\right)^{N}}\right)^{\frac{3}{4}}$$

$$\approx \left(\frac{1}{e}\right)^{\frac{3}{4}}$$

Problem5

5.

obtain iteration 1 的 weight 後需要更新資料權重 $u_n^{(1)}$

$$\epsilon_1 = \frac{\sum_{n=1}^{N} u_n^{(1)} [[y_n \neq g_1(x_n)]]}{\sum_{n=1}^{N} u_n^{(1)}} = 2\%$$

$$\lozenge_1 = \sqrt{\frac{98\%}{2\%}} = \sqrt{49\%} = 70\%$$

$$u_n^{(2)} = u_n^{(1)} * 70\% \quad , \text{ if } y_n > 0$$

$$u_n^{(2)} = u_n^{(1)} / 70\% \quad , \text{ if } y_n < 0$$

$$\Rightarrow u_n^{(1)} = \frac{1}{N}$$

$$\frac{\sum_{n: y_n > 0} u_n^{(2)}}{\sum_{n: y_n < 0} u_n^{(2)}} = \frac{\overset{49}{\cancel{98\%}} * \overset{49\%}{\cancel{70\%}} * u_n^{(1)}}{\cancel{2\%} \div \cancel{70\%} * u_n^{(1)}} = 24.01$$

\#

Problem6

6. 使用數學歸納法 prove $\frac{U_{t+1}}{U_t} = 2\sqrt{\varepsilon_t(1-\varepsilon_t)}$

consider $t=1$, $U_1 = 1$,

$$U_2 = \sum_{n=1}^{N}\left[\varepsilon_1 * \frac{\diamond_1}{N} + (1-\varepsilon_1) * \frac{1}{N\diamond_1}\right]$$

$$= \varepsilon_1 * \sqrt{\frac{1-\varepsilon_1}{\varepsilon_1}} + (1-\varepsilon_1) * \sqrt{\frac{\varepsilon_1}{1-\varepsilon_1}}$$

$$= 2\sqrt{\varepsilon_1(1-\varepsilon_1)}$$, $\frac{U_2}{U_1} = 2\sqrt{\varepsilon_1(1-\varepsilon_1)}$ 成立

consider $t=N$, 資料的權重會分別以 ε_t 與 $1-\varepsilon_t$ 的機率分成2群 $*\diamond_N$ 與 $\div\diamond_N$

$$U_{N+2} = U_{N+1}\,\varepsilon_{N+1}\,\diamond_{N+1} + U_{N+1}(1-\varepsilon_{N+1})\,\diamond_{N+1}$$

$$= U_{N+1}\left[\varepsilon_{N+1}\sqrt{\frac{1-\varepsilon_{N+1}}{\varepsilon_{N+1}}} + (1-\varepsilon_{N+1})\sqrt{\frac{\varepsilon_{N+1}}{1-\varepsilon_{N+1}}}\right]$$

$$= U_{N+1} \, 2\sqrt{\varepsilon_{N+1}(1-\varepsilon_{N+1})}$$

$$\frac{U_{N+2}}{U_{N+1}} = 2\sqrt{\varepsilon_{N+1}(1-\varepsilon_{N+1})}$$ 得证

$$\therefore \frac{U_{T+1}}{U_1} = \frac{U_2}{U_1} \cdots \frac{U_T}{U_{T-1}} \frac{U_{T+1}}{U_T}$$

$$= \prod_{t=1}^{T} 2\sqrt{\varepsilon_t (1-\varepsilon_t)}$$

Problem7

7.

we know $s_n^{t+1} = s_n^t + \alpha_t g_t(x_n)$

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

$$\frac{\partial ①}{\partial \eta} = 0$$

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

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

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

#

Problem8

8.

$$s_i^{(\ell-1)} \xrightarrow{tanh} x_i^{(\ell-1)} \xrightarrow{W_{ij}} s_j^{(\ell)} \xrightarrow{tanh}$$

$$\delta_j^{(\ell)} = \frac{\partial e_n}{\partial s_j^{(\ell)}} = \sum_{k=1}^{d^{(\ell+1)}} \frac{\partial e_n}{\partial s_k^{(\ell+1)}} \frac{\partial s_k^{(\ell+1)}}{\partial x_j^{(\ell)}} \frac{\partial x_j^{(\ell)}}{\partial s_j^{(\ell)}}$$

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

$$\delta_j^{(\ell-1)} = \sum_K \delta_K^{(\ell)} W_{jK}^{(\ell)} \tanh'(s_j)^{(\ell-1)}$$

∵ 當 S 為 0 時 tanh′ 趨向無限大

∴ 2~ℓ 層 gradient 皆非 0

而第一層的 gradient 為

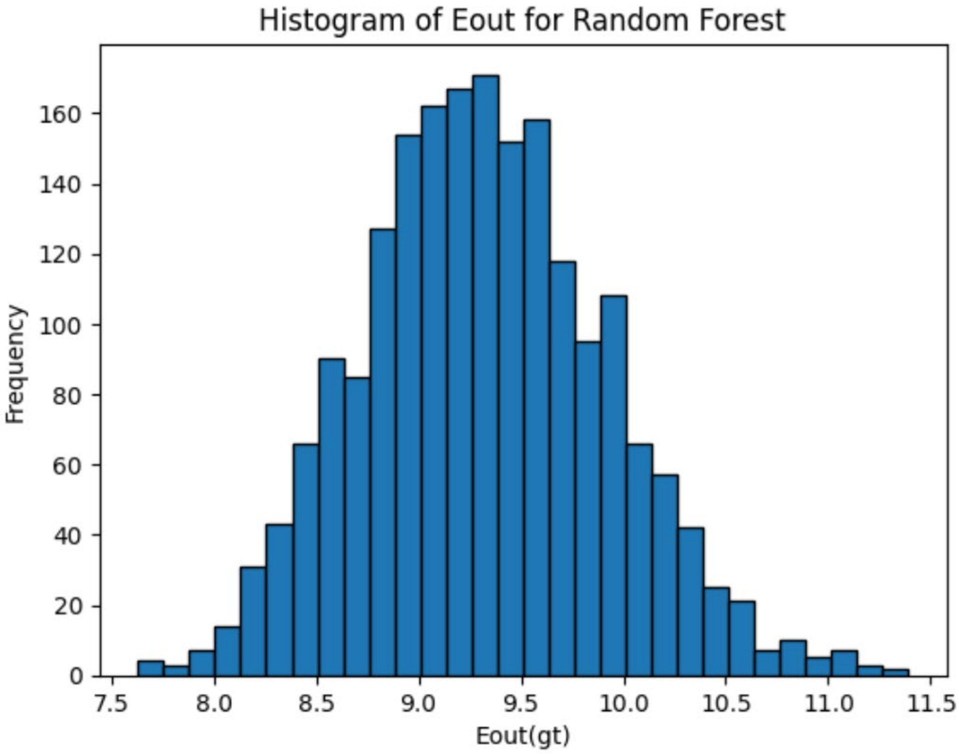
$$\delta_j^{(1)} = \sum_K \delta_K^{(2)} W_{jK}^{(2)} = 0$$

∴ 除了第一層 gradient 為 0，其他層皆非 0。

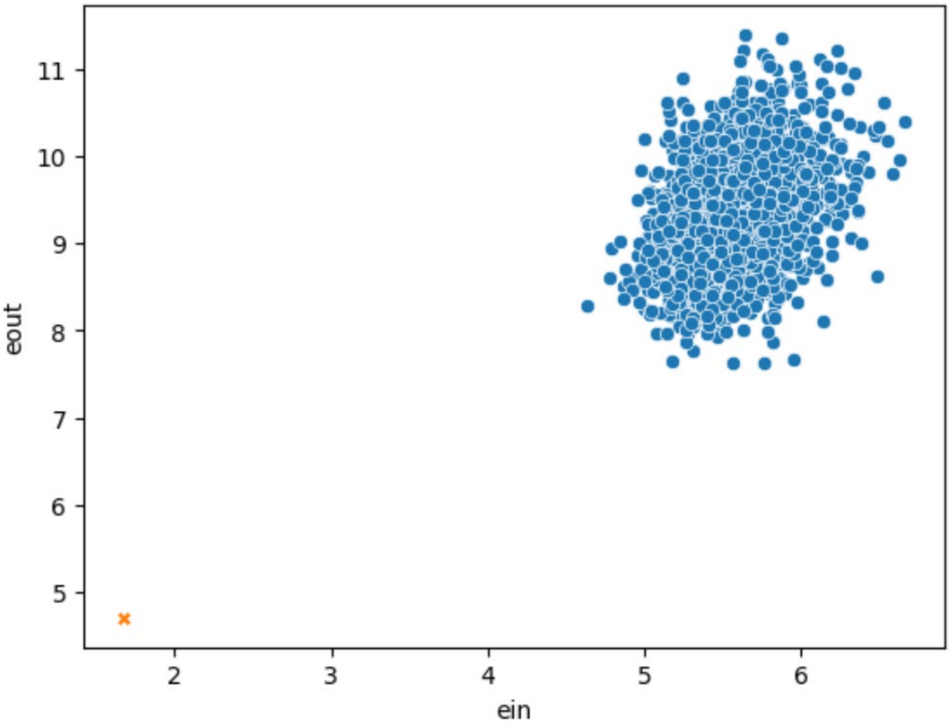
Problem9

$E_{out}(g) = 8.79181041120804$

Problem10

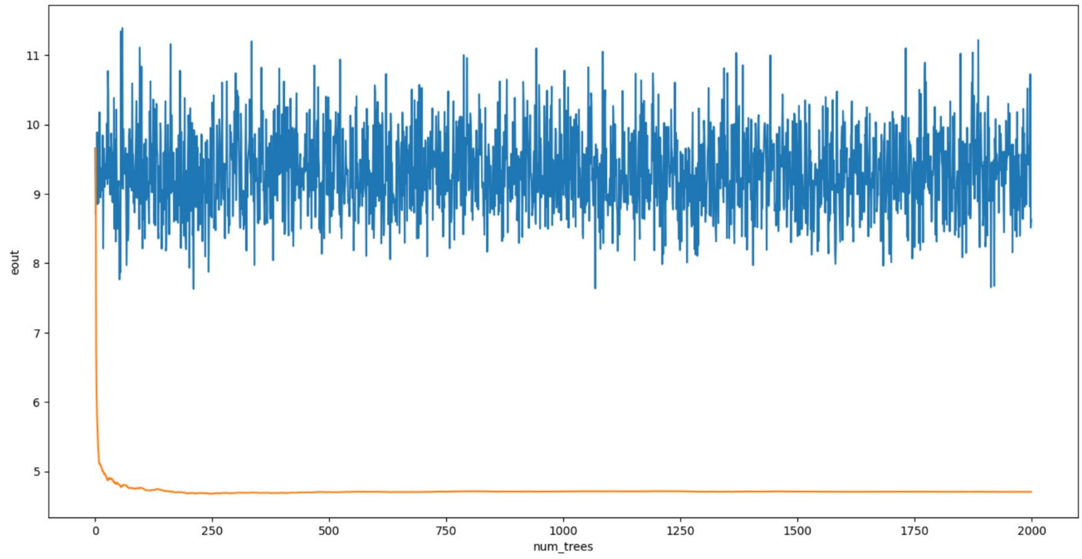


Problem11



Describe your findings: 雖然每個樹的預測都不佳 (eout~9)，但合成起來後可以得到不錯的結果 (eout~5.7)

Problem12



Describe your findings: 大概找到 20 個 trees 就收斂，因此我們不需要找太多 tree 來複雜化模型，並且可以發現 random forest 較不會發生過擬和的問題。