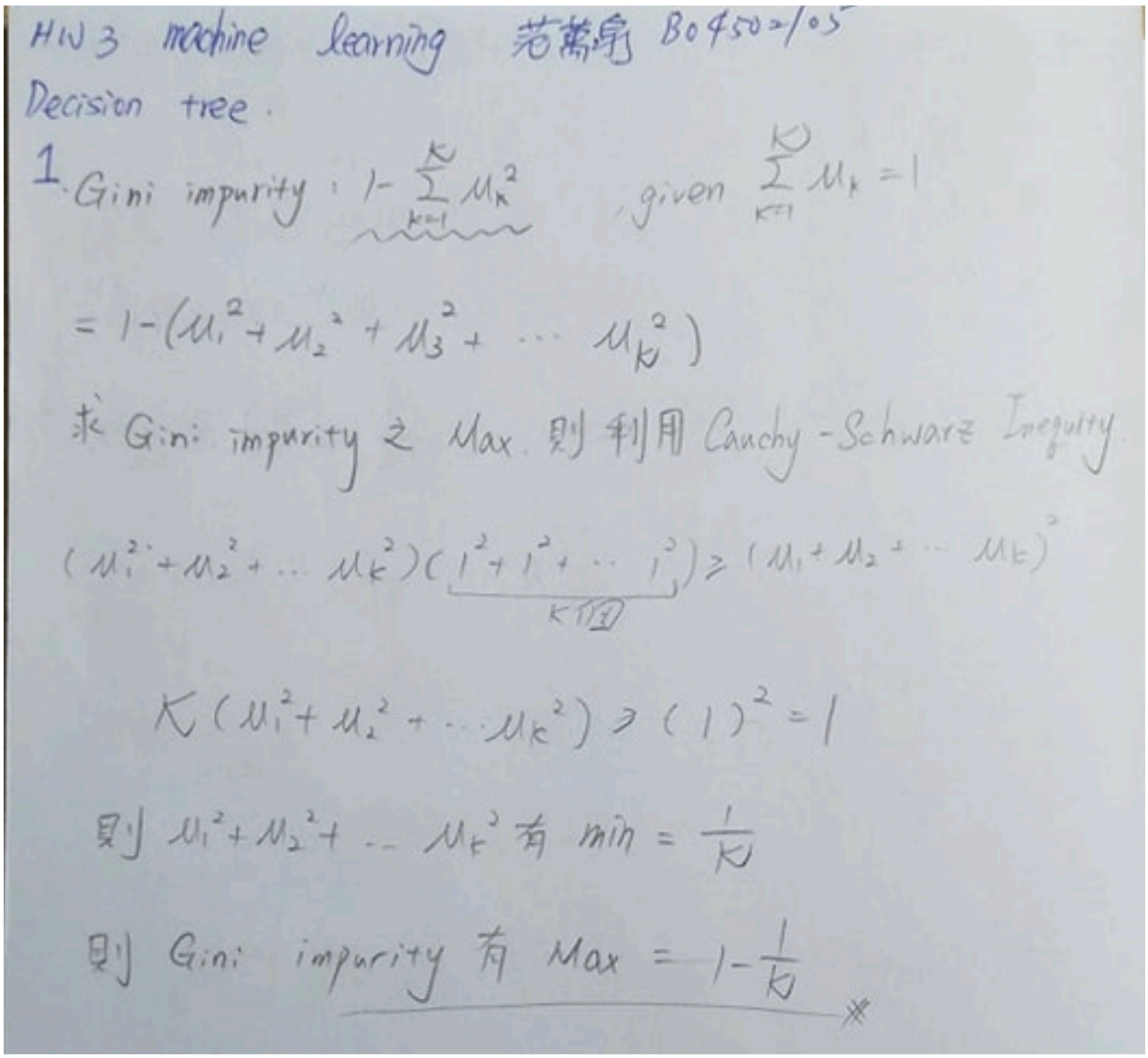
NTUEE, WAN - CYUAN FAN 范萬泉

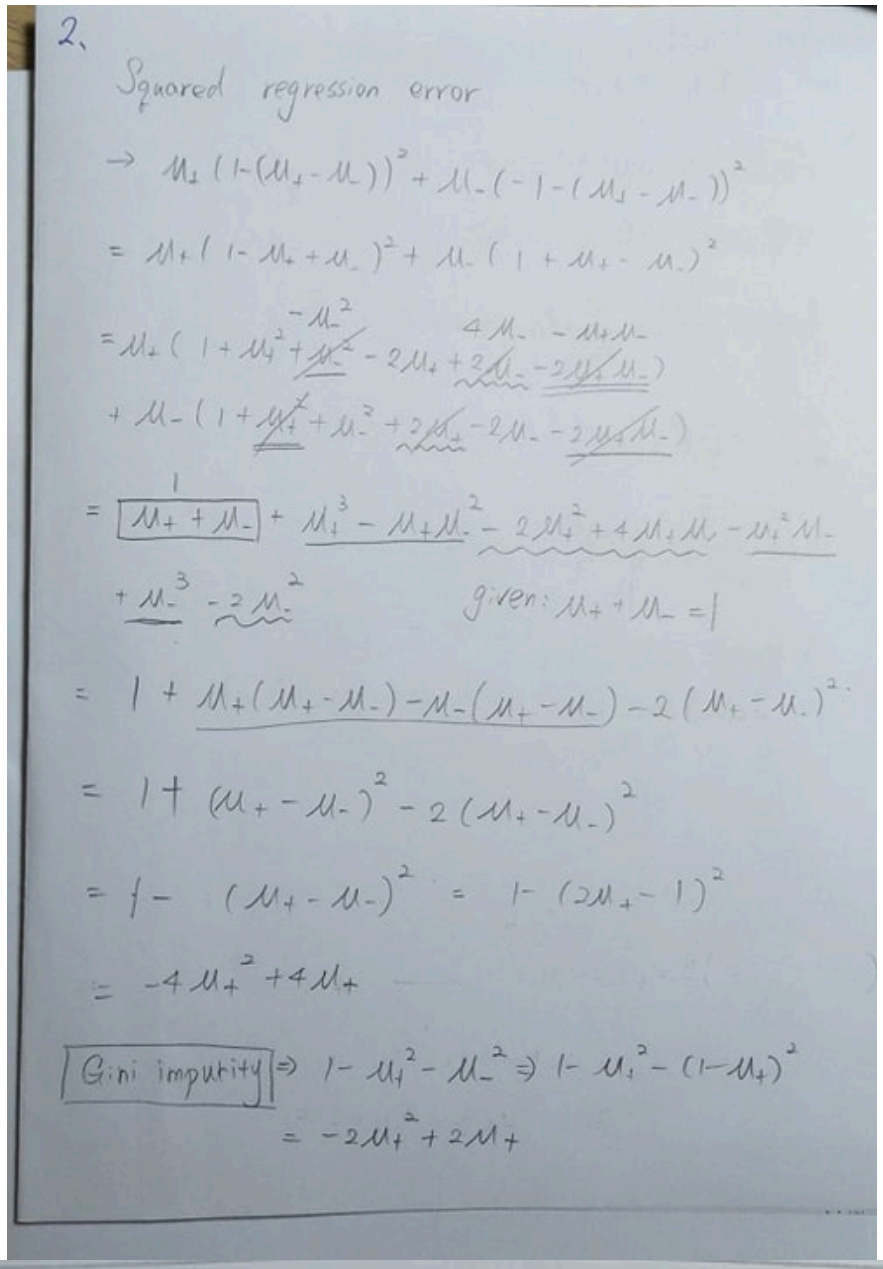
"no collaborators, with reference"

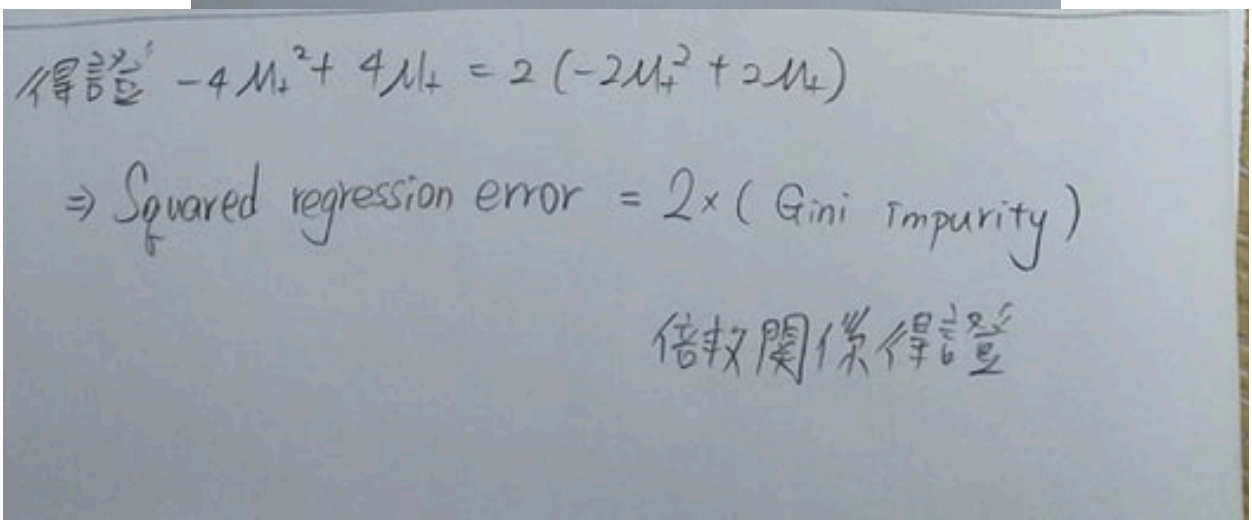
Homework 3

Problem 1 :

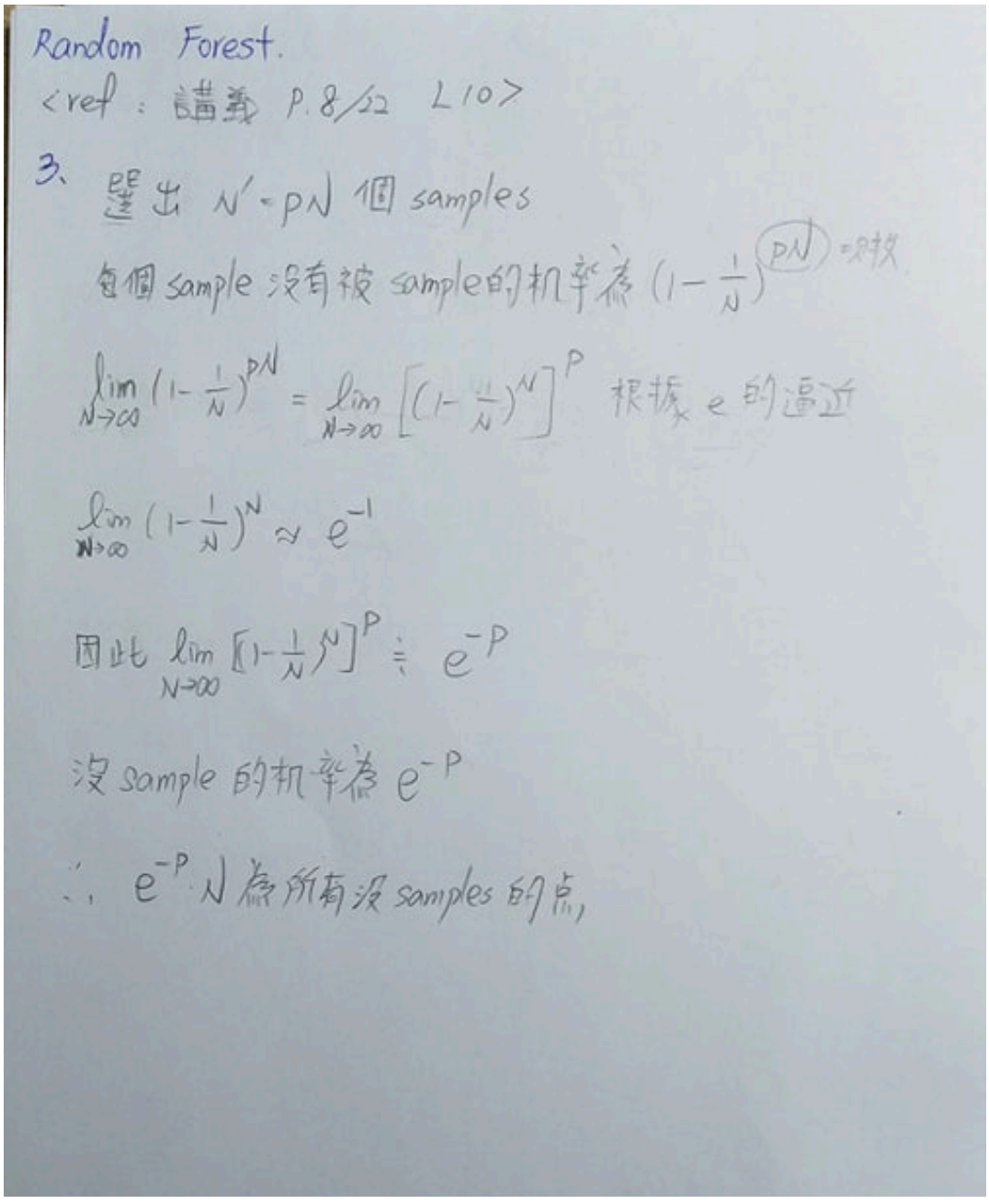


Problem 2 :

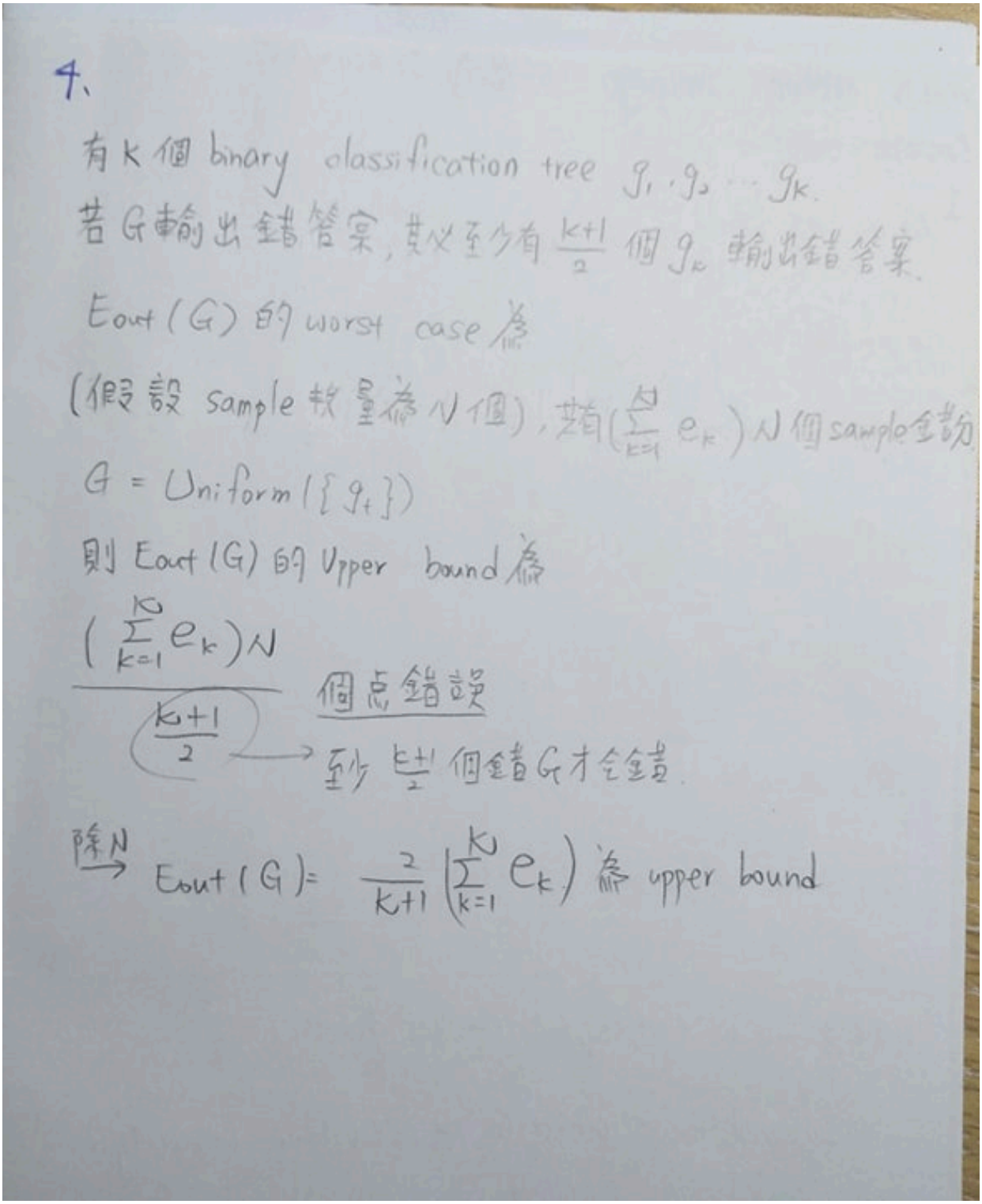

2.
Squared regression error
Gini impurity
given: $\mu_+ + \mu_- = 1$


得證 $-4\mu_+^2 + 4\mu_+ = 2(-2\mu_+^2 + 2\mu_+)$
⇒ Squared regression error = 2×(Gini Impurity)
倍数関係得證

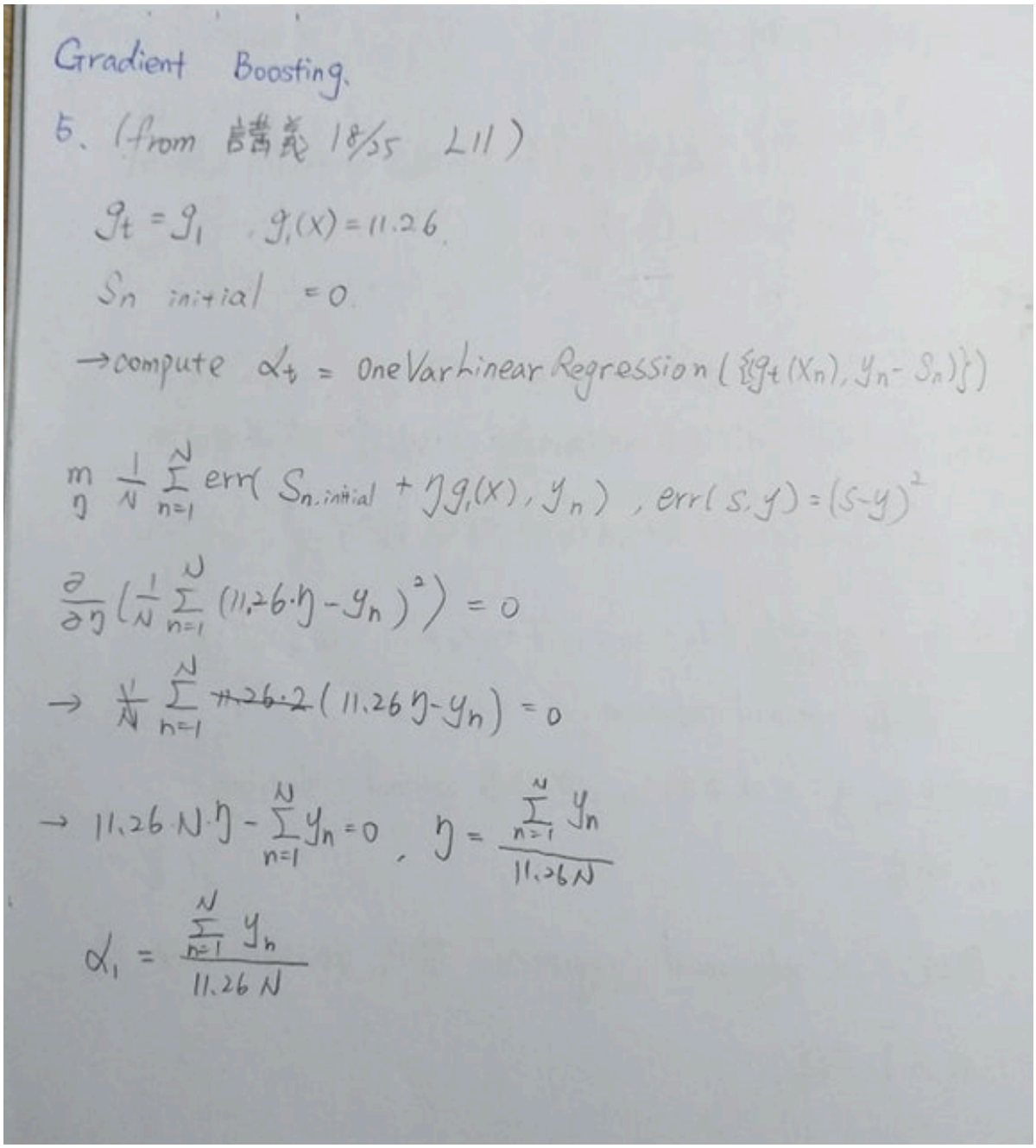
Problem 3:


Random Forest.
<ref : 講義 P.8/32 L10>
3. 選出 N'=pN 個 samples
每個 sample 沒有被 sample的机率為 $(1-\frac{1}{N})^{pN}$ =次方
$\lim_{N\to\infty}(1-\frac{1}{N})^{pN}=\lim_{N\to\infty}\left[(1-\frac{1}{N})^{N}\right]^{p}$ 根據 e 的逼近
$\lim_{N\to\infty}(1-\frac{1}{N})^{N}\approx e^{-1}$
因此 $\lim_{N\to\infty}\left[(1-\frac{1}{N})^{N}\right]^{p}\doteq e^{-p}$
沒 sample 的机率為 e^{-p}
∴ $e^{-p}\cdot N$ 為所有沒 samples 的点,

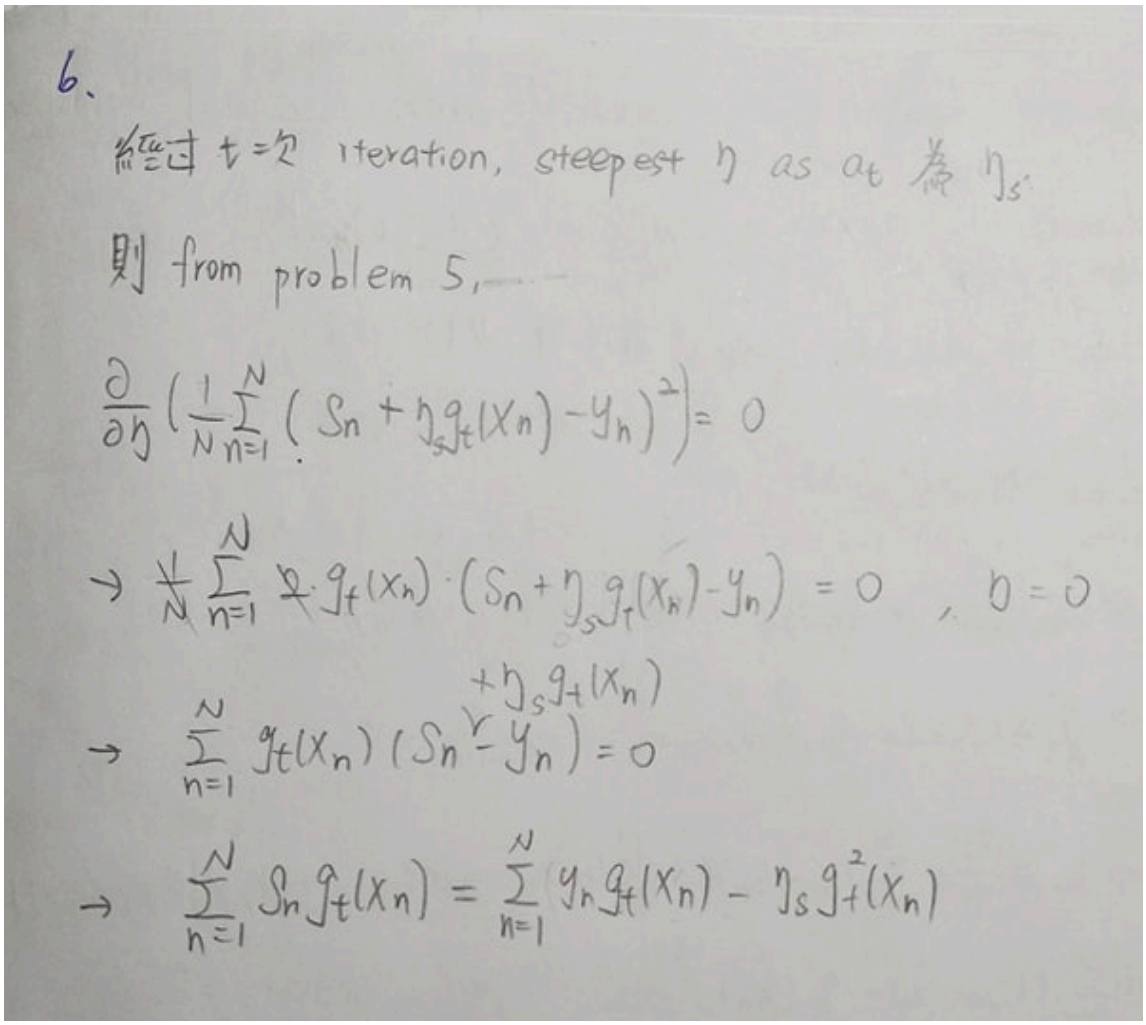
Problem 4:


4.
有K個 binary classification tree $g_1, g_2 \cdots g_K$.
若G輸出錯答案，表示至少有 $\frac{K+1}{2}$ 個 g_k 輸出錯答案.
$E_{out}(G)$ 的 worst case 為
(假設 sample 數量為 N 個)，若有 $(\sum_{k=1}^{K} e_k)N$ 個 sample 全部
$G = Uniform(\{g_t\})$
則 $E_{out}(G)$ 的 Upper bound 為
$\frac{(\sum_{k=1}^{K} e_k)N}{\frac{K+1}{2}}$ 個點錯誤
至少 $\frac{K+1}{2}$ 個錯 G 才全錯.
除N → $E_{out}(G) = \frac{2}{K+1}(\sum_{k=1}^{K} e_k)$ 為 upper bound

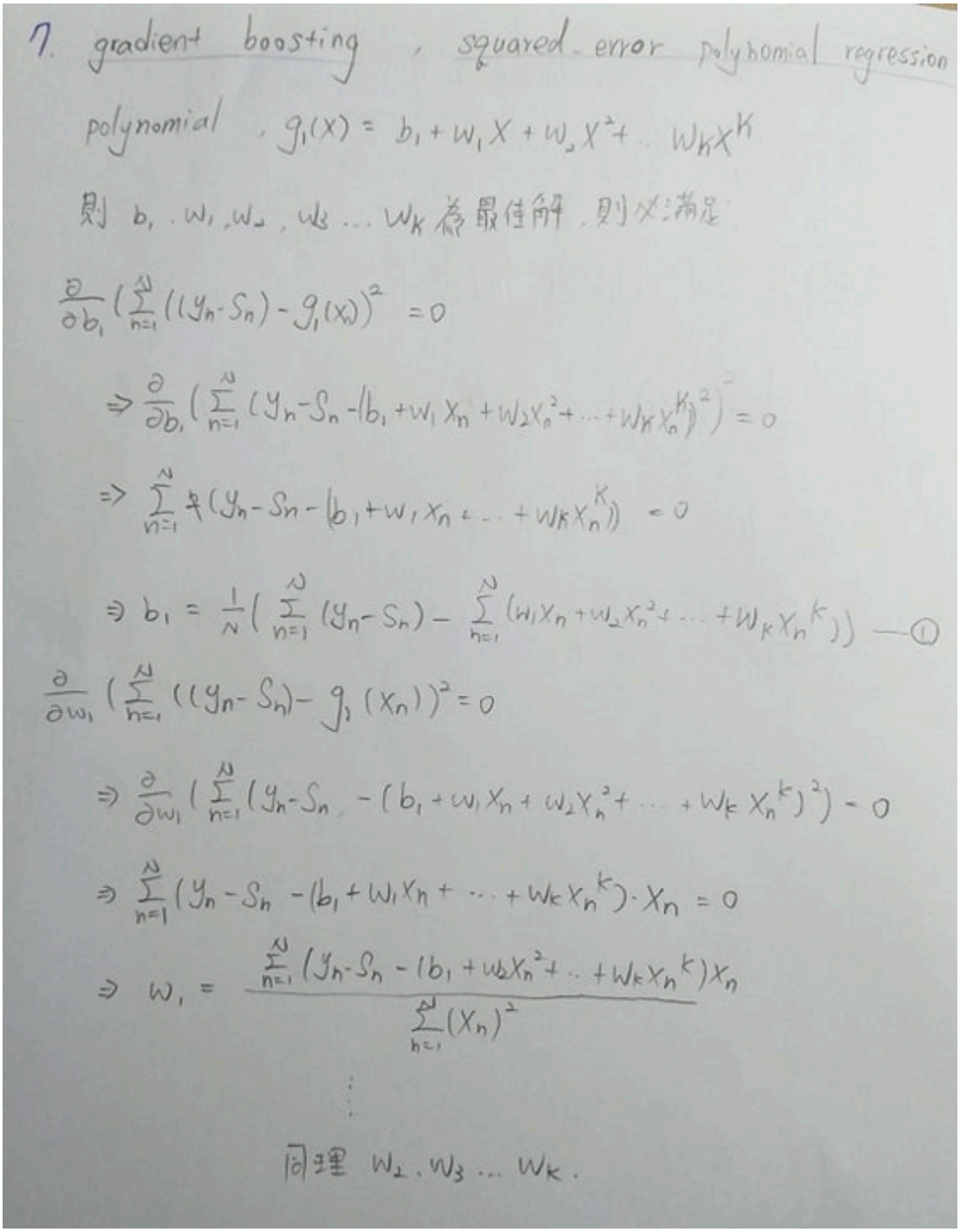
Problem 5:

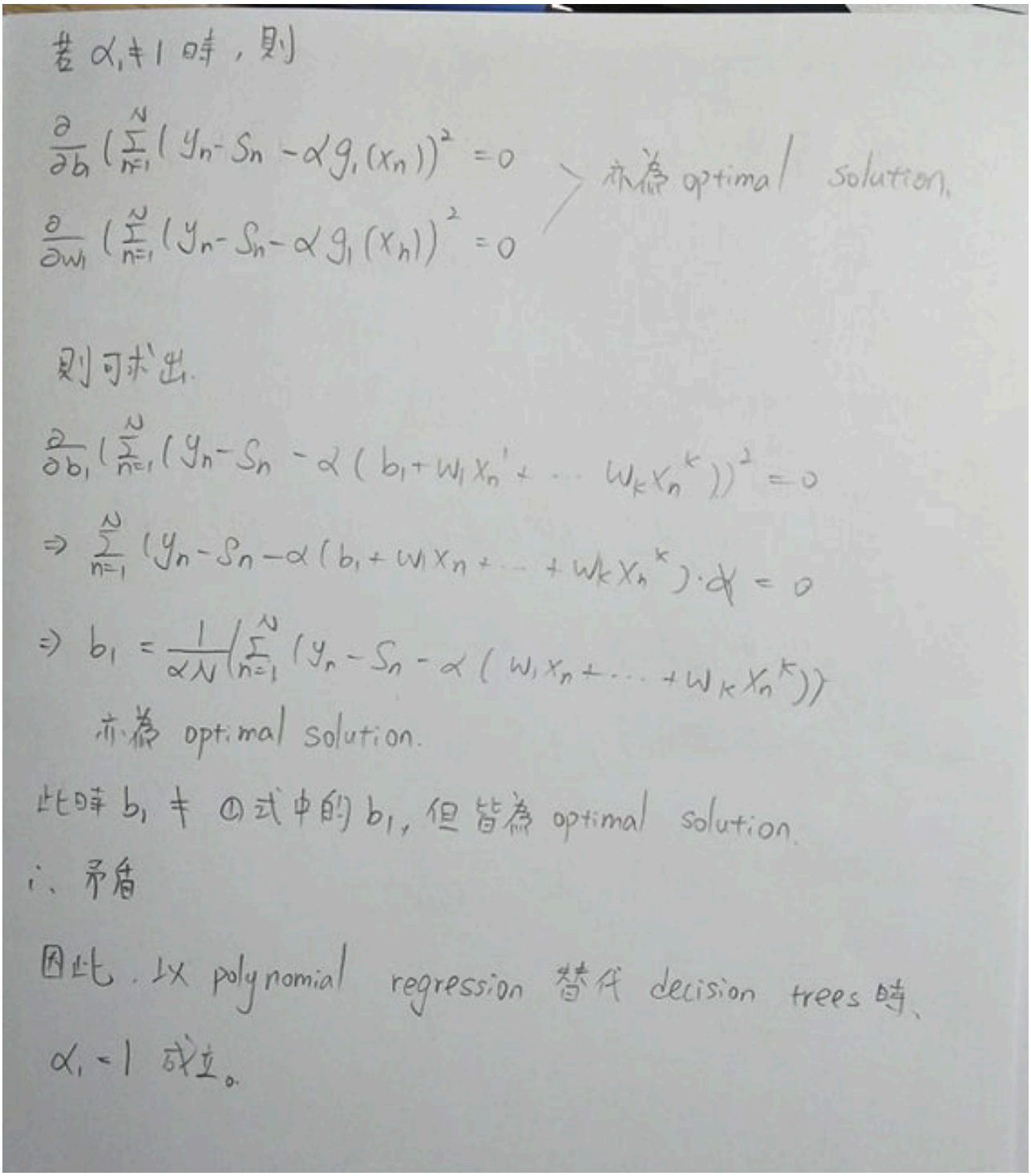

Gradient Boosting.
5. (from 課義 18/35 上11)
$g_t = g_1$, $g_1(x) = 11.26$
S_n initial $= 0$.
→compute α_t = OneVarLinearRegression ($\{(g_t(x_n), y_n - S_n)\}$)
$\min_\eta \frac{1}{N}\sum_{n=1}^{N} err(S_{n,initial} + \eta g_1(x), y_n)$, $err(s, y) = (s-y)^2$
$\frac{\partial}{\partial \eta}\left(\frac{1}{N}\sum_{n=1}^{N}(11.26\cdot\eta - y_n)^2\right) = 0$
→ $\frac{1}{N}\sum_{n=1}^{N} 11.26\cdot 2(11.26\eta - y_n) = 0$
→ $11.26N\cdot\eta - \sum_{n=1}^{N} y_n = 0$, $\eta = \frac{\sum_{n=1}^{N} y_n}{11.26N}$
$\alpha_1 = \frac{\sum_{n=1}^{N} y_n}{11.26N}$

Problem 6 :

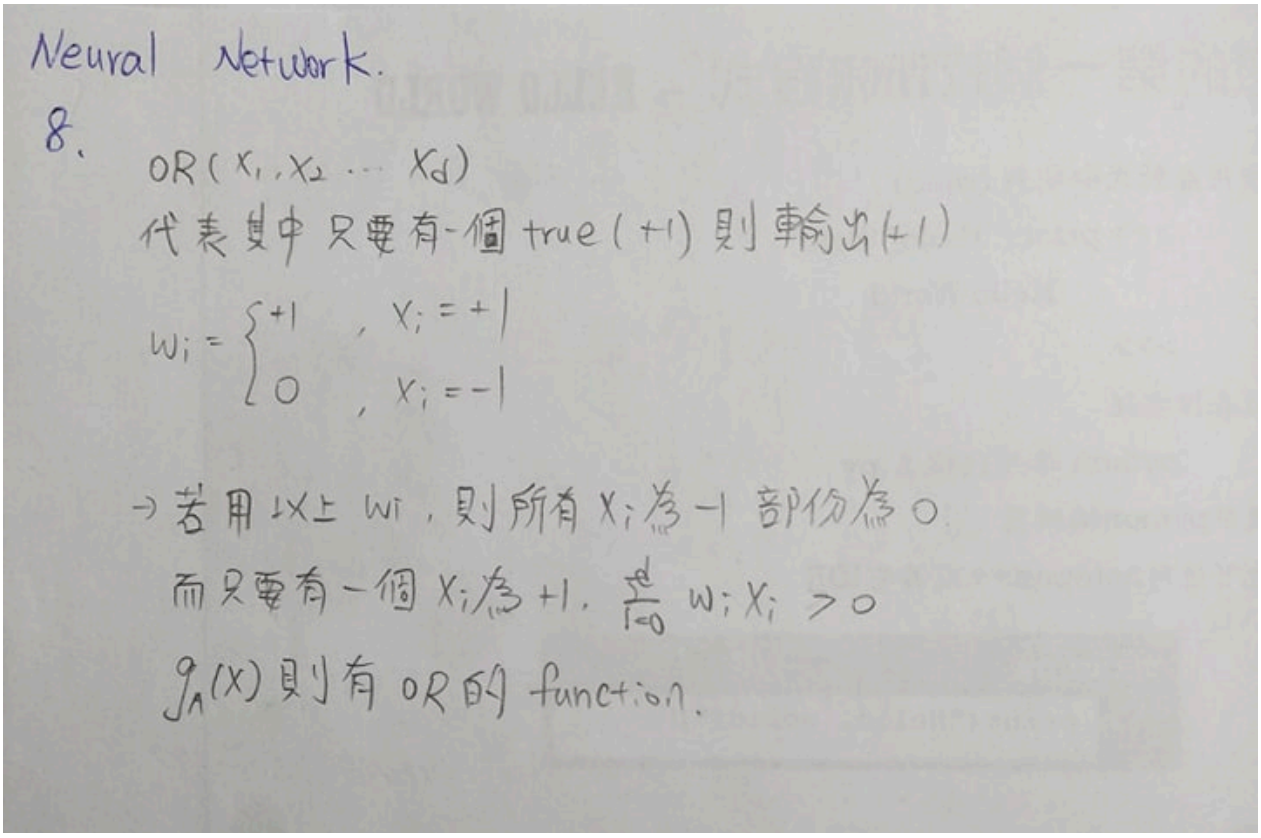

6.
經過 t=次 iteration, steepest η as a_t 為 η_s.
則 from problem 5,
$\frac{\partial}{\partial \eta}\left(\frac{1}{N}\sum_{n=1}^{N}(s_n+\eta_s g_t(x_n)-y_n)^2\right)=0$
$\rightarrow \frac{1}{N}\sum_{n=1}^{N} 2\cdot g_t(x_n)\cdot(s_n+\eta_s g_t(x_n)-y_n)=0$, $b=0$
$+\eta_s g_t(x_n)$
$\rightarrow \sum_{n=1}^{N} g_t(x_n)(s_n-y_n)=0$
$\rightarrow \sum_{n=1}^{N} s_n g_t(x_n)=\sum_{n=1}^{N} y_n g_t(x_n)-\eta_s g_t^2(x_n)$

Problem 7:

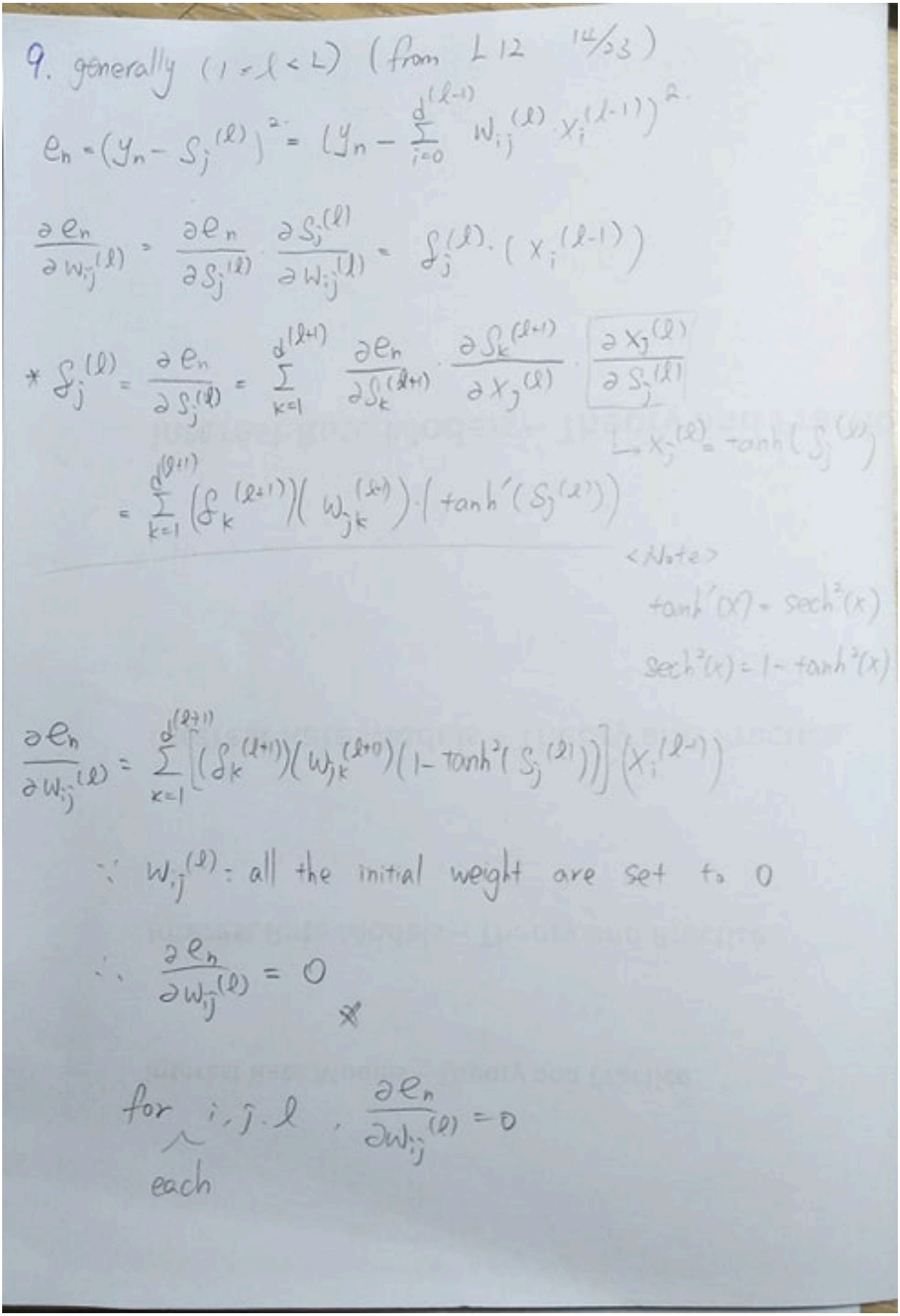

7. gradient boosting , squared_error polynomial regression
polynomial . $g_1(x) = b_1 + w_1 X + w_2 X^2 + \dots w_K X^K$
則 $b_1, w_1, w_2, w_3 \dots w_K$ 為最佳解，則必滿足
$\frac{\partial}{\partial b_1}\left(\sum_{n=1}^{N}((y_n - S_n) - g_1(x_n))^2\right) = 0$
$\Rightarrow \frac{\partial}{\partial b_1}\left(\sum_{n=1}^{N}(y_n - S_n - (b_1 + w_1 x_n + w_2 x_n^2 + \dots + w_K x_n^K))^2\right) = 0$
$\Rightarrow \sum_{n=1}^{N} 2(y_n - S_n - (b_1 + w_1 x_n + \dots + w_K x_n^K)) = 0$
$\Rightarrow b_1 = \frac{1}{N}\left(\sum_{n=1}^{N}(y_n - S_n) - \sum_{n=1}^{N}(w_1 x_n + w_2 x_n^2 + \dots + w_K x_n^K)\right)$ —①
$\frac{\partial}{\partial w_1}\left(\sum_{n=1}^{N}((y_n - S_n) - g_1(x_n))^2 = 0\right.$
$\Rightarrow \frac{\partial}{\partial w_1}\left(\sum_{n=1}^{N}(y_n - S_n - (b_1 + w_1 x_n + w_2 x_n^2 + \dots + w_k x_n^k)^2\right) = 0$
$\Rightarrow \sum_{n=1}^{N}(y_n - S_n - (b_1 + w_1 x_n + \dots + w_k x_n^k)) \cdot x_n = 0$
$\Rightarrow w_1 = \frac{\sum_{n=1}^{N}(y_n - S_n - (b_1 + w_2 x_n^2 + \dots + w_k x_n^k)) x_n}{\sum_{n=1}^{N}(x_n)^2}$
⋮
同理 $w_2, w_3 \dots w_k$.


若 $\alpha_1 \neq 1$ 時，則
$\frac{\partial}{\partial b_1}\left(\sum_{n=1}^{N}(y_n - s_n - \alpha g_1(x_n))^2 = 0\right.$
$\frac{\partial}{\partial w_1}\left(\sum_{n=1}^{N}(y_n - s_n - \alpha g_1(x_n))^2 = 0\right.$
> 亦為 optimal solution.
則可求出.
$\frac{\partial}{\partial b_1}\left(\sum_{n=1}^{N}(y_n - s_n - \alpha(b_1 + w_1 x_n^1 + \cdots w_k x_n^k))^2 = 0\right.$
$\Rightarrow \sum_{n=1}^{N}(y_n - s_n - \alpha(b_1 + w_1 x_n + \cdots + w_k x_n^k)) \cdot \alpha = 0$
$\Rightarrow b_1 = \frac{1}{\alpha N}\left(\sum_{n=1}^{N}(y_n - s_n - \alpha(w_1 x_n + \cdots + w_k x_n^k))\right.$
亦為 optimal solution.
此時 b_1 ≠ ①式中的 b_1，但皆為 optimal solution.
∴ 矛盾
因此，以 polynomial regression 替代 decision trees 時，
$\alpha_1 = 1$ 成立。

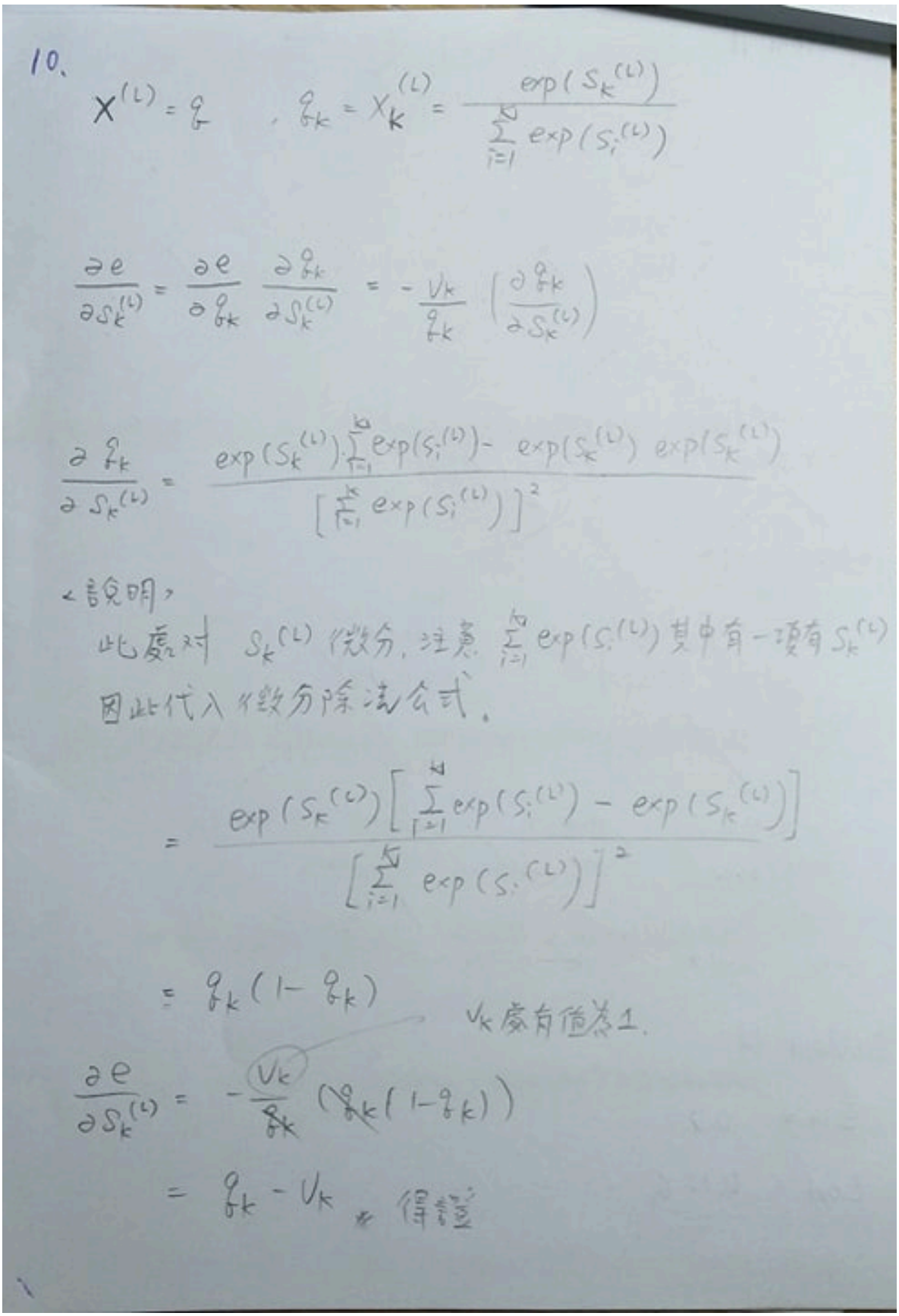
Problem 8:


Neural Network.
8.
OR(x1, x2 ··· xd)
代表其中只要有一個 true (+1) 則輸出(+1)
wi = +1 , xi = +1
0 , xi = -1
→ 若用以上 wi，則所有 xi 為 -1 部份為 0.
而只要有一個 xi 為 +1. Σ wi xi > 0
g1(x) 則有 OR 的 function.

Problem 9:


9. generally (1 ≤ l < L) (from L 12 14/33)
$e_n = (y_n - s_j^{(l)})^2 = (y_n - \sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} x_i^{(l-1)})^2$
$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} \cdot \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} \cdot (x_i^{(l-1)})$
$* \delta_j^{(l)} = \frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{(l+1)}} \frac{\partial e_n}{\partial s_k^{(l+1)}} \cdot \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} \cdot \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}}$
$\rightarrow x_j^{(l)} = \tanh(s_j^{(l)})$
$= \sum_{k=1}^{d^{(l+1)}} (\delta_k^{(l+1)})(w_{jk}^{(l+1)}) \cdot (\tanh'(s_j^{(l)}))$
<Note>
$\tanh'(x) = \mathrm{sech}^2(x)$
$\mathrm{sech}^2(x) = 1 - \tanh^2(x)$
$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \sum_{k=1}^{d^{(l+1)}} [(\delta_k^{(l+1)})(w_{jk}^{(l+1)})(1 - \tanh^2(s_j^{(l)}))](x_i^{(l-1)})$
$\because w_{ij}^{(l)}$: all the initial weight are set to 0
$\therefore \frac{\partial e_n}{\partial w_{ij}^{(l)}} = 0$ *
for each i, j, l, $\frac{\partial e_n}{\partial w_{ij}^{(l)}} = 0$

Problem 10 :



10.

$$X^{(L)} = q \quad , \quad q_k = X_k^{(L)} = \frac{\exp(S_k^{(L)})}{\sum_{i=1}^{N} \exp(S_i^{(L)})}$$

$$\frac{\partial e}{\partial S_k^{(L)}} = \frac{\partial e}{\partial q_k} \frac{\partial q_k}{\partial S_k^{(L)}} = -\frac{v_k}{q_k} \left(\frac{\partial q_k}{\partial S_k^{(L)}} \right)$$

$$\frac{\partial q_k}{\partial S_k^{(L)}} = \frac{\exp(S_k^{(L)}) \cdot \sum_{i=1}^{N} \exp(S_i^{(L)}) - \exp(S_k^{(L)}) \exp(S_k^{(L)})}{\left[\sum_{i=1}^{N} \exp(S_i^{(L)}) \right]^2}$$

<說明>

此處对 $S_k^{(L)}$ 微分，注意 $\sum_{i=1}^{N} \exp(S_i^{(L)})$ 其中有一項有 $S_k^{(L)}$

因此代入微分除法公式。

$$= \frac{\exp(S_k^{(L)}) \left[\sum_{i=1}^{N} \exp(S_i^{(L)}) - \exp(S_k^{(L)}) \right]}{\left[\sum_{i=1}^{N} \exp(S_i^{(L)}) \right]^2}$$

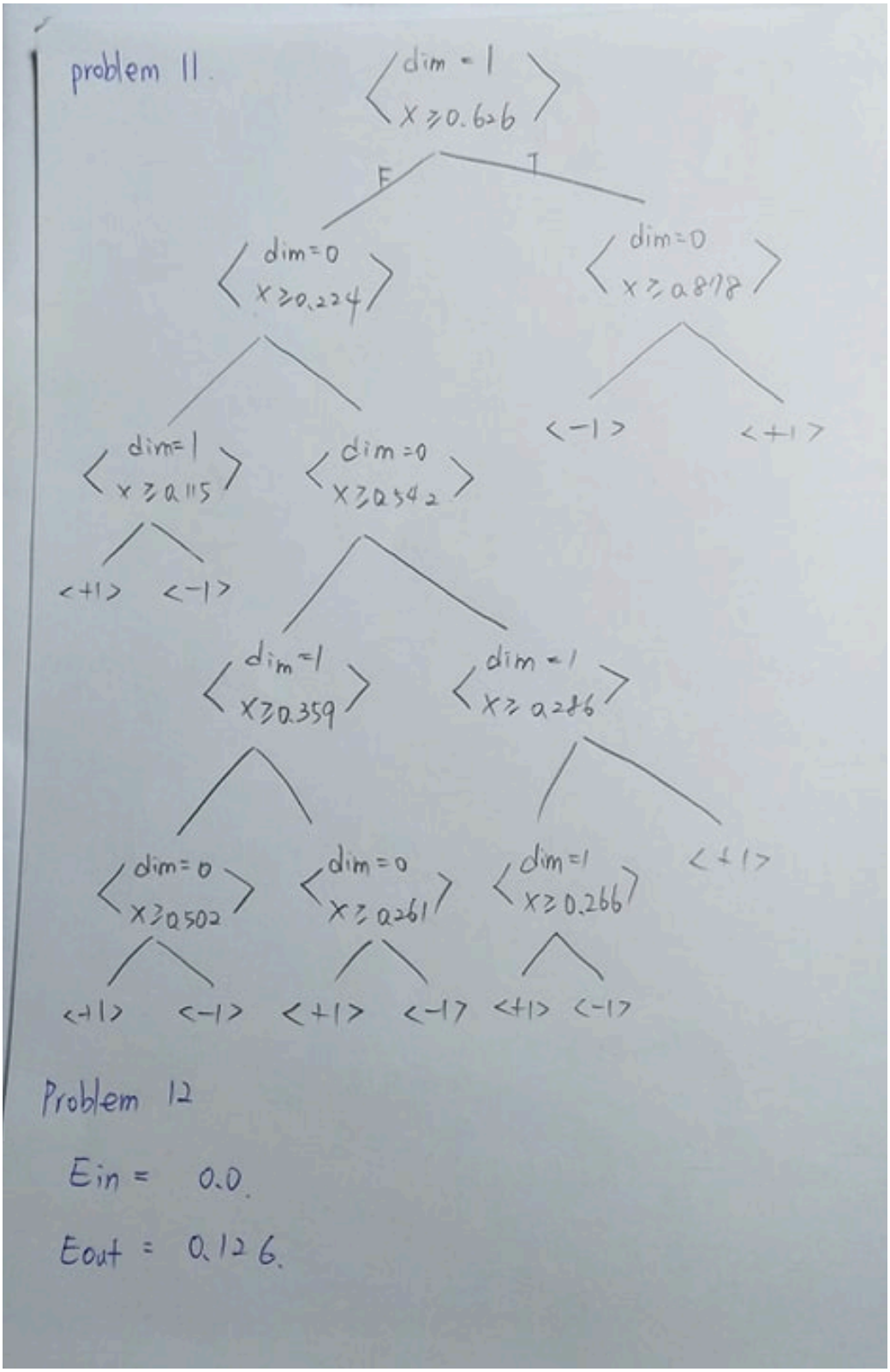
$$= q_k (1 - q_k)$$

v_k 處有值為 1.

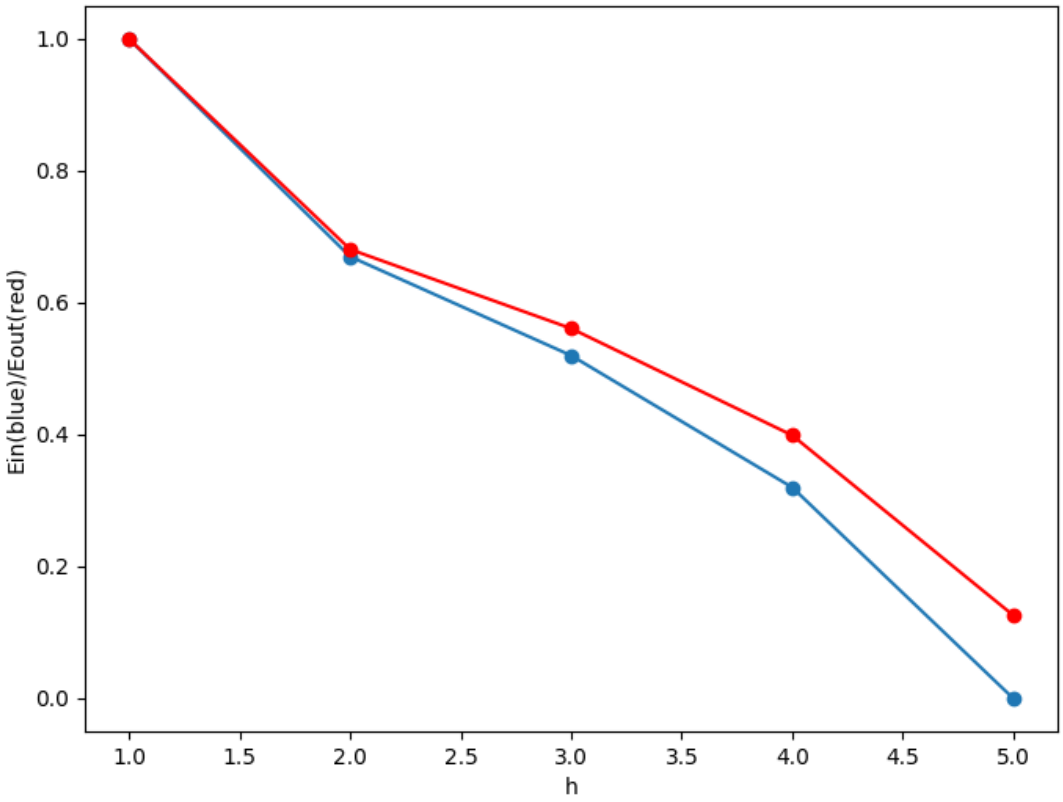
$$\frac{\partial e}{\partial S_k^{(L)}} = -\frac{v_k}{q_k} (q_k (1 - q_k))$$

$$= q_k - v_k \quad \text{得證}$$

Problem 11,12:



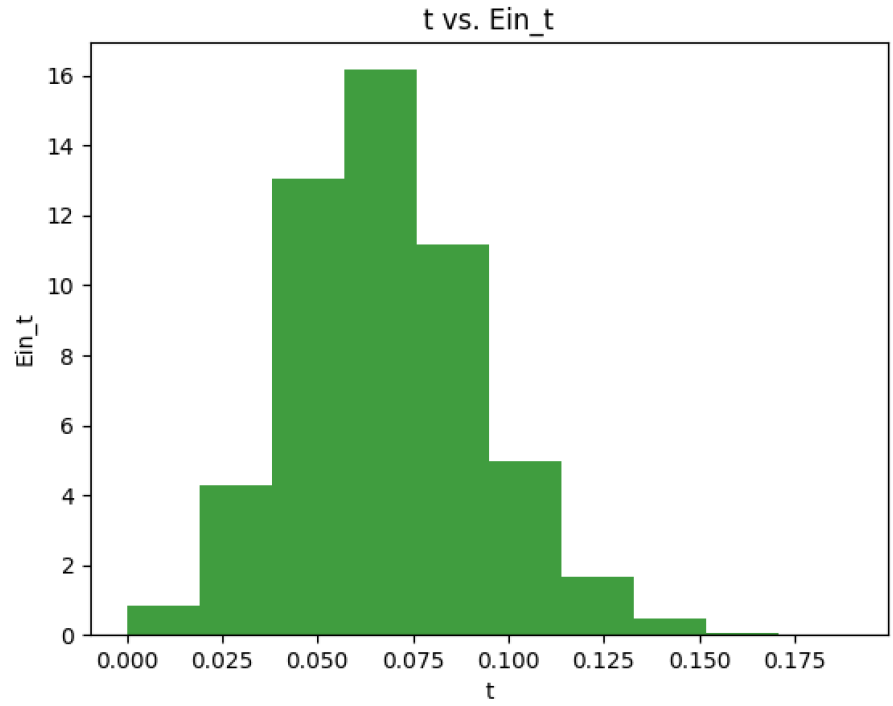
Problem 13:

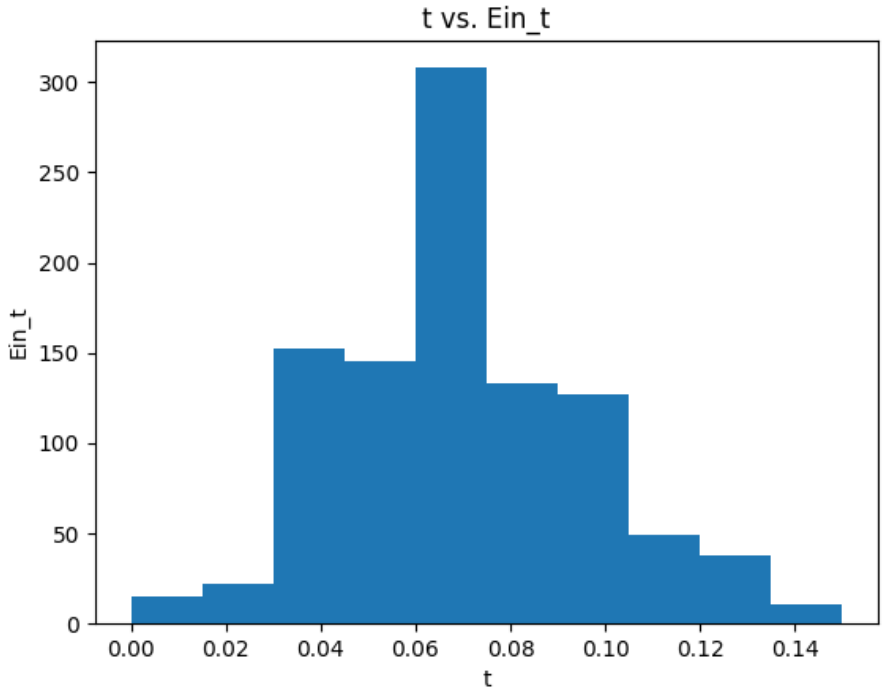


當我們增加tree的高度的時候，確實可以發現我們的正確率提高不少，而在testing set上頭做的效果比training data要來的錯誤率高一些。
總體樹高為5。

Problem 14:

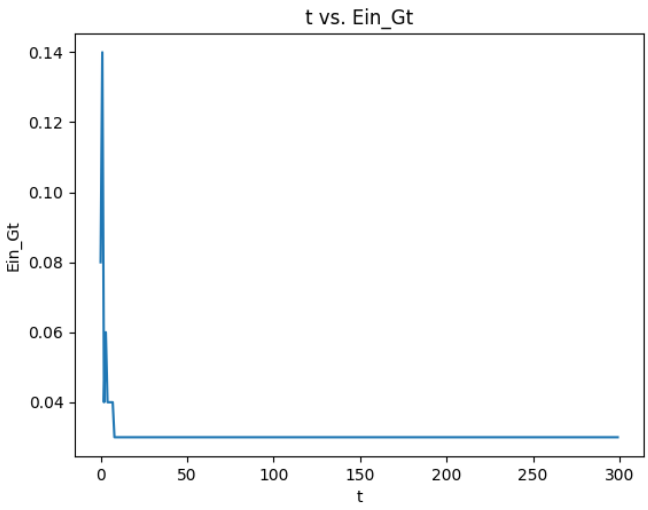
我們觀察hist上頭的變化，發現最大值大約在0.06左右的錯誤率，當T增加基本上分佈沒有太大改變，主要分布於0.06附近。

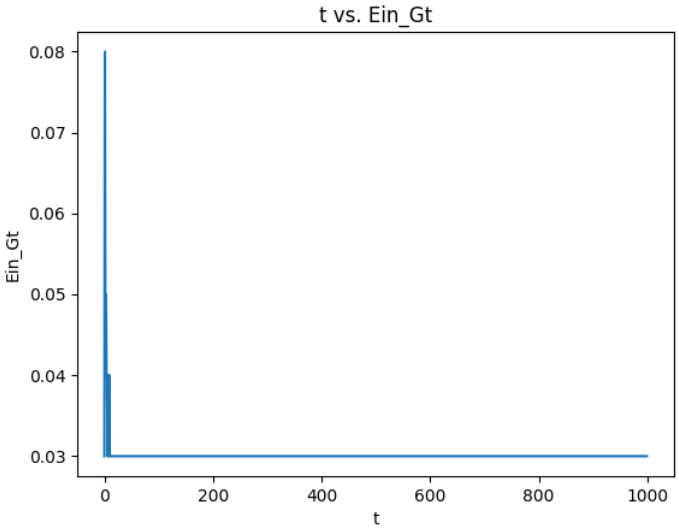


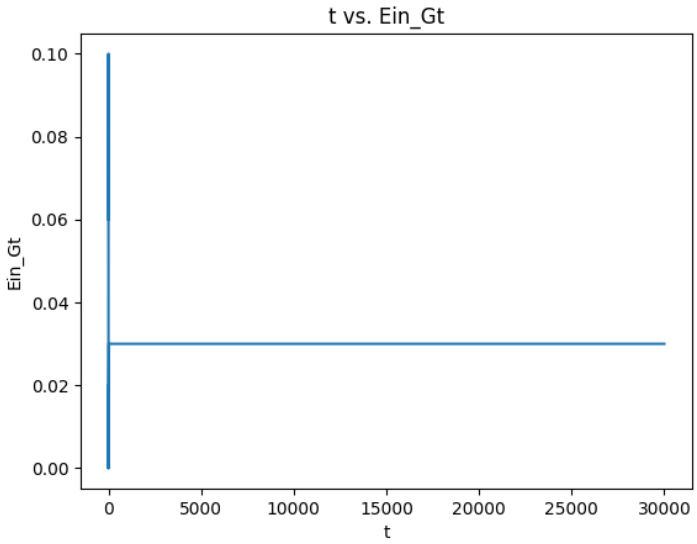


Problem 15:

以下是我們觀察T=300,T=1000,T=30000的差別，基本上後面變化已經不大了，到達極限值。







Problem 16:

我們觀察在16題所做出來的圖形發現，經過bagging的random forest雖然因為隨機取樣的關係，Ein相較於取全部data來煩而稍微上升一些，但是對於testing的部分Eout得以下降至最低點大約0.05左右，與沒有做bagging直接使用的DT Eout = 0.126要下降一些。

1000次iteration之後起伏不大，因此我們觀察較明顯起伏的前1000次觀察最低點。