WANG Disong 1155113946

第 1 題
（答題不得寫在紅綫外）

第 頁

I.
```
int isSumTree(struct TreeNode* node)
{
  if (node==Null || node->left ==Null && node->right==null)
      return 1.

  int ls, rs.
  ~~if node->left ==Null:~~
      ~~ls=0~~
  ~~else~~
      ~~ls = Sum(node->left->left)~~
  ls = Sum(node->left);
  rs = Sum(node->right);
  if (node->data== ls+rs && isSumTree(node->left) &&
                             isSumTree(node->right))
      return 1;
  return0;
}
```

where Sum function is defined as:

```
int Sum(struct TreeNode* node)
{ if (node ==Null)
      ~~returnreturn 0;~~
      return 0;
  return sum(node->left) + node->data + Sum(node->right);
}
```

WANG Disong 1155113946

第 1 题
（答题不得寫在紅綫外）

第 頁

I.

0	1	2	3	4	5	6	7	8	9	10
11	12	13		33	5		24	27	9	

← hash table ~~index~~ location

~~h(44)~~

$h(44,0) = h(44) = 44 \% 11 = 0$. examine location 0. $44 \neq 11$. fail.

~~$h(44,1) = (h(44) + 1 \cdot h'(44)) \% 11 = 5$. examine location 5. $44 \neq 5$. fail.~~

$h(44,2) = (h(44) + 2 \cdot h'(44)) \% 11 = 10$. examine location 10. null. fail. and stop.

II. ① First calculate the mediums of A_1, A_2 $\boxed{\text{denoted as } m_1, m_2 \text{ respectively}}$ if their mediums are the same, return the medium.

② else if $m_1 > m_2$. the medium is in one of the following two subarrays: $A_1[0 : m_1 \text{location}]$. from $0 \to m_1$.
$A_2[m_2 \text{location} : n_2]$ from $m_2 \to n_2$.

③ ~~if~~ else if $m_2 > m_1$. the medium is in one of the following two ~~subarra~~ subarrays: $A_1[m_1 \text{location} : n_1]$. from $m_1 \to n_1$
$A_2[0 : m_2 \text{location})$. from $0 \to m_2$.

④ Repeat the above process until the size of both subarrays is 2.

⑤ if the size of two ~~subarrays~~ subarrays is 2.
then medium is $(\max(A_1[0], A_2[0]) + \min(A_1[1], A_2[1]))/2$.

the time complexity is $O(\log(\min(n_1, n_2)))$.

WANG Disong 1155113946

第 2 題
（答題不得寫在紅綫外）

第 頁

1. Deadlock : when two or more processes are unable to proceed because each is waiting for one of the others to do something.

Starvation : when a runnable process is overlooked indefinitely by the scheduler; although it is able to proceed, it is never chosen.

II. the virtual memory address is a reference to a memory location independent of the current assignment of data to memory; a translation must be made to physical address before the memory ~~ad~~ access can be achieved; the physical memory address is the actual location in main memory.

III. a. x will be 3

b. ~~bool atomic_add(&word, value)~~

```
{
    while compare_and_swap(word, 1, word+value) is False
        /* do nothing */.
}
```

WANG Disong 1155113946

IV. LRU:

0	2	3	7	1	3	2	3	7	6	5	3	2	6	5	6
0	0	0	7	7	7	2	2	2	6	6	6	2	2	2	2
	2	2	2	1	1	1	1	7	7	7	3	3	3	5	5
		3	3	3	3	3	3	3	3	5	5	5	6	6	6
F	F	F	F	F		F		F	F	F	F	F	F	F	

FIFO:

0	2	3	7	1	3	2	3	7	6	5	3	2	6	5	6
0	0	0	7	7	7	7	3	3	3	5	5	5	6	6	6
	2	2	2	1	1	1	1	7	7	7	3	3	3	5	5
		3	3	3	3	2	2	2	6	6	6	2	2	2	2
F	F	F	F	F		F	F	F	F	F	F	F	F	F	

we can see page faults for LRU : 13

for FIFO : 14.

WANG Disong 1155113946

第 3 題
（答題不得寫在紅綫外）

第 頁

(a) I. $R_1 \leftarrow COURT \bowtie MATCH$

$R_2 \leftarrow {}_{MID}\mathcal{G}_{sum(MID)\ as\ SMID}\ R_1$

~~$Answer \leftarrow \Pi_{CID}(\sigma_{SMID \geq 100}(R_2))$~~

$Answer \leftarrow \Pi_{CID}(\sigma_{SMID \geq 100}(R_2))$

II. $R_1 \leftarrow \sigma_{PLAYER.PID = MATCH.p1\ and\ MATCH.p1wins=1}(PLAYER \times MATCH)$

$R_2 \leftarrow \sigma_{PLAYER.PID = MATCH.p2\ and\ MATCH.p1wins=0}(PLAYER \times MATCH)$

$R_3 \leftarrow R_1 \cup R_2$

~~$R_4 \leftarrow$ .PY~~

$Answer \leftarrow {}_{PLAYER.PID}\mathcal{G}_{sum(MID)\ as\ SMID}(R_3)$.

(b). I. select CID
from COURT natural join MATCH
~~where~~
group by CID
having Sum(MID) ≥ 100

II. select PLAYER.PID, Sum(MID)
from PLAYER, MATCH
where (PLAYER.PID=MATCH.p1 and MATCH.p1wins=1) or
(PLAYER.PID=MATCH.p2 and MATCH.p1wins=0)
group by PLAYER.PID.

WANG Disong 1155113946

第 3 題
（答題不得寫在紅綫外）

第 頁

(c) by using COURT as the outer relation.
it has $1000/50 = 20$ blocks

B^+-tree heighth is : $\log_{\lceil \frac{20}{2} \rceil} 50000 \approx 5$.

$\therefore$ number of block transfers :

$$20 + 1000 \times (5+1) = 6020.$$

WANG Disong 1155113946

第 4 題
（答題不得寫在紅綫外）

第 頁

(i) (a) The current Gini index: $1-(\frac{3}{8})^2-(\frac{5}{8})^2=\frac{15}{32}$

~~Gini~~ $GINI_{split-A}=\frac{3}{8}\cdot(1-(\frac{3}{3})^2-(\frac{0}{3})^2)+\frac{5}{8}\cdot(1-(\frac{2}{5})^2-(\frac{3}{5})^2=\frac{3}{10}$

$GINI_{split-B}=\frac{4}{8}\cdot(1-(\frac{2}{4})^2-(\frac{2}{4})^2)+\frac{4}{8}\cdot(1-(\frac{3}{4})^2-(\frac{1}{4})^2=\frac{7}{16}$.

$\therefore$ the Gini index gain for A: $\frac{15}{32}-\frac{3}{10}=\frac{27}{160}$

for B: $\frac{15}{32}-\frac{7}{16}=\frac{1}{32}=\frac{5}{160}$.

$\Rightarrow$ Gini index gain for A is larger.

$\therefore$ select A as splitting attribute.

(b) let C denotes the label. then $P(C=+)=\frac{5}{8}$, $P(C=-)=\frac{3}{8}$.

$P(A=1|C=+)=\frac{3+1}{5+2}=\frac{4}{7}$ $P(A=1|C=-)=\frac{0+1}{3+2}=\frac{1}{5}$

~~$P(B=0|C=+)=\frac{2+1}{5+2}=\frac{3}{7}$~~

$P(B=0|C=+)=\frac{3+1}{5+2}=\frac{4}{7}$ $P(B=0|C=-)=\frac{1+1}{3+2}=\frac{2}{5}$.

$\Rightarrow P(C=+|A=1,B=0)\propto P(A=1|C=+)P(B=0|C=+)P(C=+)=\frac{4}{7}\times\frac{4}{7}\times\frac{5}{8}=\frac{10}{49}$

$P(C=-|A=1,B=0)\propto P(A=1|C=-)P(B=0|C=-)P(C=-)=\frac{1}{5}\times\frac{2}{5}\times\frac{3}{8}=\frac{3}{100}$

$\Rightarrow X_q=(A=1,B=0)$ belongs to '+'.

WANG Disong 1155113946

第 4 題
(答題不得寫在紅綫外)

第 頁

(ii) (a) ① step0 : initial centroids (7, 50, 60).

② step1 : 3 clusters ⇒ {7, 13, 20, 25}
{30, 42, 50}
{60}.

new centroids : (16.25, $\frac{122}{3}$, 60).

③ step2 : 3 clusters ⇒ {7, 13, 20, 25} ~~30~~
{30, 42, 50}
{60}.

④ centroids are not changed. ∴ the final clusters:
{7, 13, 20, 25}   {30, 42, 50}   {60}.

(b) By using 'MIN' as the Inter-cluster similarity.

① Initially. 8 clusters : {7} {13} {20} {25} {30} {42} {50} {60}.

② {20} and {25} are closese. merge : {7} {13} {20, 25} {30} {42}, {50}, {60}

③ {20,25} and {42} are closese. merge: {7} {13} {20,25,30} {42} {50}, {60}

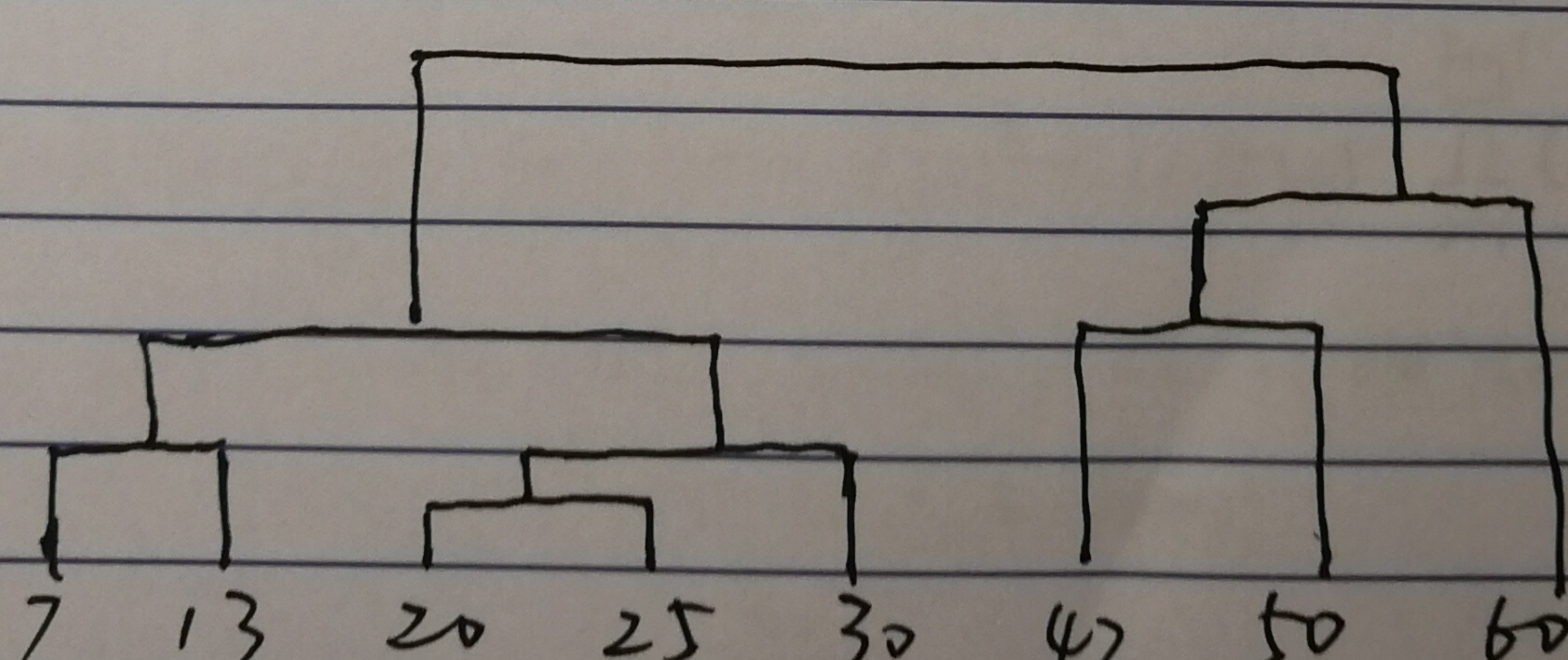
④ {7} and {13} are closese. merge: {7,13} {20,25,30}, {42} {50} {60}

⑤ {7,13} and {20,25,30} are closese. merge: {7,13,20,25,30} {42} {50} {60}

⑥ {42} and {50} are closese, merge: {7,13,20,25,30} {42,50}, {60}.

⑦ {42,50} and {60} are closese. merge: {7,13,20,25,30} {42,50,60}

⑧ final merge : {7, 13, 20, 25, 30, 42, 50, 60}



WANG Disong 1155113946

第 5 題
（答題不得寫在紅綫外）

第 頁

i. ① A document collection
② A test suite of information needs, expressible as queries.
③ A set of relevance judgements, standardly a binary assessment of either relevant or nonrelevant for each query-doc pair.

ii. precision values may decrease or increase or not change.
recall values will not decrease, it may increase or not change.

iii Because we can always get 100% recall by just returning all documents, and we can always get a 50% arithmetic mean by the same process, which means arithmetic mean is not a suitable metric.

iv. a. $g = \underset{c \in C}{\arg\max}\, P(c|d)$, where $P(c|d)$ is the posterior probability of class c given the doc d.

b. $P(c|d) = \frac{P(c,d)}{P(d)} = \frac{P(d|c)P(c)}{P(d)}$. For each doc, $P(d)$ is same.

$\therefore\ g = \underset{c \in C}{\arg\max}\, P(d|c)P(c)$

For multinomial. $g = \underset{c \in C}{\arg\max}\, \cancel{P} \prod_{i=1}^{|d|} P(t_i|c)P(c)$. t_i is ith term in d.

For Bernoulli. $g = \underset{c \in C}{\arg\max} \prod_{i=1}^{|V_1|} P(t_i|c)P(c) \cdot \prod_{j=1}^{|V_2|} (1-P(t_j|c))$

V_1 is the terms in d. V_2 is the terms that are not in d.
$V_1 + V_2$ is the vacabulary.

c. Bernoulli model. Because it is sensitive to noise features, feature selection is required, else the accuracy will be low.