Q4 (i) (a) Select A

A	+	−
1	3	0
0	2	3

$Gini(Ori) = 1-(\frac{5}{8})^2-(\frac{3}{8})^2 = \frac{15}{32} \approx 0.469$

$Gini(A) = \frac{5}{8}\times(1-(\frac{2}{5})^2-(\frac{3}{5})^2) + \frac{3}{8}\times(1-(\frac{3}{3})^2)$

$= \frac{3}{10} = 0.3$

$\Delta Gini(A) \approx 0.169$

Select B

B	+	−
1	2	2
0	3	1

$Gini(B) = \frac{4}{8}\times(1-(\frac{1}{2})^2-(\frac{1}{2})^2) + \frac{4}{8}\times(1-(\frac{3}{4})^2-(\frac{1}{4})^2)$

$= \frac{7}{16} \approx 0.438.$

$\Delta Gini(B) = 0.031$

∴ Select A

(b) Suppose $x_9 = +$.

Considering the Bayes classifier $P(x=+|A=1, B=0) = \frac{P(A=1, B=0|x=+)\cdot P(x=+)}{P(A=1, B=0)}$

$= \frac{P(A=1, B=0|x=+)\cdot P(x=+)}{P(A=1, B=0|x=+)P(x=+) + P(A=1, B=0|x=-)P(x=-)}$

$= \frac{P(A=1|x=+)P(B=0|x=+)\cdot P(x=+)}{P(A=1|x=+)P(B=0|x=+)\cdot P(x=+) + P(A=1|x=-)P(B=0|x=-)\cdot P(x=-)}$

$= \frac{P(A=1, B=0, x=+)}{P(A=1, B=0, x=+) + P(A=1, B=0, x=-)} = \frac{2+1}{2+1+1\frac{1}{2}} = \frac{3}{4}$

∴ $x_9 = +$ predicted by Bayes classifier

(ii) (a) ①: $C_1=7: \{7, 13, 20, 25\}$. $C_2=50: \{30, 42, 50\}$ $C_3=60$ $\{60\}$

② Update $C_1 = \frac{65}{4} = 16.25$ $C_2 = \frac{122}{3}$ $C_3 = 60$.

③ Cluster: $C_1=\frac{65}{4}: \{7, 13, 20, 25\}$ $C_2=\frac{122}{3}: \{30, 42, 50\}$ $C_3=60: \{60\}$.

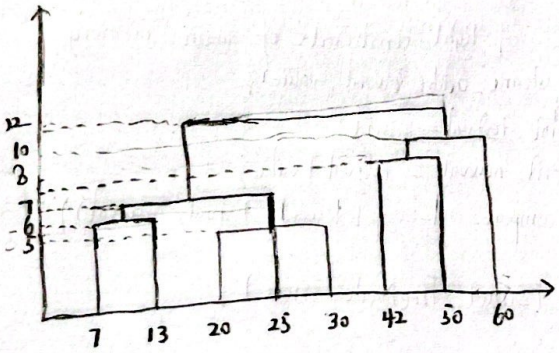
(b) ①

	7	13	20	25	30	42	50	60
7	0							
13	6	0						
20	13	7	0					
25	18	12	⑤	0				
30	23	17	10	⑤	0			
42	35	29	22	17	12	0		
50	43	37	30	25	20	8	0	
60	53	47	40	35	30	18	10	0

Merge 20 25.

	7	13	20∪25	30	42	50	60
7	0						
13	6 (3)	0					
20∪25	13	7 (4)	0				
30	23	17	(5) (2)	0			
42	35	29	17	12	0		
50	43	37	25	20	8 (5)	0	
60	53	47	35	30	18	10 (6)	0

7	13	20∪25∪30	42	50	60	
7∪13	20∪25∪30	42	50	60		
7∪13∪20∪25∪30	42	50	60			
7∪13∪20∪25∪30	42∪50	60				
7∪13∪20∪25∪30	42∪50∪60					
7∪13∪20∪25∪30∪42∪50∪60						



Q5: (i) Queries; Documents; Results of relevance of queries and documents.

(ii)

	Relevant	NonRelevant
Retrieved	tp	fp
Not Retrieved	fn	tn

More relevant documents retrieved, the precision will increase; otherwise

The Recall will increase if more documents are retrived

(iii) If we retrive all the documents to get 100% recall, we can obtain a F-measure at least 50% using the arithmetic mean.

We use the F-measure of harmonic mean, if the gap of P and R is large, the F-measure will be close to the lower one, which is reasonable.

(iv) (a) $g = \arg\max_{c \in C} P(c|d)$

(b) $g = \arg\max_{c \in C} P(c|d) = \arg\max_{c \in C} \frac{P(d|c)\,P(c)}{P(d)} = \arg\max_{c \in C} P(d|c)\,P(c) = \arg\max_{c \in C} P(d,c)$

$P(d,c)$ represents the generative process

Multinomial: $P(d|c) = P(\langle t_1, \dots, t_{n_d} \rangle | c) = \prod_{1 \le k \le n_d} P(X_k = t_k | c)$, where $\langle t_1, \dots, t_{n_d} \rangle$ is the token sequence in document d

Bernauli: $P(d|c) = P(\langle e_1, \dots, e_M \rangle | c) = \prod_{i=1}^{M} P(U_i = e_i | c)$, where $\langle e_1, \dots, e_M \rangle$ is a M-dimensional boolean vector

(c) Bernoulli Model — It is more sensitive to noise.

Q2. iv. LRU:

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

Page faults = 13

FIFO:

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

Page faults = 14

I. A set of processes is deadlock when each process in the set is blocked awaiting an event that can only be triggered by another blocked process in the set.

II. Physical memory address: address in actual RAM, main memory
Virtual memory address: address in memory on disk, that allows for effective multiprogramming and relieves the user of tight constraints of main memory

III. a. 3 → 3 b.

```
bool atomic_add(&word, value){
    do{
        int testval = *word;
        int newval = testval + value;
    }while(compare_and_swap(*word, testval, newval));
}
```

Q1. I

```
int isSumTree (struct TreeNode * node) {
    int sum = 0;
    if (node == NULL)
        return 0;
    if (node->left != NULL)
        sum = isSumTree (node->left);

    if (node->right != NULL)
        sum += isSumTree (node->right);

    if (node->value == sum)
        return sum;
    else
        print ("Not a SumTree!")
}.
```

II $h(k,i) = (h(k) + i \cdot h'(k)) \% 11$

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

0 5 empty

44: 11 5 10.

III Compare the elements in A_1 and A_2 in sequence and add the smaller one into a new array A_n as follows

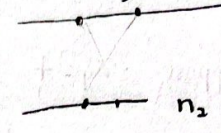
```
i ← 0 , j ← 0;
A_n ← < > ;
while A_1 ≠ NIL and A_2 ≠ NIL
do if A_1[i] <= A_2[j]
   then ADD(A_n, A_1[i])
        i ← i+1
   else if A_1[i] > A_2[j]
   then ADD(A_n, A_2[j])
        j ← j+1
a = ⌈(n_1 + n_2)/2⌉
b = ⌊(n_1 + n_2)/2⌋
return 1/2 (A_n[a] + A_n[b])
```

compare $A_1[\frac{n_1}{2}]$ with $A_2[\frac{n_2}{2}]$

if $A_1[\frac{n_1}{2}] > A_2[\frac{n_2}{2}]$

compare $A_1[\frac{n_1}{4}]$ with $A_2[\frac{3}{4}n_2]$

if $A_1[\frac{n_1}{4}] > A_2[\frac{3}{4}n_2]$



Time Complexity $(\log n_1 + \log n_2)$

Q3. (a) Answer ← $\Pi_{CID}(\sigma_{CMID \geq 100}(R_1))$

Select CID from MATCH

group by CID having count(MID) >= 100

(b) $R_1 \leftarrow \sigma_{P1Wins = True, PID = P1}(PLAYER \bowtie MATCH)$

$R_2 \leftarrow \sigma_{P1Wins = False, PID = P2}(PLAYER \bowtie MATCH)$

Answer ← $PID\,\mathcal{G}_{COUNT(MID)}(R1 \bowtie R2)$

Select PID, COUNT(MID) from

PLAYER natural join MATCH

where P1Wins = True , PID = P1.

group by PID

(c) COURT : $b = \frac{1000}{50} = 20$

MATCH : $b = \frac{50000}{40} = 1250$

$h = \lceil \log_{\frac{20}{2}}(50000) \rceil = 5$

COURT as outer relation

cost = $20 + 1000 \times (h+1) = 6020$ block transfers