S 17

Q1 (a) // Palindrome

```
int minCut (string s){
    int len = s.length();
    vector<vector<bool>> isPal(len, vector<bool>(len));

    for (int i = 0; i < len ; ++i)
    {
        for (int j=0 ; j<=i ;j++)
        {
            char ch1 = s[i], ch2 = s[j];
            if( ch1 == ch2 && (i <= j+1 || isPal[j+1][i-1]))
            {
                isPal[j][i] = true;
            }
        }
    }
```

cut X

(b) for a string with length n

construct a nxn diagonal matrix

the elements in it is more than the first element are set to be 1, others are 0

for example

	a	c	a	d	d	g	e	g
a	0							
c	1	0						
a	0	0	0					
d	1	1	1	0				
d	1	1	1	0	0			
g	1	1	1	1	1	0		
e	1	1	1	1	1	0	0	
g	1	1	1	1	1	0	1	0

s

s[1] = 1

s[2] = [0, 0]

s[3] = [1, 1, 1]

s[4] = [1, 1, 1, 0]

s[5] = [1, 1, 1, 1, 1]

s[6] = [1, 1, 1, 1, 1, 0]

s[7] = [1, 1, 1, 1, 1, 0, 1]

$O(n^2)$

(4) Precision = $\frac{TP}{TP+FP}$

IS. 17. Q4. (a) step 1: classify in $C_1 = 1100$, $C_2 = 1600$

C_1 {0, 200, 300, 900, 1100} C_2 {1600}.

step 2: update $C_1 = 500$, $C_2 = 1600$

step 3: classify in $C_1 = 500$, $C_2 = 1600$

C_1 {0, 200, 300, 900} C_2 {1100, 1600}

step 4: update $C_1 = 350$ $C_2 = 1350$

step 5: classify in $C_1 = 350$ $C_2 = 1350$

C_1 {0, 200, 300} C_2 {900, 1100, 1600}

step 6: update $C_1 = \frac{300}{3}$ $C_2 = 1200$

(b) $SSE = (0 - \frac{500}{3})^2 + (200 - \frac{500}{3})^2 + (300 - \frac{500}{3})^2 + (900-1200)^2 + (1100-1200)^2 + (1600-1200)^2$

$= 100^2 \times (\frac{25}{9} + \frac{1}{9} + \frac{16}{9} + 9 + 1 + 16)$

$= \frac{920000}{3}$

$BSS = 3 \times (\frac{500}{3} - \frac{2050}{3})^2 + 3 \times (1200 - \frac{2050}{3})^2 = 3 \times (\frac{1550}{3})^2 \times 2 = \frac{4805000}{3}$

(c) for 200: $a = \frac{1}{2} \times [\sqrt{(200-0)^2} + \sqrt{(200-300)^2}] = 150$

$b = \frac{1}{3} \times [\sqrt{(200-900)^2} + \sqrt{(200-1100)^2} + \sqrt{(200-1600)^2}] = 1000$

$S = \frac{b-a}{\max(a,b)} = 1 - \frac{a}{b} = 0.85$

for 1100: $b = \frac{1}{3} \times [\sqrt{(1100-0)^2} + \sqrt{(1100-200)^2} + \sqrt{(1100-300)^2}] = \frac{2600}{3}$

$a = \frac{1}{2} \times [\sqrt{(1100-900)^2} + \sqrt{(1100-1600)^2}] = 350$

$S = 1 - \frac{a}{b} = 1 - \frac{350}{\frac{2600}{3}} = 0.625$

(d) step 1:

d	0	200	300	900	1100	1600
0	0					
200	200	0				
300	300	100	0			
900	900	700	600	0		
1100	1100	900	800	200	0	
1600	1600	1400	1300	700	500	0

step 2:

0	200∪300	900	1100	1600
0				
200∪300	300	0		
900	900	700	0	
1100	1100	900	200	0
1600	1600	1400	700	500	0

step 3:

d	0	200∪300	900∪1100	1600
0	0			
200∪300	300	0		
900∪1100	1100	700	0	
1600	1600	1400	700	0

step 4:

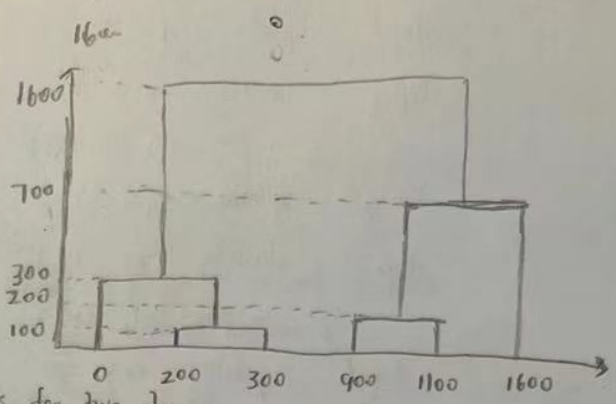
0∪200∪300	900∪1100	1600
0∪200∪300	0	
900∪1100	1100	0
1600	1600	700	0

step 5 : d 0∪200∪300 900∪1100∪1600

0∪200∪300 0

900∪1100∪1600 1600

dendrogram



Q 5. (a) (i) step1 Compute the centroids for two classes

$\mu_1 = \frac{1}{2}(d_1 + d_2) = (0, 0.4975, 0.4975, 0.707, 0)$

$\mu_2 = \frac{1}{2}(d_3 + d_4) = (0.5635, 0, 0.7085, 0.274, 0)$

(2) The boundry between two classes in Rocchio classification is the set of points with equal distance from the two centroids

The classification of d_t is determined by the minimum distance between d_t and all the centroids

(3) For the multimodal classification problem. Rocchio classification cannot classifier properly as the following example

For class 1 with the label ".", it is grouped by two clusters, the test data should be classified in ".", but the centroid of two clusters is far and "o" will be mis-classified in cluster "x"

(b) (i) true positive (TP): a sample with actual label "1" is classified in "1"

false positive (FP): a sample with actual label "0" is classified in "1"

false negative (FN): a sample with actual label "1" is classified in "0"

true negative (TN): a sample with actual label "0" is classified in "0"

Using concepts in IR

Retrived!!!

(2) $\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP}$

(3)

sample	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
label	1	1	1	1	1	0	0	1	1	1
prediction	1	1	1	1	1	1	1	1	1	1

Accuracy = 80%, but such a classifier may predict all the samples to be "1", while FPR is 100% if we have many samples with label "0", the performance will be b

(4) $\text{Precision} = \frac{TP}{TP + FP}$ $\text{Recall} = \frac{TP}{TP + FN}$

15)

	Relevant	Non relevant
Retrived	1	9
Not Retrived	0	990

the number of true relevant is 1

$Acc = \frac{1+990}{1000} = 0.991$ Recall = 100%

Precision = 10%

in such a system, to improve the performance, we prefer the precision to recall

Q2: (a) a) "mutex" ensure that the producer can't add data into full buffer and consumer can't remove datas from an empty buffer (enforce mutual exclusion)

"empty" is used to keep track of the number of empty space

"full" is used to keep track of the number of items in the buffer

(2) "Producer Process" line 8 → signal (mutex) signal(empty) × → signal(full)

b) The system spends most of its time swapping pieces rather than executing instructions

Stop the process, Increase the memory

c)

	1	2	3	4	5	3	4	1	6	7	8	9	7	8	9	7	8	9	5	4	5	4	2
	1	1	1	4	4	4	4	4	4	7	7	7	7	7	7	7	7	7	5	5	5	5	5
		2	2	2	5	5	5	1	1	1	8	8	8	8	8	8	8	8	8	4	4	4	4
			3	3	3	3	3	3	6	6	6	6	9	9	9	9	9	9	9	9	9	9	2
	F		F	F	F	F	F		F	F	F	F		F				F	F				F

page faults = 13

Q3: (a) # CUSTOMER ⋈ RENTAL = 20000 # CUSTOMER ⋈ VEHICLE = 5000×1000 = 5000000

(b) $b_c = \frac{5000}{100} = 50$ $b_R = \frac{20000}{50} = 400$

CUSTOMER as outer relation cost: $\lceil \frac{50}{4-2} \rceil \times 400 + 50$ block transfers $+ 2 \times \lceil \frac{50}{4-2} \rceil$ seeks

= 10050 block transfers + 50 seeks

RENTAL cost: $\lceil \frac{400}{4-2} \rceil \times 50 + 400$ block transfers $+ 2 \times \lceil \frac{400}{4-2} \rceil$ seeks

= 10400 block transfers + 400 seeks

(c) cost = $b_r + n_r(h+t) = 50 + 5000 \times (3+1) = 20050$