PART-2

PROBLEM-1: CLASSIFICATION BY ASSOCIATION:

1) Given: Fix Min support = 3.

TRAIN data:

O	a_1	a_2	C
O_1	1	1	1
O_2	0	0	0
O_3	0	1	0
O_4	0	0	0
O_5	1	1	1
O_6	1	1	0
O_7	0	0	0
O_8	1	0	1

Answer: Transactional Data & Supporting Calculations:

	$I_1(a_1=0)$	$I_2(a_1=1)$	$I_3(a_2=0)$	$I_4(a_2=1)$	$I_5(C=0)$	$I_6(C=1)$
1	−	+	−	+	−	+
2	+	−	+	−	+	−
3	+	−	−	+	+	−
4	+	−	+	−	+	−
5	−	+	−	+	−	+
6	−	+	−	+	+	−
7	+	−	+	−	+	−
8	−	+	+	−	−	+
Count	4	4	4	4	5	3

Step 1: Generating 1-itemset Frequent Pattern.

given min support count = 3.

L1;

(Frequent 1 itemset)

Itemset	Support count
I_1	4
I_2	4
I_3	4
I_4	4
I_5	5
I_6	3

The set of frequent 1-itemsets, L1, consists of the candidate 1-itemsets satisfying minimum support count.

Step 2: Generating 2-itemset frequent Pattern.

→ To discover the set of frequent 2-itemsets, L_2, the algorithm uses "L_1 Join L_1" to generate a candidate set of 2-itemsets, C_2 with support count

→ 2-itemsets, L_2 is then determined consisting of those candidate 2-items satisfying minimum support count.

C_2 :
(Candidate two itemsets)

Itemset	Support Count
1, 2	0
1, 3	3
1, 4	1
1, 5	4
1, 6	0
2, 3	1
2, 4	3
2, 5	1
2, 6	3
3, 4	0
3, 5	3
3, 6	1
4, 5	2
4, 6	2
5, 6	0

L_2 :
(frequent 2 item set)

Item set	support count
1, 3	3
1, 5	4
2, 4	3
2, 6	3
3, 5	3

Step 3: Generating 3-itemset Frequent Pattern.

* In order to generate C_3, we first compute L_2 Join L_2.

(Point 1) Here we use prune step to reduce size of C_3. By using property of "Apriori", we reduce the size of C_3.

C_3 : (Without Prune step) (Candidate 3 itemset)

Item set	support count
1, 3, 5	3
2, 4, 6	2

↓ (After Prune step)

Item set	Support Count
1, 3, 5)	3

L_3 :- (Frequent 3 itemset)

Item set	Support Count
1, 3, 5	3

* 3-itemset L_3, is determined consisting of those candidate 3-items satisfying minimum support count.

Further explaination w.r.t point 1 :-

↳ As mentioned above, to find C_3, we first compute

L_2 Join L_2.

$C_3 = L_2$ Join $L_2 = \{\{1, 3, 5\}, \{2, 4, 6\}\}$.

↳ Now Join step is complete.

↳ We need to check for Prune-step and check if there is any thing that needs to be removed,

↳ "Apriori property" says that all subsets of a frequent itemset must also be frequent.

↳ Take $\{1, 3, 5\}$

* The 2-item sets of it are $\{1, 3\}$, $\{1, 5\}$, $\{3, 5\}$

All of them are members of L_2

So we keep $\{1, 3, 5\}$ in C_3

↳ Take $\{2, 4, 6\}$.

* The 2-item sets of it are $\{2, 4\}$, $\{4, 6\}$, $\{2, 6\}$

But here $\{4, 6\}$ is not the member of L_2 and hence it is not frequent → violating Apriori property.

Thus we remove $\{2, 4, 6\}$ from C_3.

∴ Reduced C_3 (Candidate 3 item) is $\{1, 3, 5\}$.

↳ We now use this to determine L_3 (Candidate 3 items satisfying minimum support count.

We get $L_3 = \{1, 3, 5\}$.

We end the algorithm here as C_4 is a null set.

Step 4: Generating Association Rules from Frequent Itemsets

→ When generating classification by association rules, we take association rules of the form $(p_1 \wedge p_2 \wedge \dots p_k) \rightarrow class = c$

→ In our case, class is either I_5 or I_6

The non-empty subsets needed to create association rules

$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\},$

$\{1,3\}, \{1,5\}, \{2,4\}, \{2,6\}, \{3,5\}, \{1,3,5\}$.

↳ To create classification rules, we consider only subsets that contain class I_5 (or) I_6.

∴ Frequency set needed to form ClassificationRules is:

$$L = \{\{1,5\}, \{2,6\}, \{3,5\}, \{1,3,5\}\}$$

Short-hand Representation of Rules :-

① $1 \rightarrow 5$

② $3 \rightarrow 5$

③ $1 \nrightarrow 3 \rightarrow 5$

④ $2 \rightarrow 5$

Predicate form Representation of Rules:-

Rule 1: $a_1(x,0) \rightarrow c(x,0)$

Rule 2: $a_2(x,0) \rightarrow c(x,0)$

Rule 3: $a_1(x,0) \wedge a_2(x,0) \rightarrow c(x,0)$

Rule 4: $a_1(x,1) \rightarrow c(x,1)$

2) Testing:

	a_1	a_2	C [Test Data class]	Assigned class as per Rules	Correctly satisfied.
O_1	1	1	1	1	Yes
O_2	1	0	0	?	No
O_3	0	0	1	0	No
O_4	0	0	0	0	Yes

Record O_1: $a_1(x,1) \wedge a_2(x,1) \rightarrow c(x,1)$

from Rule 4, this is True

Record O_2: $a_1(x,1) \wedge a_2(x,0) \rightarrow c(x,0)$

As per Rule 4, c should be 1

As per Rule 2, c should be 0

Nothing can be said/concluded about this assigned class value. It is not satisfied since final class value is ambiguous.

Record O_3: $a_1(x,0)$ & $a_2(x,0) \rightarrow c(x,1)$

As per Rule 3, c should be 0.

∴ This is wrong / not satisfied.

Record O_4: $a_1(x,0)$ & $a_2(x,0) \rightarrow c(x,0)$

As per Rule 3, This is satisfied.

Predictive Accuracy $= \frac{2}{4} \times 100$

$= 50\%$.