Revision and Preparation for the Exam

Question 1

Given the following transaction record

Transaction Records	
Transaction ID	Items
#1	apple, banana, coca-cola, doughnut
#2	banana, coco-cola
#3	banana, doughnut
#4	apple, coca-cola
#5	apple, banana, doughnut
#6	apple, banana, coca-cola

a b c d

b c

b d

a c

a b d

a b c

{a:4; b:5 ; c:4; d:3}

1. Build the FP-tree using a minimum support ***min_sup*** = 2. Show how the tree evolves for each transaction.

2. With the previous transaction record, Use the Apriori algorithm on this dataset and verify that it will generate the same set of frequent itemsets with **min_sup** = 2.

3. Suppose that { **Apple, Banana, Doughnut** } is a frequent item set, derive all its association rules with

   min_confidence = 70%

1. {b:5 , a:4, c:4 , d:3}

T_1 : b a c d

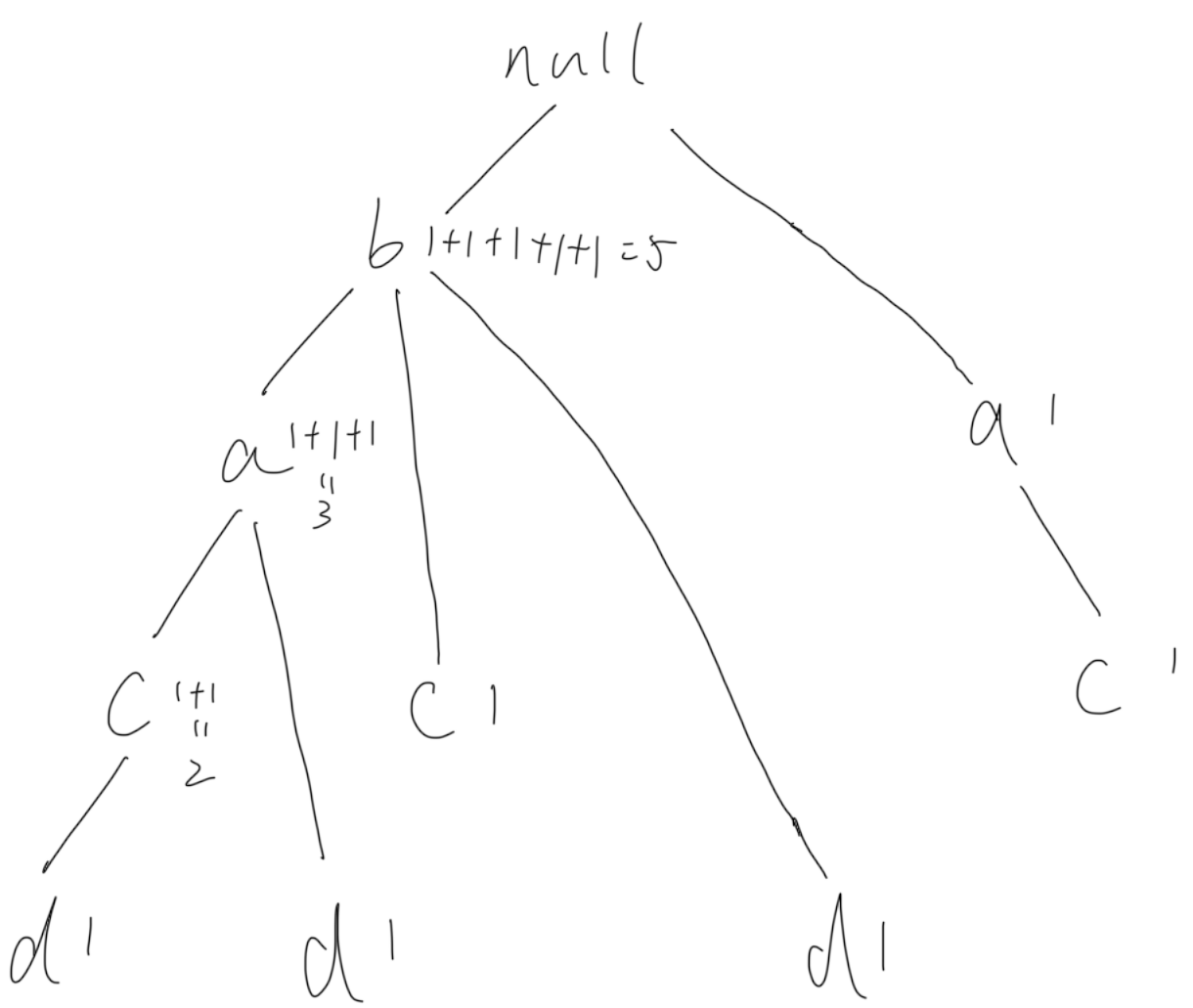
T_2 : b c

T_3 : b d

T_4 : a c

T_5 : b a d

T_6 b a c



2.

a 4
b 5
c 4
d 3

→

ab : 3
ac : 3
ad : 2
bc : 3
bd : 3

→

abc : 2
abd : 2
bcd : 1 X

↓

abcd : 1 X

a 4
b 5
c 4
d 3

ab : 3
ac 3

ad 2

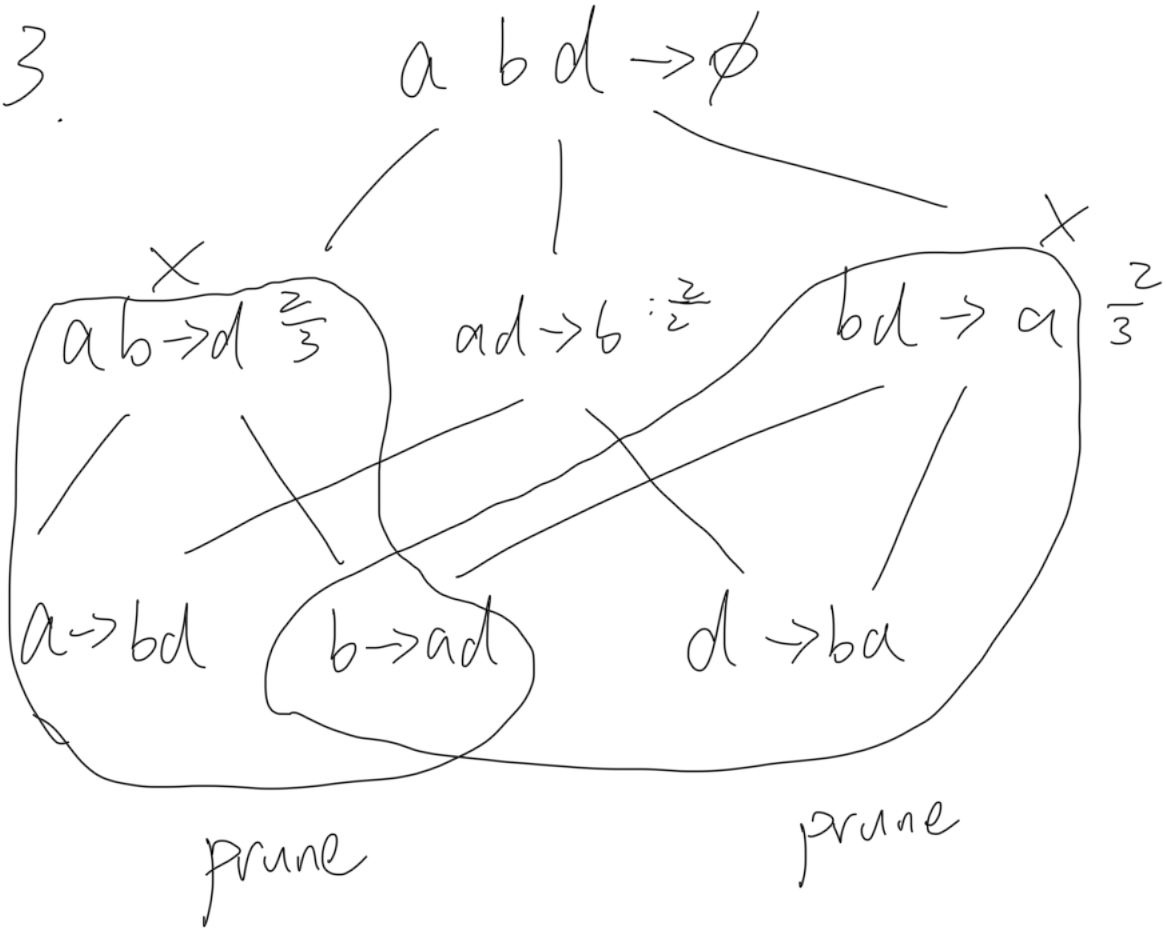
bc 3

bd 3

abc 2

abd 2

3.



so only have rule ad → b

employees { employee_id, first_name, last_name, department_id, Salary }

department { department_id, department_name }

Databases

Question 2

You have two tables, employees and departments, in your database. The employees table contains the following columns: employee_id, first_name, last_name, department_id, and salary. The departments table contains the columns: department_id and department_name. Write a SQL query to retrieve the first_name of each employee along with the department_name they belong to.

```
select first_name, department_name
from employees natural join departments
```

Question 3

Assume you have four datasets: "employees", "departments", "projects", and "assignments". Thescheme of these databases are as follows:

- The "employees" dataset contains employee information including a unique "employee_ID","employee_name", and "department_ID".
- The "departments" dataset contains department information including a unique "department_ID" and"department_name".
- The "projects" dataset contains project information including a unique "project_ID", "project_name",and "department_ID".
- The "assignments" dataset contains information about which employees are assigned to whichprojects, including the "employee_ID" and "project_ID" associated with each assignment.

employee_name, department_name

Write a SQL query to retrieve the name and department of all employees who are assigned to a projectthat is not in the same department as their own department.

↳ project's department ≠ employee's department

Question 4

A national park has created a dataset to help hikers determine if a reptile they encounter could be venomous.

	Head	Eyes	Size	Venomous	
1	Triangle	Elliptical	Small	Yes	+
2	Round	Round	Small	No	-
3	Narrow	Elliptical	Small	No	-
4	Narrow	Round	Large	No	-
5	Narrow	Elliptical	Large	Yes	+
6	Triangle	Round	Small	Yes	+
7	Narrow	Round	Large	No	-
8	Round	Elliptical	Large	No	-
9	Triangle	Elliptical	Small	Yes	+

Use Naïve Bayes to predict if the following example is venomous or not:
Head=narrow, Eyes=elliptical, Size=Large
Show the working for your calculations.

No : 5

Yes : 4

4

3. Select e.Employee_name, e.~~department~~_name d.department_name

From assignments

Inner join employees e Using (employee_ID)

Inner join projects p Using (Project_ID)

Inner join department d on e.department_ID = d.department_ID

Where e.department_ID != P.department_ID ;

4.

	No	Yes
	$\frac{5}{9}$	$\frac{4}{9}$
Head = narrow	3/5	1/4
Eyes = Eliptical	2/5	3/4
Size = large	3/5	1/4

$\frac{5}{9} \times \frac{3}{5} \times \frac{3}{5} \times \frac{2}{5} = \frac{2}{25}$

$= 0.08$ ✓

$\frac{4}{9} \times \frac{1}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{1}{48} = 0.02$

So "No"