Midterm Exam

Name:
SIN:

Problem 1. For the information system given below, find the set of rules describing C in terms of E, F, G by applying RSES algorithm.
Find the set of all coverings of C (reducts). Assume that Dom(E)={e1,e2}, Dom(F)={f1,f2,f3}, Dom(G)={g1,g2}, Dom(C)={c1,c2}.

X	E	F	G	C
x1	e1	f3	g2	c2
x2	e2	f3	g1	c1
x3	e2	f1	g1	c2
x4	e1	f2	g2	c2
x5	e2	f2	g2	c1

System S

Solution:

Discernibility Matrix

	x1	x2	x3	x4	x5
x1	-				
x2	EG	-			
x3	-	F	-		
x4	-	EFG	-	-	
x5	EF	-	GF	E	-

Discernibility Function

(E+G)(E+F)F(E+F+G)(G+F)E =FE
Reduct = {F,E}

Discovering Rules

New Discernibility Matrix

	x1	x2	x3	x4	x5
x1	-				
x2	E	-			
x3	-	F	-		
x4	-	EF	-	-	
x5	EF	-	F	E	-

EF(x1) = E(E+F) = E rule e1 -> c2 sup=2
EF(x2) = EF(E+F) = EF rule e2*f3 -> c1 sup=1
EF(x3) = F rule f1 -> c2 sup=1
~~EF(x4) = (E+F)E = E rule e1 -> c2~~
EF(x5) = (E+F)FE = EF rule e2*f2 -> c1 sup=1

Problem 2. Find the set of representative rules RR(3,75%) for the set of transactions: (A,C,D,F,H), (B,C,E,H,I), (A,B,C,E,H), (B,C,D,E,H), (A,D,E,H), (A,B,D,E,H,I) following Agrawal algorithm.

Solution.

Frequent Itemsets

A-4, B-4, C-4, D-4, E-5, F-1, H-6, I-2 (red color means NO)
AB-2, AC-2, AD-3, AE-3, AH-4, BC-3, BD-2, BE-4, BH-4, CD-2, CE-3, CH-4, DE-3, DH-4, EH-5
ADE-2, ADH-3, AEH-3, BCE-3, BCH-3, BEH-4, CEH-3, DEH-3
BCEH – 3 largest frequent itemset

Representative Itemsets (MARKED BY YELLOW COLLOR)
AD-3, AE-3, AH-4, BC-3, BE-4, BH-4, CE-3, CH-4, DE-3, DH-4, EH-5
ADH-3, AEH-3, BCE-3, BCH-3, BEH-4, CEH-3, DEH-3
BCEH – 3

Find representative rules from the largest representative itemset.

B-> CEH conf=3/4 (YES)
C->BEH conf=3/4 (YES)
E-> BCH conf = 3/5 (NO)
H->BCE conf = 3/6 (NO)
EH -> BC conf = 3/5 (NO)

Problem 3.
Discretize attributes A and B in the Decision Table T. {A, B} are classification attributes. D is the decision attribute.

X	A	B	D
x1	1	3	1
x2	8	5	2
x3	5	5	2
x4	3	8	1
x5	8	3	1
x6	3	5	2

Solution.

Dom(A): 1 3 5 8 Dom(B): 3 5 8
p1 p2 p3 q1 q2

F(x1,x2)= p1 + p2 + p3 + q1
F(x1,x3)= p1 + p2 + q1
F(x1,x6)= p1 + q1
F(x2,x4)= p2 + p3 + q2
F(x2,x5)= q1
F(x3,x4)= p2 + q2
F(x3,x5)= p3 + q1
F(x4,x6)= q2
F(x5,x6)= p2 + p3 +q1

Discretization: Dom(B) : (-,4), [4, 6.5), [6.5,-)
b1 b2 b3 labels assigned to intervals.

Discretized Table

X	A	B	D
x1	-	b1	1
x2	-	b2	2
x3	-	b2	2
x4	-	b3	1
x5	-	b1	1
x6	-	b2	2

Problem 4.
Follow DEAR2 algorithm to construct action rules reclassifying objects from class d0 to class d1 in a decision system S assuming that Table T shows all classification rules extracted from S. These classification rules (r1-r6) should be used to construct action rules. Attributes a and b are stable.

	a	b	c	d
r1		2		0
r2	2	1		1
r3		1	2	1
r4	3		1	0
r5		3	2	1
r6	3	1	2	1

Table T

Solution

Split by attribute a (smaller number of values than b)

	b	c	d
r1	2		0
r2	1		1
r3	1	2	1
r5	3	2	1

Table T(a2)

	c	d
r2		1
r3	2	1

Table T(a2,b1) - NO RULES

Table T(a2,b2) -NO RULES
Table T(a2,b3) -NO RULES

	b	c	d
r1	2		0
r3	1	2	1
r4		1	0
r5	3	2	1
r6	1	2	1

Table T(a3)

	c	d
r3	2	1
r4	1	0
r6	2	1

Table T(a3,b1)

Action Rule: a3*b1*(c1 -> c2) => (d0 -> d1)

	c	d
r1		0
r4	1	0

Table T(a3,b2) – NO RULES

	c	d
r4	1	0
r5	2	1

Table T(a3,b3)

Action Rule: a3*b3*(c1 -> c2) => (d0 -> d1)