Name:
SIN:

Midterm Exam

Problem 1. Find the set of representative rules RR(3,75%) for the set of transactions: (A,C,D,F,H,I), (B,C,D,H,E,I), (A,B,C,E,H), (A,C,D,E,I), (A,B,D,E) following Agrawal algorithm.

Solution:

A-4, B-3, C-4, D-4, E-4, F-1, H-3, I-3, AB-2, AC-3, AD-3, AE-3, AH-2, AI-2, BC-2, BD-2, BE-3, BH-2, BI-1, CD-3, CE-3, CH-3, CI-3, DE-3, DH-2, DI-3, EH-2, EI-2, HI-2, ACD-2, ACE-2, ADE-2, CDE-2, ~~CDH,~~ CDI-3, CEH-2, ~~CEI,~~ ~~CHI.~~

Representative rules from CDI-3:

C->DI conf= 3/4; D -> CI conf = ¾; I -> CD conf=3/3

Problem 2. 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	8	3	2
x2	5	5	2
x3	5	3	2
x4	3	8	1
x5	8	5	1
x6	3	5	1

Decision Table T

Solution:

Dom(A): 3 5 8 Dom(B): 3 5 8

Cuts p1 p2 q1 q2

F(x1,x4)= p1+p2+q1+q2; F(x1,x5)=q1; F(x1,x6)= p1+p2+q2;

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

F(x3,x5)=p2+q1; F(x3,x6)=p1+q1.

F(T)=

(~~p1+p2+q1+q2~~)q1(~~p1+p2+q2~~)(~~p1+q2~~)p2p1(~~p1+q1+q2~~)(~~p2+q1~~)(~~p1+q1~~)

Smallest set of cuts: q1, p1, p2.

New Domains: Dom(A)={(-4),[4,6),[6-)}, Dom(B)={(-4),[4-)}

Labels: a1 a2 a3 b1 b2

Discretized Table

X	A	B	D
x1	a3	b1	2
x2	a2	b2	2
x3	a2	b1	2
x4	a1	b2	1
x5	a3	b2	1
x6	a1	b2	1

Problem 3. Follow DEAR1 algorithm to find action rules reclassifying objects either from the class d0 to class d1 or from the class d1 to class d0 in table T. Attributes a, c are stable.

	a	b	c	d
X1	2	2	1	1
X2	1	1	1	1
X3	2	2	2	1
X4	1	3	1	0
X5	1	3	2	0
X6	2	1	2	0

Table T.

Solution.

Find classification Rules using LERS (find granules having single labels):

a1*={2,4,5}, a2*={1,3,6}; b1*={2,6}, b2*={1,3}< d1*, b3*= {4,5}<d0*;

c1*={1,2,4}, c2*={3,5,6}; d1*={1,2,3}, d0*={4,5,6}.

a1.b1*={2}<d1*, a1.c1*={2,4}, a1.c2*={5}<d0*,

a2.b1*={6}<d0*, a2.c1*={1}<d1*, a2.c2*={3,6},

b1.c1*={2}<d1*, b1.c2*={6}<d0*.

Set of rules (using Table representation):

a	b	c	d
	2		1
	3		0
1	1		1
1		2	0
2	1		0
2		1	1
	1	1	1
	1	2	0

Splitting table by decision attribute.

a	b	c
	2	
1	1	
2		1
	1	1

T(d1)

Splitting Table T(d1) by stable attribute c (has less values than a)

a	b
	2
1	1

T(d1,c?)

Splitting Table T(d1,c?) by stable attribute a

b
2

T(d1,c?,a?)

b
1

T(d1,c?,a1)

a	b
2	
	1

T(d1,c1)

Splitting Table T(d1,c1) by stable attribute a

b

T(d1,c1,a2)

b
1

T(d1,c1,a?)

a	b	c
	3	
1		2
2	1	
	1	2

T(d0)

Splitting Table T(d0) by stable attribute c (has less values than a)

a	b
	3
2	1

T(d0,c?)

Splitting Table T(d0,c?) by attribute a.

b
3

T(d0,c?,a?)

b
1

T(d0,c?,a2)

Let’s take the leaf below and build action rule from data in yellow color

b
2

T(d1,c?,a?)

a2*(b1 -> b2) => (d0 ->d1)

a	b
1	
	1

T(d0,c2)

Problem 4 Follow k-means algorithm (where k=2) to cluster objects in Table T. Chose y1, y6 as the initial seeds.

Y	M	N
y1	2	2
y2	2	8
y3	4	4
y4	8	4
y5	6	6
y6	6	4
y7	4	2

Table T

Solution.

Let's build clusters using seeds y1 and y6 based on Manhattan distance

Y	M	N
y1	2	2
y2	2	8
y7	4	2

8/3 12/3 (2.66, 4) - Center & New Seed

Y	M	N
y3	4	4
y4	8	4
y5	6	6
y6	6	4

24/4 18/4 (6, 4.5) – Center & New Seed

New clusters around these two seeds

Y	M	N
	2.66	4
y1	2	2
y2	2	8
y3	4	4
y7	4	2

12/4 16/4 (3, 4) new seed

Y	M	N
	6	4.5
y5	6	6
y4	8	4
y6	6	4

20/3 14/3 (6.67, 4.67) new seed

New clusters around these two seeds

Y	M	N
	3	4
y1	2	2
y2	2	8
y3	4	4
y7	4	2

Y	M	N
	6.67	4.67
y4	8	4
y5	6	6
y6	6	4

Since the clusters did not change, algorithm stops.