Problem 1

For the information system given below, find the set of all C-reducts and rules describing C in terms of E, F, G. Use RSES method.

Assume that Dom(E) = {e1, e2}, Dom(F) = {f1, f2, f3},
Dom(G)={g1,g2,g3}, Dom(C) = {c1, c2}.

X	E	F	G	C
x1	e2	f1	g3	c2
x2	e2	f3	g3	c1
x3	e1	f2	g2	c1
x4	e1	f1	g3	c2
x5	e2	f2	g1	c2
x6	e2	f3	g1	c2

Indiscernibility matrix

	X1	X2	X3	X4	X5	X6
X1	-					
X2	f	-				
X3	efg	-	-			
X4	-	ef	fg	-		
X5	-	fg	eg	-	-	
X6	-	g	efg	-	-	-

DF(e,f,g)= f(e+f+g)(e+f)(f+g)g(f+g)(e+g)(e+f+g) = fg
The reduct is {f, g}

Indiscernibility matrix reduced to {f, g}

	X1	X2	X3	X4	X5	X6
X1	-					
X2	f	-				
X3	fg	-	-			
X4	-	f	fg	-		
X5	-	fg	g	-	-	
X6	-	g	fg	-	-	-

Indiscernibility functions:
DF(1) = f(f+g) = f
DF(2) = f(f+g)g = fg
DF(3) = (f+g)g = g
DF(4) = f(f+g) = f
DF(5) = (f+g)g = g
DF(6) = g(f+g) = g

Rules:
X1: f1 -> c2
X2: f3g3 -> c1
X3: g2 -> c1
X4: f1 -> c2
X5: g1 -> c2
X6: g1 -> c2

So, final set of rules:
f1 -> c2 support 2
g1-> c2 support 2
f3g3 -> c1 support 1
g2 -> c1 support 1

Problem 2

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

Step 1

One element

A: 4 B:4 C:5 D:3 ~~E:2~~ F:3 H:3 ~~I:2~~

Two elements

AB:3 AC:4 ~~AD:2~~ AF:3 ~~AH:2~~

BC:4 ~~BD:2~~ ~~BF:2~~ ~~BH:2~~

CD:3 CF:3 CH:3

~~DF:2~~ ~~DH:2~~

~~FH:1~~

Three elements (you are only combining 2-element itemsets which have support minimum3)

ABC-3, ~~ABF~~ (because BF is not frequent) ACF-3, ~~CDF~~ (because DF is not frequent)

~~CDH~~ (because DH is not frequent), ~~CFH~~ (because FH is not frequent)

Please notice that you are not scanning the set of transactions for ABF, CDF, CDH, CFH because they contain subsets which are not frequent.

Representative Itemsets: ABC-3, ACF-3, CH-3, CD-3, BC-4, AC-4

Rule:

ABC

A --> BC 3/ 4 = 75%

B --> AC 3/ 4 = 75%

~~C --> AB 3/ 5 = 60%~~

BC:4

B --> C 4/4 = 100%

C --> B 4/5 = 80%

ACF

A --> CF 3/ 4 = 75%

~~C --> AF 3/ 5 = 60%~~

F --> AC 3/ 3 = 100%

CD:3

D --> C 3/3 = 100%

~~C --> D 3/5 = 60%~~

AC:4

A --> C 4/4 = 100%

C --> A 4/5 = 80%

CH:3

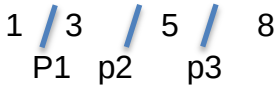
H --> C 3/3 = 100%

~~C --> H 3/5 = 60%~~

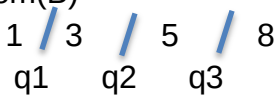
Problem 3. Discretize attributes A and B in the Decision Table below. {A, B} are classification attributes and D is the decision attribute.

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

Dom(A):



Dom(B)



Discernibility Functions

F(x1,x2) = p1+p2+p3+q3 (one of these cuts is sufficient to prevent merging x1 with x2)
F(x1,x3) = p1+p2+q2+q3
F(x1,x5) = p1+p2+p3
F(x1,x6) = q1+q2+q3
F(x2,x4) = p2+p3+q3
F(x3,x4) = p2+q2+q3
F(x4,x5) = p2+p3
F(x4,x6) = p1+q1+q2+q3

p1-4 p2-6 p3-4 q1-2 q2-4 q3-6 (we are choosing p2)

So, we still need to take care of:
F(x1,x6) = q1+q2+q3

F(x4,x6) = p1+q1+q2+q3

Minimal sets of cuts:
{p2, q1}, {p2, q2}, {p2, q3}

If we chose {p2,q3}, we get

A: (-,4] and (4,-)
B: (-,7] and (7,-)

X	A	B	D
x1	(-,4]	(7,-)	1
x2	(4,-)	(-,7]	2
x3	(4,-)	(-,7]	2
x4	(-,4]	(7,-)	1
x5	(4,-)	(7,-)	2
x6	(-,4]	(-,7]	2

Problem 4. Use tolerance relation to find classification rules in the table T assuming that d is the decision attribute.

Car	Price	Mileage	Size	Accident	d
1	*	{mid}	{full}	*	good
2	{low}	*	*	{engine}	good
3	{low}	{high}	{compact}	*	poor
4	{high}	{low}	*	{doors}	excel
5	*	*	{full}	{doors}	good
6	{low}	{mid}	{compact}	*	poor
7	{high}	{high}	{compact}	{doors}	excel

Table T

T(1) = {1, 2, 5}
T(2) = {2, 1, 3, 6}
T(3) = {3, 2}
T(4) = {4, 5}
T(5) = {5, 1,4}
T(6) = {6, 2}
T(7) = {7}

Car	Price	Mileage	Size	Accident	Nd
1	*	{mid}	{full}	*	good
2	{low}	*	*	{engine}	good and poor
3	{low}	{high}	{compact}	*	good and poor
4	{high}	{low}	*	{doors}	good and excel
5	*	*	{full}	{doors}	good and excel
6	{low}	{mid}	{compact}	*	good and poor
7	{high}	{high}	{compact}	{doors}	excel

	1	2	3	4	5	6	7
1	-						
2	NIL	-					
3	MS	-	-				
4	M	PA	PM	-			
5	Nil	A	S	-	-		
6	S	-	-	PM	S	-	
7	MS	PA	P	M	S	PM	-

(M+S)MS(P+A)A(P+M)SP Nil = Nil

RD(1) = Nil
RD(2) = Nil
RD(3) = (M+S)(P+M)SP = SP
RD(4) = M(P+A)(P+M)M = (P+A)M = PM + AM
RD(5) = Nil
RD(6) = S(P+M) = SP + SM
RD(7) = (M+S)(P+A)PMS(P+M) = PMS

Rules:

Car	Price	Mileage	Size	Accident	Nd
1	*	{mid}	{full}	*	good
2	{low}	*	*	{engine}	good and poor
3	{low}	{high}	{compact}	*	good and poor
4	{high}	{low}	*	{doors}	good and excel
5	*	*	{full}	{doors}	good and excel
6	{low}	{mid}	{compact}	*	good and poor
7	{high}	{high}	{compact}	{doors}	excel

RD(3) {S, compact}*{P, low} -> {Nd, good and poor}
RD(4) {P,high }*{M,low} ->{Nd, good and excel}
{A,doors}*{M,low} ->{Nd, good and excel}
RD(6) {S, compact}*{P, low} -> {Nd, good and poor}
{S, compact}*{M,mid} ->{Nd, good and poor}
RD(7) {P,high }*{M,high} }*{S,compact} -> {Nd, excel}

So
{S, compact}*{P, low} -> {Nd, good and poor} support 2
{P,high }*{M,low} ->{Nd, good and excel}
{A,doors}*{M,low} ->{Nd, good and excel}
{S, compact}*{M,mid} ->{Nd, good and poor}
{P,high }*{M,high} }*{S,compact} -> {Nd, excel}