**Problem 1**

For the information system Table 1 given below, find the set of all coverings of C and rules describing C in terms of E, F, G. Use either LERS or RSES method.

Assume that Dom(E) = {e1, e2}, Dom(F) = {f1, f2}, Dom(C) = {c1, c2}.

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

Table 1

**Solution using RSES Method (**Discernibility Matrix - see below**)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 | X6 |
| X1 | - |  |  |  |  |  |
| X2 | EG | - |  |  |  |  |
| X3 | EG | - | - |  |  |  |
| X4 | - | F | GF | - |  |  |
| X5 | - | E | GE | - | - |  |
| X6 | - | GF | GF | - | - | - |

Discernibility function:

 Fun(E, G, F)= (E+G)\*F\*E\*(G+F)=F\*E - covering (reduct) of C

Computing discernibility function Fun(X) for every object X.

Taking concatenation of attributes (red color) in column X1 and row X1 in the matrix above, we get Fun(X1) = E

Now, taking values of attributes E, C in row 1 of Table 1, we get rule

e2→ c2

Taking concatenation of attributes (red color) in column X2 and row X2 in the matrix above, we get Fun(X2) = EF

Now, taking values of attributes E, F, C in row 2 of Table 1, we get rule

e1\*f2→ c1

Fun(X3) = EF (taking column X3 and row X3 in the matrix above)

 e1\*f2 → c1

Fun(X4) = F (taking column X4 and row X4 in the matrix above)

 f1 → c2

Fun(X5) = E e2 → c2

Fun(X6) = F f1 → c2

**Final Rules**: e2→ c2 sup=2, e1.f2 → c1 sup=2, f1 → c2 sup=2

**Solution using LERS**

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

Looking for Reducts

C\* = {{x1, x4, x5, x6}, {x2, x3}}

E\* = {{x1, x5}, {x2, x3, x4, x6}}, F\* = {{x1, x2, x3, x5}, {x4, x6}},

G\* = {{x1, x6}, {x2, x4, x5}, {x3}}.

E\* ⊄ C\*, F\*⊄ C\*, G\* ⊄ C\*

EF\* = {{x1, x5}, {x2, x3}, {x4, x6}}⊆ C\*

EG\* ={{x1}, {x5}, {x2, x4}, {x3}, {x6}} ⊄ C\*

FG\* ={{x1}, {x2, x5}, {x3}, {x6}, {x4}} ⊄ C\*

EFG\* - cannot be extended. It contains subset which is marked YES.

So, EF is a reduct which means we need rules describing C in terms of EF.

LERS – looking for rules using reduct {E,F} (attribute G is removed)

Let’s assume that Minsup = 2; Minconf = 2/3.

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

c1\* = {x2, x3}, c2\*= {x1, x4, x5, x6}

e1\* = {x2, x3, x4, x6}, e2\* = {x1, x5}

f1\* = {x4, x6}, f2\* = {x1, x2, x3, x5}

e2 → c2 sup=2 conf=1; f1 → c2 sup=2, conf=1;

e1.f2={x2,x3}⸦c1\*, sup=2, conf=1;

e1 → c1 sup=2 , conf=2/4; f2 → c1, sup=2 conf=1/2

e1 → c2 sup=2 , conf=2/4; f2 → c2, sup=2 conf=1/2

**Problem 2**

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

A, B, C, D, E

A, B, C, E, F

A, B, C, E, H, I

B, C, D, E, F

C, D, F, H, I

A-3, B-4, C-5, D-3, E-4, F-3, H-2, I-2

AB-3, AC-3, AE-3, BC-4, BE-4, CD-3, CE-4, CF-3, DE-2, DF-2, EF-2, AD-1, AF-1, BD-2, BF-2,

ABC-3, ABE-3, ACE-3, BCE-4, ~~CDE~~, ~~CDF~~, ~~CEF~~

ABCE – 3

Representative sets:

A-3, B-4, C-5, D-3, E-4, F-3

AB-3, AC-3, AE-3, BC-4, BE-4, CD-3, CE-4, CF-3

ABC-3, ABE-3, ACE-3, BCE-4

ABCE - 3

ABCE:

A->BCE 3/3, B->ACE 3/4, C->ABE 3/5, E->ABC 3/4

**Problem 3.**

Find all certain and possible rules describing f in terms of {A, B, C, D} in Table 1. Apply LERS algorithm. Give their support and confidence.

 A B C F

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x1 |  1 |  1 |  2 |  1 |
| x2 |  2 |  1 |  2 |  0 |
| x3 |  1 |  2 |  2 |  1 |
| x4 |  2 |  1 |  1 |  1 |
| x5 |  3 |  1 |  2 |  0 |
| x6 |  3 |  2 |  1 |  1 |
| x7 |  2 |  2 |  1  |  1 |

Table 1.

**Solution.**

A1\*={x1,x3} ⊆F1\*, A2\*={x2,x4,x7}, A3\*={x5,x6}, B1\*={x1,x2,x4,x5}, B2\*={x3,x6,x7}⊆F1\*, C1\*={x4,x6,x7}⊆F1\*, C2\*={x1,x2,x3,x5}; **F1\*={x1,x3,x4,x6,x7,}, F0\*={x2,x5}**

Find rules using LERS:

A2.B1\*= {2,4}, A2.C2\*= {2}⊆F0\*, A3.B1\*= {5}⊆F0\*,

A3.C2\*= {5}⊆F0\*, B1.C2\*= {1,2,5}

A2.B1.C2\*, A3.B1.C2\* - cannot be extended

CERTAIN RULES: A1 → F1, B2 → F1, A2.C2 → F0 ……

POSSIBLE RULES:

B1.C2 → F1 conf=1/3 B1.C2 → F0 conf=2/3

A2.B1 → F1 conf= ½ A2.B1 → F0 conf = 1/2

A2 → F1 conf= 2/3 A2 -> F0 conf= 1/3

**Problem 4 (**Homework**).**

Find all certain rules describing f in terms of {a, b, c, d} in Table 2.

Apply RSES algorithm. We assume that Dom(c) = Dom(d) = {1,2}.

 a b c d f

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x1 |  1 |  1 |  2 |  1 |  1 |
| x2 |  2 |  1 |  2 |  2 |  0 |
| x3 |  1 |  2 |  2 |  1 |  1 |
| x4 |  2 |  1 |  1 |  2  |  1 |
| x5 |  1 |  1 |  1 |  2 |  0 |
| x6 |  1 |  2 |  1 |  2 |  1 |
| x7 |  2 |  2 |  1  |  2 |  1 |

Table 2.

**Solution:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 | X6 | X7 |
| X1 | **-** |  |  |  |  |  |  |
| X2 |  | **-** |  |  |  |  |  |
| X3 |  |  | **-** |  |  |  |  |
| X4 |  |  |  | **-** |  |  |  |
| X5 |  |  |  |  | **-** |  |  |
| X6 |  |  |  |  |  | **-** |  |
| X7 |  |  |  |  |  |  | **-** |

**Problem 5.**

Find the set of all rules in Table S below describing C in terms of A, F, G. Use CART (Gini Index) algorithm. Threshold for information gain = 0.2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | A | F | G | C |
| x1 | a2 | f1 | g3 | c2 |
| x2 | a1 | f2 | g1 | c1 |
| x3 | a1 | f2 | g2 | c1 |
| x4 | a1 | f1 | g1 | c2 |
| x5 | a2 | f2 | g2 | c2 |
| x6 | a1 | f2 | g3 | c2 |

Table S

**Solution**

Gen(S)= 2/6 \* 4/6= 1/3 \* 2/3 = 2/9

In CART strategy

Gen(S,A)= 4/6[(2/4)\*(2/4)] + 2/6\*0 = 4/6[1/4] = 4/24 = 1/6

Gen(S,F)= Gen(S,A)

Gen(S,G)= 2/6\*0 + 2/6[1/2 \* 1/2] + 2/6[1/2 \* 1/2] = 2/24 + 2/24 = 1/6

In-Gain(A)= 2/9 – 1/6 = 4/18 – 3/18 = 1/18

In-Gain(F)= 2/9 – 1/6 = 4/18 – 3/18 = 1/18

In-Gain(G)= 2/9 – 1/6 = 4/18 – 3/18 = 1/18

Let’s take A as the root of the tree

 A

 a1 a2 → c2 certain

|  |  |  |  |
| --- | --- | --- | --- |
| X | F | G | C |
|  x2 | f2 | g1 | c1 |
| x3 | f2 | g2 | c1 |
| x4 | f1 | g1 | c2 |
| x6 | f2 | g3 | c2 |

S1

Gin(S1)=2/4 \* 2/4 = ¼

Gin(S1,F) = ¾\*2/3\*1/3 + ¼\*0 = 6/36 = 1/6

Gin(S1,G)= ¼ \* 0 + ¼\*0 +2/4\*1/2\*1/2= 1/8

 G

g1 a1\*g2 → c1 a1\*g3 → c2

|  |  |  |
| --- | --- | --- |
| X | F | C |
|  x2 | f2 | c1 |
| x4 | f1 | c2 |

 a1\*g1\*f2 -> c1

 a1\*g1\*f1 -> c2

En(S1) = 1/2log2

En(S1, F) = ¾[-2/3log(2/3)-1/3log(1/3)] + 0 =

En(S1, G)= 0 + 0 +2/4[1/2log2] = 1/4log2 G is the winner

**Problem 6.** Discretize attributes A and B in the Decision Table below. {A, B} are classification attributes. D is the decision attribute.

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

**Solution:**

Dom(A): 1 3 5 8 Dom(B): 3 5 8

 P1 P2 P3 Q1 Q2

Discernibility Function

F(x1,x2)= Q1

F(x1,x3)= P1+P2

F(x1,x6)= P1+Q1

F(x2, x4)= P1+Q2

F(x2,x5)= P1+P2+P3

F(x3,x4)=P2+Q1+Q2

F(x3,x5)=P3+Q1

F(x4,x6)=Q2

F(x5,x6)=P2+P3

Minimal set of cuts: Q1, Q2, P2

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

 a1 a2 b1 b2 b3

Discretized Table

|  |  |  |  |
| --- | --- | --- | --- |
| X | A | B | D |
| x1 | a1 | b1 | 1 |
| x2 | a1 | b2 | 2 |
| x3 | a2 | b1 | 2 |
| x4 | a1 | b3 | 1 |
| x5 | a2 | b2 | 1 |
| x6 | a1 | b2 | 2 |