**Sample Problems for Midterm Exam**

**Problem 1.** For the information system given below, find the set of rules describing C in terms of E, F, G by applying, CART (Gini Index), RSES algorithms.

Find the set of all coverings of C (reducts) using RSES. 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 | f1 | g2 | c2 |
| x2 | e2 | f3 | g1 | c1 |
| x3 | e2 | f3 | g2 | c1 |
| x4 | e1 | f2 | g2 | c1 |
| x5 | e1 | f2 | g1 | c2 |

System S

**Solution:**

Gini(S)= 2/5 \* 3/5 = 6/25

Gini(S,E) = 3/5 \* (1/3\* 2/3) + 2/5 \*0 = 6/45

Gain(S, F) = 1/5 \* 0 + 2/5\* 1/2\* ½ + 2/5\* 0 = 2/20 = 1/10 (F is the winner)

Gain((S, G)= 2/5\* ½ \* ½ + 3/5\* 2/3\* 1/3 = 2/20 + 6/45

S(f1)

|  |  |  |  |
| --- | --- | --- | --- |
| X | E | G | C |
| x1 | e1 | g2 | c2 |

f1 -> c2

S(f2)

|  |  |  |  |
| --- | --- | --- | --- |
| X | E | G | C |
| x4 | e1 | g2 | c1 |
| x5 | e1 | g1 | c2 |

f2\*g2 -> c1 f2\*g1 -> c2

S(f3)

|  |  |  |  |
| --- | --- | --- | --- |
| X | E | G | C |
| x2 | e2 | g1 | c1 |
| x3 | e2 | g2 | c1 |

f3-> c1

Using RSES

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

Discernibility Matrix

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X1 | X2 | X3 | X4 | X5 |
| X1 | - |  |  |  |  |
| X2 | EFG | - |  |  |  |
| X3 | FE | - | - |  |  |
| X4 | F | - | - | - |  |
| X5 | - | FE | GFE | G | - |

DF(E,F,G)= (E+F+G)(F+E)FG = FG (reduct)

Red(x1)= (E+F+G)(F+E)F = F Red(x2)= (E+F+G)(F+E) = F+E

Red(x3) = (F+E)(G+F+E)= F+E, Red(x4)= FG, Red(x5) = (F+E)(E+F+G)G = FG + EG

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

f1-> c2 e2->c1 sup=2, f3->c1 sup= 2

f2\*g2 -> c1 f2.g1 -> c2 e1.f2 -> c2

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

**Solution.**

A-3, B-3, C-4, D-4, E-4, F-1, H-4, I-3

AB-1 AC-3 AD-2 AE-2 AH-2 AI-1 BC-2 BD-2 BE-3 BH-3 BI-2

CD-3 CE-3 CH-3 CI-2 DE-3 DH-3 DI-3 EH-4 EI-2 HI-2

BEH-3 CDE-2 CDH-2 CEH-3 DEH-3 DEI DHI

DEH

D-> EH conf=3/4, E->DH conf=3/4, H->DE conf=3/4

CEH

C->EH H->DE E->HD

**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 | 1 | 5 | 2 |
| x3 | 5 | 3 | 2 |
| x4 | 3 | 8 | 1 |
| x5 | 8 | 5 | 1 |
| x6 | 3 | 5 | 2 |

Decision Table T

**Solution:** Dom(A): 1 3 5 9, Dom(B): 3 5 8

 p1 p2 p3 q1 q2

**Problem 4.** Find classification rules in the table below.

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

**Problem 5.** Follow DEAR2 algorithm to extract action rules reclassifying objects from the class d0 to the class d1. Table T shows classification rules extracted from some decision system. These classification rules should be used to construct action rules. Attribute a is stable.

 a b c d

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| r1 |  2 |  2 |  1 |  0 |
| r2 |  1 |  1 |  1 |  1 |
| r3 |  2 |  1 |  2  |  1 |
| r4 |  1 |  3 |  1 |  0 |
| r5 |  1 |  3 |  2 |  1 |
| r6 |  1 |  1 |  2 |  0 |

Table T.

**Problem 6.** Follow:

1. Complete Link Technique (looks for cliques) to clusters objects in T. Show the resulting dendogram.
2. K-means (where K=2) to cluster seven objects in Table T.

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

Table T

**Solution:** Distance Matrix

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 |
| Y1 | - |  |  |  |  |  |  |
| Y2 | 2 | - |  |  |  |  |  |
| Y3 | 4 | 2 | - |  |  |  |  |
| Y4 | 8 | 6 | 4 | - |  |  |  |
| Y5 | 8 | 8 | 4 | 4 | - |  |  |
| Y6 | 6 | 4 | 2 | 6 | 2 | - |  |
| Y7 | 2 | 4 | 2 | 6 | 6 | 4 | - |

**Problem 7.** Let S=(X, {a, b, c, d}) be a decision system, where all attributes are flexible. Attribute d is the decision attribute. Find action rules reclassifying objects from the class d1 to d2 using action reducts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | a | b | c | d |
| x1 | a3 | b1 | c3 | d1 |
| x2 | a3 | b2 | c1 | d2 |
| x3 | a1 | b1 | c1 | d2 |
| x4 | a2 | b1 | c1 | d1 |
| x5 | a1 | b1 | c3 | d1 |
| x6 | a2 | b2 | c2 | d2 |

System S

**Solution:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | X2 | X3 | X6 |
| X1 | b2, c1 | a1, c1 | a2,b2,c2 |
| X4 | a3, b2 | a1 | b2,c2 |
| X5 | a3, b2, c1 | c1 | a2,b2,c2 |

R(x2)= (b2+c1)(a3+b2)(a3+b2+c1) =

 (b2+c1)(a3+b2)=b2.a3+b2+c1.a3+c1.b2 = b2+c1.a3

R(x3)=(a1+c1).a1.c1 = a1.c1 R(x6)= b2 + c2

{c1,a3}, {b2}, {a1,c1}, {c2} – Reducts

b1.(c, → c1).(a, → a3) → (d, d1 → d2) Dom= {x1,x4,x5}

a1.(b, b1 → b2) → (d, d1 → d2) Dom= {x1,x4,x5}

a2.(b, b1→ b2) → (d, d1 → d2) Dom= {x4}

a3.(b, b1 → b2) → (d, d1 → d2)

c1.(b, b1→ b2) → (d, d1 → d2)

c3.(b, b1 → b2) → (d, d1 → d2) Dom= {x1,x5}

c3.(b, b1→ b2) → (d, d1 → d2)

c3.(b, b1 → b2) → (d, d1 → d2)

b1.(a, → a1).(c, → c1) → (d, d1 → d2) Dom= {x1,x4,x5}

a1.(c, c3 → c2) → (d, d1 → d2) Dom= {x1,x4,x5}

a2.(c, c1 → c2) → (d, d1 → d2)