**Final (on Canvas): December 8 (Friday), 4:00 - 6:30pm**

**Sample Problems for Final Exam**

**Problem 1**

Assume that the matrix below represents 8 points in two-dimensional space.

|  |  |  |
| --- | --- | --- |
| Y | M | N |
| y1 | 1 | 2 |
| y2 | 2 | 4 |
| y3 | 6 | 2 |
| y4 | 1 | 8 |
| y5 | 6 | 6 |
| y6 | 1 | 4 |
| y7 | 4 | 2 |
| y8 | 5 | 3 |

Follow agglomerative strategy to find the clusters. Use Manhattan distance

/ d(yi, yj) = |Mi – Mj | + |Ni – Nj | / for objects yi, yjand the distance d(R,Q)= 1/2**⋅**d(A,Q) + 1/2**⋅**d(B,Q) - 1/2**⋅**d(A,B) between clusters R and Q, where R is formed by merging clusters A and B.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Y1 | Y2 | Y3 | Y4 | Y5 | Y6 | Y7 | Y8 |
| Y1 | x |  |  |  |  |  |  |  |
| Y2 | 3 | x |  |  |  |  |  |  |
| Y3 | 5 | 6 | x |  |  |  |  |  |
| Y4 | 6 | 5 | 11 | x |  |  |  |  |
| Y5 | 11 | 6 | 4 | 7 | x |  |  |  |
| Y6 | 2 | 1 | 7 | 4 | 7 | x |  |  |
| Y7 | 3 | 4 | 2 | 9 | 6 | 5 | x |  |
| Y8 | 5 | 4 | 2 | 9 | 4 | 5 | 2 | x |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | {Y1} | {Y2, Y6} | {Y3} | {Y4} | {Y5} | {Y7} | {Y8} |
| {Y1} | x |  |  |  |  |  |  |
| {Y2, Y6} | 3 | x |  |  |  |  |  |
| {Y3} | 5 | ? | x |  |  |  |  |
| {Y4} | 6 | ? | 11 | x |  |  |  |
| {Y5} | 11 | ? | 4 | 7 | x |  |  |
| {Y7} | 3 | ? | 2 | 9 | 6 | x |  |
| {Y8} | 5 | ? | 2 | 9 | 4 | 2 | x |

d(R,Q)= 1/2**⋅**d(A,Q) + 1/2**⋅**d(B,Q) - 1/2**⋅**d(A,B), where R = A∪B

d({Y2,Y6},{Y3})= 1/2⋅d(Y2,Y3)+1/2⋅d(Y6,Y3)-1/2⋅d({Y2},{Y6})=

(1/2)⋅6 + (1/2)⋅7 – (1/2)⋅1 = 3+3 = 6

d({Y2,Y6},{Y4})= 1/2⋅d(Y2,Y4)+1/2⋅d(Y6,Y4)-1/2⋅d({Y2},{Y6})=

(1/2)⋅5 + (1/2)⋅4 – (1/2)⋅1 = 2+2 = 4

d({Y2,Y6},{Y5})= 1/2⋅d(Y2,Y5)+1/2⋅d(Y6,Y5)-1/2⋅d({Y2},{Y6})=

(1/2)⋅6 + (1/2)⋅7 – (1/2)⋅1 = 3 + 3 = 6

d({Y2,Y6},{Y7})= 1/2⋅d(Y2,Y7)+1/2⋅d(Y6,Y7)-1/2⋅d({Y2},{Y6})=

(1/2)⋅4 + (1/2)⋅5 – (1/2)⋅1 = 2 + 2 = 4

d({Y2,Y6},{Y8})= 1/2⋅d(Y2,Y8)+1/2⋅d(Y6,Y8)-1/2⋅d({Y2},{Y6})=

(1/2)⋅4 + (1/2)⋅5 – (1/2)⋅1 = 2 + 2 = 4

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | {Y1} | {Y2, Y6} | {Y3} | {Y4} | {Y5} | {Y7} | {Y8} |
| {Y1} | x |  |  |  |  |  |  |
| {Y2, Y6} | 3 | x |  |  |  |  |  |
| {Y3} | 5 | 6 | x |  |  |  |  |
| {Y4} | 6 | 4 | 11 | x |  |  |  |
| {Y5} | 11 | 6 | 4 | 7 | x |  |  |
| {Y7} | 3 | 4 | 2 | 9 | 6 | x |  |
| {Y8} | 5 | 4 | 2 | 9 | 4 | 2 | x |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | {Y1} | {Y2, Y6} | {Y3, Y7} | {Y4} | {Y5} | {Y8} |
| {Y1} | x |  |  |  |  |  |
| {Y2, Y6} | 3 | x |  |  |  |  |
| {Y3, Y7} | ? | ? | x |  |  |  |
| {Y4} | 6 | 4 | ? | x |  |  |
| {Y5} | 11 | 6 | ? | 7 | x |  |
| {Y8} | 5 | 4 | ? | 9 | 4 | x |

d(R,Q)= 1/2**⋅**d(A,Q) + 1/2**⋅**d(B,Q) - 1/2**⋅**d(A,B), where R = A∪B

d({Y1},{Y3,Y7})= 1/2⋅d({Y1},{Y3}) + 1/2⋅d({Y1},{Y7}) – 1/2⋅d({Y3,Y7}) =

= (1/2)⋅5 + (1/2)⋅3 – (1/2)⋅2 = 4 – 1 = 3

d({Y2,Y6},{Y3,Y7})= 1/2⋅d({Y2,Y6},{Y3}) + 1/2⋅d({Y2,Y6},{Y7}) – 1/2⋅d({Y3,Y7}) =

= (1/2)⋅6 + (1/2)⋅4 – (1/2)⋅2 = 5 – 1 = 4

d({Y4},{Y3,Y7})= 1/2⋅d({Y4},{Y3}) + 1/2⋅d({Y4},{Y7}) – 1/2⋅d({Y3,Y7}) =

= (1/2)⋅11 + (1/2)⋅9 - (1/2)⋅2 = 10 – 1 = 9

d({Y5},{Y3,Y7})= 1/2⋅d({Y5},{Y3}) + 1/2⋅d({Y5},{Y7}) – 1/2⋅d({Y3,Y7}) =

= (1/2)⋅4 + (1/2)⋅6 - (1/2)⋅2 = 5 – 1 = 4

d({Y8},{Y3,Y7})= 1/2⋅d({Y8},{Y3}) + 1/2⋅d({Y8},{Y7}) – 1/2⋅d({Y3,Y7}) =

= (1/2)⋅2 + (1/2)⋅2 - (1/2)⋅2 = 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | {Y1} | {Y2, Y6} | {Y3, Y7} | {Y4} | {Y5} | {Y8} |
| {Y1} | x |  |  |  |  |  |
| {Y2, Y6} | 3 | x |  |  |  |  |
| {Y3, Y7} | 3 | 4 | x |  |  |  |
| {Y4} | 6 | 4 | 9 | x |  |  |
| {Y5} | 11 | 6 | 4 | 7 | x |  |
| {Y8} | 5 | 4 | 1 | 9 | 4 | x |

**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 8, 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 DEAR1/DEAR2 algorithm to extract action rules reclassifying objects from the class d0 to the class d1 hidden in table T. Attributes a, c are stable.

a b c d

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X1 | 2 | 2 | 1 | 0 |
| X2 | 1 | 1 | 1 | 1 |
| X3 | 2 | 1 | 2 | 1 |
| X4 | 1 | 3 | 1 | 0 |
| X5 | 1 | 3 | 2 | 1 |
| X6 | 1 | 1 | 2 | 0 |

Table T.

**Solution for DEAR2**

d0\*={x1, x4,x6} d1\*={x2,x3,x5}

a1\*={x2,x4,x5,x6} a2\*={x1,x3} b1\*={x2,x3,x6} b2\*={x1}<d0\*

b3\*={x4,x5} c1\*={x1,x2,x4} c2\*={x3,x5,x6}

a1.b1\*= {x2,x6} ~~a1.b3\*= b3\*~~ a1.c1\*= {x2,x4} a1.c2\*= {x5,x6}

a2.b1\*= {x3}<d1\* ~~a2.b3\*=~~ a2.c1\*={x1}<d0\* a2.c2\*= {x3}<d1\*

b1.c1\*= {x2}<d1\* b1.c2\*= {x3,x6} b3.c1\*={x4}<d0\* b3.c2\*={x5}<d1\*

a1.b1.c2\*= {x6}<d0\*

|  |  |  |  |
| --- | --- | --- | --- |
| a | b | c | d |
|  | B2 |  | D0 |
| A2 | B1 |  | D1 |
| A2` |  | C1 | D0 |
| A2 |  | C2 | D1 |
|  | B1 | C1 | D1 |
|  | B3 | C1 | D0 |
|  | B3 | C2 | D1 |
| A1 | B1 | C2 | D0 |

Table T

T(A1)

|  |  |  |
| --- | --- | --- |
| b | c | d |
| B2 |  | D0 |
| B1 | C1 | D1 |
| B3 | C1 | D0 |
| B3 | C2 | D1 |
| B1 | C2 | D0 |

T(A1,C1)

|  |  |
| --- | --- |
| b | d |
| B2 | D0 |
| B1 | D1 |
| B3 | D0 |

T(A1,C1,D0)

|  |
| --- |
| b |
| B2 |
| B3 |

T(A1,C1,D1)

|  |
| --- |
| b |
| B1 |

A1\*C1\*(B2 -> B1) => (D0 -> D1)

A1\*C1\*(B3->B1) => (D0 -> D1)

T(A1,C2)

|  |  |
| --- | --- |
| b | d |
| B2 | D0 |
| B3 | D1 |
| B1 | D0 |

T(A1,C2,D0)

|  |
| --- |
| b |
| B2 |
| B1 |

T(A1,C2,D1)

|  |
| --- |
| b |
| B3 |

T(A2)

|  |  |  |
| --- | --- | --- |
| b | c | d |
| B2 |  | D0 |
| B1 |  | D1 |
|  | C1 | D0 |
|  | C2 | D1 |
| B1 | C1 | D1 |
| B3 | C1 | D0 |
| B3 | C2 | D1 |

Solution for DEAR1

|  |  |  |  |
| --- | --- | --- | --- |
| a | b | c | d |
|  | B2 |  | D0 |
| A2 | B1 |  | D1 |
| A2` |  | C1 | D0 |
| A2 |  | C2 | D1 |
|  | B1 | C1 | D1 |
|  | B3 | C1 | D0 |
|  | B3 | C2 | D1 |
| A1 | B1 | C2 | D0 |

Table T

|  |  |  |
| --- | --- | --- |
| a | b | c |
| A2 | B1 |  |
| A2 |  | C2 |
|  | B1 | C1 |
|  | B3 | C2 |

Table T(d1)

|  |  |
| --- | --- |
| b | c |
| B1 |  |
|  | C2 |

Table T(d1,A2)

|  |
| --- |
| b |
| B1 |

Table T(d1,A2,C?)

|  |
| --- |
| b |
|  |

Table T(d1,A2,C2)

|  |  |
| --- | --- |
| b | c |
| B1 | C1 |
| B3 | C2 |

Table T(d1,A?)

|  |
| --- |
| b |
| B1 |

T(d1,A?,C1)

|  |
| --- |
| b |
| B3 |

T(d1,A?,C2)

|  |  |  |
| --- | --- | --- |
| a | b | c |
|  | B2 |  |
| A2` |  | C1 |
|  | B3 | C1 |
| A1 | B1 | C2 |

Table T(d0)

|  |  |
| --- | --- |
| b | c |
| B2 |  |
| B3 | C1 |

T(d0,A?)

|  |
| --- |
| b |
| B2 |

T(d0,A?,C?) \*

|  |
| --- |
| b |
| B3 |

T(d0,A?,C1)

|  |
| --- |
| b |
| B1 |

T(d1,A?,C1) \*

|  |
| --- |
| b |
| B3 |

T(d1,A?,C2)

C1.(B3 -> B1) -> (D0 -> D1)

C1.(B2 -> B1) -> (D0 -> D1)

T(d0,A1)

T(d0,A2)

**Problem 6.** Follow 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

(c, -> c1)\*(a, -> a3) => (d1 -> d2) Dom={x1,x4,x5}

~~b1\*(c, -> c1)\*(a, -> a3) => (d1 -> d2) Dom={x1,x4,x5}~~

~~b2\*(c, -> c1)\*(a, -> a3) => (d1 -> d2) Dom = 0~~

(a, -> a1)\*(c, -> c1) => (d1 -> d2) Dom={x1,x4,x5}

~~b1\*(a, -> a1)\*(c, -> c1) => (d1 -> d2) Dom = {x1,x4,x5}~~

~~b2\*(a, -> a1)\*(c, -> c1) => (d1 -> d2) Dom = Empty Set~~

(b, -> b2) => (d1 -> d2) Dom= {x1,x4,x5}

a1\*(b, b1-> b2) => (d1 -> d2) Dom={x5}

a2\*(b, b1-> b2) => (d1 -> d2) Dom={x4}

a3\*(b, b1-> b2) => (d1 -> d2) Dom={x1}

c1\*(b, b1-> b2) => (d1 -> d2) Dom={x4}

~~c2\*(b, -> b2) => (d1 -> d2)~~

c3\*(b, b1-> b2) => (d1 -> d2) Dom={x1,x5}

**Problem 8**

Systems S1, S2 are defined below:

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

System S1

|  |  |  |  |
| --- | --- | --- | --- |
|  | c | d | e |
| x2 | c1 | d2 | e1 |
| x5 | c2 | d1 | e2 |
| x6 | c2 | d1 | e2 |
| x7 | c1 | d3 | e2 |
| x8 | c3 | d1 | e1 |

System S2

Find certain and possible objects in S1 satisfying query a1\*e2

Use S2 to extract rules describing e2.

e2\*= {5,6,7}

c1\*= {2,7}, c2\*= {5,6} < e2\*, d1\*= {5,6,8}, d3\*= {7} <e2\*

c1.d1\*=0

c1->e2 conf=1/2, sup=1; c2->e2 conf=1 sup=2; d1->e2 sup=2 conf = 2/3;

d3->e2 conf=1 sup=1

a1\*e2=a1\*[c1 + c2 + d1 + d3] = a1\*c1 + a1\*c2 + a1\*d1 + a1\*d3

½ 1 2/3 1 ½ 1 2/3 1

Answer: (x1,1/2), (x3,1) (x1,2/3) -> (x1,2/3), (x3,1)

**Problem 9**: **A**ssume that {Table 1, Table 2} represents distributed information system where Table1 and Table 2 are semantically similar. Create new attribute g in Table 1 using knowledge extracted from Table 2.

a b c d f

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

Table 1.

a e c d g

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| y1 | 1 | 1 | 2 | 1 | 1 |
| y2 | 2 | 1 | 2 | 2 | 0 |
| y3 | 1 | 2 | 2 | 1 | 1 |
| y4 | 1 | 1 | 1 | 1 | 1 |
| y5 | 3 | 1 | 2 | 2 | 0 |
| y6 | 3 | 1 | 1 | 2 | 1 |
| y7 | 3 | 2 | 2 | 2 | 0 |

Table 2.

Solution: g0\*={y2,y5,y7}, g1\*={y1,y3,y4,y6}

a1\*={y1,y3,y4}<g1\*, a3\*={y5,y6,y7}, c2\*={y1,y2,y3,y5,y7}, c1\*={y4,y6}<g1\*, d1\*={y1,y3,y4}<g1\*, d2\*={y2,y5,y6,y7},

a3.c2\*={y5,y7}, a3.d2\*=a3\*,

a1->g1, s=3, c=1; a3->g1, s=1, c=1/3; c2->g1, s=2, c=2/5;

c1->g1, s=2, c=1; d1->g1, s=3, c=1; d2->g1, s=1, c=1/4

a2\*={y2} < g0\*, a3\*={y5,y6,y7} , c2\*={y1,y2,y3,y5,y7} , c1\*={y4,y6}

d2\*={y2,y5,y6,y7}, a3.c2\*={y5,y7}<g0\*, a3.d2\*=a3\*, c2.d2\*={y2,y5,y7}<g0\*,

Rules:

a2->g0 s=1, c=1; a3->g0 s=2, c=2/3; c2->g0, s=3, c=3/5; d2->g0, s=3 , c=3/4; a3.c2->g0, s=2, c=1; c2.d2->g0, s=3, c=1.

a1->g1, s=3, c=1; a3->g1, s=1, c=1/3; c2->g1, s=2, c=2/5;

c1->g1, s=2, c=1; d1->g1, s=3, c=1; d2->g1, s=1, c=1/4

a b c d f g

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x1 | 1 | 1 | 2 | 1 | 1 | (g0, 3\*3/5), (g1, 3\*1 +2\*2/5 +3\*1) |
| x2 | 3 | 1 | 2 | 2 | 0 | (g0,2\*2/3+3\*3/5+3\*3/4+2+3),(g1,1\*1/3+2\*2/5+1\*1/4) |
| x3 | 1 | 2 | 2 | 1 | 2 | (g0,3\*3/5),(g1,3+2\*2/5+3\*1) |
| x4 | 2 | 1 | 1 | 2 | 1 | (g0,1+3\*3/4)(g1,2+1\*1/4) |
| x5 | 3 | 1 | 2 | 2 | 0 | (g0,2\*2/3+3\*3/5+3\*3/4+2+3),(g1,2\*2/5+1\*1q/4) |
| x6 | 3 | 2 | 1 | 2 | 1 | (g0,2\*2/3+3\*3/4),(g1,1/3+2+1/4) |
| x7 | 2 | 2 | 1 | 2 | 2 | (g0,1+3\*3/4),(g1,2+1/4) |

Table 1.

Normalization for x1: 9/5 - 30/5+4/5=34/5; 9/43 – 34/43

a b c d f g

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x1 | 1 | 1 | 2 | 1 | 1 | (g0,9/43)(g1,34/43) |
| x2 | 3 | 1 | 2 | 2 | 0 | (g0,2\*2/3+3\*3/5+3\*3/4+2+3),(g1,1/3+2\*2/5+1/4) |
| x3 | 1 | 2 | 2 | 1 | 2 | (g0,9/43)(g1,34/43) |
| x4 | 2 | 1 | 1 | 2 | 1 | (g0,1+3\*3/4)(g1,2+1\*1/4) |
| x5 | 3 | 1 | 2 | 2 | 0 | (g0,2\*2/3+3\*3/5+3\*3/4+2+3),(g1,2\*2/5+1\*1q/4) |
| x6 | 3 | 2 | 1 | 2 | 1 | (g0,2\*2/3+3\*3/4),(g1,1/3+2+1/4) |
| x7 | 2 | 2 | 1 | 2 | 2 | (g0,13/22),(g1,9/22) |

Table 1.