www.kdd.uncc.edu

College of Computing and Informatics
University of North Carolina, Charlotte

RSES – reducts and discretization of attributes

presented by

Zbigniew W. Ras

University of North Carolina, Charlotte, NC &

Polish-Japanese Academy of Information Technology, Poland

An Example of Reducts & Core

U	Headache	Muscle pain	Temp.	Flu
U1	Yes	Yes	Normal	No
U2	Yes	Yes	High	Yes
U3	Yes	Yes	Very-high	Yes
U4	No	Yes	Normal	No
U5	No	No	High	No
U6	No	Yes	Very-high	Yes

Reduct1 = {Muscle-pain, Temp.}

U	Muscle pain	Temp.	Flu
U1,U4	Yes	Normal	No
U2	Yes	High	Yes
U3,U6	Yes	Very-high	Yes
U5	No	High	No

Reduct2 = {Headache, Temp.}

U	Headache	Temp.	Flu
U1	Yes	Normal	No
U2	Yes	High	Yes
U3	Yes	Very-high	Yes
U4	No	Normal	No
U5	No	High	No
U6	No	Very-high	Yes

CORE = {Headache, Temp} ∩ {MusclePain, Temp} = {Temp}

REDUCTS

Information System

	a	b	c	d	f
x1	0	L	0	L	0
x2	0	R	1	L	1
x3	0	L	0	L	0
x4	0	R	1	L	1
x5	1	R	0	L	2
x6	1	R	0	L	2
x7	2	S	2	H	3
x8	2	S	2	H	3

Discernibility Matrix

x1								
x2	bc							
x3	-	bc						
x4	bc	-	bc					
x5	ab	ac	ab	ac				
x6	ab	ac	ab	ac	-			
x7	abcd	abcd	abcd	abcd	abcd	abcd		
x8	abcd	abcd	abcd	abcd	abcd	abcd	-	
	x1	x2	x3	x4	x5	x6	x7	x8

Discernibility Function:

f(a, b, c, d) = (b + c) (a + b) (a + b + c + d) (a + c) =(b + c) (a + b) (a + c) =
(ba + bb + ca + cb) (a + c) = (b + ca) (a + c)
ba + bc + ca

Reducts: {b, a}, {c, a}, {c, b}

(b=L) → (f=0);
(a=0)*(b=R) → (f=1);
.....

Example of Discernibility Matrix

	a	b	c	d	E
u1	1	0	2	1	1
u2	1	0	2	0	1
u3	1	2	0	0	2
u4	1	2	2	1	0
u5	2	1	0	0	2
u6	2	1	1	0	2
u7	2	1	2	1	1

	u1	u2	u3	u4	u5	u6
u2	λ					
u3	b,c,d	b,c				
u4	b	b,d	c,d			
u5	a,b,c,d	a,b,c	λ	a,b,c,d		
u6	a,b,c,d	a,b,c	λ	a,b,c,d	λ	
u7	λ	λ	a,b,c,d	a,b	c,d	c,d

Core = {b}

Reduct1 = {b,c}

Reduct2 = {b,d}

	a	b	c	d	E
u1	1	0	2	1	1
u2	1	0	2	0	1
u3	1	2	0	0	2
u4	1	2	2	1	0
u5	2	1	0	0	2
u6	2	1	1	0	2
u7	2	1	2	1	1

	u1	u2	u3	u4	u5	u6
u2	λ					
u3	b,c,d	b,c				
u4	b	b,d	c,d			
u5	a,b,c,d	a,b,c	λ	a,b,c,d		
u6	a,b,c,d	a,b,c	λ	a,b,c,d	λ	
u7	λ	λ	a,b,c,d	a,b	c,d	c,d

Core = {b}

Reduct1 = {b,c}

Reduct2 = {b,d}

(b+c+d)b(a+b+c+d)(b+c)(b+d)(a+b+c)(a+b)(c+d) =
b(c+d) = bc + bd

Discretization

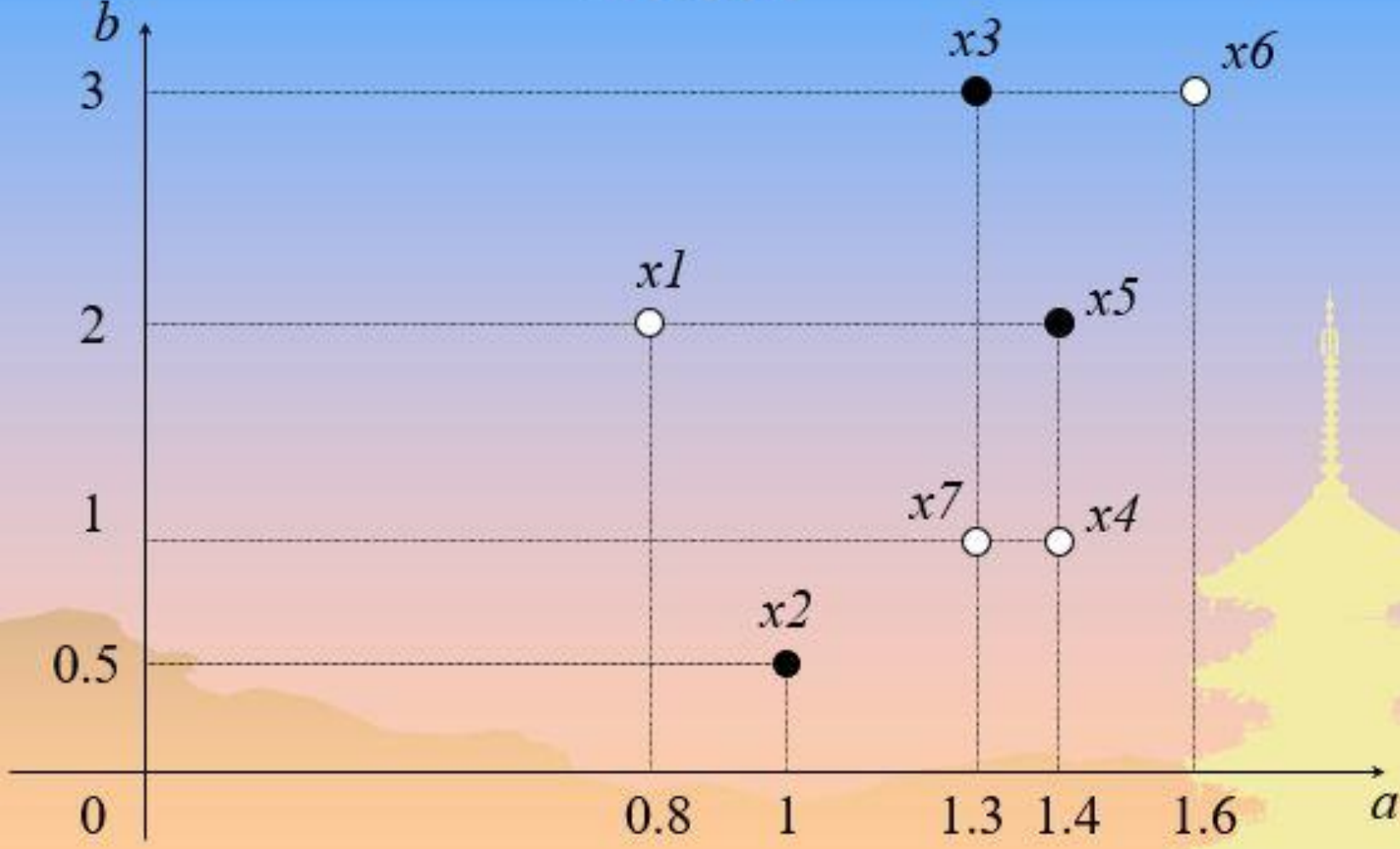
In the discretization process, we search for a set of cuts satisfying some natural conditions.

U	a	b	d
x1	0.8	2	1
x2	1	0.5	0
x3	1.3	3	0
x4	1.4	1	1
x5	1.4	2	0
x6	1.6	3	1
x7	1.3	1	1

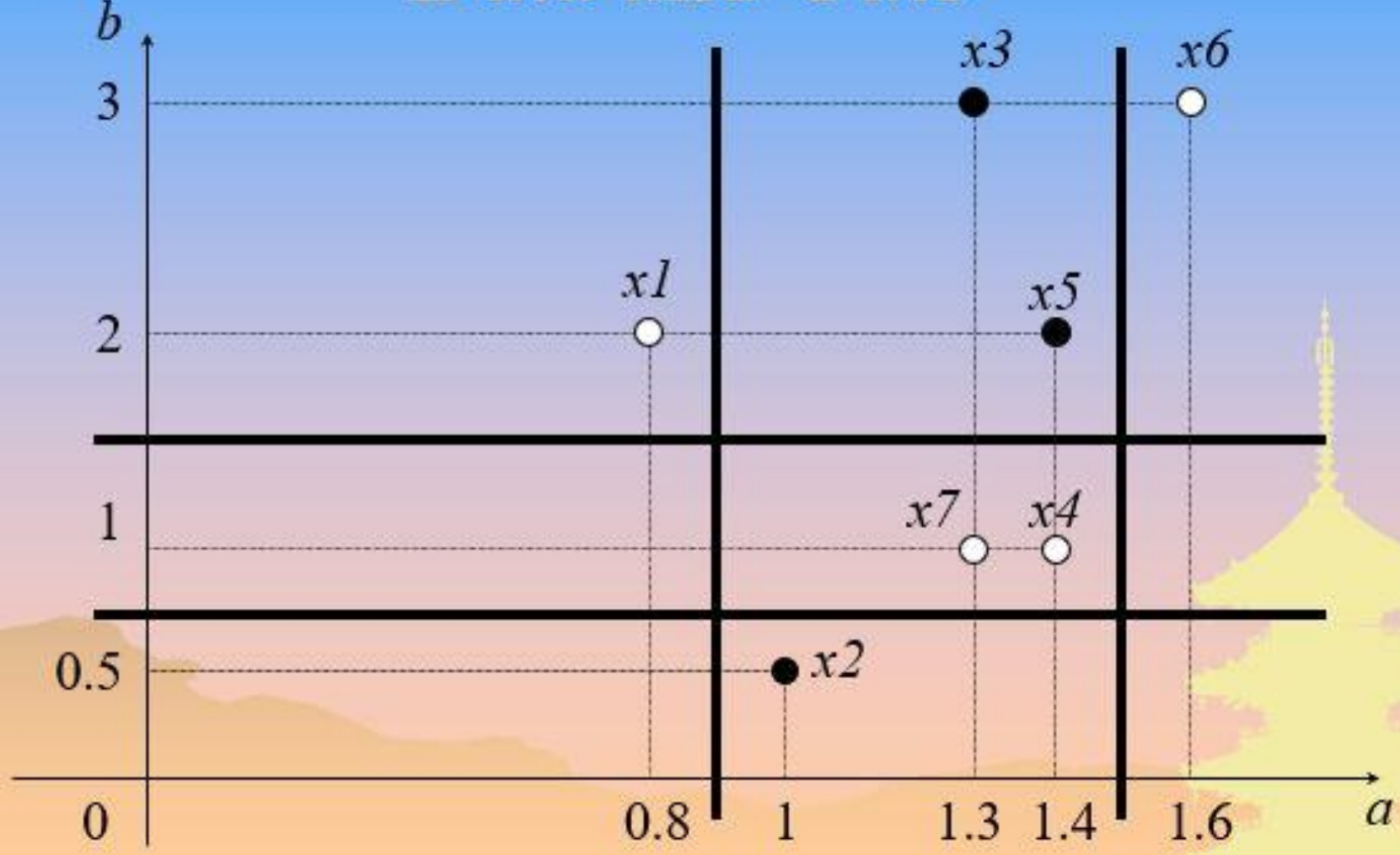
$P = \{(a, 0.9), (a, 1.5), (b, 0.75), (b, 1.5)\}$

U	a^P	b^P	d
x1	0	2	1
x2	1	0	0
x3	1	2	0
x4	1	1	1
x5	1	2	0
x6	2	2	1
x7	1	1	1

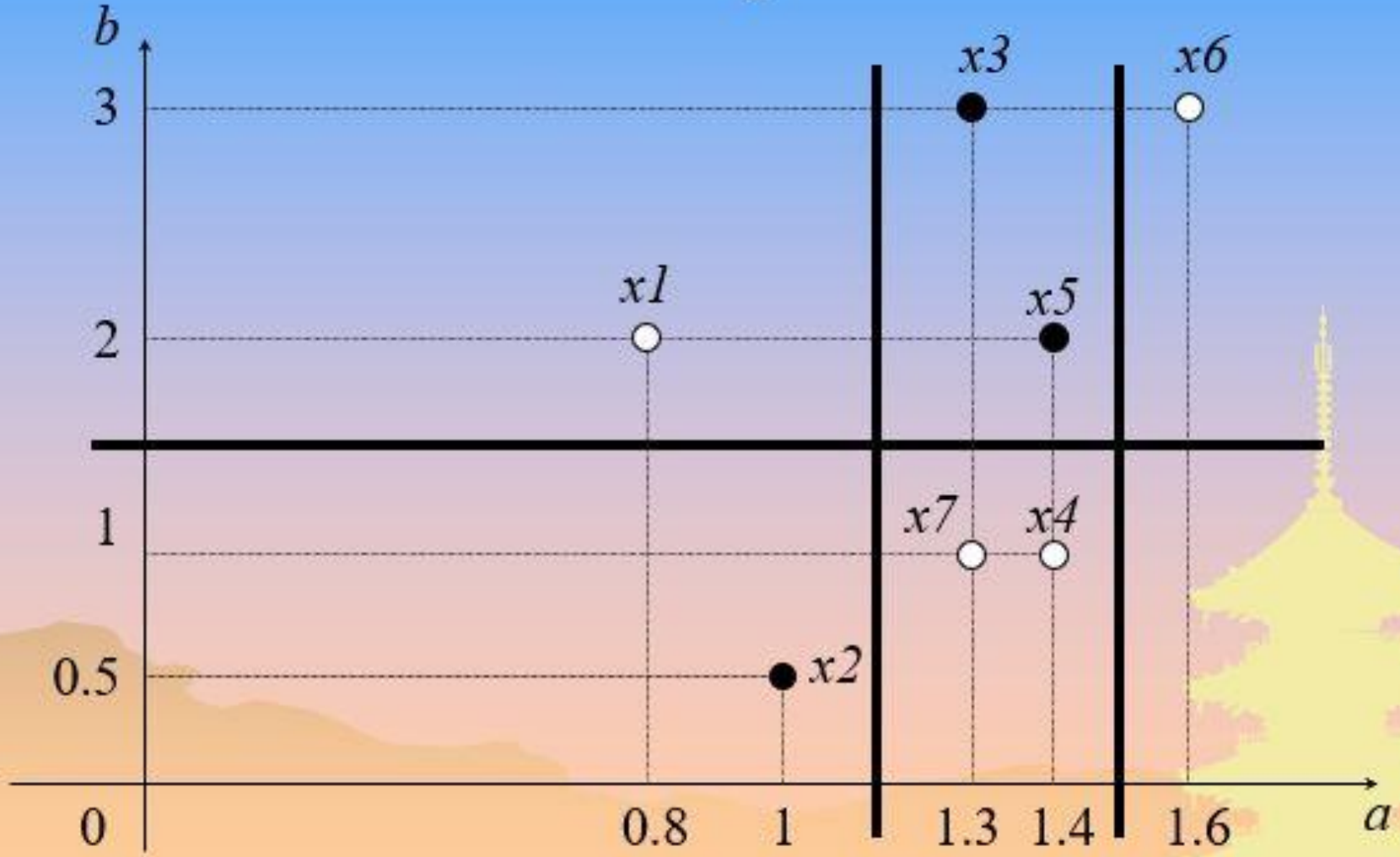
Geometrical Representation of Data



Geometrical Representation of Data and Cuts



Minimal Set Cuts for the Sample DB



Discretization of numerical attributes

X	a	b	d
x1	0.8	2	1
x2	1	0.5	0
x3	1.3	3	0
x4	1.4	1	1
x5	1.4	2	0
x6	1.6	3	1
x7	1.3	1	1

Va = [0, 2), Vb = [0, 4)

p1 p2 p3 p4

Dom(a) = {0.8, 1, 1.3, 1.4, 1.6},

q1 q2 q3

Dom(b) = {0.5, 1, 2, 3}

G(x1,x2)=p1+q1+q2
G(x1,x3)=p1+p2+q3
G(x1,x5)=p1+p2+p3
G(x2,x4)=p2+p3+q1
G(x2,x6)=p2+p3+p4+q1+q2+q3
G(x2,x7)=p2+q1
G(x3,x4)=p3+q2+q3
G(x3,x6)=p3+p4
G(x3,x7)=q2+q3
G(x4,x5)=q2
G(x5,x6)=p4+q3
G(x5,x7)=p3+q2

	P1	P 2	P 3	P4	Q1	Q2	Q3
x1,x3	1	1					1
x1,x5	1	1	1				
x2,x4							
x2,x7		1			1		
x3,x6			1	1			
x5,x6				1			1

	p(1, a)	p(2, a)	p(3, a)	p(4, a)	p(1, b)	p(2, b)	p(3, b)	d*
(x1, x2)	1	0	0	0	1	1	0	1
(x1, x3)	1	1	0	0	0	0	1	1
(x1, x5)	1	1	1	0	0	0	0	1
(x4, x2)	0	1	1	0	1	0	0	1
(x4, x3)	0	0	1	0	0	1	1	1
(x4, x5)	0	0	0	0	0	1	0	1
(x6, x2)	0	1	1	1	1	1	1	1
(x6, x3)	0	0	1	1	0	0	0	1
(x6, x5)	0	0	0	1	0	0	1	1
(x7, x2)	0	1	0	0	1	0	0	1
(x7, x3)	0	0	0	0	0	1	1	1
(x7, x5)	0	0	1	0	0	1	0	1
new	0	0	0	0	0	0	0	0

Step 1. Choose a column from B with the maximal number of occurrences of 1's.

Step 2. Delete from B the column chosen in Step 2 and all rows marked in this column by 1.

Step 3. If B is non-empty go to step 2 else Stop.

Questions?

Thank You