Methods of Knowledge Representation Using Rough Sets

葉建華

jhyeh@mail.au.edu.tw

http://jhyeh.csie.au.edu.tw/

Outline

- Introduction
- Basic terms
- Set approximation
- Approximation of family of sets
- Analysis of decision tables

Introduction

- No two objects are identical (real world)
  - How about similar? What is similar?
- A sufficiently large number of their features (attributes) => a sufficiently great accuracy
- We don’t need such details often
- Decrease precision of description (features)
  - “Indiscernible” (not distinguishable) begins
  - E.g. # inhabitants of cities

Notations and Symbols

- ***U***: the universe of discourse
  - The set of all objects which constitute the area of our interest
- $\boldsymbol{x_j}$: *j-th* element of ***U***
  - Each object of ***U*** may have specific features
- ***Q***: the interesting set of object features of ***U***
- $\boldsymbol{q_i}$: *i-th* feature of ***Q***
- $\boldsymbol{V_q}$: the set of values that the feature ***q*** can take
  - $\boldsymbol{v_q^x}$: the value of feature ***q*** of the object ***x***
  - $\mathbf{v}^x = \left[v_{q_1}^x, v_{q_2}^x, \ldots, v_{q_n}^x\right]$

Outline

- Introduction
- Basic terms
- Set approximation
- Approximation of family of sets
- Analysis of decision tables

Def: Information System

- $SI = \langle U, Q, V, f \rangle$
- Set of all possible values of features: $V = \bigcup_{q \in Q} V_q$
- Information function: $f : U \times Q \to V$
- $v_q^x = f(x, q)$, $f(x, q) \in V_q$, or $v_q^x = f_x(q)$
- $f_x : Q \to V$

Example 1: Used Car Dealer

- 10 used cars, 10 objects

$$U = \{x_1, x_2, ..., x_{10}\}$$

- 4 features of each car
  - # doors, horsepower, color, make

$$Q = \{q_1, q_2, q_3, q_4\}$$
$$= \{\text{number of doors, horsepower, colour, make}\}$$

Used Car: Features

Object (U)	Number of doors (q_1)	Horsepower (q_2)	Colour (q_3)	Make (q_4)
x_1	2	60	blue	Opel
x_2	2	100	black	Nissan
x_3	2	200	black	Ferrari
x_4	2	200	red	Ferrari
x_5	2	200	red	Opel
x_6	3	100	red	Opel
x_7	3	100	red	Opel
x_8	3	200	black	Ferrari
x_9	4	100	blue	Nissan
x_{10}	4	100	blue	Nissan

$$V_{q_1} = \{2, 3, 4\},$$

$$V_{q_2} = \{60, 100, 200\},$$

$$V_{q_3} = \{\text{black, blue, red}\},$$

$$V_{q_4} = \{\text{Ferrari, Nissan, Opel}\}$$

Example 2: Real Numbers

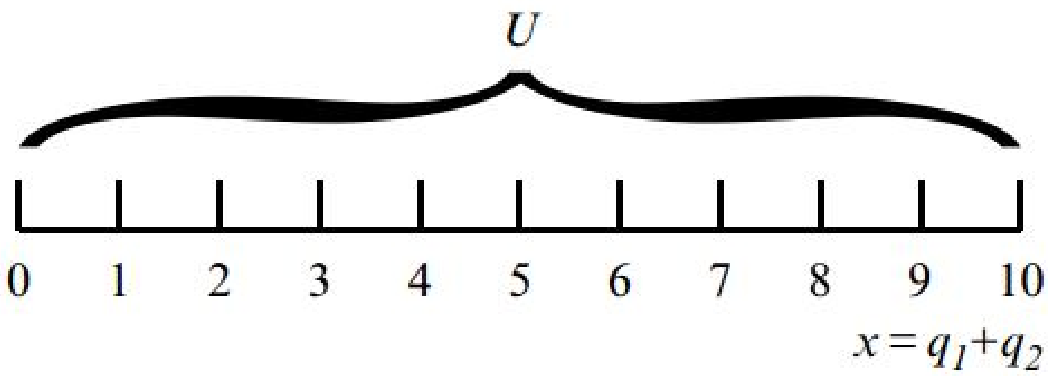
$x \in U$

$$U = [0, 10)\,.$$

- q_1: integral part,
- q_2: decimal part

$$Q = \{q_1, q_2\}$$

$$x = \{q_1, q_2\}$$



- Ent(.): integral part

$$V_{q_1} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$f_x(q_1) = \mathrm{Ent}(x)$$

$$V_{q_2} = [0; 1)$$

$$f_x(q_2) = x - \mathrm{Ent}(x)$$

Example 3: 2-dim. Space

$$U = \{\mathbf{x} = [x_1; x_2] \in [0;10) \times [0;10)\}$$

$$Q = \{q_1, q_2, q_3, q_4\}$$

$$f_{\mathbf{x}}(q_1) = \text{Ent}(x_1),$$

$$f_{\mathbf{x}}(q_2) = x_1 - \text{Ent}(x_1),$$

$$f_{\mathbf{x}}(q_3) = \text{Ent}(x_2),$$

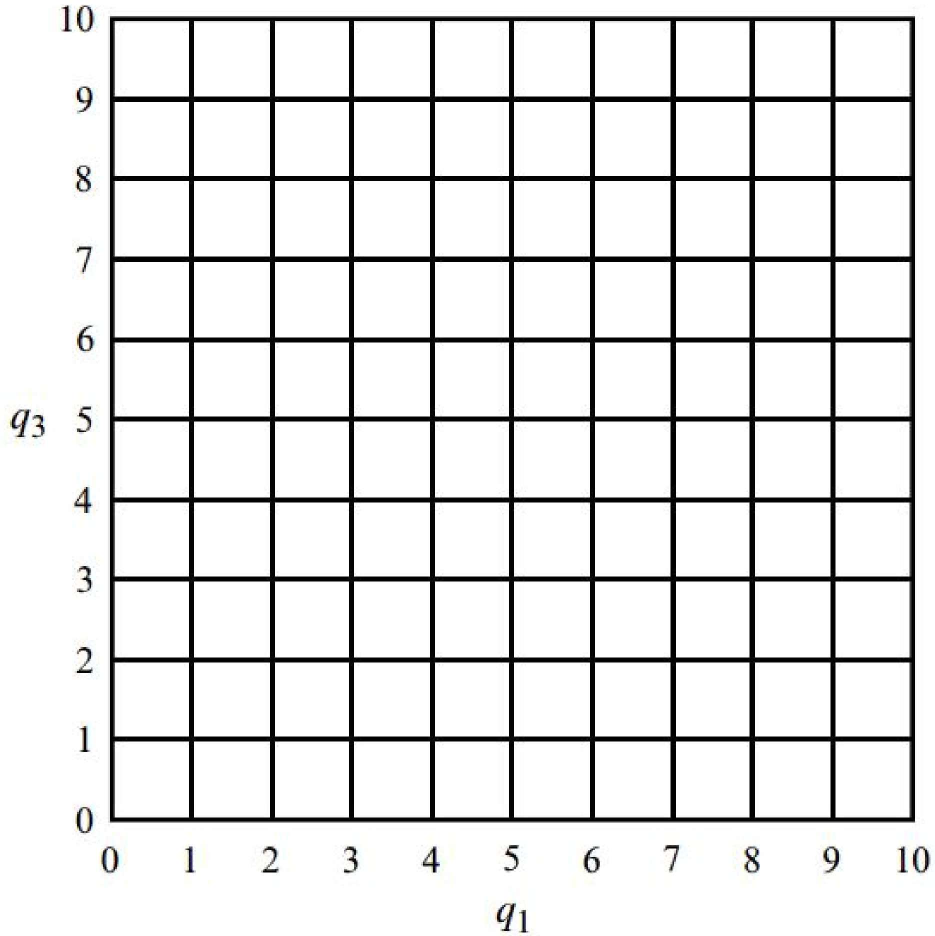
$$f_{\mathbf{x}}(q_4) = x_2 - \text{Ent}(x_2)$$

$$V_{q_1} = \{0,1,2,3,4,5,6,7,8,9\}$$

$$V_{q_2} = [0;1)$$

$$V_{q_3} = \{0,1,2,3,4,5,6,7,8,9\}$$

$$V_{q_4} = [0;1)$$



Decision Table

- Decision table: a special case of information system

$$DT = \langle U, C, D, V, f \rangle$$

- Conditional features ***C***
- Decision features ***D***
  - Originally, feature set is ***Q***
- ℓ : decision rule $f_l : C \times D \to V$
- The information with relation to the rules

$$R^l : \textbf{IF } c_1 = v^l_{c_1} \textbf{ AND } c_2 = v^l_{c_2} \textbf{ AND...AND } c_{n_c} = v^l_{c_{n_c}} \textbf{ THEN } d_1 = v^l_{d_1}$$
$$\textbf{AND } d_2 = v^l_{d_2} \textbf{ AND} \ldots \textbf{AND } d_{n_d} = v^l_{d_{n_d}}$$

Example 4: Used Car Dealer

$$C = \{c_1, c_2, c_3\} = \{q_1, q_2 . q_3\}$$
$$= \{\text{number of doors, horsepower, colour}\}$$

$$D = \{d_1\} = \{q_4\} = \{\text{make}\}$$

- Rules

R^1 : **IF** $c_1 = 2$ **AND** $c_2 = 60$ **AND** $c_3 =$ blue**THEN** $d_1 =$ Nissan

R^2 : **IF** $c_1 = 2$ **AND** $c_2 = 100$ **AND** $c_3 =$ black **THEN** $d_1 =$ Nissan

...

R^{10} : **IF** $c_1 = 4$ **AND** $c_2 = 100$ **AND** $c_3 =$ blue **THEN** $d_1 =$ Nissan

Example of Decision Table

Rule (l)	Number of doors (c_1)	Horsepower (c_2)	Colour (c_3)	Make (d_1)
1	2	60	blue	Opel
2	2	100	black	Nissan
3	2	200	black	Ferrari
4	2	200	red	Ferrari
5	2	200	red	Opel
6	3	100	red	Opel
7	3	100	red	Opel
8	3	200	black	Ferrari
9	4	100	blue	Nissan
10	4	100	blue	Nissan

P-indiscernible

$x_1, x_b \in U$

$P \subseteq Q$

$\forall_q \in P, \;\; f_{x_a}(q) = f_{x_b}(q)$

$$P\text{-}indiscernibility\ relation\ \left(x_a, \widetilde{P} x_b\right)$$

$$x_a \widetilde{P} x_b \iff \forall_q \in P; \; f_{x_a}(q) = f_{x_b}(q)$$

where $x_a, x_b \in U$, $P \subseteq Q$.

- $\widetilde{P}$ relation is reflexive, symmetrical and transitive, and thus it is a relation of equivalence

Equivalence Class

$$[x_a]_{\widetilde{P}} = \left\{ x \in U : x_a \widetilde{P} x \right\}$$

- P^*: the family of all equivalence classes of the relation P in the space U(called the quotient of set U by relation P)
  - Or $U/\widetilde{P}$

Example 5: Used Car Dealer

$$C = \{c_1, c_2, c_3\} = \{q_1, q_2.q_3\}$$
$$= \{\text{number of doors, horsepower, colour}\}$$

C-indiscernibility $\widetilde{C}$

$$[x_1]_{\widetilde{C}} = \{x_1\},$$
$$[x_2]_{\widetilde{C}} = \{x_2\},$$
$$[x_3]_{\widetilde{C}} = \{x_3\},$$
$$[x_4]_{\widetilde{C}} = [x_5]_{\widetilde{C}} = \{x_4, x_5\},$$
$$[x_6]_{\widetilde{C}} = [x_7]_{\widetilde{C}} = \{x_6, x_7\},$$
$$[x_8]_{\widetilde{C}} = \{x_8\},$$
$$[x_9]_{\widetilde{C}} = [x_{10}]_{\widetilde{C}} = \{x_9, x_{10}\}$$

Object (U)	Number of doors (q_1)	Horsepower (q_2)	Colour (q_3)	Make (q_4)
x_1	2	60	blue	Opel
x_2	2	100	black	Nissan
x_3	2	200	black	Ferrari
x_4	2	200	red	Ferrari
x_5	2	200	red	Opel
x_6	3	100	red	Opel
x_7	3	100	red	Opel
x_8	3	200	black	Ferrari
x_9	4	100	blue	Nissan
x_{10}	4	100	blue	Nissan

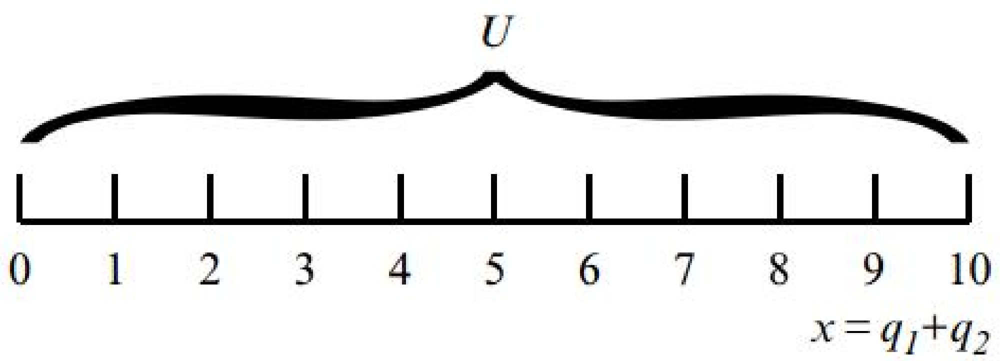
Example 6: Real Numbers

Q-indiscernibility relation

$$P_1 = \{q_1\}$$

$$U = [0, 10)\,.$$

$$P_2 = \{q_2\}$$



$$[0]_{P_1} = [0;1)\,,$$

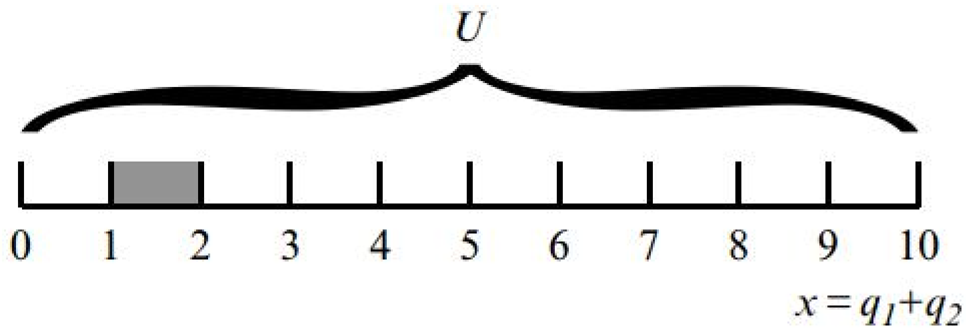
$$[1]_{P_1} = [1;2)\,,$$

$$\ldots$$

$$[9]_{P_1} = [9;10)$$

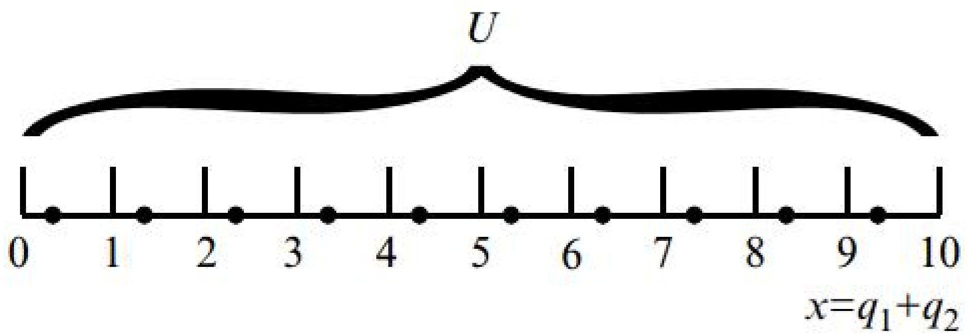
$$P_1^* = \{[0;1);[1;2);[2;3);[3;4);[4;5);[5;6);[6;7);[7;8);[8;9);[9;10)\}$$

Equivalence Class: P1



Example of equivalence class $[1]_{\tilde{P}_1}$

- What about this?



Equivalence Class: P2

$$[x]_{P_2} = \{\widehat{x} \in U : \widehat{x} - \mathrm{Ent}\,(\widehat{x}) = x - \mathrm{Ent}\,(x)\}\,.$$

$$\begin{aligned} P_2^* &= \left\{[x]_{P_2} = \{\widehat{x} \in U : \widehat{x} - \mathrm{Ent}\,(\widehat{x}) = x - \mathrm{Ent}\,(x)\} : x \in [0;1)\right\} \\ &= \left\{[x]_{P_2} = \{\widehat{x} \in U : \widehat{x} - \mathrm{Ent}\,(\widehat{x}) = x\} : x \in [0;1)\right\}. \end{aligned}$$

Example 7: 2-dim. Space

- 100 equivalence classes

$$P = \{q_1, q_3\}$$

$$[\mathbf{x}]_{\tilde{P}} = \{\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) \in U\colon \mathrm{Ent}\,(\hat{x}_1) = \mathrm{Ent}\,(x_1) \wedge \mathrm{Ent}\,(\hat{x}_2) = \mathrm{Ent}\,(x_2)\}$$

$$U = \{\mathbf{x} = [x_1; x_2] \in [0; 10) \times [0; 10)\}$$

$$Q = \{q_1, q_2, q_3, q_4\}$$

$$f_{\mathbf{x}}(q_1) = \mathrm{Ent}\,(x_1),$$

$$f_{\mathbf{x}}(q_2) = x_1 - \mathrm{Ent}\,(x_1),$$

$$f_{\mathbf{x}}(q_3) = \mathrm{Ent}\,(x_2),$$

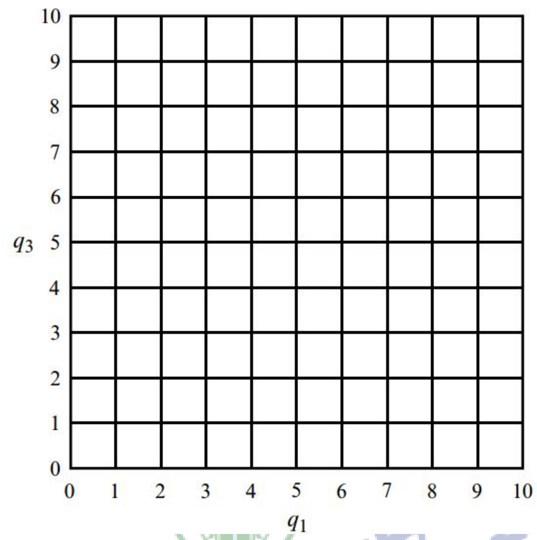
$$f_{\mathbf{x}}(q_4) = x_2 - \mathrm{Ent}\,(x_2)$$

$$V_{q_1} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

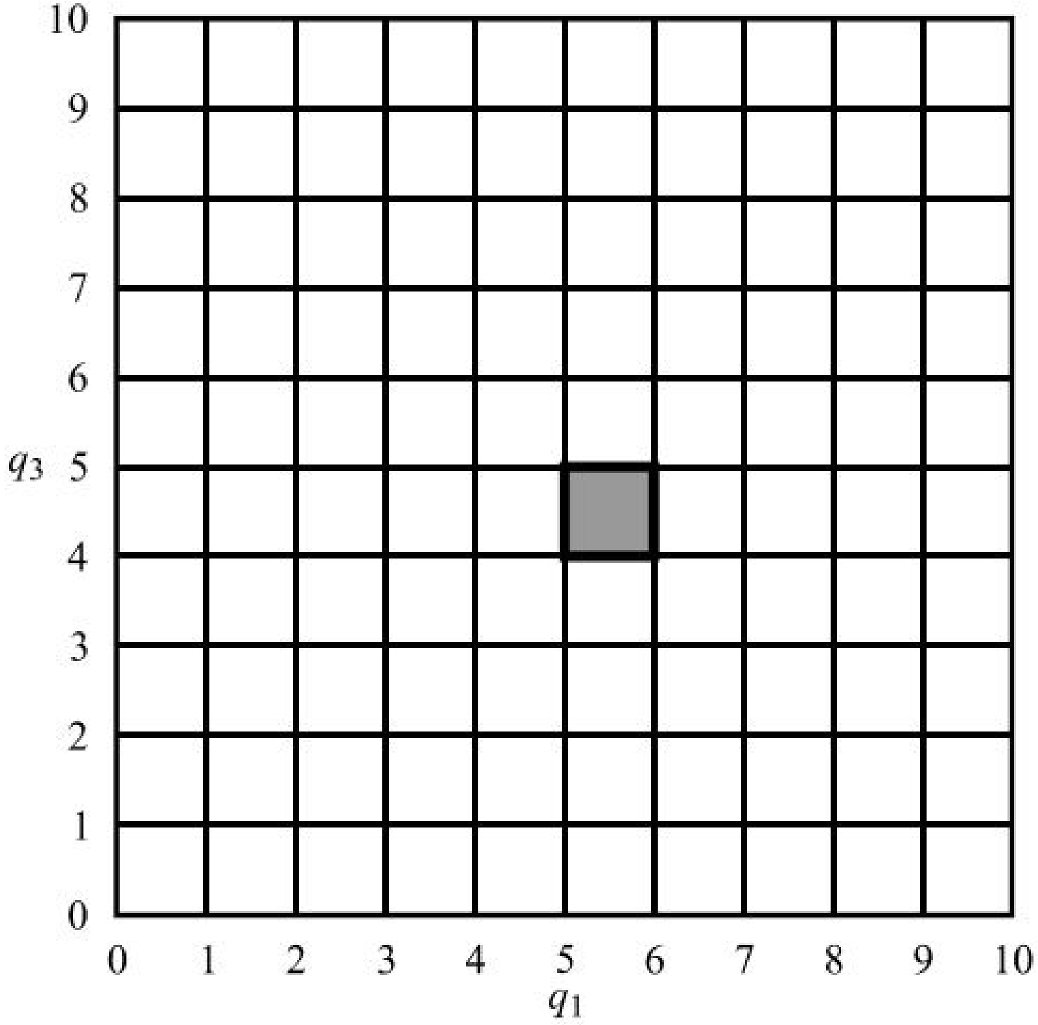
$$V_{q_2} = [0; 1)$$

$$V_{q_3} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$V_{q_4} = [0; 1)$$



Equivalence Class Example



Example of equivalence class $[(5;4)]_{\tilde{P}}$

Equivalence Class

$$
\begin{aligned}
P^* &= \left\{ [\mathbf{x}]_{\widetilde{P}} = \{ \widehat{\mathbf{x}} = (\widehat{x}_1; \widehat{x}_2) \in U : \operatorname{Ent}(\widehat{x}_1) = \operatorname{Ent}(x_1) \wedge \operatorname{Ent}(\widehat{x}_2) \right. \\
&= \operatorname{Ent}(x_2) : \mathbf{x} = (x_1; x_2) ; x_1; x_2 = 0; \ldots ; 9 \} \\
&= \left\{ [\mathbf{x}]_{\widetilde{P}} = \{ \widehat{\mathbf{x}} = (\widehat{x}_1; \widehat{x}_2) \in U : \operatorname{Ent}(\widehat{x}_1) = x_1 \wedge \operatorname{Ent}(\widehat{x}_2) = x_2 \} : \right. \\
&\qquad \left. \mathbf{x} = (x_1; x_2) ; x_1; x_2 = 0; \ldots ; 9 \right\} .
\end{aligned}
$$

真理大學 Aletheia University

Outline

- Introduction
- Basic terms
- Set approximation
- Approximation of family of sets
- Analysis of decision tables

P-lower Approximation

$$\underline{\widetilde{P}}X = \{x \in U : [x]_{\widetilde{P}} \subseteq X\}$$

is called $\widetilde{P}$-*lower approximation* of the set $X \subseteq U$.

Example 8: Used Car Dealer

- 3 sets by car maker

  (Ferrari, Nissan, Opel)

$$X_{\mathrm{F}} = \{x_3, x_4, x_8\},$$

$$X_{\mathrm{N}} = \{x_2, x_9, x_{10}\},$$

$$X_{\mathrm{O}} = \{x_1, x_5, x_6, x_7\}$$

Object (U)	Number of doors (q_1)	Horsepower (q_2)	Colour (q_3)	Make (q_4)
x_1	2	60	blue	Opel
x_2	2	100	black	Nissan
x_3	2	200	black	Ferrari
x_4	2	200	red	Ferrari
x_5	2	200	red	Opel
x_6	3	100	red	Opel
x_7	3	100	red	Opel
x_8	3	200	black	Ferrari
x_9	4	100	blue	Nissan
x_{10}	4	100	blue	Nissan

$$C = \{c_1, c_2, c_3\} = \{q_1, q_2.q_3\}$$
$$= \{\text{number of doors, horsepower, colour}\}$$

$\widetilde{C}$-lower approximation

only classes $[x_3]_{\widetilde{C}}$ and $[x_8]_{\widetilde{C}}$ are the subsets of the set X_{F},

$$\underline{\widetilde{C}}X_{\mathrm{F}} = \{x_3\} \cup \{x_8\} = \{x_3, x_8\}\,.$$

Sets $[x_2]_{\widetilde{C}}$ and $[x_9]_{\widetilde{C}} = [x_{10}]_{\widetilde{C}}$ are subsets of the set X_{N},

$$\underline{\widetilde{C}}X_{\mathrm{N}} = \{x_2\} \cup \{x_9, x_{10}\} = \{x_2, x_9, x_{10}\}\,.$$

Sets $[x_1]_{\widetilde{C}}$ and $[x_6]_{\widetilde{C}} = [x_7]_{\widetilde{C}}$ are subsets of the set X_{O},

$$\underline{\widetilde{C}}X_{\mathrm{O}} = \{x_1\} \cup \{x_6, x_7\} = \{x_1, x_6, x_7\}\,.$$

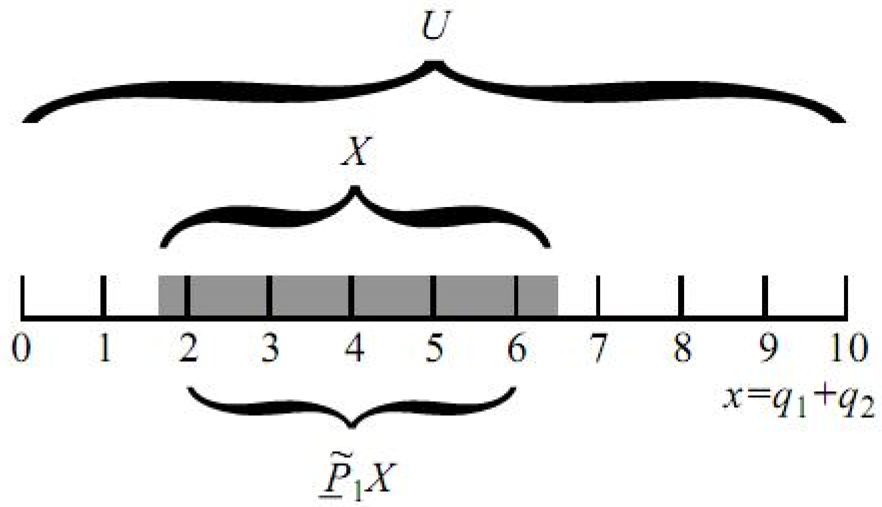
Example 9: Real Numbers

$X = [1,75;\ 6,50]$

$\widetilde{P}_1$ and $\widetilde{P}_2$-lower approximation

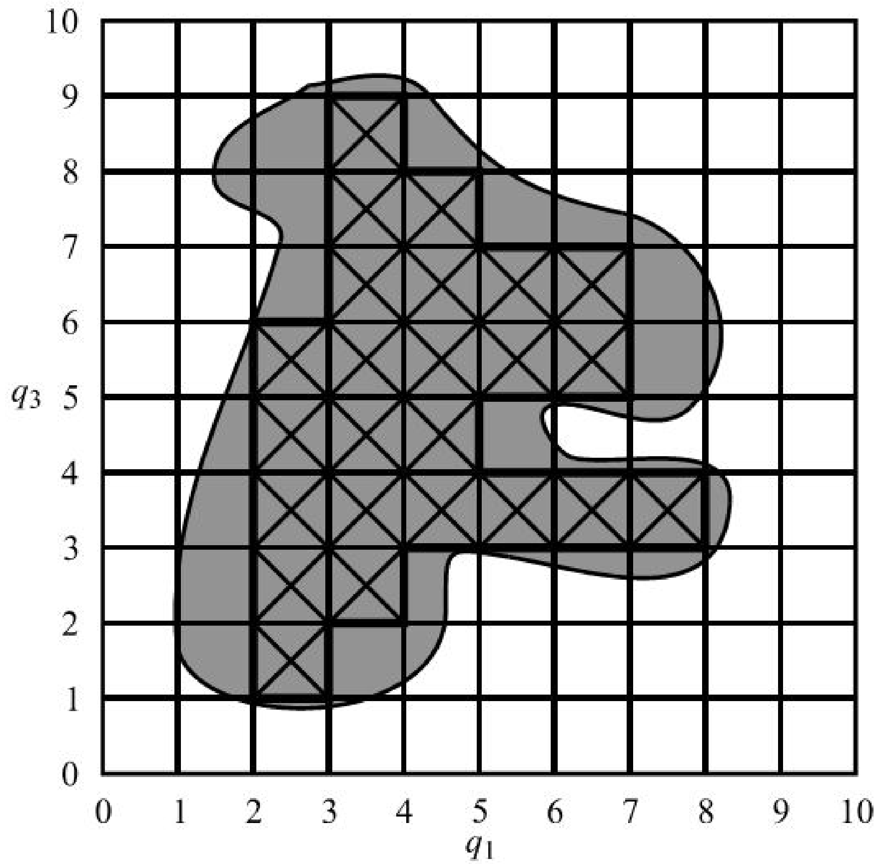
$$\underline{\widetilde{P}_1}X = [2]_{P_1} \cup [3]_{P_1} \cup [4]_{P_1} \cup [5]_{P_1} = [2,6)$$

$$\underline{\widetilde{P}_2}X = \varnothing$$



Example 10: 2-dim. Space

- $\widetilde{P}$-lower approximation of the set X
- 25 equivalence classes



Lower approximation in two-dimensional universe of discourse

P-upper Approximation

$$\overline{\widetilde{P}}X = \{x \in U : [x]_{\widetilde{P}} \cap x \neq \varnothing\}$$

is called $\widetilde{P}$-*upper approximation* of the set $X \subseteq U$

Example 11

- For example 8:

$\widetilde{C}$-upper approximation

$$\overline{\widetilde{C}}X_{\mathrm{F}} = \{x_3\} \cup \{x_4, x_5\} \cup \{x_8\} = \{x_3, x_4, x_5, x_8\}$$

$$\overline{\widetilde{C}}X_{\mathrm{N}} = \{x_2\} \cup \{x_9, x_{10}\} = \{x_2, x_9, x_{10}\}$$

$$\overline{\widetilde{C}}X_{\mathrm{O}} = \{x_1\} \cup \{x_4, x_5\} \cup \{x_6, x_7\} = \{x_1, x_4, x_5, x_6, x_7\}$$

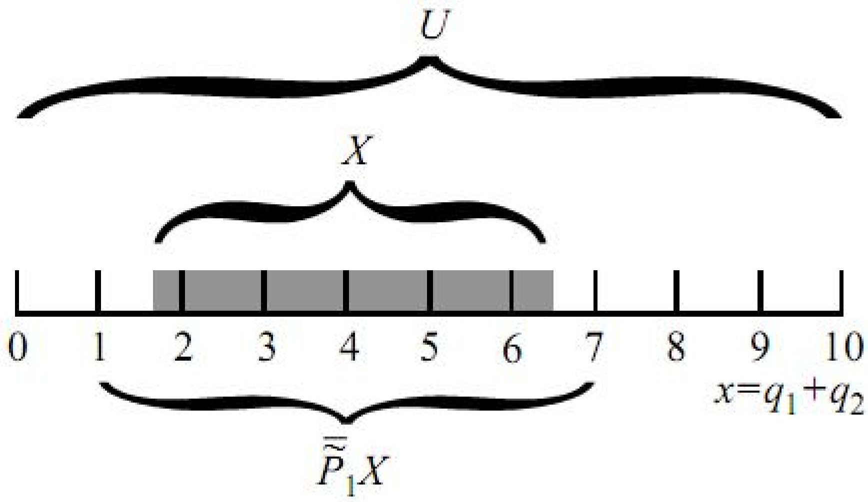
Example 12

- For example 9:

$\widetilde{P}_1$ and $\widetilde{P}_2$-upper approximation

$$\overline{\widetilde{P}}_1 X = [1]_{\widetilde{P}_1} \cup [2]_{\widetilde{P}_1} \cup [3]_{\widetilde{P}_1} \cup [4]_{\widetilde{P}_1} \cup [5]_{\widetilde{P}_1} \cup [6]_{\widetilde{P}_1} = [1,7)$$

$$\overline{\widetilde{P}}_2 X = U$$

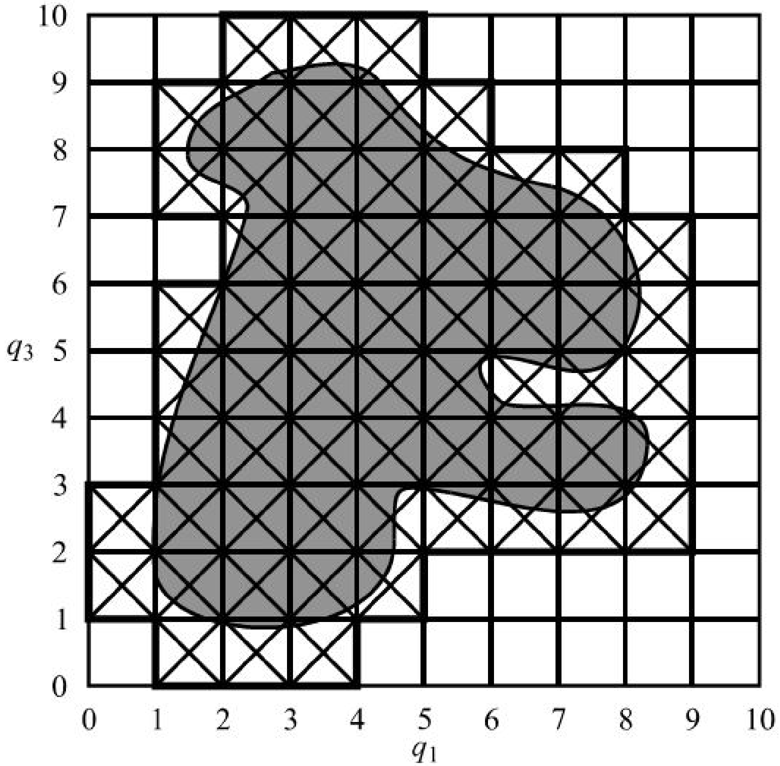


Upper approximation in one-dimensional universe of discourse

Example 13

- For example 10:

$\widetilde{P}_1$ and $\widetilde{P}_2$-upper approximation



Upper approximation in two-dimensional universe of discourse

Positive/Boundary Region

- The positive region of the set X is equal to its lower approximation

$$\tilde{P}\text{-}\mathit{positive\ region}\quad \mathrm{Pos}_{\tilde{P}}(X) = \underline{\tilde{P}}X$$

- Boundary region

$$\tilde{P}\text{-}\mathit{boundary\ region}\quad \mathrm{Bn}_{\tilde{P}}(X) = \overline{\tilde{P}}X \setminus \underline{\tilde{P}}X$$

真理大學
Aletheia University

Example 14: Used Car Dealer

- Boundary region

$$
\begin{aligned}
\mathrm{Bn}_{\widetilde{C}}\left(X_{\mathrm{F}}\right) &= \overline{\widetilde{C}} X_{\mathrm{F}} \setminus \underline{\widetilde{C}} X_{\mathrm{F}} \\
&= \left\{x_3, x_4, x_5, x_8\right\} \setminus \left\{x_3, x_8\right\} = \left\{x_4, x_5\right\},
\end{aligned}
$$

$$
\mathrm{Bn}_{\widetilde{C}}\left(X_{\mathrm{N}}\right) = \overline{\widetilde{C}} X_{\mathrm{N}} \setminus \underline{\widetilde{C}} X_{\mathrm{N}} = \varnothing,
$$

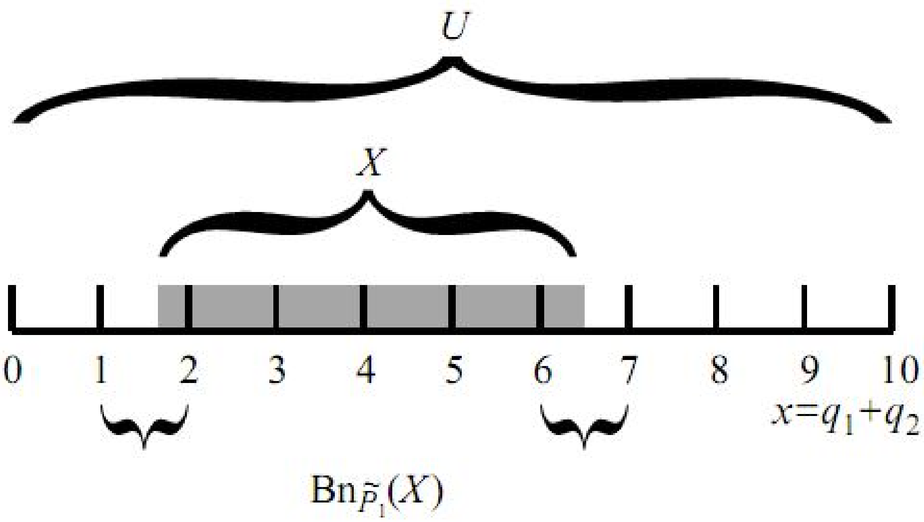
$$
\mathrm{Bn}_{\widetilde{C}}\left(X_{\mathrm{O}}\right) = \overline{\widetilde{C}} X_{\mathrm{O}} \setminus \underline{\widetilde{C}} X_{\mathrm{N}} = \left\{x_4, x_5\right\}.
$$

Example 15:

- Boundary region

$$\mathrm{Bn}_{\tilde{P}_1}(X) = \overline{\tilde{P}_1}X \setminus \underline{\tilde{P}_1}X$$
$$= [1;7) \setminus [2;6) = [1;2) \cup [6;7),$$

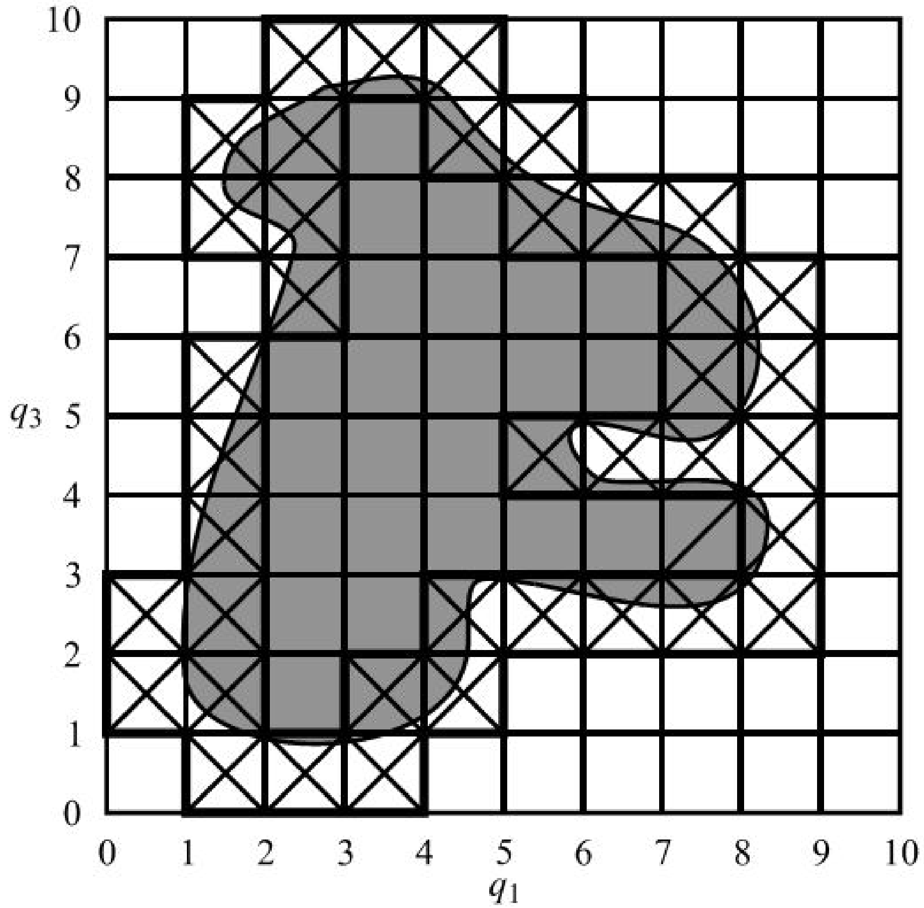
$$\mathrm{Bn}_{\tilde{P}_2}(X) = \overline{\tilde{P}_2}X \setminus \underline{\tilde{P}_2}X$$
$$= U \setminus \varnothing = U.$$



Boundary region in one-dimensional universe of discourse

Example 16: 2-dim. Space

- Boundary region



Boundary region in two-dimensional universe of discourse

Negative Region

$\widetilde{P}$-*negative region* of the set X

$$\mathrm{Neg}_{\widetilde{P}}(X) = U \setminus \overline{\widetilde{P}}X$$

Example 17: Used Car Dealer

- Negative region

$$\mathrm{Neg}_{\tilde{C}}\left(X_{\mathrm{F}}\right) = U \setminus \overline{\tilde{C}}X = \{x_1, x_2, x_6, x_7, x_9, x_{10}\},$$

$$\mathrm{Neg}_{\tilde{C}}\left(X_{\mathrm{N}}\right) = U \setminus \overline{\tilde{C}}X = \{x_1, x_3, x_4, x_5, x_6, x_7, x_8\},$$

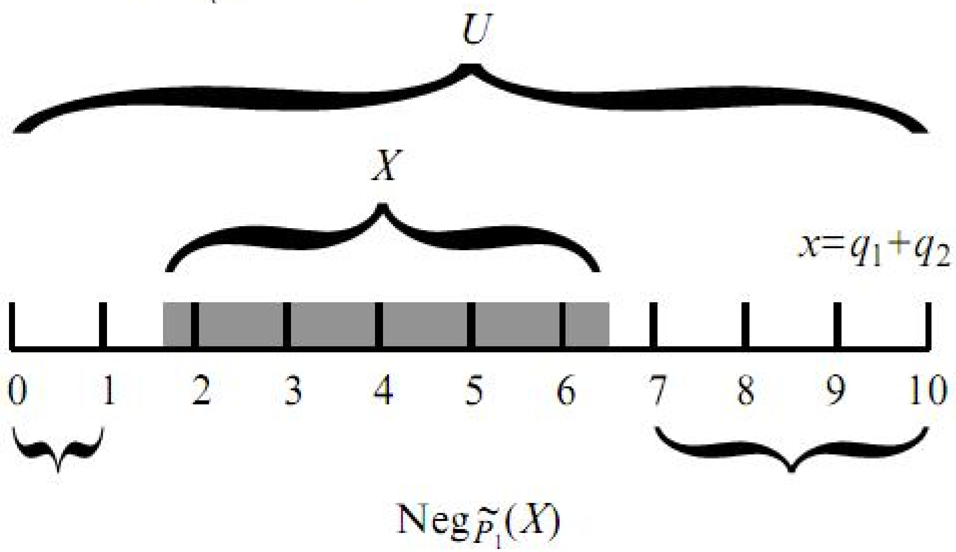
$$\mathrm{Neg}_{\tilde{C}}\left(X_{\mathrm{O}}\right) = U \setminus \overline{\tilde{C}}X = \{x_2, x_3, x_8, x_9, x_{10}\}.$$

Example 18: Real Number

- Negative region

$$\begin{aligned}\mathrm{Neg}_{\tilde{P}_1}(X) &= U \setminus \overline{\tilde{P}}_1 X \\ &= U \setminus [1;7) = [0;1) \cup [7;10)\,,\end{aligned}$$

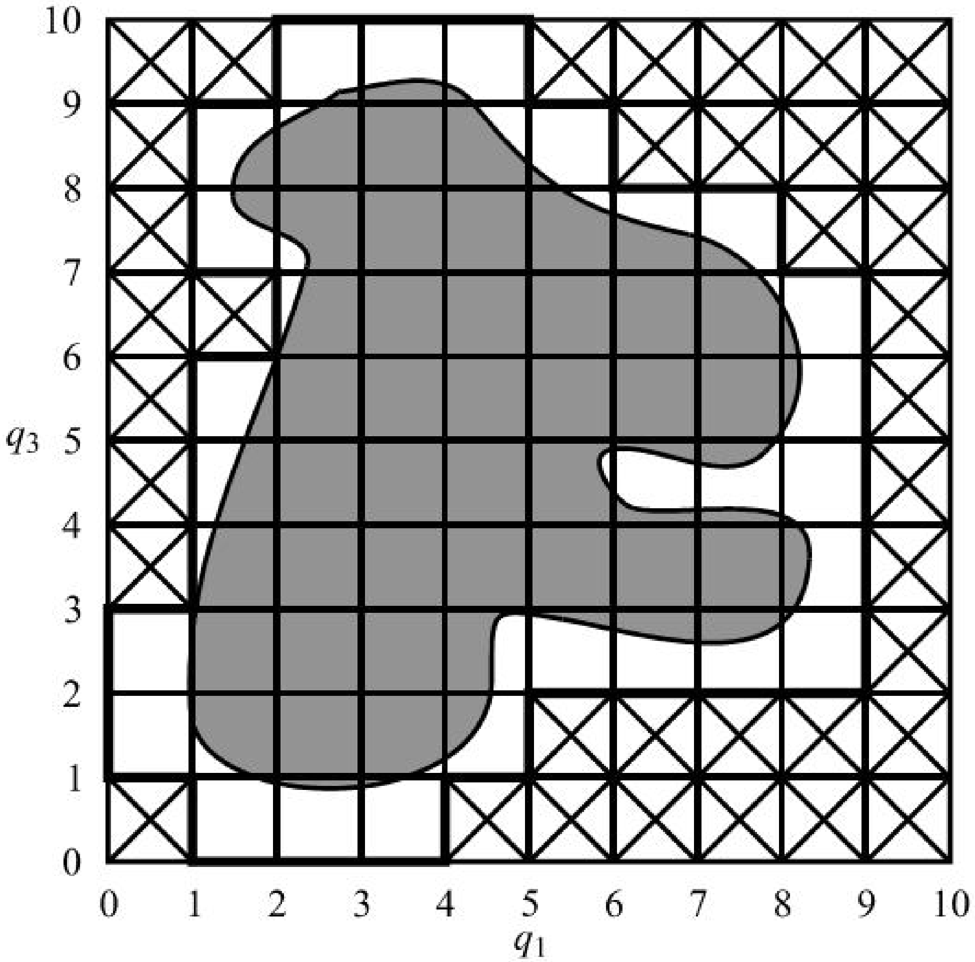
$$\begin{aligned}\mathrm{Neg}_{\tilde{P}_2}(X) &= U \setminus \overline{\tilde{P}}_2 X \\ &= U \setminus U = \varnothing\end{aligned}$$



Negative region in one-dimensional universe of discourse

Example 19

- Negative region



Negative region in two-dimensional universe of discourse

Exactly/Rough Set

- The lower and upper approximation are equal

$\widetilde{P}$-*exactly set* $\quad \underline{\widetilde{P}}X = \overline{\widetilde{P}}X$

- Rough set:

$\widetilde{P}$-rough set $\quad \underline{\widetilde{P}}X \neq \overline{\widetilde{P}}X.$

Definable Set

The set X is called

a) roughly $\widetilde{P}$*-definable set*, if

$$\begin{cases} \underline{\widetilde{P}}X \neq \varnothing \\ \overline{\widetilde{P}}X \neq U, \end{cases}$$

b) internally $\widetilde{P}$*-non definable set*, if

$$\begin{cases} \underline{\widetilde{P}}X = \varnothing \\ \overline{\widetilde{P}}X \neq U, \end{cases}$$

c) externally $\tilde{P}$*-non definable set*, if

$$\begin{cases} \underline{\widetilde{P}}X \neq \varnothing \\ \overline{\widetilde{P}}X = U, \end{cases}$$

d) totally $\widetilde{P}$*-non definable set*, if

$$\begin{cases} \underline{\widetilde{P}}X = \varnothing \\ \overline{\widetilde{P}}X = U. \end{cases}$$

Accuracy of Approximation

$\widetilde{P}$-*accuracy* of approximation of the set X

$$\mu_{\widetilde{P}}(X) = \frac{\overline{\overline{\underline{\widetilde{P}}X}}}{\overline{\overline{\overline{\widetilde{P}}X}}}$$

Example 20: Used Car Dealer

$\tilde{C}$-accuracy of sets X_{F}, X_{N} and X_{O}

$$\mu_{\tilde{C}}(X_{\mathrm{F}}) = \frac{\overline{\overline{\underline{\tilde{C}}X_{\mathrm{F}}}}}{\overline{\overline{\overline{\tilde{C}}X_{\mathrm{F}}}}} = \frac{2}{4} = 0.5,$$

$$\mu_{\tilde{C}}(X_{\mathrm{N}}) = \frac{\overline{\overline{\underline{\tilde{C}}X_{\mathrm{N}}}}}{\overline{\overline{\overline{\tilde{C}}X_{\mathrm{N}}}}} = \frac{3}{3} = 1, \qquad \tilde{C}\text{-exact set.}$$

$$\mu_{\tilde{C}}(X_{\mathrm{O}}) = \frac{\overline{\overline{\underline{\tilde{C}}X_{\mathrm{O}}}}}{\overline{\overline{\overline{\tilde{C}}X_{\mathrm{O}}}}} = \frac{3}{5} = 0.6.$$

Outline

- Introduction
- Basic terms
- Set approximation
- Approximation of family of sets
- Analysis of decision tables

Approximation of Family of Sets

$\mathrm{X} = \{X_1, X_2, ..., X_n\}$

$\widetilde{P}$-*lower approximation* of the family of sets X

$$\underline{\widetilde{P}}\mathrm{X} = \left\{\underline{\widetilde{P}}X_1, \widetilde{P}X_2, ..., PX_n\right\}$$

Example 21: Used Car Dealer

$$
\begin{aligned}
\mathrm{X} &= \{X_{\mathrm{F}}, X_{\mathrm{N}}, X_{\mathrm{O}}\} \\
&= \{\{x_3, x_4, x_8\}, \{x_2, x_9, x_{10}\}, \{x_1, x_5, x_6, x_7\}\}
\end{aligned}
$$

$$
\begin{aligned}
\underline{\widetilde{C}}X &= \left\{\underline{\widetilde{C}}X_{\mathrm{F}}, \underline{\widetilde{C}}X_{\mathrm{N}}, \underline{\widetilde{C}}X_{\mathrm{O}}\right\} \\
&= \{\{x_3, x_8\}, \{x_2, x_9, x_{10}\}, \{x_1, x_6, x_7\}\}
\end{aligned}
$$

Upper Approximation of Family of Sets

$\widetilde{P}$-*upper approximation of family* of sets X

$$\overline{\widetilde{P}}\mathrm{X} = \left\{ \overline{\widetilde{P}}\mathrm{X}_1, \overline{\widetilde{P}}\mathrm{X}_2, \ldots, \overline{\widetilde{P}}\mathrm{X}_n \right\}$$

Example 22: Used Car Dealer

$\tilde{C}$-upper approximation of the family of sets X

$$\overline{\tilde{C}}\mathrm{X} = \left\{\overline{\tilde{C}}\mathrm{X_F}, \overline{\tilde{C}}\mathrm{X_N}, \overline{\tilde{C}}\mathrm{X_O}\right\}$$

$$= \{\{x_3, x_4, x_5, x_8\}, \{x_2, x_9, x_{10}\}, \{x_1, x_4, x_5, x_6, x_7\}\}$$

Positive Region of Family of Sets

$\widetilde{P}$-*positive region of family* of the sets X

$$\mathrm{Pos}_{\widetilde{P}}(\mathrm{X}) = \bigcup_{X_i \in \mathrm{X}} \mathrm{Pos}_{\widetilde{P}}(X_i)$$

Example 23: Used Car Dealer

$\widetilde{C}$-positive region of family of sets X

$$\mathrm{Pos}_{\widetilde{C}}(\mathrm{X}) = \mathrm{Pos}_{\widetilde{C}}(X_{\mathrm{F}}) \cup \mathrm{Pos}_{\widetilde{C}}(X_{\mathrm{N}}) \cup \mathrm{Pos}_{\widetilde{C}}(X_{\mathrm{O}})$$

$$= \{x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{10}\}.$$

Other Definitions

$\widetilde{P}$-*boundary region of family* of the sets X

$$\mathrm{Bn}_{\widetilde{P}}(\mathrm{X}) = \bigcup_{X_i \epsilon \mathrm{X}} \mathrm{Bn}_{\widetilde{P}}(X_i)$$

$\widetilde{P}$-*negative region of family* of the sets X

$$\mathrm{Neg}_{\widetilde{P}}(\mathrm{X}) = U \setminus \bigcup_{X_i \epsilon \mathrm{X}} \overline{\widetilde{P}} X_i$$

$\widetilde{P}$-*quality of approximation of family* of sets X

$$\gamma_{\widetilde{P}}(\mathrm{X}) = \frac{\overline{\overline{\mathrm{Pos}_{\widetilde{P}}(\mathrm{X})}}}{\overline{\overline{U}}}$$

$\widetilde{P}$-*accuracy of approximation of family* of sets X

$$\beta_{\widetilde{P}}(X) = \frac{\overline{\overline{\mathrm{Pos}_{\widetilde{P}}(\mathrm{X})}}}{\sum_{X_i \epsilon \mathrm{X}} \overline{\overline{\overline{\widetilde{P}} X_i}}}$$

Outline

- Introduction
- Basic terms
- Set approximation
- Approximation of family of sets
- Analysis of decision tables

The Rough Set Theory

- Introduces the notion of dependency between features of the information system

Dependence Degree

- Dependence degree of set of attributes ***P*$_2$** on the set of attributes ***P*$_1$** :

$$P_1, P_2 \subseteq Q$$

$$k = \gamma_{\tilde{P}_1}\left(P_2^*\right) \quad \tilde{P}\textit{-quality of approximation of family}$$

$$\gamma_{\tilde{P}}(\mathrm{X}) = \frac{\overline{\overline{\mathrm{Pos}_{\tilde{P}}(\mathrm{X})}}}{\overline{\overline{U}}}$$

- $P_1 \xrightarrow{k} P_2$: the set of attributes ***P*$_2$** depends on the set of attributes ***P*$_1$** to the degree ***k*** < 1
- $P_1 \rightarrow P_2$ on ***k*** = 1

Deterministic v.s. Nondeterministic

- Deterministic
  - For 2 rules:

$$\forall_{\substack{l_a,l_b=1,\ldots,N \\ l_a \neq l_b}} : \forall_{c \in C} \; f_{l_a}(c) = f_{l_b}(c) \rightarrow \forall_{d \in D} \; f_{l_a}(d) = f_{l_b}(d)$$

- Nondeterministic
  - For 2 rules:

$$\exists_{\substack{l_a,l_b \\ l_a \neq l_b}} : \forall_{c \in C} \; f_{l_a}(c) = f_{l_b}(c) \rightarrow \exists_{d \in D} \; f_{l_a}(d) \neq f_{l_b}(d)$$

- The decision table is well defined => all its rules are deterministic

Example 24: Used Car Dealer

- Dependency degree

$$k = \gamma_{\widetilde{C}}(D^*) = \frac{\overline{\overline{\mathrm{Pos}_{\widetilde{C}}(D^*)}}}{\overline{\overline{U}}}$$

$$D^* = \{X_{\mathrm{F}}, X_{\mathrm{N}}, X_{\mathrm{O}}\}$$

Rule (l)	Number of doors (c_1)	Horsepower (c_2)	Colour (c_3)	Make (d_1)
1	2	60	blue	Opel
2	2	100	black	Nissan
3	2	200	black	Ferrari
4	2	200	red	Ferrari
5	2	200	red	Opel
6	3	100	red	Opel
7	3	100	red	Opel
8	3	200	black	Ferrari
9	4	100	blue	Nissan
10	4	100	blue	Nissan

$$\begin{aligned}
\mathrm{Pos}_{\widetilde{C}}(D^*) &= \mathrm{Pos}_{\widetilde{C}}(X_{\mathrm{F}}) \cup \mathrm{Pos}_{\widetilde{C}}(X_{\mathrm{N}}) \cup \mathrm{Pos}_{\widetilde{C}}(X_{\mathrm{O}}) \\
&= \underline{\widetilde{C}}X_{\mathrm{F}} \cup \underline{\widetilde{C}}X_{\mathrm{N}} \cup \underline{\widetilde{C}}X_{\mathrm{O}} \\
&= \{x_3, x_8\} \cup \{x_2, x_9, x_{10}\} \cup \{x_1, x_6, x_7\} \\
&= \{x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{10}\}.
\end{aligned}$$

$$k = \frac{\overline{\overline{\{x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{10}\}}}}{\overline{\overline{\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}}}} = \frac{8}{10}$$

Dependency Degree

- k < 1 means “not well defined”

$$C \xrightarrow{0.8} D$$

- We cannot unambiguously infer on the membership of objects of the space ***U*** to the particular sets $\boldsymbol{X_F}$, $\boldsymbol{X_N}$ and $\boldsymbol{X_O}$

- What if we remove nondeterministic rules: 4 and 5?

$$\gamma_{\tilde{C}}(D^*) = \frac{\overline{\overline{\{x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{10}\}}}}{\overline{\overline{\{x_1, x_2, x_3, x_6, x_7, x_8, x_9, x_{10}\}}}} = 1$$

  – Well defined!

Well-defined Decision Table

A well-defined decision table (after removing non-deterministic rules)

Rule (l)	Number of doors (c_1)	Horsepower (c_2)	Colour (c_3)	Make (d_1)
1	2	60	blue	Opel
2	2	100	black	Nissan
3	2	200	black	Ferrari
6	3	100	red	Opel
7	3	100	red	Opel
8	3	200	black	Ferrari
9	4	100	blue	Nissan
10	4	100	blue	Nissan

Another Way to Well-define?

- How?

Attribute Independency

The set of attributes $P_1 \subset Q$ is *independent*

for each $P_2 \subset P_1$ the inequality $\widetilde{P}_1 \neq \widetilde{P}_2$ occurs

The set of attributes $P_1 \subseteq Q$ is *independent with respect to* the set of attributes $P_2 \subseteq Q$ (P_2-*independent*)

for each $P_3 \subset P_1$ $\quad \mathrm{Pos}_{\widetilde{P}_1}(P_2^*) \neq \mathrm{Pos}_{\widetilde{P}_3}(P_2^*)$

Otherwise, the set P_1 is P_2-*dependent.*

Reduct and Indispensable

Every independent set $P_2 \subset P_1$ for which $\widetilde{P}_2 = \widetilde{P}_1$ is called *the reduct* of the set of attributes $P_1 \subseteq Q$.

Every P_2-independent set $P_3 \subset P_1$ for which $\widetilde{P}_3 = \widetilde{P}_1$ is called *the relative reduct* of a set of attributes $P_1 \subseteq Q$ with respect to P_2 (the so-called P_2-reduct).

The attribute $p \in P_1$ is *indispensable from* P_1, if for $P_2 = P_1 \setminus \{p\}$, the equation $\widetilde{P}_2 \neq \widetilde{P}_1$ holds. Otherwise, the attribute p is *dispensable.*

Core

The set of all indispensable attributes from the set P is called a *core* of P, which is notated as follows:

$$\mathrm{CORE}\,(P) = \left\{ p \in P : \widetilde{P}' \neq \widetilde{P}, P' = P \backslash \{p\} \right\}.$$

Normalized Coefficient of Significance

The normalized coefficient of significance of subset of the set of conditional attributes $C' \subset C$ is expressed by the following formula

$$\sigma_{(C,D)}\left(C'\right) = \frac{\gamma_{\tilde{C}}\left(D^*\right) - \gamma_{\tilde{C}''}\left(D^*\right)}{\gamma_{\tilde{C}}\left(D^*\right)},$$

where $C'' = C \setminus C'$.

Approximation Error

Any given subset of the set of conditional attributes $C' \subset C$ is called a rough D-reduct of the set of attributes C, and *the approximation error* of this reduct is defined as follows:

$$\varepsilon_{(C,D)}(C') = \frac{\gamma_{\tilde{C}}(D^*) - \gamma_{\tilde{C}'}(D^*)}{\gamma_{\tilde{C}}(D^*)}.$$

Applications of Rough Set Theory

- Decision rules
  - Rules generation and removal of redundant data
- Classification

Application Example: Used Car Dealer

Original decision table (before reduction)

Number of doors (c_1)	Horsepower (c_2)	Colour (c_3)	Fuel (c_4)	Upholstery (c_5)	Rims (c_6)	Make (d_1)
2	60	blue	Ethyl gasoline	woven fabric	steel	Opel
2	100	black	Diesel oil	woven fabric	steel	Nissan
2	200	black	Ethyl gasoline	leather	Al	Ferrari
2	200	red	Ethyl gasoline	leather	Al	Ferrari
2	200	red	Ethyl gasoline	woven fabric	steel	Opel
3	100	red	Diesel oil	leather	steel	Opel
3	100	red	gas	woven fabric	steel	Opel
3	200	black	Ethyl gasoline	leather	Al	Ferrari
4	100	blue	gas	woven fabric	steel	Nissan
4	100	blue	Diesel oil	woven fabric	Al	Nissan

Results

Decision table after removing redundant data

Number of doors (c_1)	Horsepower (c_2)	Colour (c_3)	Fuel (c_4)	Upholstery (c_5)	Rims (c_6)	Make (d_1)
	60					Opel
		black	Diesel oil			Nissan
	200			leather		Ferrari
	200			leather		Ferrari
		red			steel	Opel
		red			steel	Opel
		red			steel	Opel
	200			leather		Ferrari
4						Nissan
4						Nissan

IF rims is steel **AND** colour is red **THEN** make is Opel
IF horsepower is 60 **THEN** make is Opel
IF doors is 4 **THEN** make is Nissan
IF colour is black **AND** fuel is Diesel oil **THEN** make is Nissan
IF horsepower is 200 **AND** upholstery is leather **THEN** make is Ferrari