Rough Set

Rough Set

- **Definition (*Information System) Let A = (A1, A2, A3, …, Ak) be a non-empty finite set of attributes and U = {(a1, a2, ..., ak)} be a non-empty finite set of k-tuples, termed as the objects. V (Ai) denote the set of values for the attributes Ai. Then an information system is defined as an ordered pair I (U, A) such that for all i = 1, 2, ..., k there is a function fi*
- ***fi : U → V(Ai)***
- **This means, every object in the set *U has an attribute value for every element in the set A. The set U is called the universe of the information system.***

Information System

#	Roll No.	Physics	Chemistry	Mathematics
1	1	82	90	98
2	2	80	96	100
3	3	63	62	68
4	4	70	92	100
5	5	54	51	36
6	6	92	94	90
7	7	10	12	0

An information system augmented with a decision attribute

Roll No.	Physics	Chemistry	Mathematics	PCM %age	Admitted
1	82	90	98	90	Yes
2	80	96	100	92	Yes
3	63	62	68	64.3	No
4	70	92	100	87.3	Yes
5	54	51	36	47	No
6	92	94	90	92	Yes
7	10	12	0	7.3	No

(Decision System) A ***decision system D (U, A, d) is an information system I (U, A) augmented with a special attribute d $\notin$ A, known as the decision attribute.***
The decision system shown in Table has the decision attribute *Admitted that has binary values 'Yes' or 'No'. These values are based on certain rules which guide the decision. On a closer scrutiny into Table, this rule maybe identified as If PCM %age is greater than or equal to 87.3 then Admitted = Yes, else Admitted = No.*

INDISCERNIBILITY

- Decision systems have the capacity to express knowledge about the underlying information system. However, a decision table may contain redundancies such as indistinguishable states or superfluous attributes. In Table, the attributes *Physics, Chemistry and Mathematics are unnecessary to take decision about admittance so long as the aggregate percentage is available. The decision attributes in decision systems are generated from the conditional attributes. These conditional attributes share common properties as clarified in the subsequent examples. However, before we go on to discuss these issues, we need to review the concept of equivalence relation.*

Definition 5.5 (*Indiscernibility*) *Let I = (U, A) be an information system where U = {(a1, …, ak)} is the non-empty finite set of k-tuples known as the objects and U = {A1, …, Ak} is a non-empty finite set of attributes. Let P ⊆ A be a subset of the attributes. Then the set of P-indiscernible objects is defined as the set of objects having the same set of attribute values.*

INDI (P) = {(x, y), x, y ∈ U | ∀a ∈ A, x (a) = y(a)}

Personnel profiles

Name	Gender	Nationality	Complexion	Mother-tongue	Profession
Amit	M	Indian	Dark	Hindi	Lawyer
Bao	M	Chinese	Fair	Chinese	Teacher
Catherine	F	German	Fair	German	Journalist
Dipika	F	Indian	Fair	Hindi	Journalist
Lee	M	Chinese	Dark	Chinese	Lawyer

Consider the profiles of a set of persons as shown Table. Let *P = {Gender, Complexion, Profession}* ⊆ *A = {Gender, Nationality, Complexion, Mother-tongue, Profession}. From the information system shown in Table 5.4 Catherine and Dipika are P-indiscernible as both are fair complexioned lady journalists. Similarly, Amit and Lee are also P-indiscernible. Hence, INDI (P) = {{Catherine, Dipika}, {Amit, Lee}}. On the other hand, the set of P-indiscernible objects with respect to P = {Gender, Complexion} happens to be INDI (P) = {{Amit, Lee}, {Bao}, {Catherine, Dipika}}.*

- Table presents an information system regarding various features of three types of cars, viz., Car A, B and C. Unlike the other tables, the attributes are arranged here in rows while the objects are along the columns.

Table Car features

Features	Car A	Car B	Car C
Power Door Locks	Yes	Yes	No
Folding Rear Seats	No	Yes	No
Rear Wash Wiper	Yes	Yes	Yes
Tubeless Tyres	Yes	Yes	Yes
Remote Boot	Yes	No	Yes
Steering Adjustment	No	No	Yes
Rear Defroster	Yes	No	Yes
Seating Capacity	4	5	4
Mileage (in km/litre)	18	18	16
Max. Speed (in km/h)	160	160	180

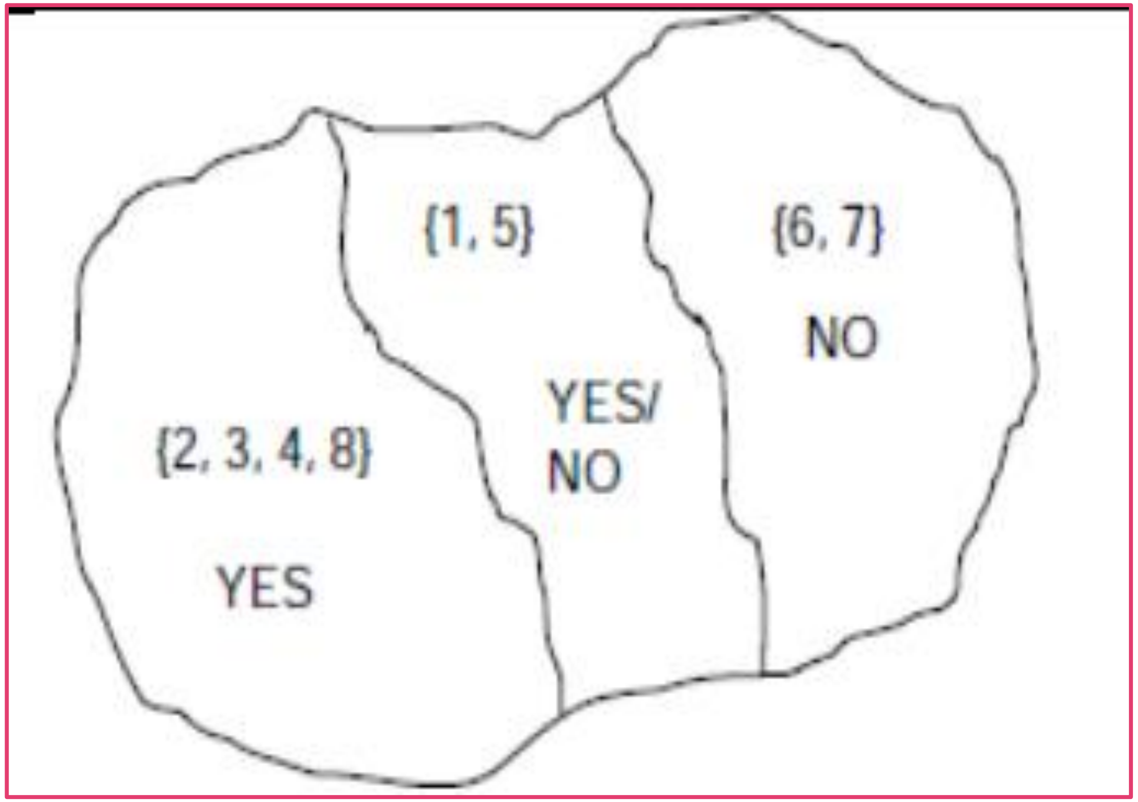
Here, A = {*Power Door Locks, Folding Rear Seat, Rear Wash Wiper, Tubeless Tyres, Remote Boot, Steering Adjustment, Rear Defroster, Seating Capacity, Mileage, Max. Speed*}, U = {*Car A, Car B, Car C*}. Let us consider the three subsets of attributes M = {*Mileage, Max. Speed*}, R = {*Rear Wash Wiper, Remote Boot, Rear Defroster*} and L = {*Power Door Locks, Steering Adjustment*}. Then $IND_I(M)$ = {{*Car A, Car B*}, {*Car C*}}, $IND_I(R)$ = {{*Car A, Car C*}, {*Car B*}}, $IND_I(L)$ = {{*Car A, Car B*}, {*Car C*}}.

Indiscernibility is an equivalence relation and an indiscernibility relation partitions the set of objects in an information system into a number of equivalence classes. The set of objects B-indiscernible from x is denoted as $[x]_B$. For example, if B = {*Folding Rear Seats, Rear Wash Wiper, Tubeless Tyres, Remote Boot, Rear Defroster*}, then $[Car\ A]_B$ = {*Car A, Car C*}. However, if F = {*Power Door Locks, Steering Adjustments, Mileage, Max. Speed*}, then $[Car\ A]_F$ = {*Car A, Car B*}.

SET APPROXIMATIONS

- In a decision system, the indiscernibility equivalence relation partitions the universe U into a number of subsets based on identical values of the outcome attribute.
- Such partitions are crisp and have clear boundaries demarcating the area of each subset. However, such crisp boundaries might not always be possible. For example consider the decision system presented in Table.
- It consists of age and activity information of eight children aged between 10 to 14 months. The outcome attribute 'Walk' has the possible values of YES or NO depending on whether the child can walk or not.
- A closer observation reveals that it is not possible to crisply group the pairs (Age, Can Walk) based on the outcome into YES / NO categories.
- The problem arises in case of entries 1 and 5 where the ages of the children are same but the outcomes differ. Therefore, it is not possible to decisively infer whether a child can walk or not on the basis of its age information only.

#	Age (in months)	Can Walk
1	12	No
2	14	Yes
3	14	Yes
4	13	Yes
5	12	Yes
6	10	No
7	10	No
8	13	Yes



- The situation is depicted in Fig.. Objects 2, 3, 4 and 8 belong to the class that can be described by the statement 'If age is 13 or 14 months then the child can walk'.
- Similarly, objects 6 and 7 define a class corresponding to the rule 'If age is 10 months then the child can not walk'.
- However, objects 1 and 5 are on the boundary region in the sense that though both of them correspond to children of age 12 years, their 'Can Walk' information is NO in case of object 1 and YES in case of object 5.
- It is under such circumstances that the concept of rough sets comes into the picture and informally we may say that 'Sets which consist objects of an information system whose membership cannot be ascertained with certainty or any measure of it are called rough sets.
- Formally, rough sets are defined in terms of lower and upper approximations.

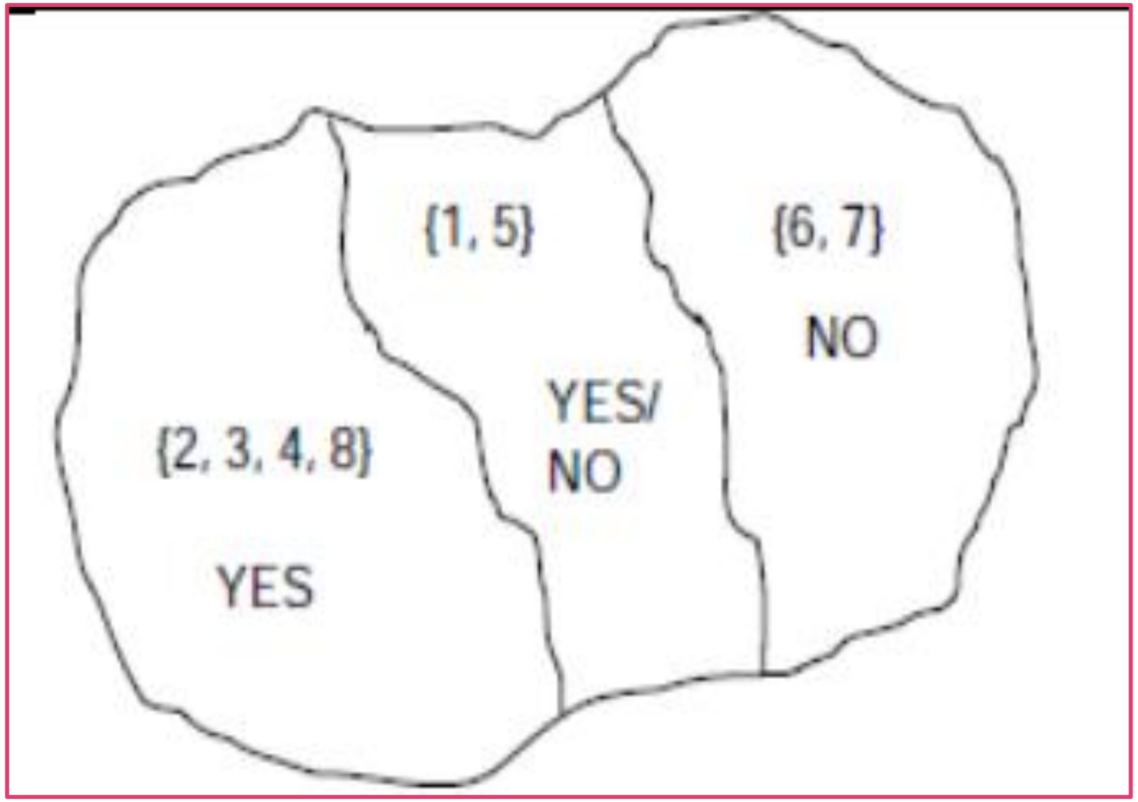
Definition . (*Lower and Upper Approximations*) Let $I = (U, A)$ be an information system and $B \subseteq A$ is a subset of attributes and $X \subseteq U$ is a set of objects. Then

B-lower approximation of $X = \underline{B}(X) = \{x \mid [x]_B \subseteq X\}$
B-upper approximation of $X = \overline{B}(X) = \{x \mid [x]_B \cap X \neq \emptyset\}$
The objects that comply with the condition and fall in $\underline{B}(X)$ are classified with certainty as members of set X, while, those objects that comply with and therefore belong to $\overline{B}(X)$ are classified as *possible* members.

Definition . (*Boundary Region*) The set $BN_B(X) = \overline{B}(X) - \underline{B}(X)$ is called the *B-boundary region* of X.The B-boundary region of X consists of those objects which we cannot decisively classify as inside or outside the set X on the basis of the knowledge of their values of attributes in B. If a set has a nonempty boundary region, it is said to be a rough set.

Definition (*Outside Region*) The set $U - \overline{B}X$ is called the *B-outside region* of X. The *B-outside region* of X consists of elements that are classified with certainty as not belonging to X on the basis of knowledge in B.

#	Age (in months)	Can Walk
1	12	No
2	14	Yes
3	14	Yes
4	13	Yes
5	12	Yes
6	10	No
7	10	No
8	13	Yes



With reference to the information system presented in Table ..., let $W = \{y \mid Can\ Walk\ (y) = Yes\} = \{2, 3, 4, 5, 8\}$. Now, the set of *Age*-indiscernible objects of U, $IND_{Age}(U) = \{\{1, 5\}, \{2, 3\}, \{4, 8\}, \{6, 7\}\}$. Hence the sets of the *Age*-indiscernible objects for various objects are $[1]_{Age} = [5]_{Age} = \{1, 5\}$, $[2]_{Age} = [3]_{Age} = \{2, 3\}$, $[4]_{Age} = [8]_{Age} = \{4, 8\}$, $[6]_{Age} = [7]_{Age} = \{6, 7\}$. Thus, assuming $B = \{Age\}$ we have

B-lower approximation of W: $\underline{B}W = \{2, 3, 4, 8\}$

B-upper approximation of W: $\overline{B}W = \{1, 2, 3, 4, 5, 8\}$

B-boundary region of W: $BN_B(W) = \{1, 5\}$

B-outside region of W: $U - \overline{B}W = \{6, 7\}$

As $BN_B(W) = \{1, 5\} \neq \emptyset$, W is a rough set with respect to knowledge about walking.

PROPERTIES OF ROUGH SETS

- **Rough sets, defined as above in terms of the lower and upper approximations, satisfy certain properties.**

1. $\underline{B}(X) \subseteq X \subseteq \overline{B}(X)$
2. $\underline{B}(\varnothing) = \overline{B}(\varnothing) = \varnothing;\ \underline{B}(U) = \overline{B}(U) = U$
3. $\overline{B}(X \cup Y) = \overline{B}(X) \cup \overline{B}(Y)$
4. $\underline{B}(X \cap Y) = \underline{B}(X) \cap \underline{B}(Y)$
5. $\underline{B}(X \cup Y) \supseteq \underline{B}(X) \cup \underline{B}(Y)$
6. $\overline{B}(X \cap Y) \subseteq \overline{B}(X) \cap \overline{B}(Y)$
7. $X \subseteq Y \rightarrow \underline{B}(X) \subseteq \underline{B}(Y)$ and $\overline{B}(X) \subseteq \overline{B}(Y)$
8. $\underline{B}(U - X) = U - \overline{B}(X)$
9. $\overline{B}(U - X) = U - \underline{B}(X)$
10. $\underline{B}\,\underline{B}(X) = \overline{B}\,\underline{B}(X) = \underline{B}(X)$
11. $\overline{B}\,\overline{B}(X) = \underline{B}\,\overline{B}(X) = \overline{B}(X)$

ROUGH MEMBERSHIP

- **Rough sets are also described with the help of rough membership of individual elements. The membership of an object x to a rough set X with respect to knowledge in B is expressed**

as $\mu_X^B(x)$. Rough membership is similar, but not identical, to fuzzy membership. It is defined as

$$\mu_X^B(x) = \frac{|[x]_B \cap X|}{|[x]_B|}$$

Obviously, rough membership values lie within the range 0 to 1, like fuzzy membership values.

$$\mu_X^B : U \rightarrow [0,1]$$

The rough membership function may as well be interpreted as the conditional probability that *x*belongs to *X* given *B*. It is the degree to which *x* belongs to *X* in view of information about *x* expressed by *B*. The lower and upper approximations, as well as the boundary regions, can be defined in terms of rough membership function.

$$\underline{B}(X)=\{x \in U \mid \mu_X^B(x)=1\}$$

$$\overline{B}(X)=\{x \in U \mid \mu_X^B(x)>0\}$$

$$BN_B(X)=\{x \in U \mid 0< \mu_X^B(x)<1\}$$

Let us again consider the information system presented in Table . $W = \{y \mid Can\ Walk\,(y) = Yes\} = \{2, 3, 4, 5, 8\}$ and $B = \{Age\}$. In Example we have found $IND_{Age}(U) = \{\{1, 5\}, \{2, 3\}, \{4, 8\}, \{6, 7\}\}$, $\underline{B}W = \{2, 3, 4, 8\}$, $\overline{B}W = \{1, 2, 3, 4, 5, 8\}$, $BN_B(W) = \{1, 5\}$, and $U - \overline{B}W = \{6, 7\}$. Moreover, $[1]_{Age} = [5]_{Age} = \{1, 5\}$, $[2]_{Age} = [3]_{Age} = \{2, 3\}$, $[4]_{Age} = [8]_{Age} = \{4, 8\}$, $[6]_{Age} = [7]_{Age} = \{6, 7\}$. Now

$$\mu_W^B(1) = \frac{|[1]_B \cap W|}{|[1]_B|} = \frac{|\{1,5\} \cap \{2,3,4,5,8\}|}{|\{1,5\}|} = \frac{1}{2}$$

Similarly,

$$\mu_W^B(5) = \frac{1}{2}, \mu_W^B(2) = \mu_W^B(3) = \mu_W^B(4) = \mu_W^B(8) = 1, \mu_W^B(6) = \mu_W^B(7) = 0.$$

Properties of Rough Membership

The properties listed below are satisfied by rough membership functions. These properties either follow from the definition or are easily provable.

1. $\mu_X^B(x)=1$ *iff* $x \in \underline{B}(X)$
2. $\mu_X^B(x)=0$ *iff* $x \in U-\overline{B}(X)$
3. $0 < \mu^B(x) < 1$ *iff* $x \in BN_B(X)$
4. $\mu_{U-X}^B(x)=1-\mu_X^B(x)$
5. $\mu_{X\cup Y}^B(x) \geq max\ (\mu_X^B(x), \mu_Y^B(x))$
6. $\mu_{X\cap Y}^B(x) \leq min\ (\mu_X^B(x), \mu_Y^B(x))$

Definition (*Rough Sets*) Given an information system $I = (U, A)$, $X \subseteq U$ and $B \subseteq A$, roughness of X is defined as follows.

1. Set X is *rough* with respect to B if $\underline{B}(X) \neq \overline{B}(X)$ or $\overline{B}(X) - \underline{B}(X) \neq \phi$
2. Set X is *rough* with respect to B if there exist $x \in U$ such that $0 < \mu_X^B(x) < 1$.

Based on the properties of set approximations and the definition of indiscernibility, four basic classes of rough sets are defined.

We can further characterize rough sets in terms of the accuracy of approximation, defined as

$$\alpha_B(X) = \frac{\underline{B}(X)}{\overline{B}(X)}$$

It is obvious that $0 \leq \alpha_B(X) \leq 1$. If $\alpha_B(X) = 1$, the set X is *crisp* with respect to B, otherwise, if $\alpha_B(X) < 1$, then X is rough with respect to B.

Definition (*Dependency*) Let $I = (U, A)$ be an information system and $B_1, B_2 \in A$ are sets of attributes. B_1 is said to be totally dependent on attribute B_2 if all values of attribute B_1 are uniquely determined by the values in B_2. This is denoted as $B_2 \Rightarrow B_1$.

Set approximations and rough membership) Table presents a decision system for a number of individuals seeking loan from a bank and the bank's decision in this regard. The conditional attributes are *Gender, Age, Income, Car* (indicating whether the applicant owns a car or not), *Defaulter* (whether the applicant is a defaulter in paying of a previous loan) and their valid attribute values are {*Male, Female*}, {*Middle-aged, Young-adult, Aged*}, {*High, Medium, Low*}, {*Yes, No*}, and {*Yes, No*} respectively. The decision attribute is *Loan Granted* with value set {*Yes, No*}.

Let B = {*Age, Income, Car*} $\subset A$ = {*Gender, Age, Income, Car, Defaulter*} be a set of attributes. Then

1. Compute $IND_B(I)$
2. If $X = \{x \in U \mid \textit{Loan Granted}(x) = \textit{Yes}\}$ then compute B-lower and B-upper approximations of X and determine if X is rough in terms of knowledge in B.
3. Calculate $\mu_X^B(x)$ for each $x \in U$.

Table Loan applicants' data set

#	Name	Gender (G)	Age (A)	Income (I)	Car (C)	Defaulter (D)	Loan Granted
1	Tony	M	Middle-aged	High	Yes	Yes	No
2	Vinod	M	Middle-aged	High	No	No	Yes
3	Sheela	F	Young-adult	High	Yes	No	Yes
4	Kete	F	Aged	Low	No	No	No
5	Nina	F	Middle-aged	Middle	Yes	No	Yes
6	Sandip	M	Aged	High	No	No	No
7	Mita	F	Young-adult	High	No	No	Yes
8	Bob	M	Young-adult	High	Yes	Yes	No
9	Bill	M	Middle-aged	Middle	No	No	Yes
10	Martha	F	Middle-aged	Middle	No	No	Yes
11	Bruce	M	Middle-aged	High	Yes	No	Yes
12	Gogo	M	Aged	Low	No	No	No

Solution The computations are shown below.

1. $IND_B(I) = \{\{1, 11\}, \{2\}, \{3, 8\}, \{4, 12\}, \{5\}, \{6\}, \{7\}, \{9, 10\}\}$

2. $X = \{x \in U \mid Loan\ Granted\ (x) = Yes\} = \{2, 3, 5, 7, 9, 10, 11\}$

$\underline{B}\,(X) = \{2, 5, 7, 9, 10\}$, $\overline{B}\,(X) = \{1, 2, 3, 5, 7, 8, 9, 10, 11\}$ and

$BN_B\,(X) = \{1, 8\} \neq \phi$.

$\therefore$ X is rough with respect to knowledge of the attributes $\{Age, Income, Car\}$ (Fig.).

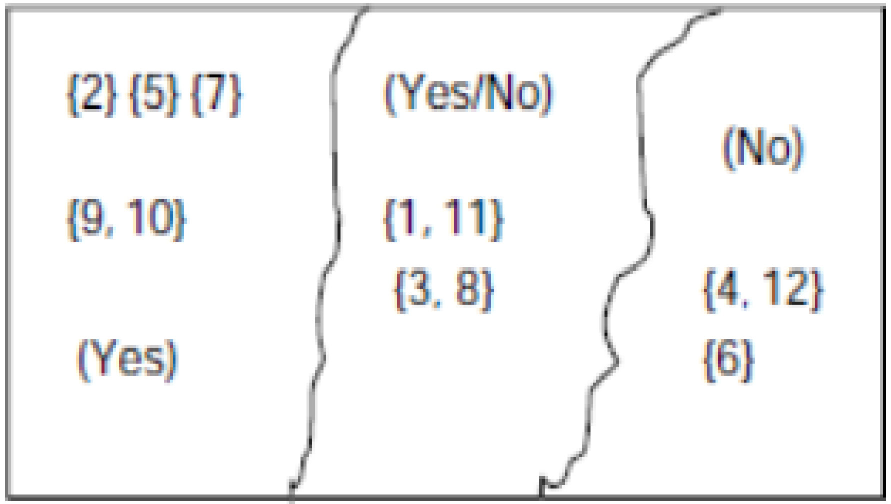


Fig. Set approximations

Computations of the rough membership values for individual elements are shown below.

$$\mu_X^B(1) = \mu_X^B(11) = \frac{|\{1,11\} \cap \{2,3,5,7,9,10,11\}|}{|\{1,11\}|} = \frac{1}{2}$$

Similarly,

$$\mu_X^B(3) = \mu_X^B(8) = \frac{|\{3,8\} \cap \{2,3,5,7,9,10,11\}|}{|\{3,8\}|} = \frac{1}{2}$$

$\mu_X^B(2) = \mu_X^B(5) = \mu_X^B(7) = \mu_X^B(9) = \mu_X^B(10) = 1$, and

$\mu_X^B(4) = \mu_X^B(6) = \mu_X^B(12) = 0.$

It will be Continued....