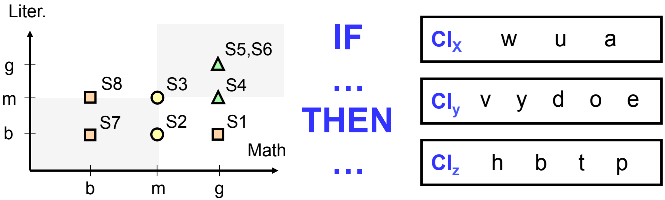


Rough Set Based Decision Support
Dominance-based Rough Set Approach

Miłosz Kadziński

Institute of Computing Science
Poznan University of Technology, Poland

Sorting (Ordinal Classification)

How to assign alternatives to pre-defined and ordered decision classes?

sorting

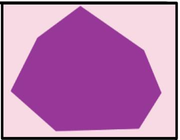
C_3	w	u	a		
C_2	v	y	d	o	e
C_1	h	b	t	p	

↓ C_3=good ↓ C_2=med. C_1=bad

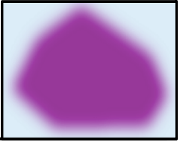
C_3=leading C_2=average C_1=poor

- **Data mining**: a process of ***discovering patterns*** in data sets
- **Classification**: Find a model for predicting *class* (decision) attribute as a function of the values of other attributes
- **Financial management:** credit risk assessment; business failure prediction; stock evaluation
- **Marketing:** customer satisfaction measurement
- **Medicine:** diagnosis (e.g., the severity of illness),
- **Education:** Parametric evaluation of research units: A^+, A, B, B+, and C
- *As opposed to standard classification, the classes are ordered in terms of preference*
- *As opposed to clustering, the classes are pre-defined*

Crisp and Vague Concepts (Sets)



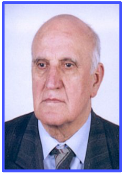
- Use the concepts of "**sets**" and "**classes**" interchangeably
- **Mathematics** requires that all mathematical notions, including "sets" must be **exact**
  - Otherwise, precise reasoning would be impossible
- In classical set theory, a set is uniquely determined by its elements
  - Every element must be uniquely classified as belonging to the set or not
  - In mathematics, the notion of a set is crisp (precise) one
  - E.g., the set of odd numbers is crisp since every integer is either odd or even
- German logician Gottlob Frege: "*the concept must have sharp boundary*"
- In the last decades, philosophers and computer scientists have become interested in **vague (imprecise) concepts**
- The notion of *beautiful paintings* is vague: we are unable to classify all paintings uniquely as beautiful and not beautiful; some paintings remain in a doubtful area



Vague Concepts in Rough Set Approach

- *Almost all concepts that we use in natural language are vague*
- Common sense reasoning must be based on **vague concepts**
- Vagueness is often associated with the **boundary region**, i.e., the existence of objects which cannot be uniquely classified relative to a set of its complement
- Vagueness related to **insufficient specificity**: a lack of feasible searching methods for a set of features adequately describing concepts

- **Rough set approach** was proposed by Pawlak (1982) for dealing with imperfect knowledge, vague concepts, and inconsistency of data
- Fundamental importance in AI and cognitive sciences
- Final aim of RST = discovering knowledge from data for multi-attribute decision making = finding (classification) patterns, based on concepts of rough sets, agreeing with situations described by data



Zdzisław Pawlak

Z. Pawlak. Rough sets. *International Journal of Parallel Programming*. **11** (5): 341–356, 1982

Objects, Attributes, Information and Decision Systems

An **information system (*U, Q*)** = a data table containing rows labeled by objects (alternatives), columns labeled by attributes (criteria), and entries are criteria values

We distinguish a partition of the sets of attributes into two classes: **condition** or **decision** side of the description, corresponding to input or output

- ***C*** = a set of condition attributes (V_q – condition attribute values for $q \in \boldsymbol{C}$)
- ***D*** = a set of decision attributes (V_q – decision attribute values for $q \in \boldsymbol{D}$)
  - *For this lecture, we will assume that **D** is a singleton*

The sets condition and decision attributes are disjoint: $\boldsymbol{C} \cap \boldsymbol{D} = \varnothing$

The tuple (U, C, D) is called a **decision system** or **decision table**

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	medium	medium	**good**
S5	good	good	**good**
S6	good	good	**good**
S7	bad	bad	**bad**
S8	bad	medium	**bad**

A decision table (*U, C, D*) concerning a few examples of students described in terms of their results in **Math (M)** and **Literature (L)**, and **General achievement = Decision (D)**

U = a finite set of objects (*universe*) $\quad U = \{\text{S1, S2, S3, S4, S5, S6, S7, S8}\}$

Q = a set of attributes $\quad Q = \{\text{M, L, D}\} \quad C = \{\text{M, L}\} \quad D = \{\text{D}\}$

$\boldsymbol{V_q}$ = a set of values $\quad V_M = V_L = V_D = \{\text{bad, medium, good}\}$

$V_C = \prod_{q=1}^{|C|} V_q$ condition attribute space $\quad V_D = \prod_{q=1}^{|D|} V_q$ decision attribute space

$x_q = g_q(x)$ = evaluation of object x on criterion q

Illustrative Study - Students

The data describing a given decision situation include either observations of **DM's past decisions** in the same decision context, or **examples of decisions** consciously elicited by the DM on the demand of an analyst (holistic judgments = easy and natural preferences)

- **Indirect preference information** incorporating the value system of the DM
- AI and Machine Learning: learning from examples
- Aim: discovering a law relating the evaluation with the classification decision
- The classification model has to be discovered by inductive learning from data and can be applied to other objects

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	medium	medium	**good**
S5	good	good	**good**
S6	good	good	**good**
S7	bad	bad	**bad**
S8	bad	medium	**bad**

Reasoning about preference ordered data

- Criteria = attributes associated with preference orders
  - bad < medium < good
- Classification decision with preference ordered classes
  - bad < medium < good
- Monotonic relationship between evaluations of objects on criteria and the preference ordered value of decision
- An improvement in one criterion should not worsen the classification decision, while the other criteria values are unchanged

Treating Inconsistency

Past decisions or example decisions may be **inconsistent**

The inconsistency may come from various sources:

- **Missing attributes/criteria** in the description of objects (opinion of the student's tutor expressed only verbally during an assessment)
- **Unstable preferences of decision makers** (e.g., the assessment committee members changed their view on the influence of Math (M) on the General assessment (D))

Handling inconsistency is of crucial importance prior to induction of preference (knowledge) model

- **Possible solutions**: treat inconsistencies as noise or errors to be eliminated; amalgamate with consistent data by some averaging operator
- **Desired solution**: identify inconsistencies and treat them explicitly

Indiscernibility Principle in IRSA

Key messages underlying the **Classical Rough Set Approach (CRSA)** =
= **Indiscernibility Rough Set Approach (IRSA)**

- Ignore preference orders in the values sets of condition and decision attributes
- ***Indiscernibility principle***:

  if x and y are **indiscernible** with respect to **all relevant attributes** ($x_q = y_q$ for all $q \in C$), then x should be **classified to the same class (set)** as y ($x_r = y_r$ for all $r \in D$)
- Due to a lack of information or knowledge, objects that are the same in terms of all relevant attributes may be assigned to various sets (classes)
- $I_P(x)$, $I_R(x)$ – equivalence classes including x for $P \subseteq C$ and $R \subseteq C$
- A set of all indiscernible objects is called an **elementary set** and forms a basic **granule (atom) of knowledge** about the universe
- Granules are the basic building blocks of our knowledge
- Due to the granularity of knowledge, some objects cannot be discerned and appear the same

St.	M	L	D
S3	medium	medium	**medium**
S4	medium	medium	**good**
S5	good	good	**good**
S6	good	good	**good**

$P = \{M, L\}$

$I_P(S3) = I_P(S4) = \{S3, S4\}$ $I_P(S5) = I_P(S6) = \{S5, S6\}$

$\{S3, S4\}_P$ $\{S5, S6\}_P$

$R = \{D\}$

$I_R(S3) = \{S3\}$ $I_R(S4) = I_R(S5) = I_R(S6) = \{S4, S5, S6\}$

$\{S3\}_R$ $\{S4, S5, S6\}_R$

Lower and Upper Approximations in IRSA

- Classific. patterns are functions representing granules $I_R(x)$, $R \subseteq D$, by granules $I_P(x)$, $P \subseteq C$
- Rough sets (classes) cannot be characterized in terms of information about their elements
- Any rough set is replaced by a pair of precise sets, called the **lower and upper approximations** of the set
- The **lower approximation** of set X with respect to $P \subseteq C$ is composed of all elementary sets included in X using P (are certainly in X in view of P)

$$\boxed{\underline{P}(X) = \{x \in U : I_P(x) \subseteq X\}}$$

- The **upper approximation** of set X with respect to $P \subseteq C$ is composed of all elementary sets which have a non-empty intersection with X using P (are possibly in X in view of P)

$$\boxed{\overline{P}(X) = \bigcup_{x \in X} I_P(x)}$$

St.	M	L	D
S3	medium	medium	**medium**
S4	medium	medium	**good**
S5	good	good	**good**
S6	good	good	**good**

$$P = \{\mathrm{M}, \mathrm{L}\} \qquad I_P(\mathrm{S3}) = I_P(\mathrm{S4}) = \{\mathrm{S3}, \mathrm{S4}\} \qquad I_P(\mathrm{S5}) = I_P(\mathrm{S6}) = \{\mathrm{S5}, \mathrm{S6}\}$$

$$X = \mathrm{good} = \{\mathrm{S4}, \mathrm{S5}, \mathrm{S6}\}$$

$$\underline{P}(\mathrm{good}) = \{\mathrm{S5}, \mathrm{S6}\}$$

$$\overline{P}(\mathrm{good}) = \{\mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}\}$$

Class "good" includes students S5 and S6 certainly and students S3 and S4 possibly

Dominance Principle vs. IRSA

- IRSA ignores preference orders in the values sets of condition and decision attributes
- $\succeq_q$ is a weak preference relation on U with respect to criterion $q \in C \cup D$
- $x_q \succeq_q y_q$ means x_q is at least as good as y_q with respect to criterion q
- x dominates y with respect to $P \subseteq C$ ($x\ D_P\ y$; x P-dominates y) if $x_q \succeq_q y_q$ for all $q \in C$
- D_P is transitive and reflexive; it is a partial preorder
- IRSA does not detect inconsistency w.r.t. dominance (the basic principle in decision analysis)
- ***Dominance principle*** (monotonicity constraints)

  If x is **at least as good** as y with respect to **all relevant criteria** (x dominates y),
  then x **should be classified at least as good as** y

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	medium	medium	**good**

- S3 dominates S4 while being assigned to a worse class
- S3 and S4 are indiscernible, so this inconsistency is handled by IRSA
- S1 dominates S2 while being assigned to a worse class
- S1 and S2 are not indiscernible, so this inconsistency is not handled by IRSA

Dominance-based Rough Set Approach

„The procedure of induction consists in accepting as true the simplest law that can be reconciled with our experiences” (L. Wittgenstein, *Tractatus Logico-Philosophicus*, 6.363)

- This simplest law is just monotonicity and, therefore, the inductive discovery of preference model (rules) can be seen as a specific way of dealing with monotonicity
- Dominance-based Rough Set Approach (DRSA) permits data structuring w.r.t. possible violation of dominance prior to rule induction
- In order to handle monotonic dependency between conditions and decision, we will reason in terms of class unions rather than individual classes (sets):
  - **upward union of classes**, t=2,...,m (*„at least”* class Cl_t)
  - **downward union of classes**, t=1,...,m-1 (*„at most”* class Cl_t)

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	good	medium	**good**
S5	good	good	**good**
S6	good	good	**good**
S7	bad	bad	**bad**
S8	bad	medium	**bad**

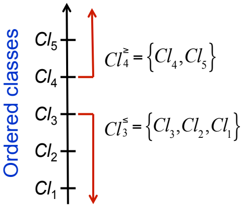
$$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$$

$$Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$$

$Cl_t^{\geq}$ and $Cl_t^{\leq}$ are positive and negative dominance cones in decision space reduced to a single dimension

For our problem:

- At most bad
- At most medium
- At least medium
- At least good



Dominance Cones

D_P – dominance relation (partial preorder) in condition space, $P \subseteq C$

Granules of knowledge are **dominance cones** in condition space:

P-dominating set (positive cone)

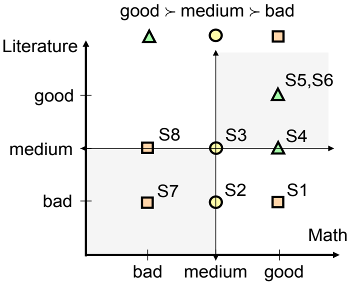
$$\boxed{D_P^+(x) = \{y \in U : yD_P x\}}$$

P-dominated set (negative cone)

$$\boxed{D_P^-(x) = \{y \in U : xD_P y\}}$$

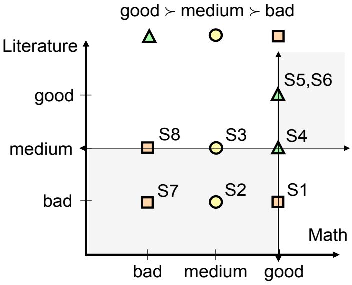
Classification patterns to be discovered are functions representing granules $Cl_t^{\geq}$, $Cl_t^{\leq}$ by granules $D_P^+(x)$, $D_P^-(x)$

St.	M	L	D
S1	g	b	**b**
S2	m	b	**m**
S3	m	m	**m**
S4	g	m	**g**
S5	g	g	**g**
S6	g	g	**g**
S7	b	b	**b**
S8	b	m	**b**



$$D_P^+(\mathrm{S3}) = \{x \in U :\ xD_P\mathrm{S3}\} = \{\mathrm{S3, S4, S5, S6}\}$$

$$D_P^-(\mathrm{S3}) = \{x \in U :\ \mathrm{S3}D_P x\} = \{\mathrm{S2, S3, S7, S8}\}$$

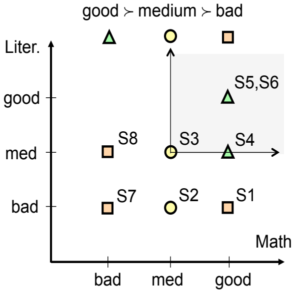


$$D_P^+(\mathrm{S4}) = \{x \in U :\ xD_P\mathrm{S4}\} = \{\mathrm{S4, S5, S6}\}$$

$$D_P^-(\mathrm{S4}) = \{x \in U :\ \mathrm{S4}D_P x\} = \{\mathrm{S1, S2, S3, S4, S7, S8}\}$$

Lower Approximation (1)

- The **lower approximation** of the class union consists of all objects which **definitely** belong to the class union
- The **lower approximation** of *upward class union* $Cl_t^{\geq}$ with respect to $P \subseteq C$ is the set of all objects that can **for certain** be classified to $Cl_t^{\geq}$ using P (are certainly in $Cl_t^{\geq}$ in view of P)
  - x belongs to $Cl_t^{\geq}$ and all objects P-dominating x ($D_P^+(x)$) belong to $Cl_t^{\geq}$
- Formally, the P-lower approximation of class union $Cl_t^{\geq}$ is a set of objects in universe U whose **positive dominance cones $D_P^+(x)$ for $P \subseteq C$ are contained in class union $Cl_t^{\geq}$**
- Iterate over objects contained in $Cl_t^{\geq}$, and gather these whose positive dominance cones are entirely contained in $Cl_t^{\geq}$
- The lower approximation of $Cl_t^{\geq}$ = objects that are dominated only by objects from $Cl_t^{\geq}$



$$\underline{P}\left(Cl_t^{\geq}\right) = \left\{x \in U : D_P^+(x) \subseteq Cl_t^{\geq}\right\}$$

$$D_P^+(S2) = \{S1, S2, S3, S4, S5, S6\}$$

$$D_P^+(S5) = \{S5, S6\}$$

$$D_P^+(S3) = \{S3, S4, S5, S6\}$$

$$D_P^+(S6) = \{S5, S6\}$$

$$D_P^+(S4) = \{S4, S5, S6\}$$

$$\underline{P}\left(Cl_{medium}^{\geq}\right) = \left\{x \in U : D_P^+(x) \subseteq Cl_{medium}^{\geq}\right\}$$

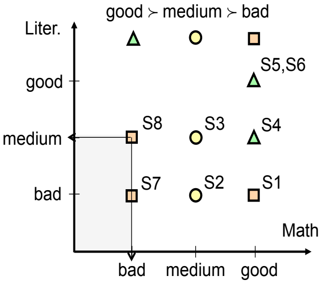
$$\underline{P}\left(Cl_{good}^{\geq}\right) = \left\{x \in U : D_P^+(x) \subseteq Cl_{good}^{\geq}\right\}$$

$$\underline{P}\left(Cl_{medium}^{\geq}\right) = \{S3, S4, S5, S6\}$$

$$\underline{P}\left(Cl_{good}^{\geq}\right) = \{S4, S5, S6\}$$

Lower Approximation (2)

- The lower approximation of the class union consists of all objects which **definitely** belong to the class union
- The **lower approximation** of *downward class union* $Cl_t^{\leq}$ with respect to $P \subseteq C$ is the set of all objects that can **for certain** be classified to $Cl_t^{\leq}$ using P (are certainly in $Cl_t^{\leq}$ in view of P)
  - x belongs to $Cl_t^{\leq}$ and all objects P-dominated by x ($D_P^-(x)$) belong to $Cl_t^{\leq}$
- Formally, the P-lower approximation of class union $Cl_t^{\leq}$ is a set of objects in universe U whose **negative dominance cones $D_P^-(x)$ for $P \subseteq C$ are contained in class union $Cl_t^{\leq}$**
- Iterate over objects contained in $Cl_t^{\leq}$, and gather these whose negative dominance cones are entirely contained in $Cl_t^{\leq}$
- The lower approximation of $Cl_t^{\leq}$ = objects that dominate only objects from $Cl_t^{\leq}$



$$\underline{P}\left(Cl_t^{\leq}\right) = \left\{x \in U : D_P^-(x) \subseteq Cl_t^{\leq}\right\}$$

$$D_P^-(S1) = \{S1, S2, S7\} \qquad D_P^-(S3) = \{S2, S3, S7, S8\} \qquad D_P^-(S8) = \{S7, S8\}$$

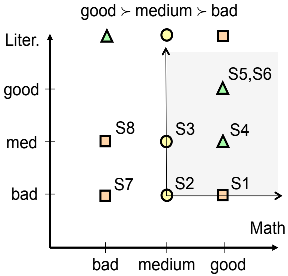
$$D_P^-(S2) = \{S2, S7\} \qquad D_P^-(S7) = \{S7\}$$

$$\underline{P}\left(Cl_{bad}^{\leq}\right) = \left\{x \in U : D_P^-(x) \subseteq Cl_{bad}^{\leq}\right\} = \{S7, S8\}$$

$$\underline{P}\left(Cl_{medium}^{\leq}\right) = \left\{x \in U : D_P^-(x) \subseteq Cl_{medium}^{\leq}\right\} = \{S1, S2, S3, S7, S8\}$$

Upper Approximation (1)

- The upper approximation of the class union consists of all objects which **possibly** belong to the class union
- The **upper approximation** of *upward class union* $Cl_t^{\geq}$ with respect to $P \subseteq C$ is the set of all objects that can **possibly** be classified to $Cl_t^{\geq}$ using P (are possibly in $Cl_t^{\geq}$ in view of P)
  - x belongs to $Cl_t^{\geq}$
  - x does not belong to $Cl_t^{\geq}$, but it P-dominates object y belonging to $Cl_t^{\geq}$ ($x \in D_P^-(y)$, $y \in Cl_t^{\geq}$)
- Formally, the P-upper approximation of class union $Cl_t^{\geq}$ is a set of objects in universe U whose negative dominance cones $D_P^-(x)$ for $P \subseteq C$ have a non-empty intersection with class union $Cl_t^{\geq}$
- Iterate over objects contained in $Cl_t^{\geq}$, and gather their entire positive dominance cones $D_P^+(x)$
- The upper approximation of $Cl_t^{\geq}$ = objects that dominate at least one object from $Cl_t^{\geq}$



$$\overline{P}\left(Cl_t^{\geq}\right) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x) = \left\{x \in U : D_P^-(x) \cap Cl_t^{\geq} \neq \emptyset\right\}$$

$$D_P^+(\text{S2}) = \{\text{S1, S2, S3, S4, S5, S6}\} \qquad D_P^+(\text{S3}) = \{\text{S3, S4, S5, S6}\}$$

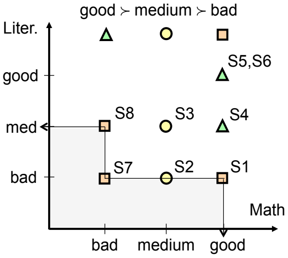
$$D_P^+(\text{S4}) = \{\text{S4, S5, S6}\} \qquad D_P^+(\text{S5}) = \{\text{S5, S6}\} \qquad D_P^+(\text{S6}) = \{\text{S5, S6}\}$$

$$\overline{P}\left(Cl_{medium}^{\geq}\right) = \bigcup_{x \in Cl_{medium}^{\geq}} D_P^+(x) = \{\text{S1, S2, S3, S4, S5, S6}\}$$

$$\overline{P}\left(Cl_{good}^{\geq}\right) = \bigcup_{x \in Cl_{good}^{\geq}} D_P^+(x) = \{\text{S4, S5, S6}\}$$

Upper Approximation (2)

- The upper approximation of the class union consists of all objects which **possibly** belong to the class union
- The **upper approximation** of *downward class union* $Cl_t^{\leq}$ with respect to $P \subseteq C$ is the set of all objects that can **possibly** be classified to $Cl_t^{\leq}$ using P (are possibly in $Cl_t^{\leq}$ in view of P)
  - x belongs to $Cl_t^{\leq}$
  - x does not belong to $Cl_t^{\leq}$, but it is P-dominated by object y belonging to $Cl_t^{\leq}$ ($x \in D_P^+(y)$, $y \in Cl_t^{\leq}$)
- Formally, the P-upper approximation of class union $Cl_t^{\leq}$ is a set of objects in universe U whose positive dominance cones $D_P^+(x)$ for $P \subseteq C$ have a non-empty intersection with class union $Cl_t^{\leq}$
- Iterate over objects contained in $Cl_t^{\leq}$, and gather their entire negative dominance cones $D_P^-(x)$
- The upper approximation of $Cl_t^{\leq}$ = objects that are dominated by at least one object from $Cl_t^{\leq}$



$$\overline{P}\left(Cl_t^{\leq}\right) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x) = \left\{x \in U : D_P^+(x) \cap Cl_t^{\leq} \neq \emptyset\right\}$$

$D_P^-(S1) = \{S1, S2, S7\}$ | $D_P^-(S2) = \{S2, S7\}$ | $D_P^-(S3) = \{S2, S3, S7, S8\}$

$D_P^-(S7) = \{S7\}$ | $D_P^-(S8) = \{S7, S8\}$

$$\overline{P}\left(Cl_{bad}^{\leq}\right) = \bigcup_{x \in Cl_{bad}^{\leq}} D_P^-(x) = \{S1, S2, S7, S8\}$$

$$\overline{P}\left(Cl_{medium}^{\leq}\right) = \bigcup_{x \in Cl_{medium}^{\leq}} D_P^-(x) = \{S1, S2, S3, S7, S8\}$$

Boundary

- The boundary region of the rough set is the difference between the upper and lower approximations
- The **boundary region** of class union $Cl_t^{\geq}$ ($Cl_t^{\leq}$) with respect to $P \subseteq C$ is the set of all objects which belong to $Cl_t^{\geq}$ ($Cl_t^{\leq}$) using P with some ambiguity
- DRSA expresses vagueness by employing a boundary region of a class union

$$Bn_P\left(Cl_t^{\geq}\right) = \overline{P}\left(Cl_t^{\geq}\right) - \underline{P}\left(Cl_t^{\geq}\right), \quad Bn_P\left(Cl_{t-1}^{\leq}\right) = \overline{P}\left(Cl_{t-1}^{\leq}\right) - \underline{P}\left(Cl_{t-1}^{\leq}\right)$$

$$\underline{P}\left(Cl_{good}^{\geq}\right) = \overline{P}\left(Cl_{good}^{\geq}\right) = \{S4,\ S5,\ S6\}$$

$$Bn_P\left(Cl_{good}^{\geq}\right) = \varnothing$$

$$\underline{P}\left(Cl_{medium}^{\geq}\right) = \{S3,\ S4,\ S5,\ S6\} \quad \overline{P}\left(Cl_{medium}^{\geq}\right) = \{S1,\ S2,\ S3,\ S4,\ S5,\ S6\}$$

$$Bn_P\left(Cl_{medium}^{\geq}\right) = \{S1,\ S2\}$$

$$\underline{P}\left(Cl_{bad}^{\leq}\right) = \{S7,\ S8\} \quad \overline{P}\left(Cl_{bad}^{\leq}\right) = \{S1,\ S2,\ S7,\ S8\}$$

$$Bn_P\left(Cl_{bad}^{\leq}\right) = \{S1,\ S2\}$$

$$\underline{P}\left(Cl_{medium}^{\leq}\right) = \overline{P}\left(Cl_{medium}^{\leq}\right) = \{S1,\ S2,\ S3,\ S7,\ S8\}$$

$$Bn_P\left(Cl_{medium}^{\leq}\right) = \varnothing$$

- If the boundary region is empty, it means that the class union is **crisp**
- If the boundary region is non-empty, the class union is **rough** (inexact)
- A non-empty boundary region means that our knowledge about the class union is not sufficient to define the set precisely

Inclusion and Complementarity Properties

Rough set is a tuple composed of two crisp sets, the **lower and upper approximations**

If the object is certainly contained in $Cl_t^{\ge}$ ($Cl_t^{\le}$), then it is also possibly contained in $Cl_t^{\ge}$ ($Cl_t^{\le}$)

Inclusion property of rough approximation:

$$\underline{P}\left(Cl_t^{\le}\right) \subseteq Cl_t^{\le} \subseteq \overline{P}\left(Cl_t^{\le}\right)$$
$$\underline{P}\left(Cl_t^{\ge}\right) \subseteq Cl_t^{\ge} \subseteq \overline{P}\left(Cl_t^{\ge}\right)$$

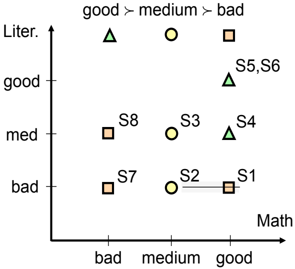
$$\underline{P}\left(Cl_{bad}^{\le}\right) = \{S7, S8\} \quad Cl_{bad}^{\le} = \{S1, S7, S8\} \quad \overline{P}\left(Cl_{bad}^{\le}\right) = \{S1, S2, S7, S8\}$$

Complementarity property of rough approx.:

$$\underline{P}\left(Cl_t^{\le}\right) = U - \overline{P}\left(Cl_{t-1}^{\ge}\right)$$
$$\underline{P}\left(Cl_t^{\ge}\right) = U - \overline{P}\left(Cl_{t-1}^{\le}\right)$$

$$\underline{P}\left(Cl_{bad}^{\le}\right) = \{S7, S8\} \qquad \overline{P}\left(Cl_{medium}^{\ge}\right) = \{S1, S2, S3, S4, S5, S6\}$$

If the object is certainly contained in $Cl_t^{\ge}$, then it is impossible that it could belong to $Cl_{t-1}^{\le}$



Identity of boundaries: if the object belongs with ambiguity to $Cl_t^{\ge}$, then it also belongs with ambiguity to $Cl_{t-1}^{\le}$

$$Bn_P\left(Cl_t^{\ge}\right) = Bn_P\left(Cl_{t-1}^{\le}\right)$$

$$Bn_P\left(Cl_{bad}^{\le}\right) = Bn_P\left(Cl_{medium}^{\ge}\right) = \{S1, S2\}$$

Need for Data Mining with Rules

- There is an increasing gap between generating data and their understanding
- **Data mining** is extracting information from a data set and transforming it to a comprehensive structure for further use
  - Data mining is the process of discovering patterns in data
- Pattern – rule, trend, phenomenon, regularity, anomaly, hypothesis, function, etc.
  - Patterns should be true, non-trivial, potentially useful, and understandable

- Class union representations have limited practical use because they provide no insight for deciding whether novel objects can be deemed as their members
  - Intentional description of classes is desired
- **Decision rule approach** using a set of „*if …, then …*" decision rules
  - „*People make decisions by searching for rules that provide good justification of their choices*" (Slovic, 1975)
- Rules are classification patterns discovered from data in the table
- Decision rules ensure the minimal representation of knowledge in decision tables
- Rules describe the scope of various classes

IF
…
THEN
…

The **decision rule** is an expression of the form $\Phi \Rightarrow \Psi$ (*if* Φ, *then* Ψ)

- Φ is a **condition** part of the rule (predecessor of the rule)
- Ψ is a **decision** part of the rule (successor of the rule)

The condition part is a conjunction ($\wedge$) of conditions in the form:

$x_{qj} \geq r_{qj}$ (the evaluation of alternative x on attribute q_j is at least equal to v_{qj})

$x_{qj} \leq r_{qj}$ (the evaluation of alternative x on attribute q_j is at most equal to v_{qj})

The decision part suggests the assignment to one or disjunction ($\vee$) of classes:

$x \in Cl_t^{\geq}$ (x is assigned to class at least Cl_t); $x \in Cl_t^{\leq}$ (x is assigned to class at most Cl_t)

$x \in Cl_{t-1} \vee Cl_t \vee Cl_{t+1}$ (x is assigned to class Cl_{t-1} or Cl_t or Cl_{t+1})

St.	M	L	D
S1	g	b	b
S2	m	b	m
S3	m	m	m
S4	g	m	g
S5	g	g	g
S6	g	g	g
S7	b	b	b
S8	b	m	b

The objects supporting the rule satisfy its condition and decision parts

if M $\preceq$ bad, ***then*** student $\preceq$ **bad** (support: **S7**, **S8**)

if M $\preceq$ medium, ***then*** student $\preceq$ **medium** (support: **S2**, **S3**, **S7**, **S8**)

if L $\preceq$ bad, ***then*** student $\preceq$ **medium** (support: **S1**, **S2**, **S7**)

if M $\succeq$ medium $\wedge$ L $\succeq$ medium, ***then*** student $\succeq$ **medium** (support: **S3**, **S4**, **S5**, **S6**)

if M $\succeq$ good $\wedge$ L $\succeq$ medium, ***then*** student $\succeq$ **good** (support: **S4**, **S5**, **S6**)

if M $\succeq$ medium $\wedge$ L $\preceq$ bad, ***then*** student **bad** *or* **medium** (support: **S1**, **S2**)

Types of Decision Rules Induced from Approximations

Rules identify values that drive DM's decisions: each rule is a scenario of a causal relationship between evaluations on a subset of criteria and a comprehensive judgment

- The condition part of all types of rules is a conjunction of conditions of type "at least" or "at most"

Certain (discriminant) **decision rule** supported by objects from the lower approximation of $Cl_t^{\geq}$ or $Cl_t^{\leq}$

$$\text{if } x_{q1} \succeq r_{q1} \text{ and } x_{q2} \succeq r_{q2} \text{ and } \dots \text{ and } x_{qp} \succeq r_{qp} \text{ then } x \in Cl_t^{\geq}$$

$$\text{if } x_{q1} \preceq r_{q1} \text{ and } x_{q2} \preceq r_{q2} \text{ and } \dots \text{ and } x_{qp} \preceq r_{qp} \text{ then } x \in Cl_t^{\leq}$$

Possible decision rule supported by objects from the upper approximation of of $Cl_t^{\geq}$ or $Cl_t^{\leq}$ (partly discriminant rule)

$$\text{if } x_{q1} \succeq r_{q1} \text{ and } x_{q2} \succeq r_{q2} \text{ and } \dots \text{ and } x_{qp} \succeq r_{qp} \text{ then (possibly) } x \in Cl_t^{\geq}$$

$$\text{if } x_{q1} \preceq r_{q1} \text{ and } x_{q2} \preceq r_{q2} \text{ and } \dots \text{ and } x_{qp} \preceq r_{qp} \text{ then (possibly) } x \in Cl_t^{\leq}$$

Approximate decision rule supported by objects in class union boundaries, belonging to $Cl_t \cup Cl_s \cup \dots \cup Cl_u$ (i.e., classes to which belong inconsistent objects supporting this rule)

$$\text{if } x_{q1} \succeq r_{q1} \text{ and } \dots \text{ and } x_{qk} \succeq r_{qk} \text{ and } x_{qk+1} \preceq r_{qk+1} \text{ and } \dots \text{ and } x_{qp} \preceq r_{qp}$$
$$\text{then } x \in Cl_t \text{ or } x \in Cl_s \text{ or } \dots \text{ or } x \in Cl_u$$

where $\{q_1, q_2, \dots, q_p\} \subseteq C,\ (r_{q1}, r_{q2}, \dots, r_{qp}) \in V_{q1} \times V_{q2} \times \dots \times V_{qp}$

Measures Characterizing Rule in System $S = (U, C, D)$

Consider a decision rule: $\Phi \Rightarrow \Psi$

Support is the number of objects satisfying the condition and decision parts of the rule

$$\sup_S(\Phi,\Psi) = card\left(\left\|\Phi \wedge \Psi\right\|_S\right)$$

Strength is the percentage of objects in *U* supporting the rule

$$\sigma_S(\Phi,\Psi) = \frac{card\left(\left\|\Phi \wedge \Psi\right\|_S\right)}{card(U)}$$

Certainty (or confidence) factor of the rule is the ratio of support and the number of objects consistent with the condition part: *how many objects satisfying the conditions, meet the conclusion*

$$cer_S(\Phi,\Psi) = \frac{card\left(\left\|\Phi \wedge \Psi\right\|_S\right)}{card\left(\left\|\Phi\right\|_S\right)}$$

Coverage factor of the rule is the ratio of support and the number of objects consistent with the decision part: *how many objects satisfying the conclusion, meet the conditions*

$$\mathrm{cov}_S(\Phi,\Psi) = \frac{card\left(\left\|\Phi \wedge \Psi\right\|_S\right)}{card\left(\left\|\Psi\right\|_S\right)}$$

Relation between certainty, coverage, and Bayes theorem

$$cer_S(\Phi,\Psi) = \Pr(\Psi \mid \Phi) = \frac{\Pr(\Phi \wedge \Psi)}{\Pr(\Phi)}$$

$$\mathrm{cov}_S(\Phi,\Psi) = \Pr(\Phi \mid \Psi) = \frac{\Pr(\Phi \wedge \Psi)}{\Pr(\Psi)}$$

Values of Measures Characterizing Rules

St.	M	L	D
S1	g	b	**b**
S2	m	b	**m**
S3	m	m	**m**
S4	g	m	**g**
S5	g	g	**g**
S6	g	g	**g**
S7	b	b	**b**
S8	b	m	**b**

if M $\preceq$ bad ***then*** st $\preceq$ bad		$\|\|\Phi\|\|$	**S7,S8**	$\|\|\Psi\|\|$	**S1,S7,S8**	$\|\|\Phi\wedge\Psi\|\|$	**S7, S8**
sup	2	**σ**	2/8	**cer**	2/2	**cov**	2/3

if M $\succeq$ good $\wedge$ L $\succeq$ medium ***then*** st $\succeq$ good				$\|\|\Phi\|\|$	**S4,S5,S6**	
$\|\|\Psi\|\|$	**S4,S5,S6**			$\|\|\Phi\wedge\Psi\|\|$	**S4,S5,S6**	
sup	3	**σ**	3/8	**cer**	3/3	**cov** 3/3

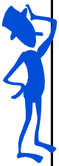
if L $\preceq$ medium ***then*** student $\preceq$ medium				$\|\|\Phi\|\|$	**S1,S2,S3,S4,S7,S8**	
$\|\|\Psi\|\|$	**S1,S2,S3,S7,S8**			$\|\|\Phi\wedge\Psi\|\|$	**S1,S2,S3,S7,S8**	
sup	5	**σ**	5/8	**cer**	5/6	**cov** 5/5

In general, we want to construct rules that:

- are supported by numerous objects (then, the discovered pattern is more universal)
- are (reasonably) certain (so that the conclusions follow the conditions with high certainty)
- have high coverage (so that the class (union) is described concisely, with a low number of rules)
- have few conditions (so that the rules are easy to understand)

Minimal Conjunction of Conditions and Local Coverage

B	a non-empty (lower or upper) approximation for $Cl_t^{\geq}$ (the set positive of objects) (can be also class union boundary or (lower or upper) approximation for $Cl_t^{\leq}$)
q	a condition attribute in C with domain V_q
$(q \succeq r_q)$	condition for condition part of the rule, $q \in C$, $r_q \in V_q$
$t = (q \succeq r_q)$	condition $(q \succeq r_q)$ such that there exists $x \in B : x_q = r_q$
$[t]$	a set of objects satisfying the condition $(q \succeq r_q)$
T	a set of pairs t (attribute, value) called conjunction, creating the rule's condition part: $(q_1 \succeq r_{q1}) \wedge (q_2 \succeq r_{q2}) \wedge \ldots \wedge (q_p \succeq r_{qp})$
$[T]$	a set of objects satisfying the conjunction T (a set of objects covered by T)



Set B depends on the conjunction T of pairs (attribute, value) *iff* $\emptyset \neq [T] = \cap_{t \in T} [t] \subseteq B$

Conjunction T is minimal *iff* B depends on T, and eliminating any pair t from T implies that B does not depend on T

$\boldsymbol{T}$ is a set of conjunctions T serving as a **local coverage of set B** *iff*

- Each conjunction $T \in \boldsymbol{T}$ is minimal
- $\cup_{T \in \boldsymbol{T}} [T] = B$ ($\boldsymbol{T}$ is complete covering all objects in B)
- An exclusion of any conjunction T from $\boldsymbol{T}$ makes it non-complete, *i.e.*, $\cup_{T \in \boldsymbol{T}} [T] \subset B$ ($\boldsymbol{T}$ is non-redundant; sometimes, formally defined as $\boldsymbol{T}$ has the smallest number of members)

The DOMLEM Algorithm - Pseudocode

Conjunction $T \in \boldsymbol{T}$ for $B = \underline{C}(Cl_t^{\geq})$ underlies a **certain** decision rule:

if $(q_1 \succeq r_{q1}) \wedge (q_2 \succeq r_{q2}) \wedge \ldots \wedge (q_p \succeq r_{qp})$, ***then*** x belongs to $Cl_t^{\geq}$

Conjunction $T \in \boldsymbol{T}$ for $B = \overline{C}(Cl_t^{\geq})$ underlies a **possible** decision rule:

if $(q_1 \succeq r_{q1}) \wedge (q_2 \succeq r_{q2}) \wedge \ldots \wedge (q_p \succeq r_{qp})$, ***then*** x possibly belongs to $Cl_t^{\geq}$

Conjunction $T \in \boldsymbol{T}$ for $B = Bn_C(Cl_t^{\geq})$ underlies an **approximate** decision rule:

if $(q_1 \succeq r_{q1}) \wedge \ldots \wedge (q_2 \succeq r_{q2}) \wedge (q_{k+1} \preceq r_{qk+1}) \wedge \ldots \wedge (q_p \preceq r_{qp})$, ***then*** x belongs to Cl_t or … or Cl_u

DOMLEM

- DOMLEM finds **a single local covering *T*** for set *B*, which can be lower or upper approximation of class union $Cl_t^{\geq}$ ($Cl_t^{\leq}$) or the boundary of $Cl_t^{\geq}$ ($Cl_t^{\leq}$)
- DOMLEM is run for a given set *B*, resulting in the minimal complete set of minimal decision rules covering all objects from set *B*
  - It is not tailored to the lower or upper approximations or boundaries
  - Focusing on inducing the set of conjunctions, it works, in the same way, irrespective of the interpretation of set *B*
  - The interpretation of rules (certain, possible, or approximate) depends on *B*

The DOMLEM Algorithm - Pseudocode

input: a set B, ***output***: a single local covering T of set B;
begin
$G := B$; (a set of objects not covered by the conjunction $T \in \mathbf{T}$
$\mathbf{T} := \emptyset$; (a set of conjunctions holding the current coverage of set $B \setminus G$)
while $G \neq \emptyset$ (there are still some objects to be covered)
  begin
    $T := \emptyset$; (conjunction of conditions; a candidate for condition part of the rule)
    $T(G) := \{t = (q \succeq r_q) : \exists x \in G: x_q = r_q\}$ (a set of candidate conditions for not covered objects)
    while $T = \emptyset$ ***or*** $[T] \not\subseteq B$ (the conjunction needs to cover only objects from B)
      begin
        select $t \in T(G)$ such that $|\{t\} \cap G| / |\{t\}|$ is maximum; if a tie occurs, select a pair such that $|\{t\} \cap G|$ is maximum; if another tie occurs, select the first pair;
        $T := T \cup t$; (add condition t to the conjunction T}
        $G := G \cap [t]$; (reduce the set of objects generating new elementary conditions)
        $T(G) := \{w = (q \succeq r_q) : \exists x \in G: x_q = r_q\}$; (update a set of candidate conditions)
        $T(G) := T(G) - T$; (eliminate the already used conditions from the list)
    end (while $[T] \not\subseteq B$)
    for each elementary condition $t \in T$ ***do***
      if $[T - \{t\}] \subseteq B$ ***then*** $T := T - t$; (eliminate redundant conditions)
    $\mathbf{T} := \mathbf{T} \cup T$; (add the minimal conjunction to the set of conjunctions)
    $G := B - \cup_{T \in \mathbf{T}} [T]$; (update the set of objects to be covered)
  end (***while*** $G \neq \emptyset$)
  for each conjunction $T \in \mathbf{T}$ ***do***
    if $\cup_{K \in \mathbf{T} \setminus T} [K] = B$ ***then*** $\mathbf{T} := \mathbf{T} - T$; (eliminate redundant conjunctions)
create rules R based on all conjunctions $T \in \mathbf{T}$;
end (procedure)

DOMLEM for Inducing Certain Rules For "At least good"

C-lower approximation of class "**at least good student**": B = {**S4**, **S5**, **S6**};

$G = B$ = {**S4**, **S5**, **S6**}; $\boldsymbol{T} := \varnothing$;

$T(G)$ = {M $\succeq$ good, L $\succeq$ med, L $\succeq$ good}

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	good	medium	**good**
S5	good	good	**good**
S6	good	good	**good**
S7	bad	bad	**bad**
S8	bad	medium	**bad**

	t	$\lvert[t] \cap G\rvert$	$\lvert[t]\rvert$	$\lvert[t] \cap G\rvert/\lvert[t]\rvert$
$T(G)$	M $\succeq$ good	3	4	3/4
	L $\succeq$ med	3	5	3/5
	L $\succeq$ good	2	2	2/2

We select t = (L $\succeq$ good); $T = T \cup t$ = {(L $\succeq$ good)}
with $[T]$ = {**S5, S6**} that is a subset of B.

We cannot eliminate any condition from conjunction T.

$\boldsymbol{T} := \boldsymbol{T} \cup T$ = {{(L $\succeq$ good)}};

$G := B - \cup_{T \in \boldsymbol{T}} [T]$ = {**S4, S5, S6**} – {**S5, S6**} = {**S4**}

$T(G)$ = {M $\succeq$ good, L $\succeq$ med}

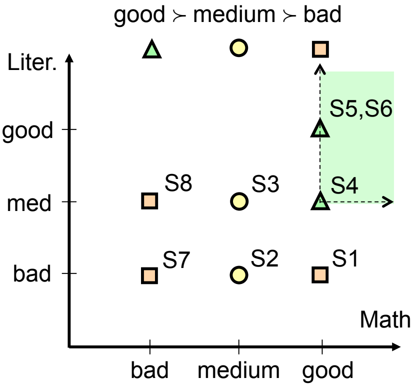
	t	$\lvert[t] \cap G\rvert$	$\lvert[t]\rvert$	$\lvert[t] \cap G\rvert/\lvert[t]\rvert$
$T(G)$	M $\succeq$ good	1	4	1/4
	L $\succeq$ med	1	5	1/5

We select t = (M $\succeq$ good); T = {(M $\succeq$ good)}
with $[T]$ = {**S1**, **S4**, **S5**, **S6**} that is not a subset of B.

$G = G \cap [t]$ = {**S4**}; $T(G)$ = {(L $\succeq$ med)}; *No choice.* $T = T \cup t$ = {(M $\succeq$ good), (L $\succeq$ med)}
with $[T]$ = {**S4, S5, S6**} that is a subset of B. We cannot eliminate any condition from conjunction T.

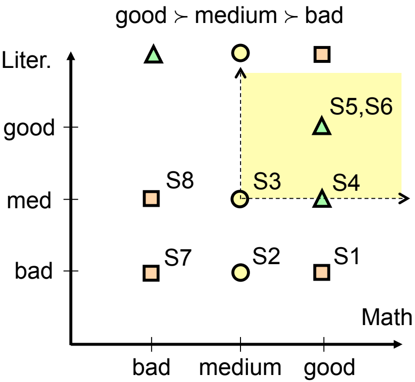
$\boldsymbol{T} := \boldsymbol{T} \cup T$ = {{(L $\succeq$ good)}, {(M $\succeq$ good), (L $\succeq$ med)}}
$G := B - \cup_{T \in \boldsymbol{T}} [T]$ = {**S4, S5, S6**} – {**S4, S5, S6**} = $\varnothing$. Terminate. We can eliminate the first conjunction from $\boldsymbol{T}$. Terminate with rule: "***if*** (M $\succeq$ good) $\wedge$ (L $\succeq$ med), ***then*** student at least good".

Example Certain Rules For Upward Class Unions



$$\underline{P}\left(Cl_{good}^{\geq}\right) = \{S4,\ S5,\ S6\}$$

if M $\succeq$ good $\wedge$ L $\succeq$ medium,
then student $\succeq$ **good**



$$\underline{P}\left(Cl_{medium}^{\geq}\right) = \{S3,\ S4,\ S5,\ S6\}$$

if M $\succeq$ medium $\wedge$ L $\succeq$ medium,
then student $\succeq$ **medium**

DOMLEM for Inducing Certain Rules For "At most med."

C-lower approximation of class "**at most medium student**": B = {**S1**, **S2**, **S3**, **S7**, **S8**}

$\boldsymbol{T} := \varnothing$; $G = B$ = {**S1**, **S2**, **S3**, **S7**, **S8**}; $T(G)$ = {M $\preceq$ good, M $\preceq$ med, M $\preceq$ bad, L $\preceq$ med, L $\preceq$ bad}

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	good	medium	**good**
S5	good	good	**good**
S6	good	good	**good**
S7	bad	bad	**bad**
S8	bad	medium	**bad**

	t	$\|[t] \cap G\|$	$\|[t]\|$	$\|[t] \cap G\|/\|[t]\|$
$T(G)$	M $\preceq$ good	5	8	5/8
	M $\preceq$ med	4	4	4/4
	M $\preceq$ bad	2	2	2/2
	L $\preceq$ med	5	6	5/6
	L $\preceq$ bad	3	3	3/3

We select t = (M $\preceq$ med)*;* $T = T \cup t$ = {(M $\preceq$ med)}
with $[T]$ = {**S2**, **S3**, **S7**, **S8**} that is a subset of B.
We cannot eliminate any condition from conjunction T.

$\boldsymbol{T} := \boldsymbol{T} \cup T$ = {{(M $\preceq$ med)}};

$G := B - \cup_{T \in \boldsymbol{T}} [T]$ = {**S1**, **S2**, **S3**, **S7**, **S8**} – {**S2**, **S3**, **S7**, **S8**} = {**S1**}

$T(G)$ = {M $\preceq$ good, L $\preceq$ bad}

	t	$\|[t] \cap G\|$	$\|[t]\|$	$\|[t] \cap G\|/\|[t]\|$
$T(G)$	M $\preceq$ good	1	8	1/8
	L $\preceq$ bad	1	3	1/3

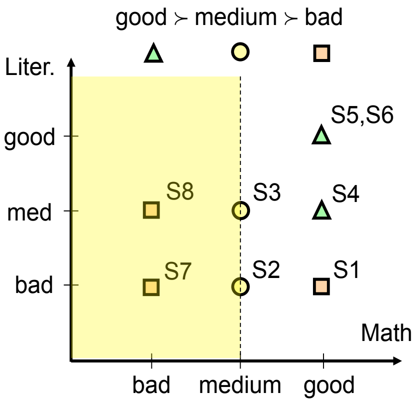
We select t = (L $\preceq$ bad)*;* $T = T \cup t$ = {(L $\preceq$ bad)}
with $[T]$ = {**S1**, **S2**, **S7**} that is a subset of B.
We cannot eliminate any condition from conjunction T.

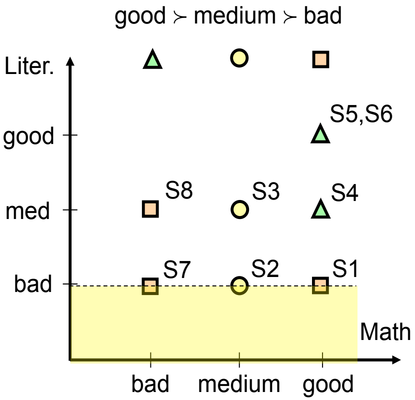
$\boldsymbol{T} := \boldsymbol{T} \cup T$ = {{(M $\preceq$ med)}, {(L $\preceq$ bad)}};

$G := B - \cup_{T \in \boldsymbol{T}} [T]$ = {**S1**, **S2**, **S3**, **S7**, **S8**} – {**S1**, **S2**, **S3**, **S7**, **S8**} = $\varnothing$. Terminate.

We cannot eliminate any conjunction from $\boldsymbol{T}$. Terminate with rules:
"***if*** (M $\preceq$ med), ***then*** student at most medium" and "***if*** (L $\preceq$ bad), ***then*** student at most medium".

Example Certain Rules For Downward Class Unions



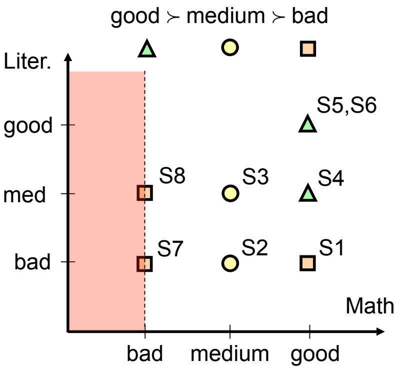


$$\underline{P}\left(Cl^{\leq}_{medium}\right) = \{S1, S2, S3, S7, S8\}$$

if M $\preceq$ medium, ***then*** student $\preceq$ **medium**

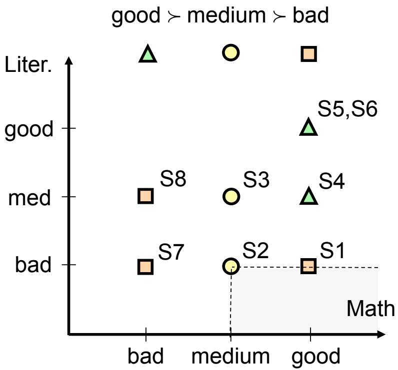
if L $\preceq$ bad, ***then*** student $\preceq$ **medium**

Example Certain and Approximate Rules



$$\underline{P}\left(Cl^{\leq}_{bad}\right) = \{S7,\ S8\}$$

if M $\preceq$ bad, ***then*** student $\preceq$ **bad**



$$Bn_P\left(Cl^{\geq}_{medium}\right) = \{S1,\ S2\} \quad Bn_P\left(Cl^{\leq}_{bad}\right) = \{S1,\ S2\}$$

if M $\succeq$ medium $\wedge$ L $\preceq$ bad,

then student **bad** *or* **medium**

All Possible and Approximate Rules

St.	M	L	D
S1	good	bad	**bad**
S2	medium	bad	**medium**
S3	medium	medium	**medium**
S4	good	medium	**good**
S5	good	good	**good**
S6	good	good	**good**
S7	bad	bad	**bad**
S8	bad	medium	**bad**

The minimal set of possible rules generated by DOMLEM

if M $\succeq$ good $\wedge$ L $\succeq$ medium, ***then*** *possibly* student $\succeq$ good	**S4, S5, S6**
if M $\succeq$ medium, ***then*** *possibly* student $\succeq$ **medium**	**S1, S2, S3, S4, S5, S6**
if M $\preceq$ medium, ***then*** *possibly* student $\preceq$ **medium**	**S2, S3, S7, S8**
if L $\preceq$ bad, ***then*** *possibly* student $\preceq$ **medium**	**S1, S2, S7**
if L $\preceq$ bad, ***then*** *possibly* student $\preceq$ **bad**	**S1, S2, S7**
if M $\preceq$ bad, ***then*** *possibly* student $\preceq$ **bad**	**S7, S8**

The minimal set of approximate rules generated by DOMLEM

if M $\succeq$ medium $\wedge$ L $\preceq$ bad, ***then*** student **bad** *or* **medium**	**S1, S2**

$\overline{P}\left(Cl^{\geq}_{good}\right) = \{S4, S5, S6\}$ $\overline{P}\left(Cl^{\geq}_{medium}\right) = \{S1, S2, S3, S4, S5, S6\}$ $\overline{P}\left(Cl^{\leq}_{medium}\right) = \{S1, S2, S3, S7, S8\}$

$\overline{P}\left(Cl^{\leq}_{bad}\right) = \{S1, S2, S7, S8\}$ $Bn_P\left(Cl^{\geq}_{medium}\right) = \{S1, S2\}$ $Bn_P\left(Cl^{\leq}_{bad}\right) = \{S1, S2\}$

Quality of Approximation of Classification

For every $P \subseteq C$, the objects being consistent in the sense of the dominance principle with all upward and downward class unions are called *P*-correctly classified

Quality of approximation *of classification* $\mathbf{Cl}=\{Cl_t,\ t=1,\dots m\}$ by criteria $P \subseteq C$ measures the quality of approximation of all decision classes on the universe *U*

- Measure of precision of the entire decision table (the quality of knowledge that can be extracted from the decision table)
- The ratio between the number of *P*-correctly classified objects and the number of all objects

St.	M	L	D
S1	g	b	**b**
S2	m	b	**m**
S3	m	m	**m**
S4	g	m	**g**
S5	g	g	**g**
S6	g	g	**g**
S7	b	b	**b**
S8	b	m	**b**

$$\gamma_P(\mathbf{Cl}) = \frac{\left|U - \bigcup_{t\in\{2,\dots,m\}} Bn_P\left(Cl_t^{\geq}\right)\right|}{|U|} = \frac{\left|U - \bigcup_{t\in\{1,\dots,m-1\}} Bn_P\left(Cl_t^{\leq}\right)\right|}{|U|}$$

$$Bn_P\left(Cl_{medium}^{\leq}\right) = Bn_P\left(Cl_{good}^{\geq}\right) = \varnothing \qquad Bn_P\left(Cl_{bad}^{\leq}\right) = Bn_P\left(Cl_{medium}^{\geq}\right) = \{\mathrm{S1}, \mathrm{S2}\}$$

$$\gamma_P(\mathbf{Cl}) = \frac{8-2-0}{8} = \frac{6}{8}$$

Quality of approximation will be essential for defining the concept of reduct

Slightly Greater Example

St.	Math	Physics	Literat.	*D*
S1	good	medium	bad	**bad**
S2	medium	medium	bad	**medium**
S3	medium	medium	medium	**medium**
S4	good	good	medium	**good**
S5	good	medium	good	**good**
S6	good	good	good	**good**
S7	bad	bad	bad	**bad**
S8	bad	bad	medium	**bad**

Quality of approximation $\gamma_P(\mathbf{Cl}) = \frac{6}{8}$

$$P = \{\mathrm{M}, \mathrm{Ph}, \mathrm{L}\}$$

$$\underline{P}\left(Cl^{\geq}_{good}\right) = \{\mathrm{S4}, \mathrm{S5}, \mathrm{S6}\}$$

$$\underline{P}\left(Cl^{\geq}_{medium}\right) = \{\mathrm{S3}, \mathrm{S4}, \mathrm{S5}, \mathrm{S6}\}$$

$$\underline{P}\left(Cl^{\leq}_{bad}\right) = \{\mathrm{S7}, \mathrm{S8}\}$$

$$\underline{P}\left(Cl^{\leq}_{medium}\right) = \{\mathrm{S1}, \mathrm{S2}, \mathrm{S3}, \mathrm{S7}, \mathrm{S8}\}$$

$$Bn_P\left(Cl^{\leq}_{bad}\right) = Bn_P\left(Cl^{\geq}_{medium}\right) = \{\mathrm{S1}, \mathrm{S2}\}$$

$$Bn_P\left(Cl^{\leq}_{medium}\right) = Bn_P\left(Cl^{\geq}_{good}\right) = \varnothing$$

Monotonicity Property

Monotonicity property with respect to the cardinality of $P' \subseteq P \subseteq C$

- With the constrained set of criteria, the lower approximations do not become greater (they may become narrower), and the upper approximations do not become narrower (may become more extensive)

$$\underline{P'}(X) \subseteq \underline{P}(X)$$

$$\overline{P'}(X) \supseteq \overline{P}(X)$$

St.	M	Ph	L	*D*
S1	g	m	b	**b**
S2	m	m	b	**m**
S3	m	m	m	**m**
S4	g	g	m	**g**
S5	g	m	g	**g**
S6	g	g	g	**g**
S7	b	b	b	**b**
S8	b	b	m	**b**

$$P'' = \{\text{M}\}$$

$$\underline{P''}(Cl^{\geq}_{good}) = \emptyset$$

$$\underline{P''}(Cl^{\geq}_{medium}) = \emptyset$$

$$\underline{P''}(Cl^{\leq}_{bad}) = \{\text{S7, S8}\}$$

$$\underline{P''}(Cl^{\leq}_{medium}) = \{\text{S2, S3, S7, S8}\}$$

$$\gamma_{P''}(\mathbf{Cl}) = \frac{2}{8}$$

$$P' = \{\text{M, Ph}\}$$

$$\underline{P'}(Cl^{\geq}_{good}) = \{\text{S4, S6}\}$$

$$\underline{P'}(Cl^{\geq}_{medium}) = \{\text{S4, S6}\}$$

$$\underline{P'}(Cl^{\leq}_{bad}) = \{\text{S7, S8}\}$$

$$\underline{P'}(Cl^{\leq}_{medium}) = \{\text{S2, S3, S7, S8}\}$$

$$\gamma_{P'}(\mathbf{Cl}) = \frac{4}{8}$$

$$P = \{\text{M, Ph, L}\}$$

$$\underline{P}(Cl^{\geq}_{good}) = \{\text{S4, S5, S6}\}$$

$$\underline{P}(Cl^{\geq}_{medium}) = \{\text{S3, S4, S5, S6}\}$$

$$\underline{P}(Cl^{\leq}_{bad}) = \{\text{S7, S8}\}$$

$$\underline{P}(Cl^{\leq}_{medium}) = \{\text{S1, S2, S3, S7, S8}\}$$

$$\gamma_{P}(\mathbf{Cl}) = \frac{6}{8}$$

Reduct

- We often face the question of whether we can remove some data from a data table and still, preserve its basic properties
  - *Are there criteria that are more important to the knowledge represented in granules of knowledge*?
- Let *C, D* ⊆ Q be sets of condition and decision attributes
- We say that *P'* ⊆ *C is a D-**reduct*** (**reduct with respect to *D***) of *C* ($RED_D(C)$), if *P'* is a **minimal subset** of *C* such that the **quality of approximation for *P'* is the same as for *C***

$$\gamma_{P'}(Cl) = \gamma_C(Cl)$$

- The quality of approximation needs to be **maintained the same as for the entire set of condition attributes**
  - Quality of approximation depends on the cardinalities of lower approximations
  - The dominance cones may be modified, but the lower approximations cannot
- Minimal subset *P'* ⊆ *C* means that there is no proper subset of *P'* (*P"* ⊂ *P'*) for which the property would be satisfied, i.e., **no criterion can be removed from the reduct without changing lower approximations**
- It means that for $C = \{C_1, C_2, C_3, C_4, C_5\}$, it is possible that $\{C_1, C_2\}$ and $\{C_2, C_3, C_4\}$ are reducts (if only none of their proper subsets maintains the quality of approximation)

Reducts - Example

- The basic algorithm for identifying the reducts is called an **additive filter**
  - We start with singletons = subsets composed of individual criteria
  - In the case some reduct is identified, we eliminate all its supersets from the list of potential candidates (reduct needs to be minimal!)
  - We continue with pairs, triplets, …, until the list of candidates is empty
- There is always at least one reduct (in the worst case, it is equal to the set of all condition attributes *C*)
- The problem of finding the minimal subsets (**reducts**) of criteria that can describe all the class unions in the given data set is **NP-hard**

St.	M	Ph	L	*D*
S1	g	m	b	**b**
S2	m	m	b	**m**
S3	m	m	m	**m**
S4	g	g	m	**g**
S5	g	m	g	**g**
S6	g	g	g	**g**
S7	b	b	b	**b**
S8	b	b	m	**b**

$C = P = \{\mathrm{M, Ph, L}\}$ $\gamma_P(\mathbf{Cl}) = \gamma_C(\mathbf{Cl})$ $\boxed{\gamma_P(\mathbf{Cl}) = 6/8}$

Are there any additional troublesome dominance relations?

$\boxed{\gamma_M(\mathbf{Cl}) = 2/8}$ e.g., (S1, S3)

$\boxed{\gamma_{Ph}(\mathbf{Cl}) = 4/8}$ e.g., (S1, S3)

$\boxed{\gamma_L(\mathbf{Cl}) = 2/8}$ e.g., (S2, S7)

$\boxed{\gamma_{M,Ph}(\mathbf{Cl}) = 4/8}$ e.g., (S1, S5)

Non-reducts = do not maintain the quality of approximation

$\boxed{\gamma_{M,L}(\mathbf{Cl}) = 6/8}$ without Phys. not more inconsistencies

$\boxed{\gamma_{Ph,L}(\mathbf{Cl}) = 6/8}$ without Math not more inconsistencies

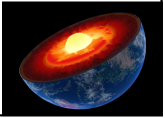
Reducts = maintain the quality of approximation and are minimal

$\boxed{\gamma_{M,Ph,L}(\mathbf{Cl}) = 6/8}$

Non-reduct = it is not minimal

Core

- The reduct of an information system may not be unique
- There may be many sets of attributes that preserve the quality of approximation
- *D*-**core** (**core with respect to *D***) is the **intersection of all the *D*-reducts**
  - Each element of the core belongs to some reduct
- The core is the most important subset of criteria since none of its elements can be removed without affecting the classification power of attributes
  - The core may be thought of as the sets of necessary attributes (necessary for the class structure to be represented)
- When there is only one reduct, it is equivalent to the core
- The core can be empty (e.g., for $C = \{C_1, C_2, C_3, C_4, C_5\}$ with two reducts $\{C_1, C_2\}$ and $\{C_3, C_4, C_5\}$, their intersection is empty)



$$\boxed{CORE_D(C) = \bigcap RED_D(C)}$$

$P = \{\text{M, Ph, L}\}$

$RED_D^1(P) = \{\text{M, L}\}$

$RED_D^2(P) = \{\text{Ph, L}\}$

Intersection of reducts {M, L} and {Ph, L} gives the core {L}

$CORE_D(P) = \{\text{L}\}$

Classification - Matching Rules

St.	M	Ph	L	*D*
S1	g	m	b	**b**
S2	m	m	b	**m**
S3	m	m	m	**m**
S4	g	g	m	**g**
S5	g	m	g	**g**
S6	g	g	g	**g**
S7	b	b	b	**b**
S8	b	b	m	**b**

A set of decision rules in terms of *C*={M, Ph, L} representing preferences is more concise (e.g., less rules for class union "at most medium"):

R1: ***if*** M $\succeq$ good & L $\succeq$ medium, ***then*** student $\succeq$ good {S4,S5,S6}

R2: ***if*** M $\succeq$ medium & L $\succeq$ medium, ***then*** student $\succeq$ medium {S3,S4,S5,S6}

R3: ***if*** Ph $\preceq$ medium & L $\preceq$ medium ***then*** student $\preceq$ medium {S1,S2,S3,S7,S8}

R4: ***if*** M $\preceq$ bad, ***then*** student $\preceq$ bad {S7,S8}

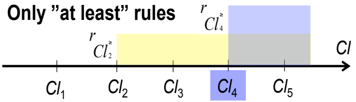
R5: ***if*** M $\succeq$ medium & L $\preceq$ bad, ***then*** student bad or medium {S1,S2}

St.	M	Ph	L	*D*
S9	g	m	g	**?**
S10	b	m	b	**?**
S11	m	m	g	**?**
S12	g	m	m	**?**

St.	Set of activated decision rules	Set of non-activated decision rules
S9	**R1:** $\succeq$ good, **R2:** $\succeq$ medium	**R3**, **R4**, **R5**
S10	**R3:** $\preceq$ medium, **R4:** $\preceq$ bad	**R1**, **R2**, **R5**
S11	**R2:** $\succeq$ medium, **R3:** $\preceq$ medium	**R1**, **R4**, **R5**
S12	**R1:** $\succeq$ good, **R2:** $\succeq$ medium, **R3:** $\preceq$ medium	**R4**, **R5**

Classification - Standard Algorithm

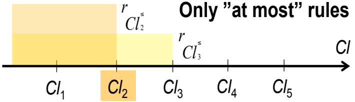
Only "at least" rules



The lowest class of upward union of recommendations

St.	Set of activated decision rules	Recommendation
S9	**R1:** $\succeq$ good, **R2:** $\succeq$ medium	good

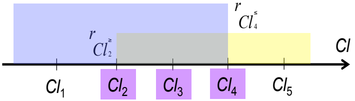
Only "at most" rules



The highest class of downward union of recommendations

St.	Set of activated decision rules	Recommendation
S10	**R3:** $\preceq$ medium, **R4:** $\preceq$ bad	bad

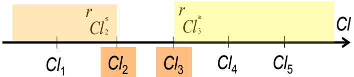
Overlaping "at least" and "at most" rules



The range between the lowest class of upward union of recommendations and the highest class of downward union of recommendations

St.	Set of activated decision rules	Recommendation
S11	**R2:** $\succeq$ medium, **R3:** $\preceq$ medium	[medium, medium]

Disjoint "at least" and "at most" rules



The range between the lowest class of upward union of recommendations and the highest class of downward union of recommendations

St.	Set of activated decision rules	Recommendation
S12	**R1:** $\succeq$ good, **R2:** $\succeq$ medium, **R3:** $\preceq$ medium	[medium, good]

Classification - Advanced Algorithm

Consider **rules** $\varphi_1 \rightarrow \psi_1, \ldots, \varphi_k \rightarrow \psi_k$ matching alternative x

$$\left(\psi_j = Cl_s^{\geq} \ \text{ or } \ \psi_j = Cl_q^{\leq}, \ s,q \in \{1,\ldots,m\}, \ j = 1,\ldots,k\right)$$

$R_t(x)=\{j\!: \ Cl_t \in \psi_j, j=1,\ldots,k\}$, $\quad R_{\neg t}(x)=\{j\!: \ Cl_t \notin \psi_j, j=1,\ldots,k\}$

$\|\varphi_j\|, \|\psi_j\|$ are sets of alternatives with property φ_j (condition part), ψ_j (decision part), $j=1,\ldots,k$

For classified alternative x, the score is calculated for each candidate class Cl_t, $t=1,\ldots,m$

$$\boxed{score(Cl_t, x) = score^+(Cl_t, x) - score^-(Cl_t, x)}$$

The **positive** and **negative scores** represent the credibility and relative strength of all rules covering object x and suggesting its assignment to class Cl_t or class different than Cl_t

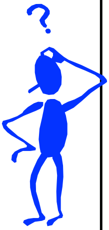
$$score^+(Cl_t, x) = \frac{\left|\bigcup_{j \in R_t(x)} \left(\|\varphi_j\| \cap Cl_t\right)\right|^2}{\left|\bigcup_{j \in R_t(x)} \|\varphi_j\|\right| \times |Cl_t|} \qquad score^-(Cl_t, x) = \frac{\left|\bigcup_{j \in \neg R_t(x)} \left(\|\varphi_j\| \cap \|\psi_j\|\right)\right|^2}{\left|\bigcup_{j \in \neg R_t(x)} \|\varphi_j\|\right| \times \left|\bigcup_{j \in \neg R_t(x)} \|\psi_j\|\right|}$$

Recommendation: $x \rightarrow Cl_t$ where $Cl_t = \underset{t \in \{1,\ldots,m\}}{\arg\max}\left(score(Cl_t, x)\right)$

J. Błaszczyński, S. Greco, R. Słowiński: Multi-criteria classification – a new scheme for application of dominance-based decision rules. *European Journal Operational Research*, 181 (2007) 1030-1044

Dominance-based Rough Set Approach - Summary (1)

DRSA



Preference information:

- Indirect: a set of classification examples on a subset of reference objects (alternatives) described in terms of a set of criteria
- Often, information comes from available data (observations, situations) rather than DM's indirect holistic judgments
- In general, the number of decision attributes can be greater than one

Preference model:

- A set of *if* … *then* … decision rules capturing classification patterns in data
- Easy-to-understand formulation and straightforward interpretation of results

Technique:

- Efficient rule induction algorithms for finding hidden patterns in data
- A minimal set of minimal certain, possible, and/or approximate decision rules based on the lower and upper approximations of class unions or class boundaries

Recommendation:

- Class assignments for all alternatives, including those unseen at the stage of data structuring and rule induction (i.e., non-reference alternatives)

S. Greco, B. Matarazzo, R. Słowiński: Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129 (2001) no.1, 1-47

Dominance-based Rough Set Approach - Summary (2)

- Accepting preference information in the form of **classification examples**
- **No need for preliminary or additional information** about data like probability distribution, external parameters, models, functions, grades, or subjective memberships
- **Structuring** potentially **inconsistent data**
- Characterization of decision classes (even in case of inconsistency) in terms of chosen criteria by lower and upper approximations
- A measure of the **quality of approximation** indicating how good the chosen set of criteria is for approximating the classification
- **Reduction of knowledge** contained in the table to the description by relevant criteria belonging to reducts (feature reduction, where criteria that do not contribute to the classification of the given set can be identified and removed)
- The core of attributes indicating **indispensable criteria**
- Decision rules are the only aggregation operators that: give an account of the most complex **interactions** among criteria, **accept ordinal evaluation scales**, and do not convert ordinal evaluations into cardinal ones
- Particularly useful for **feature selection, feature extraction, data reduction, decision rule induction, and pattern extraction**

DRSA - Example Applications

- **Mobile Emergency Triage** (MET) **System**: triage disposition for presentations of acute pain (abdominal and scrotal pain, hip pain) with or without complete clinical information: observation, triage, or discharge (Michalowski et al., 2005)
- **Medical applications**: complications after open-heart operations, colon cancer surgery, prostate cancer treatment, breast cancer treatment, pediatric hip surgery
- **Airline market**: rules determining customer attitudes and loyalties; develop strategies to acquire customers / retain highly valued ones (Liou and Tzeng, 2010)
- **Fraud prediction** and the development of a fraud prevention system in auto loan applications (Błaszczyński et al., 2021)
- **Consumer attitude analysis** to insects as a novel food (Roma et al., 2020)