Interpretability and Explainability Facets of Data Analytics: Symbols and Information Granules

Witold Pedrycz

Department of Electrical & Computer Engineering
University of Alberta, Edmonton, AB, Canada
and
Systems Research Institute
Polish Academy of Sciences, Warsaw, Poland

wpedrycz@ualberta.ca

Agenda

Introductory notes: interpretability and explainability

Levels of abstraction, information granules

Principles of Granular Computing

Design of information granules

Rule-based architectures and hierarchy of interpretability

Relational factorization

Design and analysis of concepts

Concluding comments

Explainability

Explainability and interpretabilty facets -motivation

Human-centricity of constructs
General Data Protection Regulation (EU) – the right of explanation; understand the behavior of AI systems

Decision processes realized in autonomous environments and of high level of criticality

Awareness of possible adversarial samples used to deceive AI systems
Adversarial machine learning

Interpretability and explainability (1)

Interpretability

Formation of results that are easily comprehended by the user resulting in semantically sound quantified and actionable findings.

Examples:

low value of sales

high likelihood of temperature in range [10, 12C]

…

Level of abstraction

Interpretability and explainability (2)

Explainability

modeling faculties to:

produce knowledge about relationship existing in data/models and help explain and audit prediction/classification results in response to issues of regulatory or fairness nature

support “what-if” analysis.

support traceability of the reasoning (inference) process.

why did the model produce a particular prediction?

why weren’t other decisions made?

Abstraction and information granularity as a prerequisite

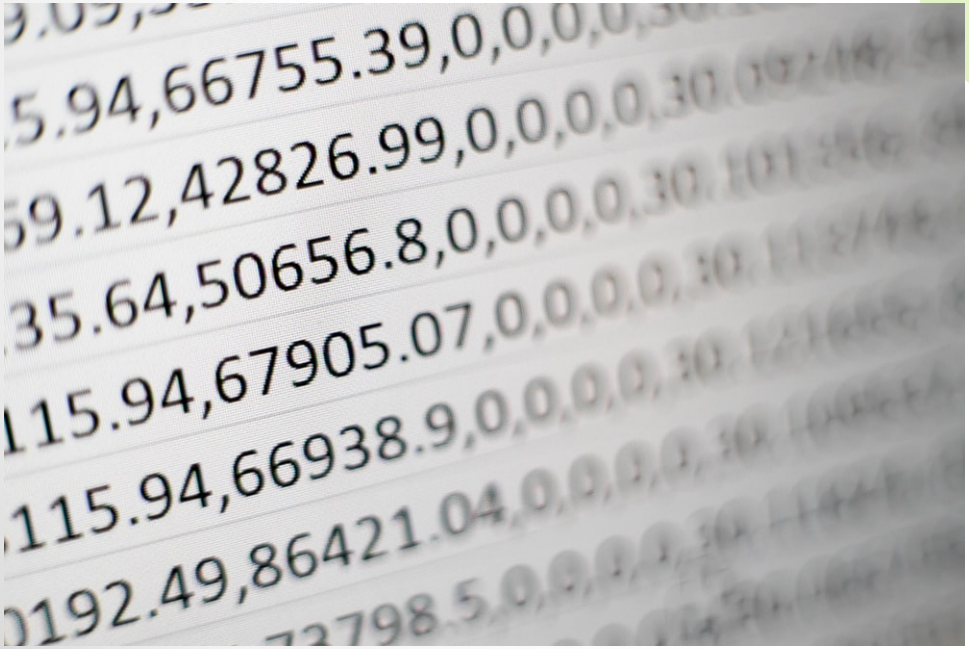
Flexibility in coping with overwhelmed amount of details

View at systems at suitable level of abstraction (granularity)

Conciseness and stability of results

Logic fabric of resulting constructs

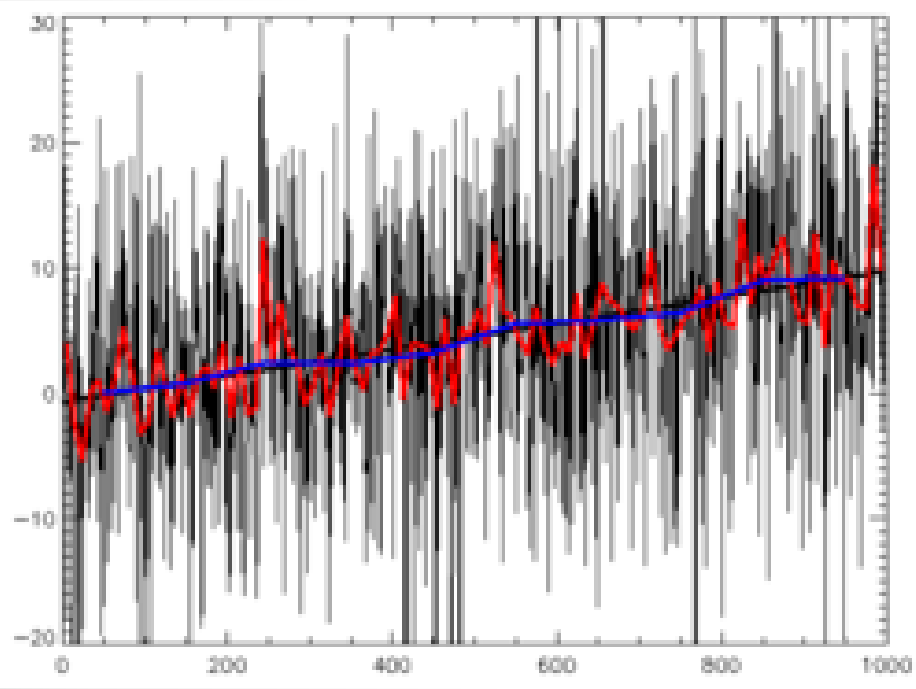
Abstraction: from data to information granules



low values reported in recent weeks

Rapidly changing trend

Abstraction: from data to information granules



High oscillations of signal

slowly increasing trend

From data to information granules: blind men and an elephant

Information granules

Realization of abstraction at a suitable level

Coping with problem complexity

Various formal frameworks and related processing environments

Information granularity

Our ability to conceptualize the world at different granularities and to switch among these granularities is fundamental to our intelligence and flexibility.

It enables us to map the complexities of the world around us into simple theories that are computationally tractable to reason in.

J. R. Hobbs, Proc. IJCAI, 1985

Information granules: formalization

Plethora of formal settings:

Sets, interval computing
Fuzzy sets
Probabilities
Rough sets
Random sets
shadowed sets
Intuitionistic sets
Soft sets
Hesitant sets
Probabilistic sets
Information granules of higher order and higher type
...

associated formal processing frameworks

Information granules: key characteristics and challenges

Variable level of abstraction (granularity, specificity)

Representation of original data and assessment of quality of representation (granulation-degranulation)

Reflection of semantics of data

Reliance on data and accommodation of auxiliary source of knowledge

Information granules: generative and discriminative aspects

Generative models

description of structure of data, knowledge representation, can engage auxiliary sources of knowledge

Discriminative models

formation of classifiers, predictors, association models

Design of information granules

Clustering

From data to clusters and information granules

Results of clustering – partition matrix, prototypes, proximity matrix...

Principle of justifiable granularity

Principle of justifiable information granularity

Construct information granule so that is it

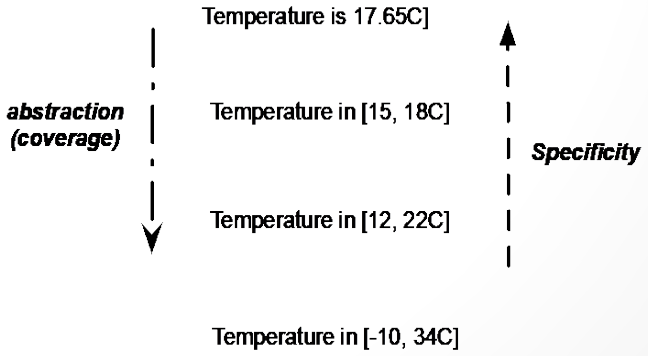
experimentally justified – legitimized by available experimental evidence (e.g., numeric data)

semantically sound – exhibits clearly defined meaning (semantics)- can be easily interpreted and becoming distinguishable from others.

Principle of justifiable granularity: Experimental justification and semantic soundness

coverage $cov(Y) = \text{card}\{y_k \mid y_k \in Y\}/n$

specificity $\text{sp}(Y) = \text{g}(\text{length}(Y)) = 1 - \frac{|bound_2 - - bound_1|}{range}$



$V = cov(B_i)sp(B_i)$

determine information granule so that V becomes maximized

Principle of justifiable granularity: one-dimensional case

One-dimensional data: $x_1, x_2, ..., x_N$

Construct information granule A =[a, b]

Coverage. cov(A)= card{x_k | x_k in [a, b]}

Specificity sp(A) = g(size(A)) g – monotonically decreasing function g({z})=1

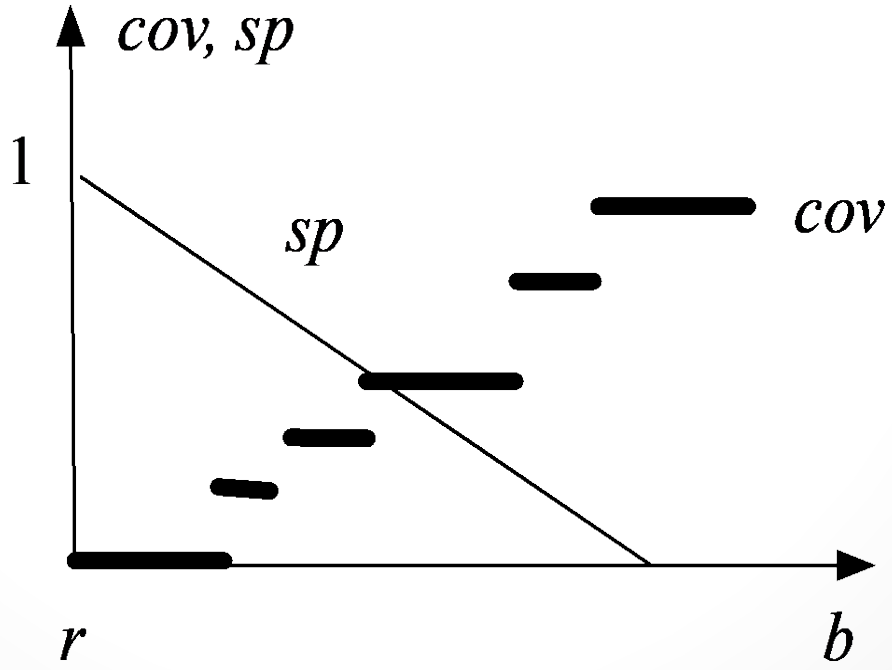
Two phase development:

1.Numeric representative of data, v

2. Optimization of a and b; $\max_b$[cov([v,b]sp([v,b])

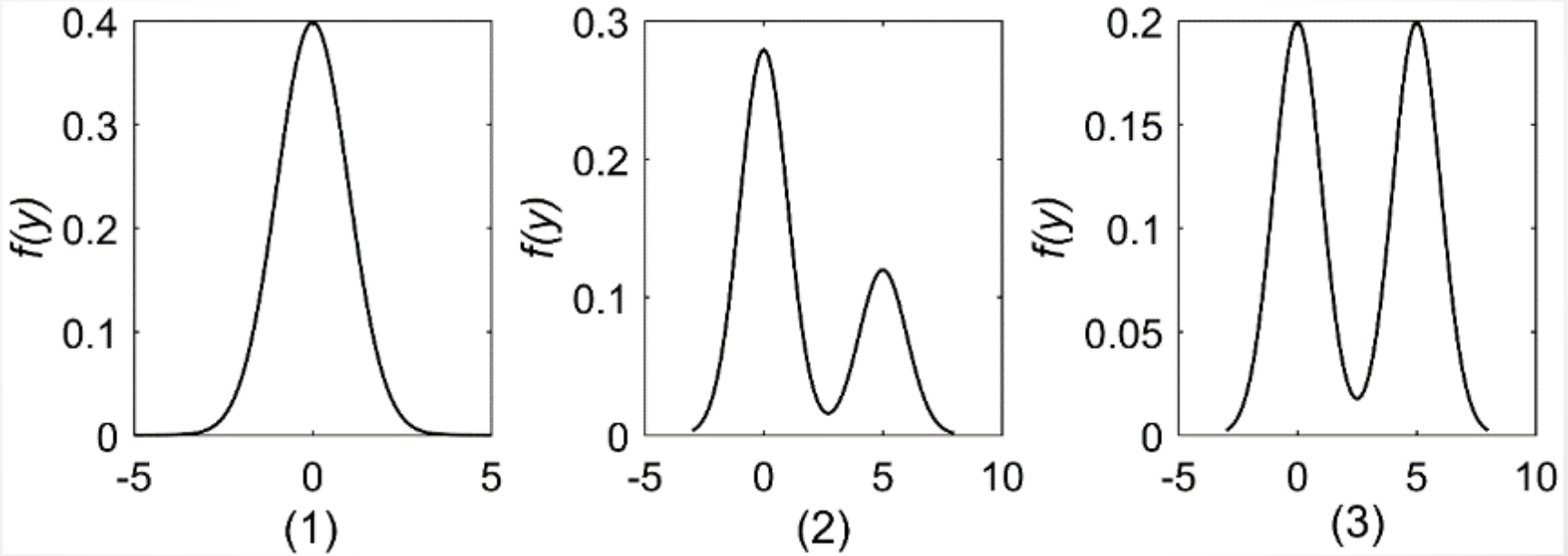
$\max_a$[cov([a,v]sp([a,v])

Principle of justifiable granularity: optimization



Principle of justifiable granularity: example (1)

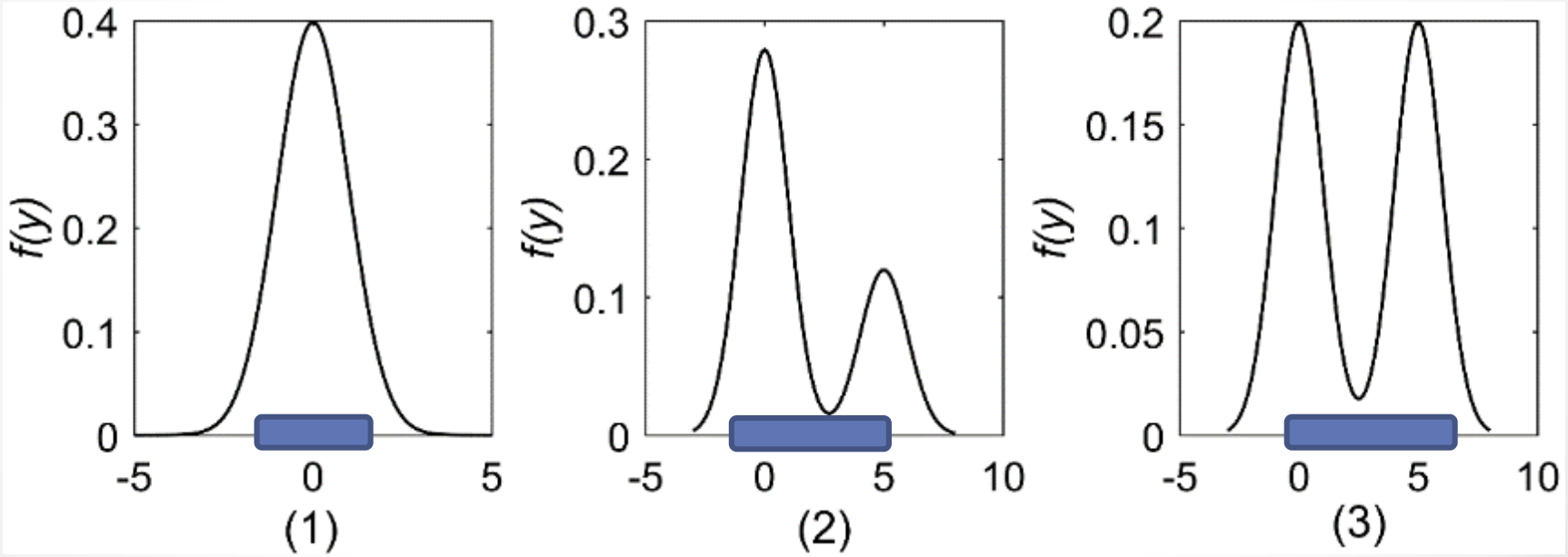
normally distributed data N(0,1) and N(5, 1)



N(0,1) 0.7N(0,1)+0.3N(5,1) 0.5N(0,1) +0.5N(5,1)

Principle of justifiable granularity: example (2)

normally distributed data N(0,1) and N(5, 1)



V =0.47

V = 0.37

V =0.36

Principle of justifiable granularity: generalizations

information granules for multidimensional data

auxiliary sources of knowledge
(classification and regression problems)

adversarial environment

collaborative development

Principle of justifiable granularity: multidimensional data

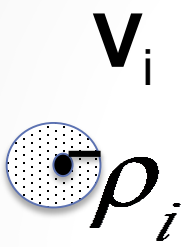
Coverage $\mathrm{cov}(\rho) = card\{\boldsymbol{x}_k \,|||\, \boldsymbol{x}_k - \boldsymbol{v}_i \,||^2 \le n\rho^2\}$

$$\mathrm{cov}(\rho) = \sum_{k=1:\|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2 \le n\rho^2}^{N} u_{ik}$$

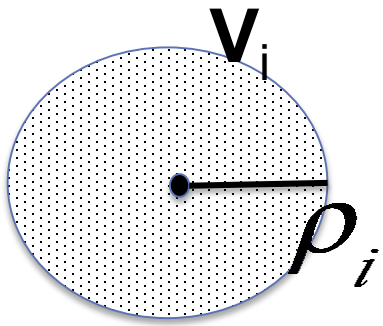
Specificity $sp(\rho) = 1 - \rho$

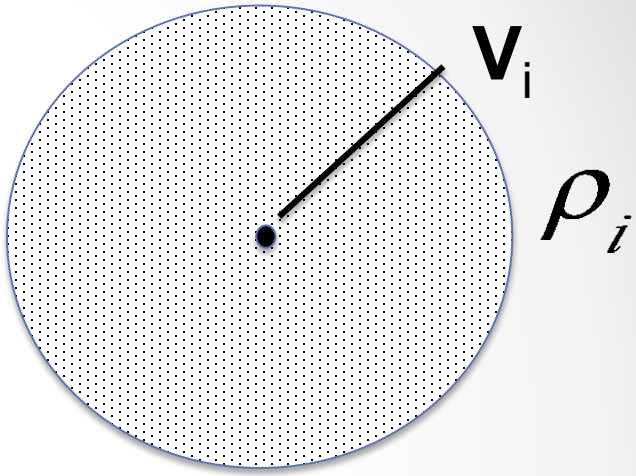
$$\rho_{opt} = \arg Max_{\rho}[\mathrm{cov}(\rho) * sp(\rho)]$$

Coverage and semantics



Low coverage
High specificity





High coverage
Low specificity

Ellipsoidal information granules

$$\| \boldsymbol{x} - \boldsymbol{v}_i \|^2_{M_i} = (\boldsymbol{x} - \boldsymbol{v}_i)^T M_i (\boldsymbol{x} - \boldsymbol{v}_i)$$

Mahalanobis distance

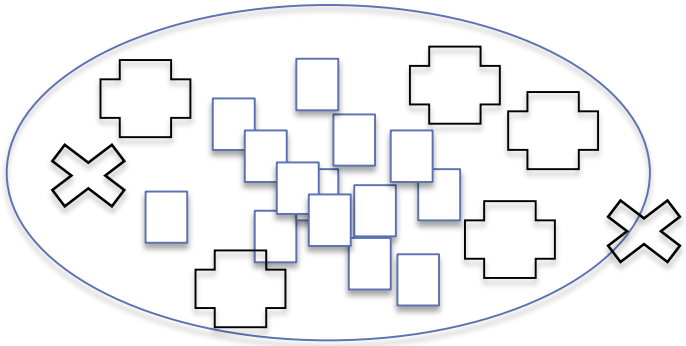
$$\mathrm{cov}(\rho_i) = card\{\boldsymbol{x}_k \,|\| \boldsymbol{x}_k - \boldsymbol{v}_i \|^2_{M_i} \le n\rho^2\}$$

$$sp(\rho_i) = 1 - \rho_i$$

$$\rho_{i,opt} = \arg_{\rho_i}[\mathrm{cov}(\rho_i) * sp(\rho_i)]$$

Accommodation of discriminatory type of knowledge

Classification problem- entropy of patterns forming information granule



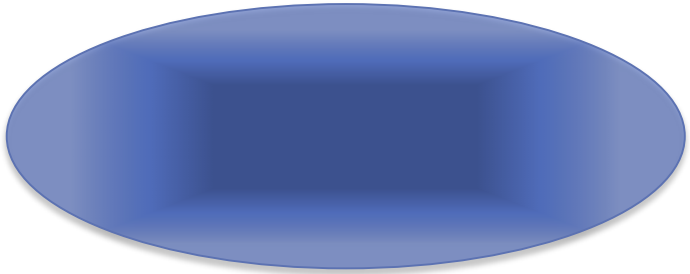
$$\mathrm{cov}(\rho) = \sum_{k=1:\|x_k - v_i\|^2 \le \rho^2} u_{ik} \qquad \mathrm{cov}(\rho) = card\{x_k \,|\|\, x_k - v_i \,\|^2 \le \rho^2\}$$

Modified coverage $\mathrm{cov}'(\rho) = \mathrm{cov}(\rho) * (1 - h(\omega))$

h(.) – entropy function

Incorporation of discriminatory type of knowledge

Regression problem- involvement of output variable



$$\mathrm{cov}(\rho) = \sum_{k=1:\|x_k - v_i\|^2 \le \rho^2}^{N} u_{ik} \qquad \mathrm{cov}(\rho) = card\{x_k \,|\|\, x_k - v_i \,\|^2 \le \rho^2\}$$

Modified coverage $\mathrm{cov}'(\rho) = \mathrm{cov}(\rho) * \exp(-\beta\sigma_y^2), \beta > 0$

Adversarial information granule

untargeted adversarial attack

x' – adversarial data from $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$

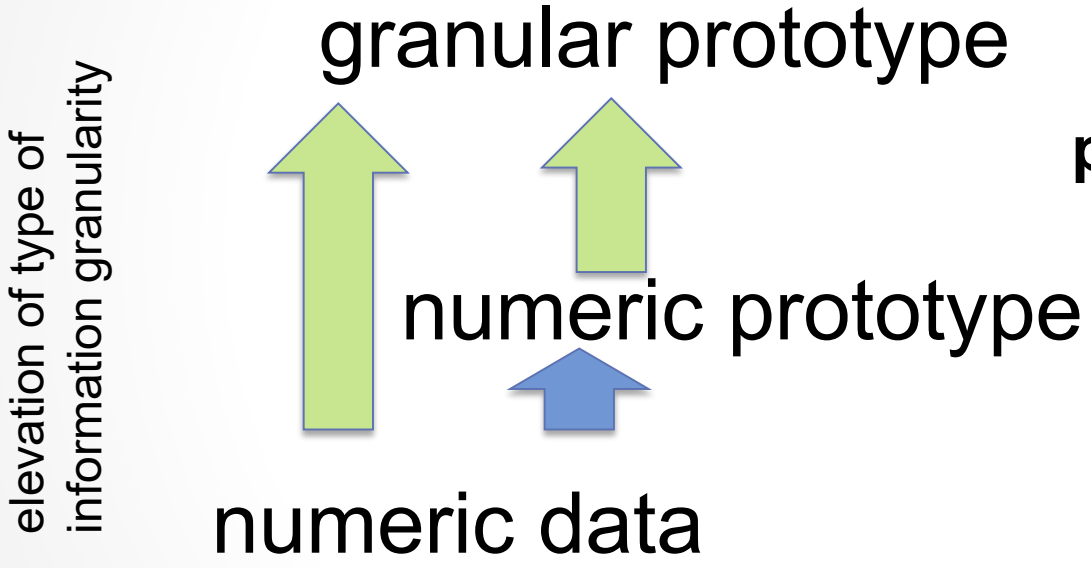
$||\mathbf{x}-\mathbf{x}'||$ small and $f(\mathbf{x})-f(\mathbf{x}')$ different (f – classifier)

$A(\mathbf{x}', \rho)$ – information granule, solution to the optimization problem

$$V(\rho) = \frac{\sum_{\boldsymbol{x_k}:||\boldsymbol{x}_k-\boldsymbol{x}'||\leq n\rho^2} |f(\boldsymbol{x}_k) - f(\boldsymbol{x}')|}{\sum_{\boldsymbol{x_k}:||\boldsymbol{x}_k-\boldsymbol{x}'||\leq n\rho^2} ||\boldsymbol{x}_k - \boldsymbol{x}'||}$$

$\rho_{max} = \arg \mathrm{Max}_{\rho}\, V(\rho)$

Principle of justifiable granularity



principle of justifiable granularity

Rule-based architectures

One among broadly studied knowledge representation schemes of AI

Collection of “if-then” statements with information granules forming condition/conclusion parts

Application-oriented models (~1970):

- Expert systems, knowledge-based systems (MYCIN...)
- Transparency
- Explanation mechanisms
- Knowledge acquisition bottleneck

Interpretability feature of rule-based architectures: déjà vu (DARPA, 2017)

Rule-based architectures main categories

relational form of rules

-If condition is A_i then conclusion is B_i

$i=1, 2,...,c$

A_i, B_i – information granules (e.g., fuzzy sets)

functional form of rules

-If condition is A_i then conclusion is $f_i(x; a_i)$

$i=1, 2,...,c$

Multidimensional rule-based models

n-input variables

Condition part

-if x_1 is A_1 & x_2 is B_1 &... x_n is Z_1 then...

Expert-driven development:

tedious, yet interpretable?
curse of dimensionality,
knowledge acquisition bottleneck

Data-driven development:

clustering
multidimensional, difficult to interpret

Rule-based computing

mapping from $\boldsymbol{R}^n$ to $\boldsymbol{R}$

$$\text{If } \boldsymbol{x} \text{ is } A_i \text{ then } y = f_i(\boldsymbol{x};\, \boldsymbol{a}_i)$$

$$\hat{y} = \sum_{i=1}^{c} A_i(\boldsymbol{x}) f_i(\boldsymbol{x};\boldsymbol{a}_i)$$

Rule-based computing

$$\hat{y} = \sum_{i=1}^{c} A_i(\boldsymbol{x}) f_i(\boldsymbol{x}; \boldsymbol{a}_i)$$

Design of the rules

Condition part: construction of information granules A_i
clustering, usually Fuzzy C-Means (FCM)

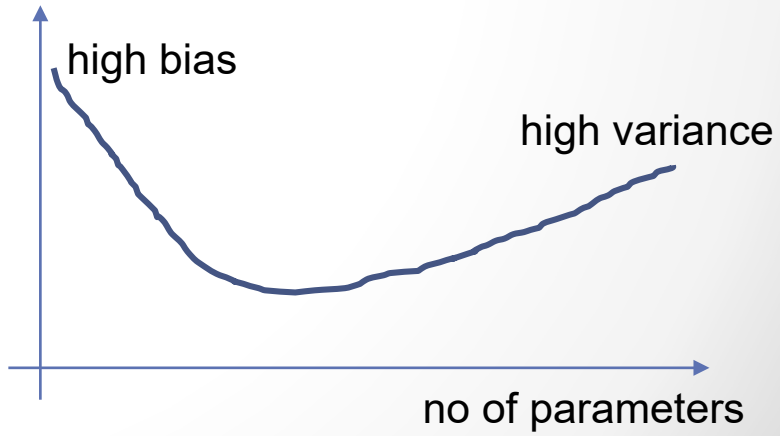
Conclusion part: selection of local function f_i (linear)
estimation of parameters

Rule-based architectures: interpretability

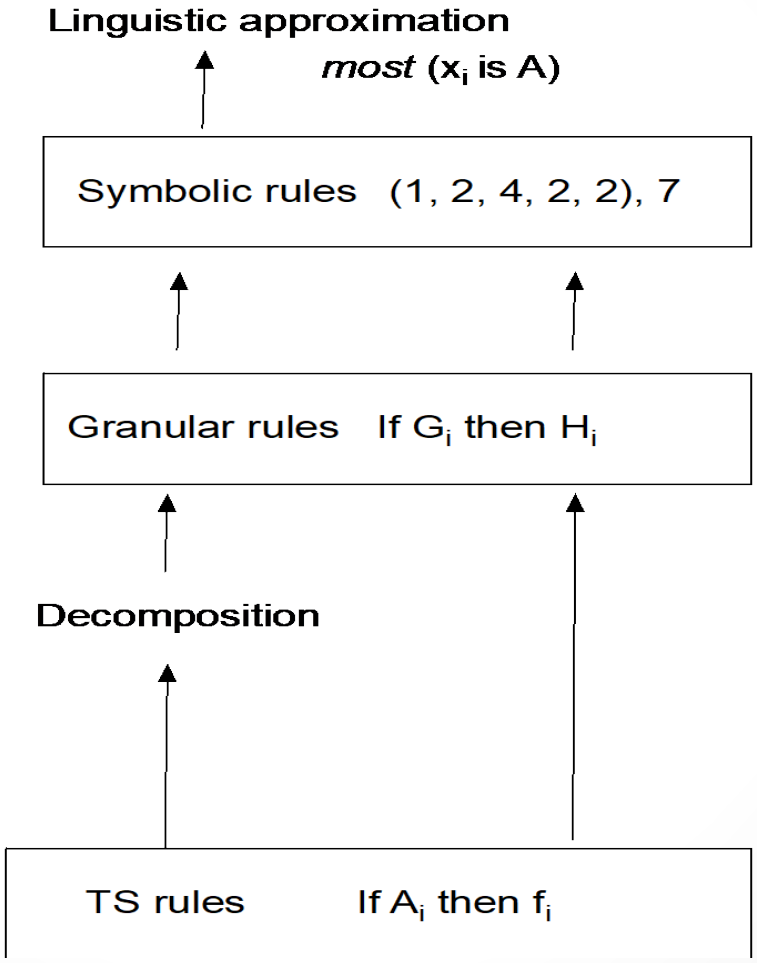
Interpretability –quantification with the use of
-number of rules
-number of variables in condition part
-complexity of local functions
...

interpretability –accuracy

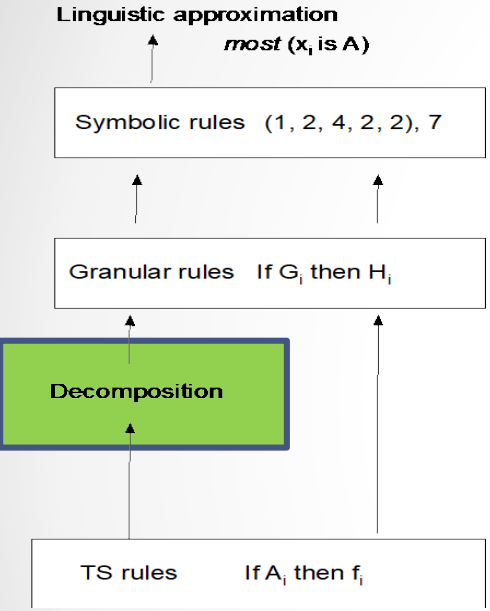
bias –variance dilemma



Interpretability in rule-based models: hierarchy



Interpretability in rule-based models: decomposition



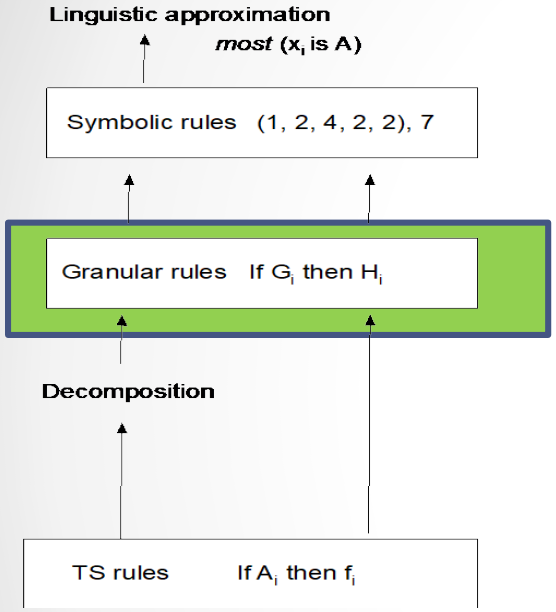
Decomposition of rule conditions

$$A_i(x_j) = \frac{1}{\Sigma_{l=1}^{c}\left(\frac{||x_j - \boldsymbol{v}_{ij}||}{||x_j - \boldsymbol{v}_{lj}||}\right)^{2/(m-1)}}$$

$$\tilde{A}(\boldsymbol{x}) = \min(A_{i1}(x_1), A_{i2}(x_2), \ldots, A_{in}(x_n))$$

$$D = \frac{1}{Nc}\Sigma_{i=1}^{c} \quad \Sigma_{k=1}^{N} |A_i(\boldsymbol{x}_k) - \tilde{A}(\boldsymbol{x}_k)|$$

Interpretability in rule-based models: granular rules (1)



Granular conditions

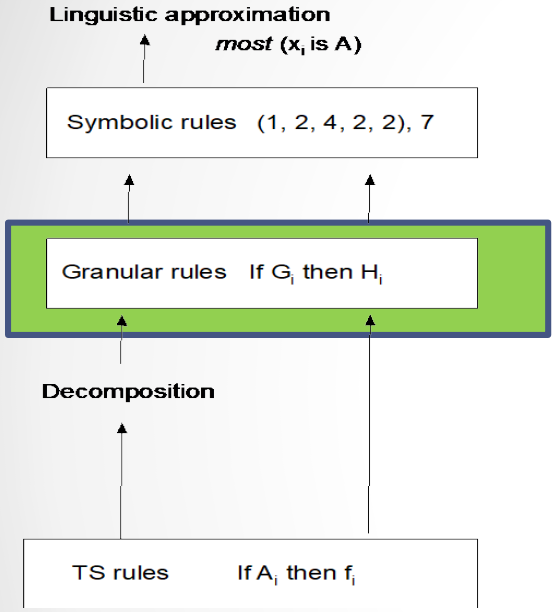
$$cov(G_i) = \sum_{\boldsymbol{x}_k \in \mathbf{G}} u_{ik}$$

$$sp(G_i) = 1 - \rho_i$$

$$\mathbf{G} = \{\boldsymbol{x}_k \mid \|\boldsymbol{x}_k - \boldsymbol{v}_i\| \leq n\rho_i\},\ \rho_i \in [0,1].$$

$$\rho_{i,\text{opt}} = \arg\min_{\rho i} [cov(G_i) sp(G_i)]$$

Interpretability in rule-based models: granular rules (2)



Granular conclusions

$$\mathbf{T}_i=\{target_k,\ k=1,2,\ldots,N \mid\ target_k=f_i(\boldsymbol{x}_k,\boldsymbol{a}_i),\ \boldsymbol{x}_k\in G_i\}$$

$$\bar{t}_i=\Sigma_{k:\mathrm{target}_k\in\mathbf{T}_i} u_{ik}target_k \,/\, \Sigma_{k:target_k\in\mathbf{T}_i} u_{ik}$$

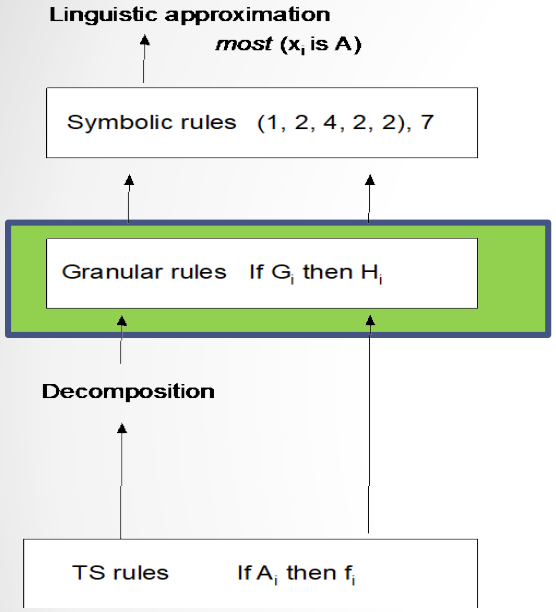
$$\mathbf{T}_i^{+}=\{target_k \mid target_k\in\mathbf{T}_i \text{ and } target_k\geq\bar{t}_i\}$$

$$\mathrm{cov}([\bar{t}_i,b_i])=\Sigma_{target\in T_i^{+}} u_{ik}$$

$$\mathrm{sp}([\bar{t}_i,b_i])=1-\frac{b_i-\bar{t}_i}{y_{max}-\bar{t}_i}$$

$$b_{i,\mathrm{opt}}=\arg\max_{bi}\{\,cov([\bar{t}_i,b_i])\ sp([\bar{t}_i,b_i])\}$$

Interpretability in rule-based models: granular rules (3)



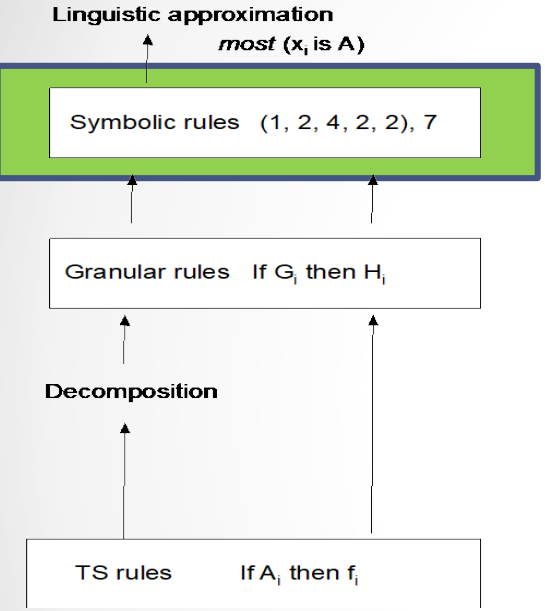
Variability of the rule

$$\xi_i = sp(H_i)/sp(G_i)$$

$$(x = low, y = large)$$

$$(x = low, y = large, \text{variability} = medium)$$

Interpretability in rule-based models: symbolic rules



One-dim projections of prototypes ordered linearly

Symbols

1, 2,3...

small, medium, large...

positive large, positive medium...

Rule – sequence of integers

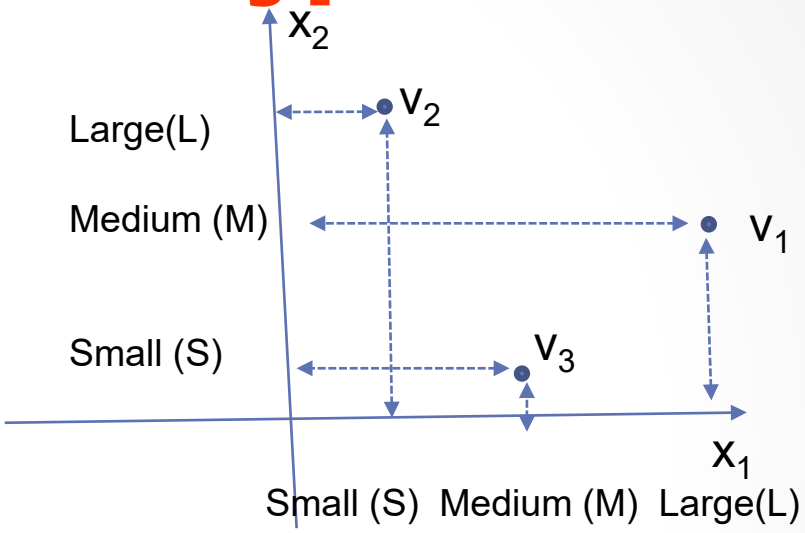
$$[i_1\ i_2 ... i_c]\, j_i$$

Combination of rules

[5, 1, 6, 8, 5] 6
[4, 9, 5, 7,1] 5

[{4 or 5}, {1 or 9}, {5 or 6}, {7 or 8}, {1 or 5})] 5

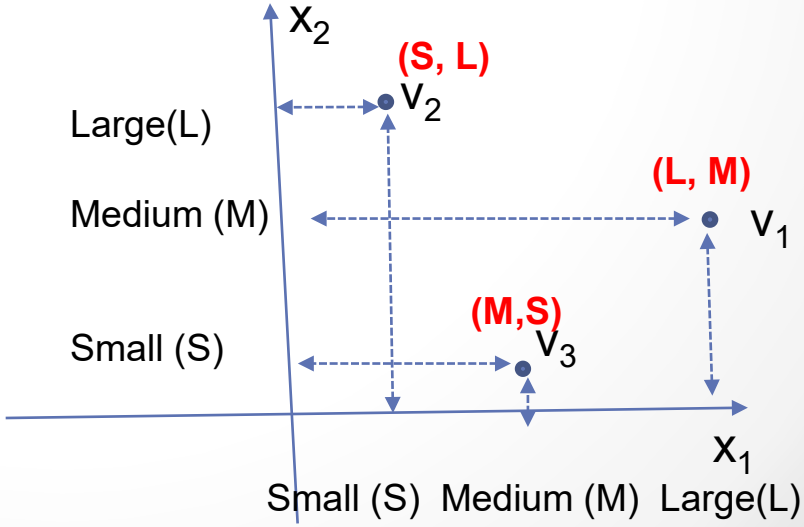
Symbol-oriented interpretation of FCM prototypes



Duality of information granule:

Numeric description

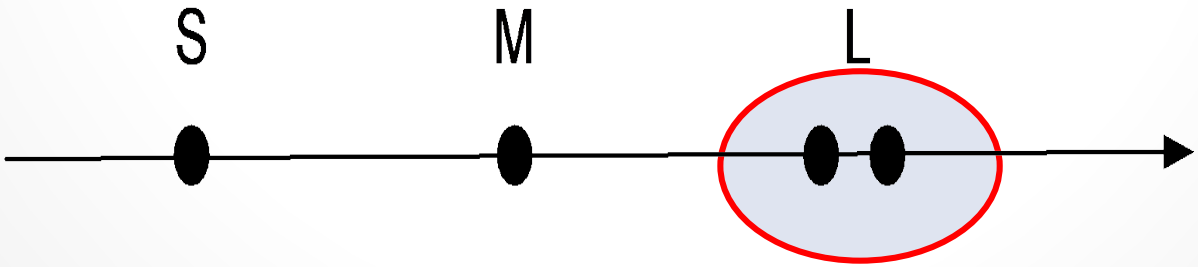
Symbolic view



Symbolic perspective: Semantic indistiguishability

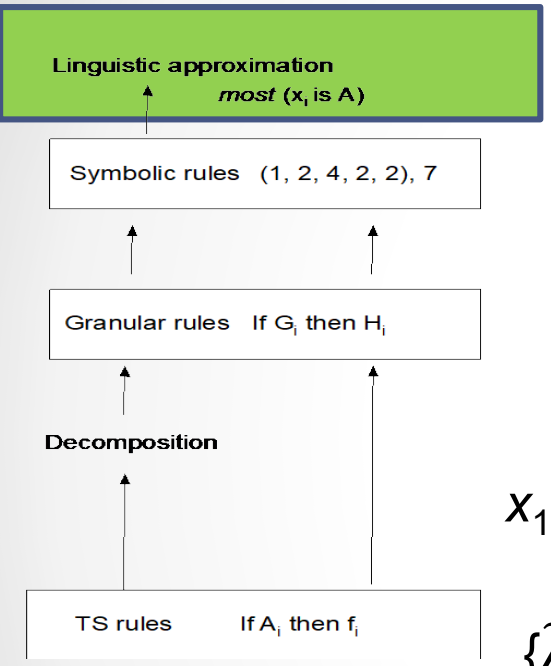
Categorical view at projections of prototypes

semantic indistinguishability of projected prototypes



Interpretability in rule-based models: linguistic approximation

Kacprzyk,& Zadrozny



(x_1 is *low* and x_2 is *medium* and x_3 is *low* and x_4 is *medium*)

most(x_is are *low*).

x_1 is A_{i1} and x_2 is A_{i2} and ... and x_n is A_{in} $\approx$*most* (x_is are $\tilde{\mathbf{A}}$}

$\{\tilde{A}_1,\tilde{A}_2,...\tilde{A}_n\}$ fuzzy sets exhibiting the same semantics

$$L(i)=\frac{1}{Nn}\Sigma_{j=1}^{n} \quad \Sigma_{k=1}^{N} |A_{ij}\left(x_{kj}\right)-\tilde{A}_j(x_{kj})|$$

$$i_{\text{opt}} = \arg\min_{i=1,2,..,c} L(i_0)$$

-if *most* (x_is are i_{opt})$|_{Lopt}$ then j

Examples

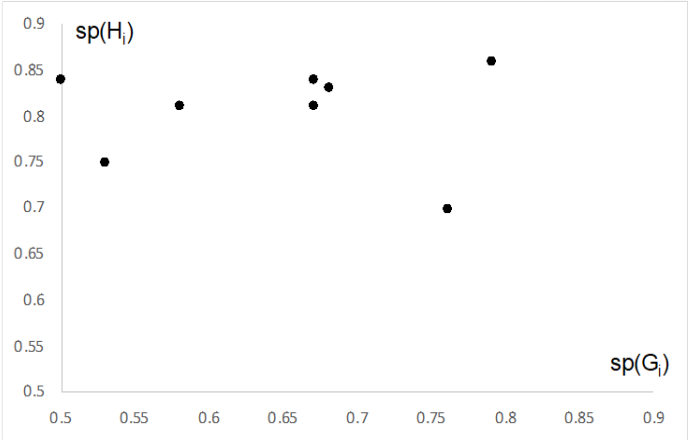
Boston housing data (UCI)
c=5

Rule_1: [5 2 4 3 4 2 4 1 4 4 4 1 5] 1
Rule_2: [1 5 1 1 1 5 1 5 1 2 1 5 1] 5
Rule_3: [3 3 3 5 3 3 3 3 3 3 3 3 3] 3
Rule_4: [2 4 2 4 2 4 2 4 2 1 2 4 2] 4
Rule_5: [4 1 5 2 5 1 5 2 5 5 5 2 4] 2

Rule_3: [3 3 3 5 3 3 3 3 3 3 3 3 3] 3	$most(x_i\text{s are } 3)_{0.000}$
Rule_4: [2 4 2 4 2 4 2 4 2 1 2 4 2] 4	$most(x_i\text{s are } 2)_{0.109}$
Rule_1:[5 2 4 3 4 2 4 1 4 4 4 1 5] 1.	$most(x_i\text{s are } 4)_{0.114}$
Rule_2:[1 5 1 1 1 5 1 5 1 2 1 5 1] 5	$most(x_i\text{s are } 1)_{0.148}$
Rule_5: [4 1 5 2 5 1 5 2 5 5 5 2 4] 2	$most(x_i\text{s are } 5)_{0.165}$

Examples

Boston housing data (UCI)
c=9



[5 5 5 9 5 4 5 5 5 5 5 5 5] 4
[6 4 6 8 6 2 6 4 6 6 6 4 6] 4
[4 6 4 7 4 7 4 6 4 3 3 6 4] 6
[3 7 3 6 3 6 3 7 3 2 4 7 3] 6
[7 2 8 3 9 3 8 3 8 8 8 3 7] 3
[2 8 2 5 2 8 2 8 2 1 2 8 2] 8
[8 3 9 4 7 1 9 1 7 7 7 2 8] 2
[1 9 1 2 1 9 1 9 1 4 1 9 1] 9
[9 1 7 1 8 5 7 2 9 9 9 1 9] 1

$most(x_i\text{s are } 5)_{0.009}$
$most(x_i\text{s are } 6)_{0.028}$
$most(x_i\text{s are } 4)_{0.034}$
$most(x_i\text{s are } 3)_{0.058}$
$most(x_i\text{s are } 7)_{0.077}$
$most(x_i\text{s are } 2)_{0.086}$
$most(x_i\text{s are } 7)_{0.094}$
$most(x_i\text{s are } 1)_{0.132}$
$most(x_i\text{s are } 9)_{0.145}$

Examples

[8 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 5] 5
[9 6 6 6 6 6 6 6 6 6 6 6 6 6 4 6 7 6] 5
[1 9 1 1 1 1 1 1 1 1 1 1 1 1 9 1 1 1] 1
[6 8 7 7 7 7 7 7 7 7 7 7 7 3 7 8 7] 5
[7 3 4 4 4 4 4 4 4 4 4 4 4 6 4 5 4] 5
[3 5 2 2 2 2 2 2 2 2 2 2 2 8 2 3 2] 2
[5 7 8 8 8 8 8 8 8 8 8 8 8 2 8 9 8] 9
[4 2 3 3 3 3 3 3 3 3 3 3 3 7 3 4 3] 3
[2 1 9 9 9 9 9 9 9 9 9 9 9 1 9 2 9] 4

$most(x_i\text{s are } 5)_{0.004}$
$most(x_i\text{s are } 6)_{0.015}$
$most(x_i\text{s are } 1)_{0.017}$
$most(x_i\text{s are } 7)_{0.018}$
$most(x_i\text{s are } 4)_{0.020}$
$most(x_i\text{s are } 2)_{0.021}$
$most(x_i\text{s are } 8)_{0.021}$
$most(x_i\text{s are } 3)_{0.027}$
$most(x_i\text{s are } 9)_{0.036}$

(c) Parkinsons

(a) Abalone

(b) airfoil

[9 7 4 1] 2

$st(x_i\text{s are } 1)_{0.107}$
$most(x_i\text{s are } 7)_{0.141}$

(d) Cycle power

Enhancements of interpretability of rule-based models

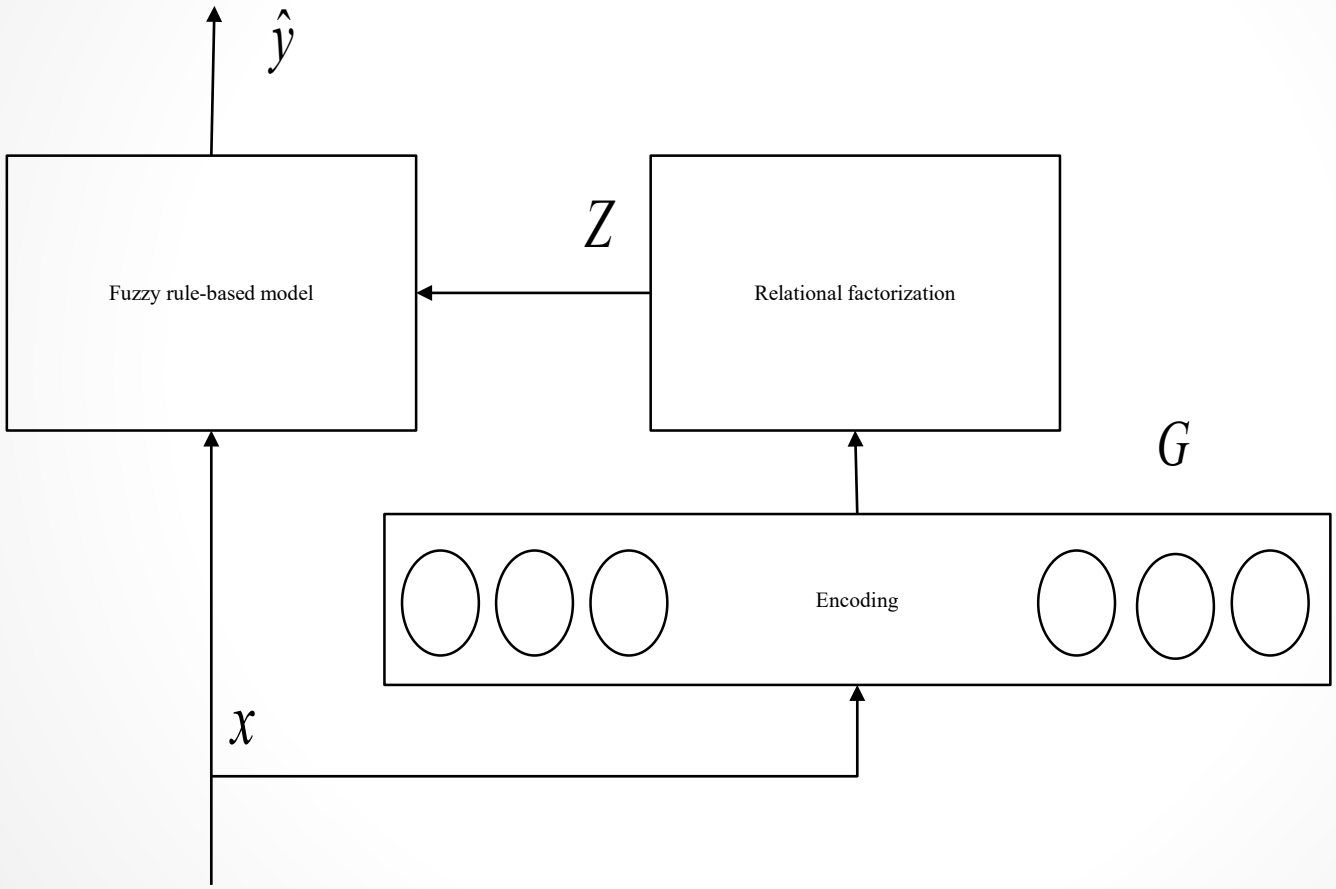
if $\boldsymbol{x}$ is A_i then $y = f_i(\boldsymbol{x}; \boldsymbol{a}_i)$

$$\hat{y} = \sum_{i=1}^{c} A_i(\boldsymbol{x}) f_i(\boldsymbol{x}, \boldsymbol{a}_i)$$

$$\tilde{y} = \sum_{i=1}^{c} L[A_i(\boldsymbol{x})] f_i(\boldsymbol{x}, \boldsymbol{a}_i)$$

$$\tilde{y} = G^{-1} \sum_{i=1}^{c} A_i(\boldsymbol{x}) B_i(\boldsymbol{y})\}$$

Rule-based models with relational factorization



Relational factorization

Nonnegative matrix factorization: dimensionality reduction

Relational factorization: interpretable dimensionality reduction

Decompose X

$$\hat{X} = rel(Y, R)$$

dim (X)= N by n
dim(Y)= N by p **p<<n**
dim(R)= p by c

Interpretable composition operators (s-t, t-s)

Relational factorization: details

$$\hat{X} = Y \circ R$$

$$\hat{x}_{ij} = \underset{l=1}{\overset{p}{S}}(y_{il}\ t\ r_{lj})$$

$$x_{ij} = [(y_{i1}\ and\ r_{1j})] or\ \ldots\ or\ [(y_{ip}\ and\ r_{pj})]$$

$$x_{ij} = y_{i1}|_{r_{1j}}\ and\ y_{i2}|_{r_{2j}}\ and \ldots and\ \ y_{ip}|_{r_{pj}}$$

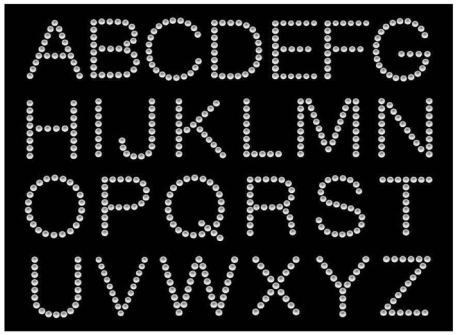
$$\hat{X} = Y \cdot R$$

$$\hat{x}_{ij} = \underset{l=1}{\overset{p}{T}}(y_{il}\ s\ r_{lj})$$

$$Q = \sum_{i=1}^{N}\sum_{j=1}^{n}(x_{ij} - \hat{x}_{ij})^2$$

$\text{Min}_{Y,\,R}\, Q$

Relational factorization: Example(1)



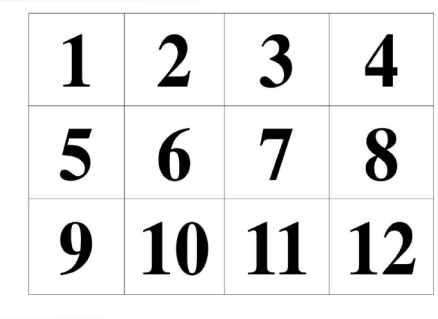
Original
680x870; p =15

Non-negative matrix factorization, MSE =0.0222

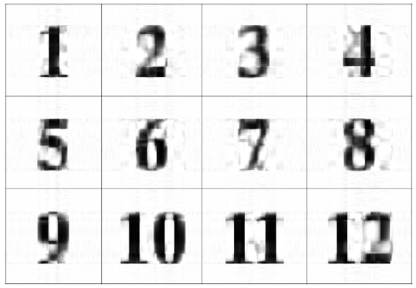
s-t composition, MSE =0.0218

t-s composition, MSE =0.0163

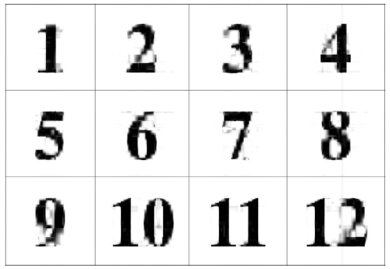
Relational factorization: Example(2)



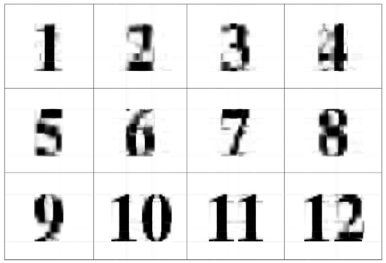
Original 786x1206, p=20



Non-negative matrix factorization, MSE =0.0101



s-t composition, MSE =0.0056



t-s composition, MSE =0.0085

Concepts

concepts – descriptors of data forming their representation at some suitable level of abstraction

used as building blocks to form a blueprint of ensuing modeling faculties

Design carried out a suitable level of abstraction – information granularity

Synergistic involvement of data and domain knowledge (human-centricity)

Logic fabric of architectures of concepts

Conditional clustering

Context – formalizing domain knowledge through information granules
focused clustering

Clustering – capturing structure of data

Conditional clustering

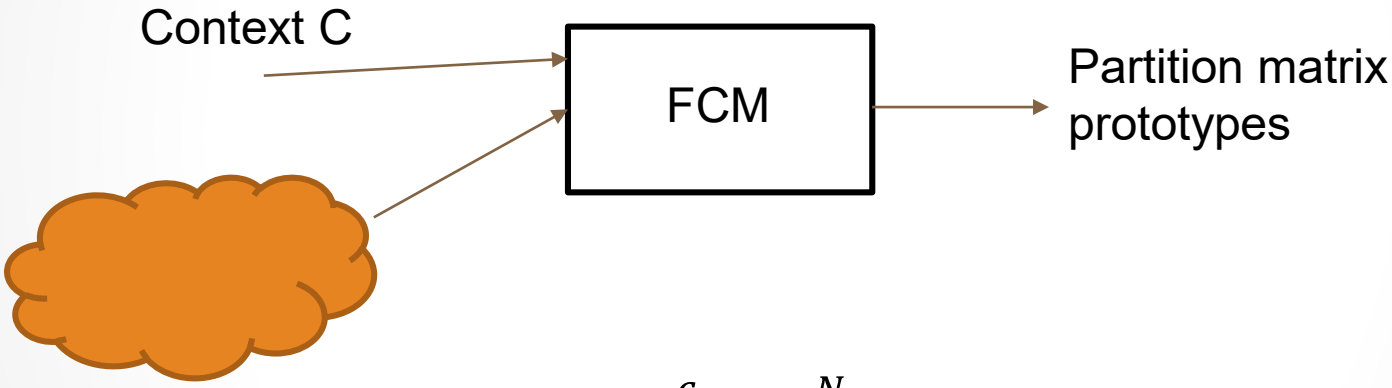
Context – formalizing domain knowledge through information granules
focused clustering

Clustering – capturing structure of data

Conditional Fuzzy C-Means (FCM)

data ($\mathbf{x}_k$, y_k), c-number of clusters

context C (information granule) C(y_k)-concept



Objective function $$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 ||x_k - v_i||^2$$

partition matrix U=[u_{ik}], i=1,2,..., c; k=1,2,...,N,
prototypes $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, ..., $\boldsymbol{v}_c$

Conditional Fuzzy C-Means (FCM)

data ($\mathbf{x}_k$, y_k), context C (information granule) $C(y_k)$

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 ||x_k - v_i||^2$$

$$0 < \sum_{i=1}^{c} u_{ik} < N, \; i = 1,2, \ldots, c$$

Optimization

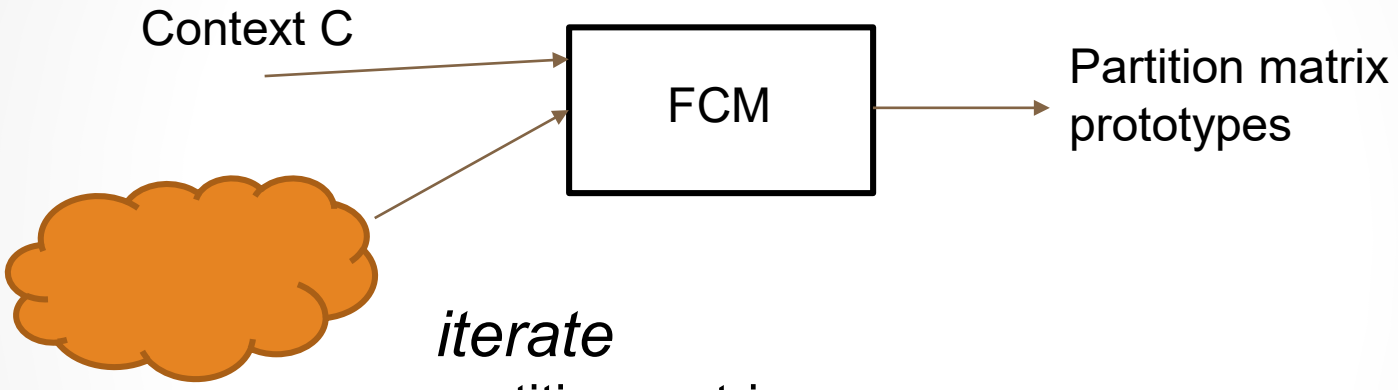
Minimize Q with respect to partition matrix and prototypes

$$\text{Min}_{\,U,v1,v2,\ldots,vc}\, Q$$

$$\sum_{i=1}^{c} u_{ik} = C(y_k) \qquad 0 < \sum_{i=1}^{c} u_{ik} < N, \; i = 1,2, \ldots, c$$

Conditional FCM: optimization

data ($\mathbf{x}_k$, y_k), context C (information granule) $C(y_k)$



iterate

partition matrix

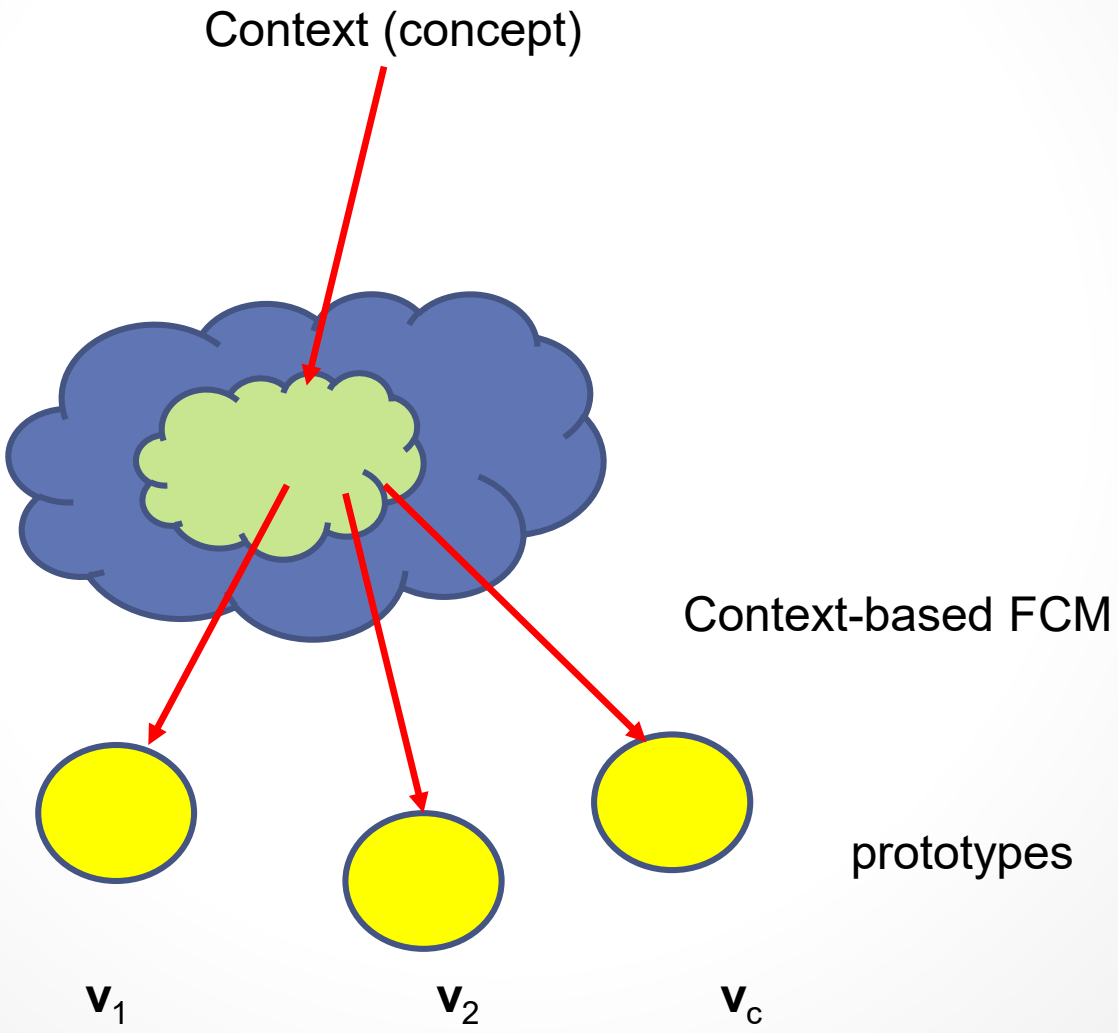
$$u_{ik} = \frac{C(y_k)}{\sum_{j=1}^{c} \left(\frac{||x_k - v_i||^2}{||x_k - v_j||^2} \right)}$$

prototypes

$$v_i = \sum_{k=1}^{N} u_{ik}^2 x_k / \sum_{k=1}^{N} u_{ik}^2$$

until termination condition satisfied

Concept: architecture development



Concept selection

Concept= information granule C in space Y

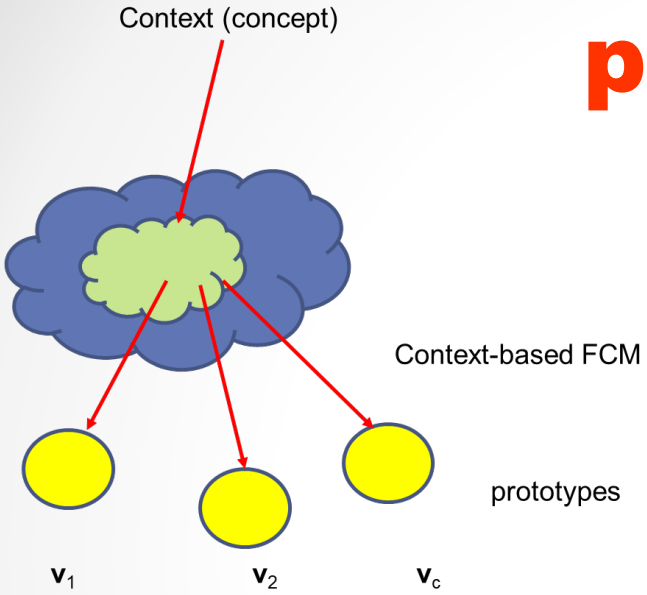
Interval
Fuzzy set
Rough set
…

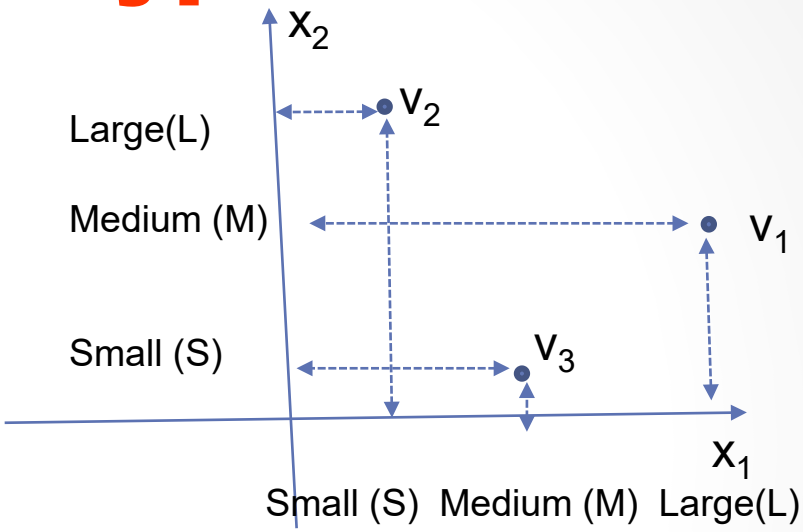
Formalization of domain knowledge-focused view at data

Coverage requirement (sufficient data support):

$$\sum_k C(y_k) > \alpha$$

Symbol-oriented interpretation of prototypes

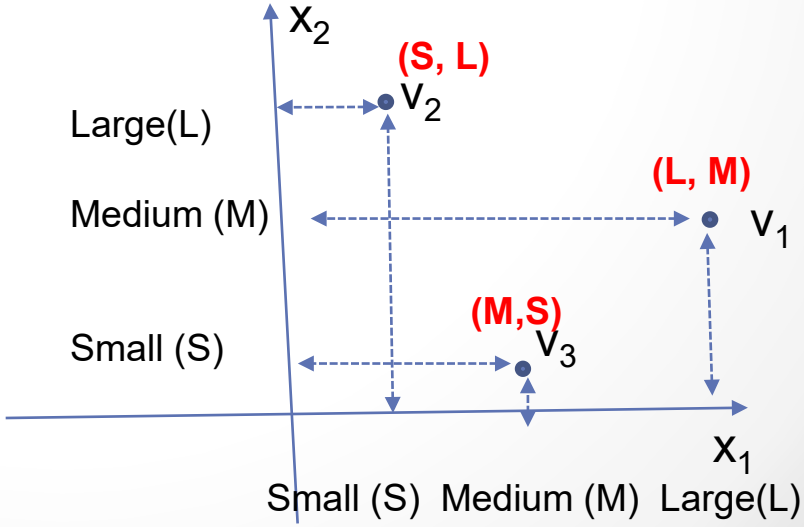




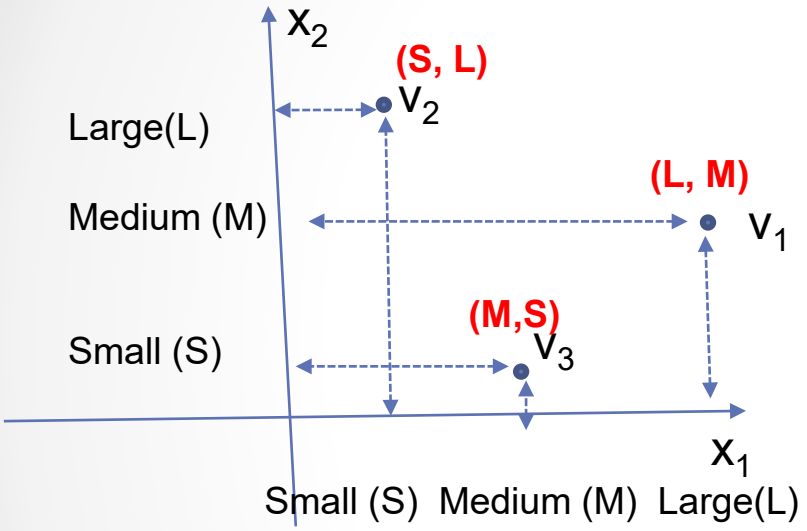
Duality of information granule:

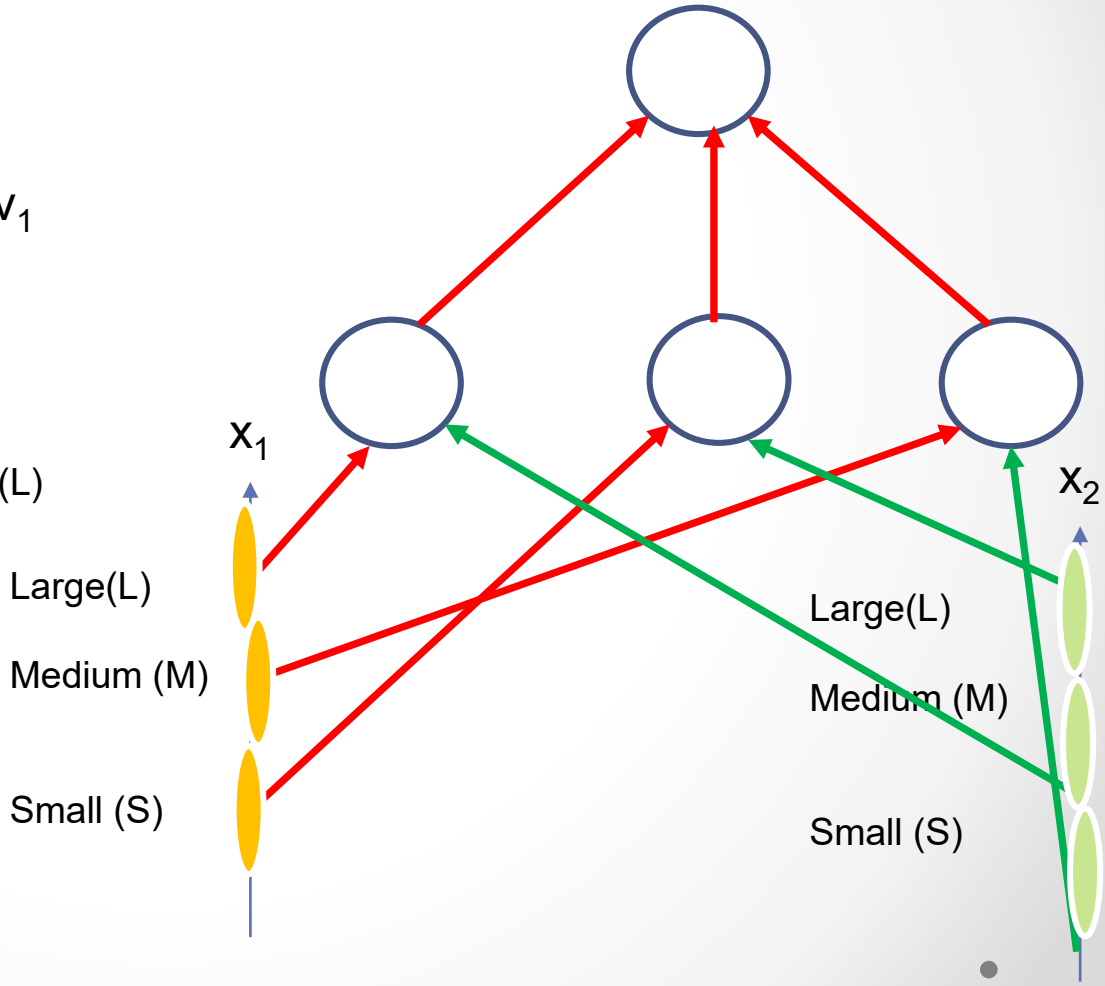
Numeric description

Symbolic view

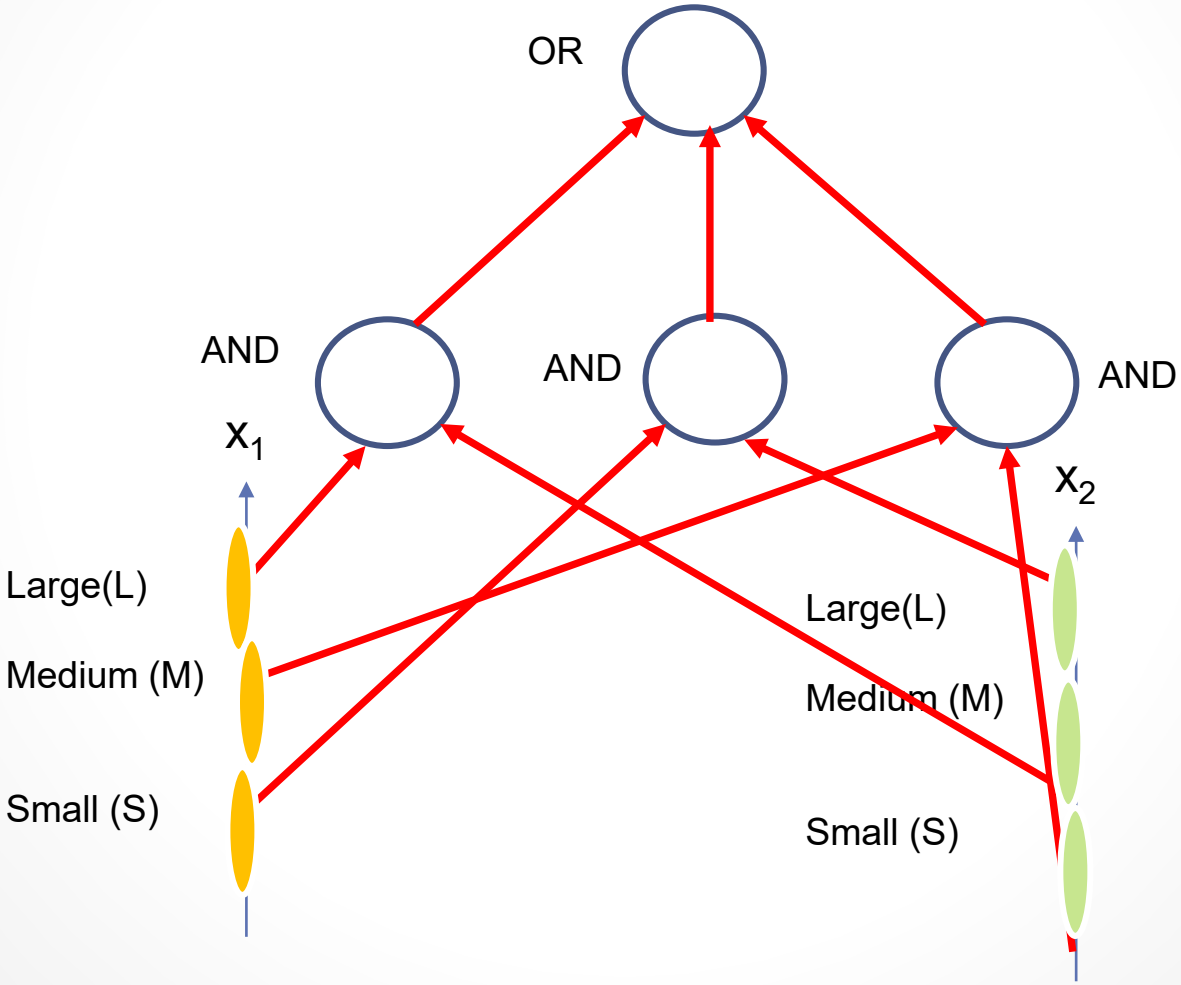


Architectural considerations: Logic network

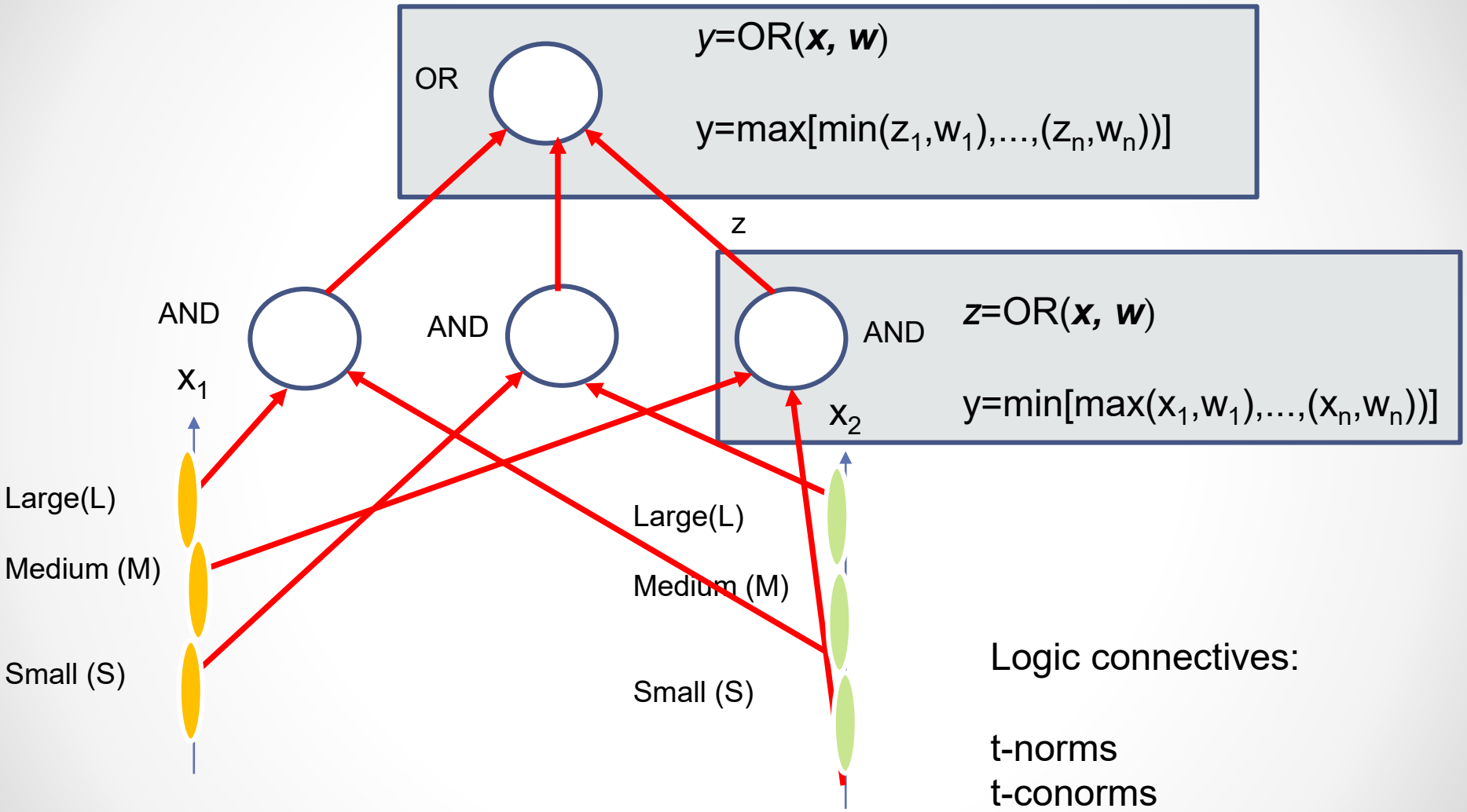




Logic network



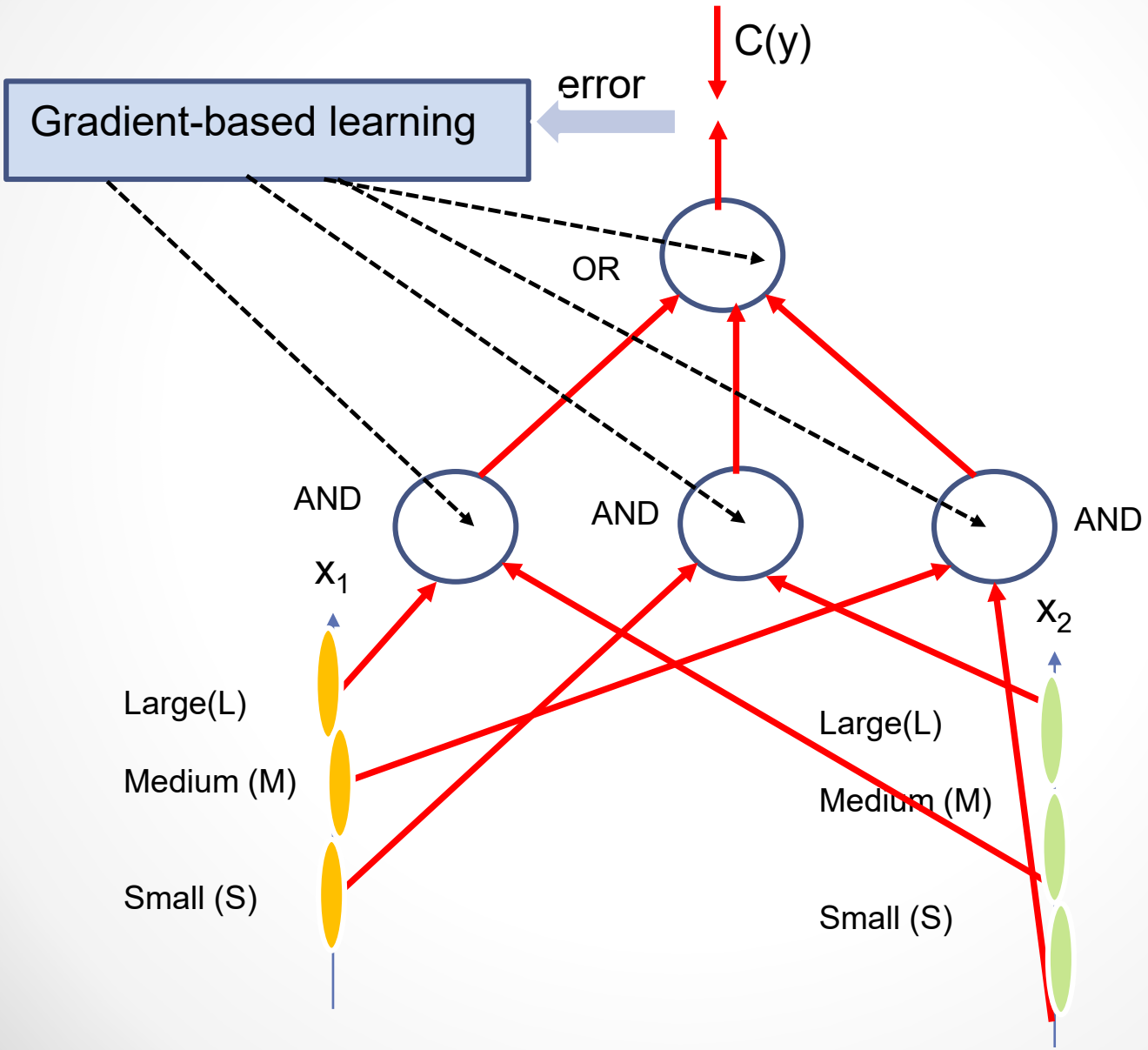
Logic network (3)



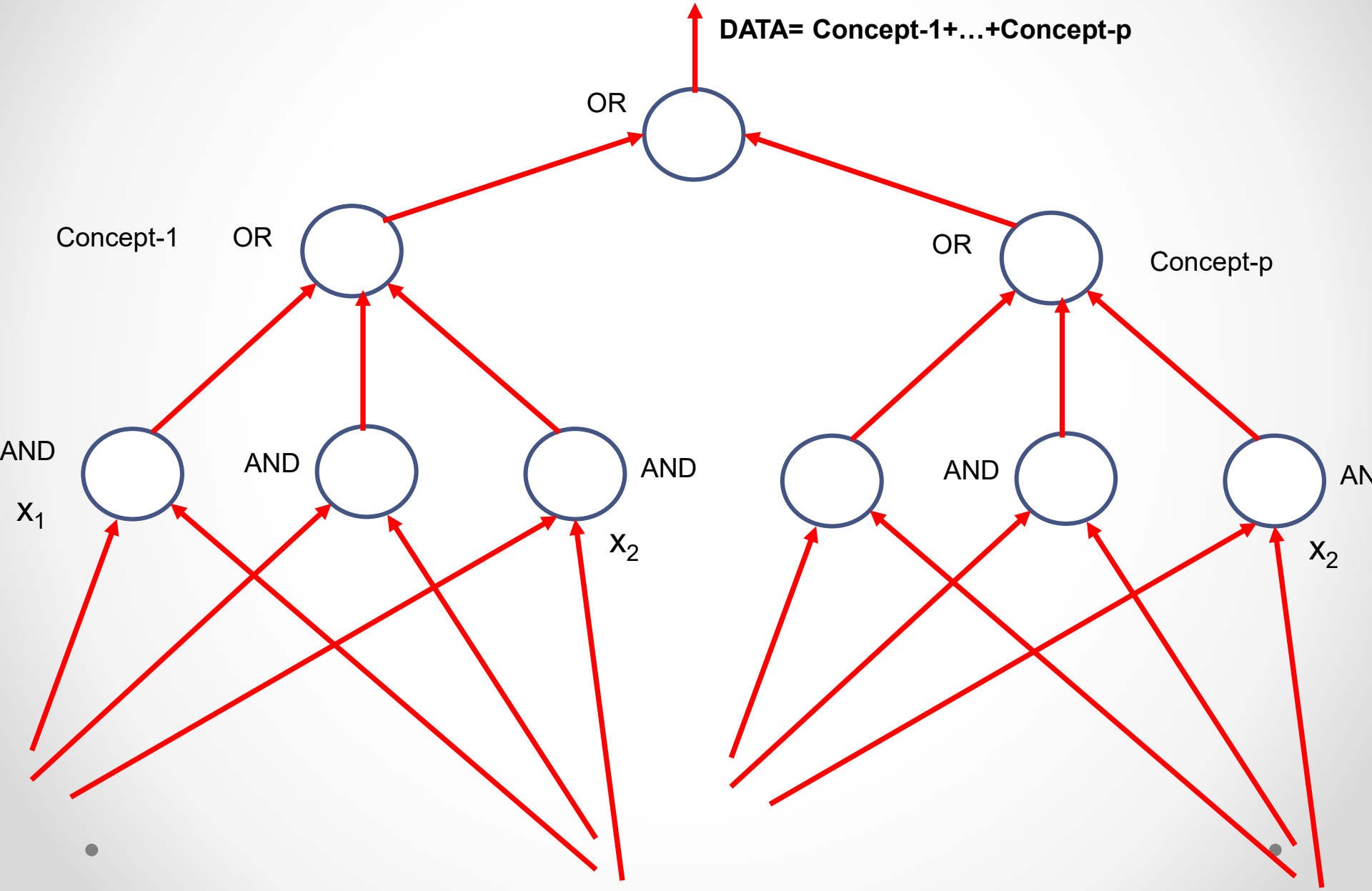
Logic connectives:

t-norms
t-conorms

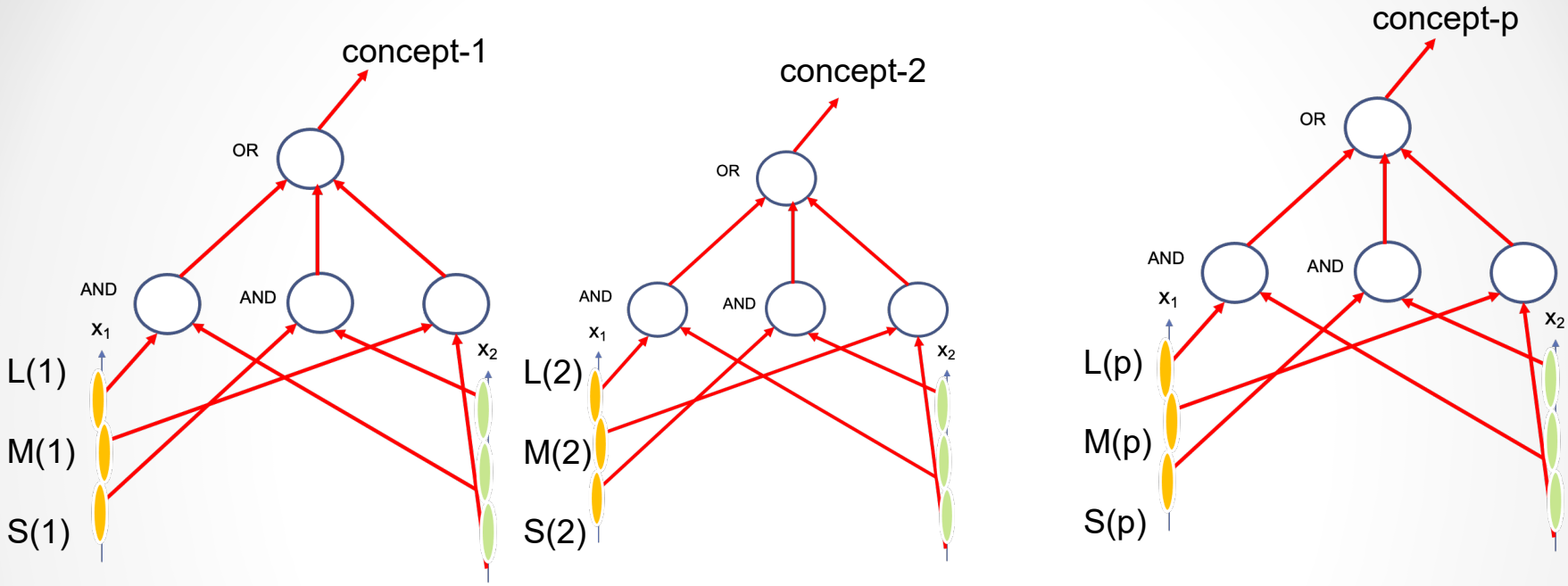
Logic network - learning



Data description through concepts



Network of concepts



clustering prototypes $\{\mathbf{v}_i[1]\}_{c1}$, $\{\mathbf{v}_i[2]\}_{c2,...,}$ $\{\mathbf{v}_i[p]\}_{cp}$

resulting prototypes: $\mathbf{v}_i$~, i=1, 2,..., c and symbolic representation S, M, L...

expressing original prototypes with the aid of $\mathbf{v}_i$~

Double-level clustering

clustering prototypes $\{\mathbf{v}_i[1]\}_{c1}$, $\{\mathbf{v}_i[2]\}_{c2,...,}$ $\{\mathbf{v}_i[p]\}_{cp}$

resulting prototypes: $\mathbf{v}_i^{\sim}$, i=1, 2,..., c and symbolic representation S, M, L...

expressing original prototypes with the aid of $\mathbf{v}_i^{\sim}$

$\mathbf{v}_j[ii] \rightarrow \mathbf{v}^{\sim}_{i0}$ $\qquad\qquad$ $i_0 = \arg\min \|\mathbf{v}_j[ii] - \mathbf{v}^{\sim}_i\|$

Conclusions

Centrality of information granules and Granular Computing in the realization of interpretability requirements

Construction of information granules

Logic-oriented development of rules, concepts and dimensionality reduction