Knowledge-Data Environment of Machine Learning

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

Introduction: Data and Machine Learning; concepts and key challenges

Data and knowledge in Machine Learning

Knowledge representation

Knowledge-data Machine Learning: Architectures and Learning

Conclusions

Centrality of Data in Machine Learning

Models of Machine Learning (ML) as **data-driven** constructs

$$\mathcal{D} \rightarrow M_{\mathcal{D}}$$

- Credibility of $M\mathcal{D}$ associated with presence and mechanisms of inductive reasoning
- Going beyond the scope of data $\mathcal{D}$ - open issues
- Learning realized from scratch
- Loss function focused predominantly on optimization of prediction/classification performance
- Inherently black-box nature of $M\mathcal{D}$
- Possible data attacks

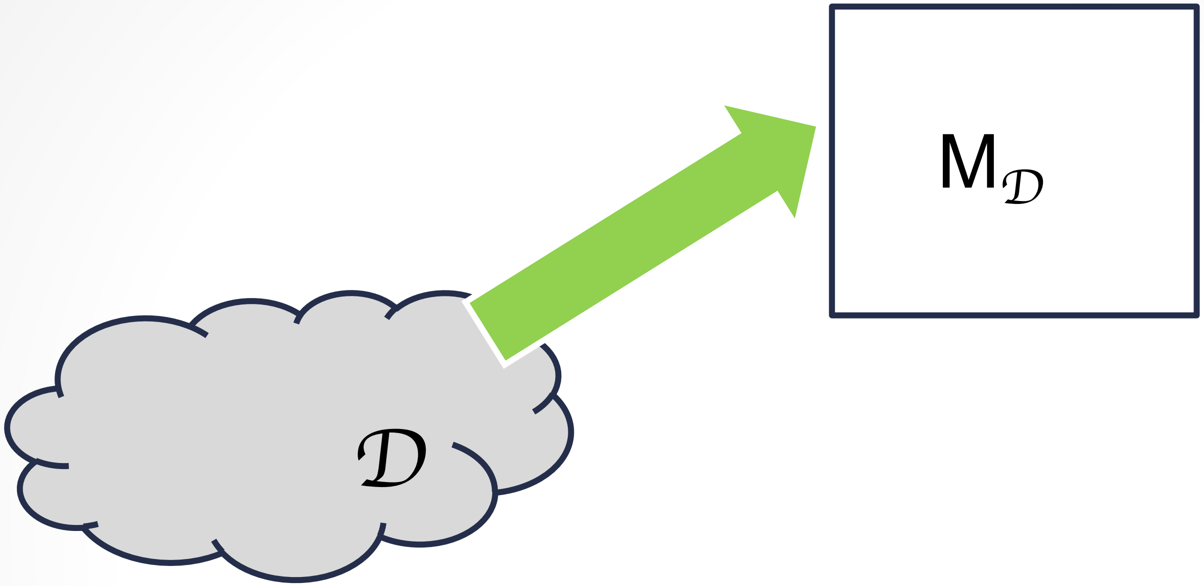
Machine Learning: Challenges

credibility (confidence)

interpretability, explainability, and transparency

computational sustainability

Data in Machine Learning

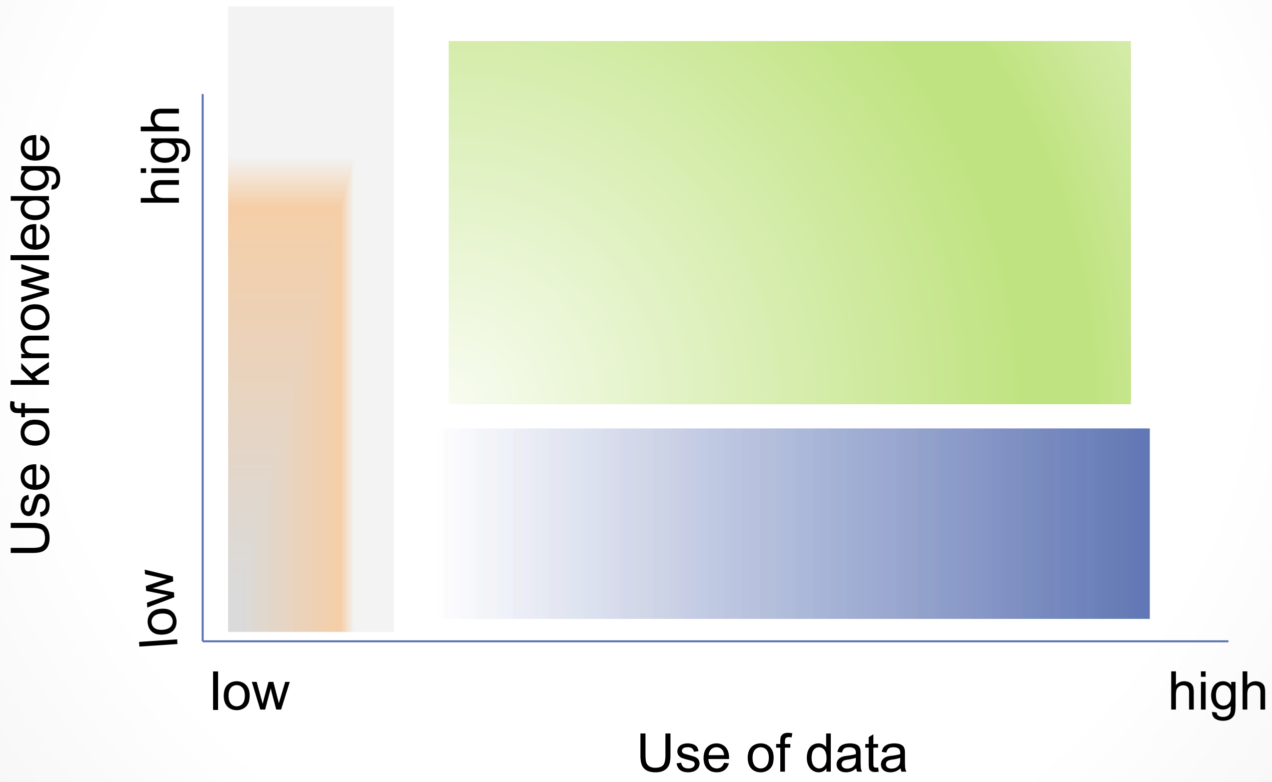


$\mathcal{D}= \{(\mathbf{x}_k, t_k)\}$, k=1,2,...,N

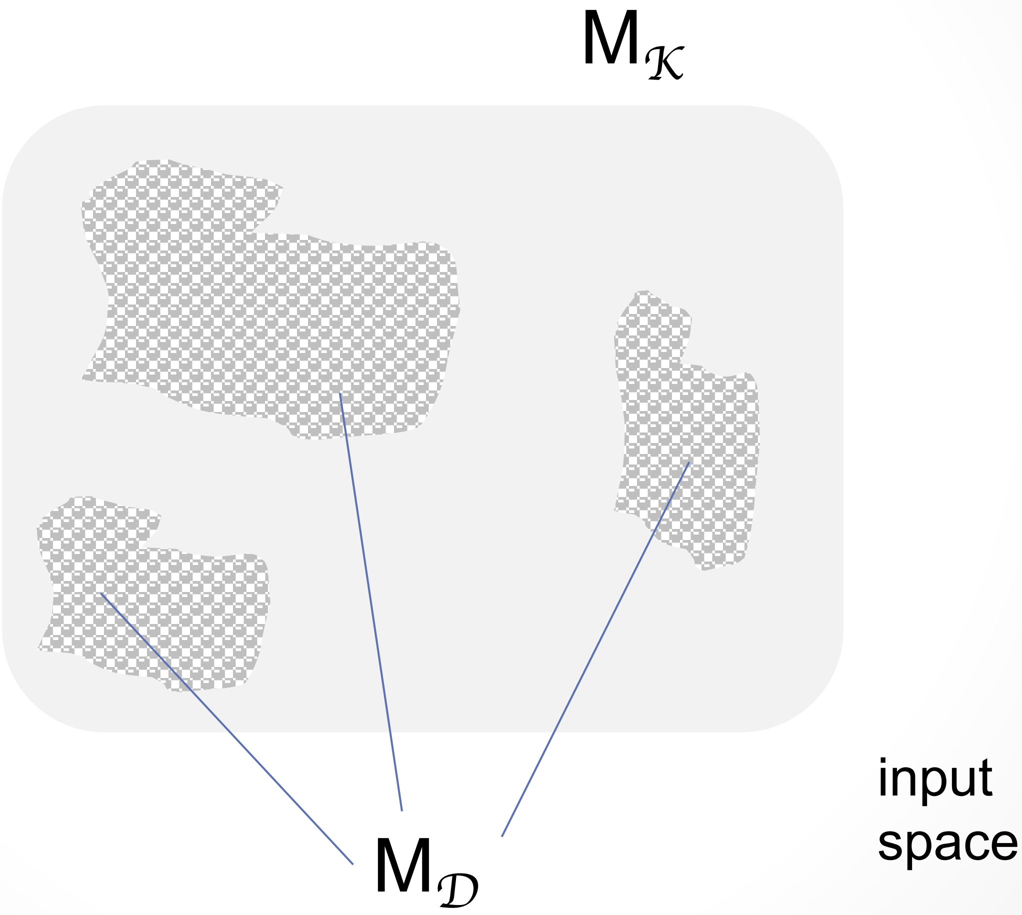
Loss function

$$L_{\mathcal{D}} = \sum_{\mathcal{D}} ||t_k - M\mathcal{D}(x_k)|| + \lambda R(\mathcal{D})$$

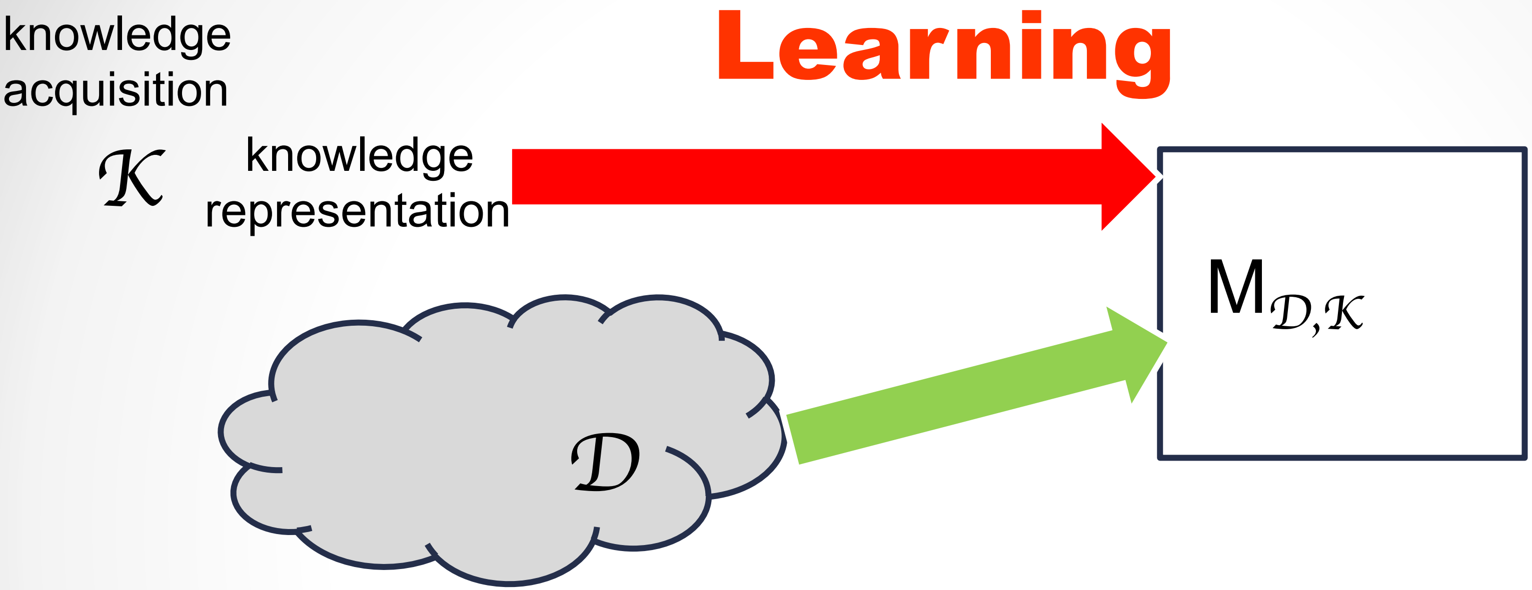
Data and knowledge in Machine Learning



Data and knowledge in Machine Learning



Data and Knowledge in Machine Learning



Data and knowledge

$\mathcal{D} = \{(\mathbf{x}_k, t_k)\}$, k=1,2,...,M

$\mathcal{K}$

Loss function

$$L_{\mathcal{D},\mathcal{K}} = \sum_{\mathcal{D}} ||t_k - M_{\mathcal{D},\mathcal{K}}(x_k)|| + \lambda_1 R(\mathcal{D}) + \lambda_2 R(\mathcal{K})$$

Knowledge Representation

Knowledge in Machine Learning: Research Agenda

Origin and taxonomy of knowledge

Knowledge representation

Realization of unified knowledge –data environment of Machine Learning framework

Efficient accommodation of knowledge

Learning schemes

Knowledge: origin and taxonomy

Scientific knowledge
Universal laws of physics, chemistry, ...
Physics-informed ML

World knowledge
Facts from everyday life; intuitive and validated by human reasoning (subsumes linguistics) and validated through empirical studies; levels of abstraction (information granules)

Expert knowledge
Common knowledge, held by a particular group of experts; levels of abstraction (information granules)

Physics -informed Machine Learning (1)

physics –oriented knowledge

$$g(\mathbf{x}, y) = 0$$

Example

$$g: \frac{\partial y}{\partial t} + y\frac{\partial y}{\partial x} - 0.2\frac{\partial^2 y}{\partial^2 x} = 0$$

Newton’s law of motion, Maxwell’s law of electromagnetics
Conservation law (mass, moment, energy...)

Physics -informed Machine Learning (2)

Data $\mathcal{D} = \{(\mathbf{x}_k, y_k)\}$, k=1, 2,..., N

M- ML model $M(\mathbf{x}_k, \mathbf{w})$

$g(\mathbf{x}, y) = 0$

Loss function

$$L=\sum_{\mathcal{D}} \|M(\mathbf{x}_k,\mathbf{w})-y_k\|^2+\lambda \sum_{\mathcal{D}0} \|g(\mathbf{x}_k,M(\mathbf{x}_k,\mathbf{w}))\|^2$$

Commonly encountered regularization term in ML

Knowledge representation

Algebraic equations

Differential equations

Simulation results

Spatial invariances (translations and rotations)

Logic rules and rule-based models

Knowledge graphs

Relations and relational calculus

Semantic networks

Frames with default assignments

...

Knowledge representation

Knowledge expressed at the higher level of abstraction than the one being realized by numeric entities

The central role of information granules

Symbolic-subsymbolic (numeric) perspective: Duality of information granule

semantics ⟷ Numeric description (parameters)

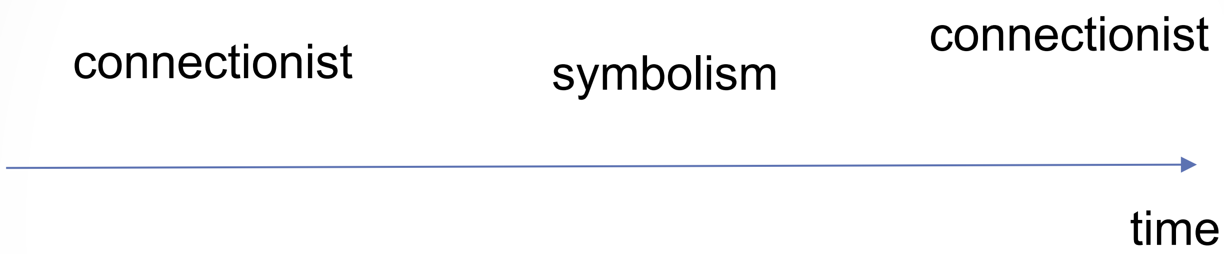
Symbols and symbol-oriented processing
L,M, S...
S+M=L
Not(S)=L
...
S, M, L,-L,...

number-oriented processing
parameters of characteristic functions
Membership functions

fuzzy set

Numeric parameters of Triangular membership function
T(x; 0, 4, 7)

Symbolic versus connectionist pursuits in AI



... our purely numeric connectionist networks are inherently deficient in abilities to reason well; or purely symbolic logical systems are inherently deficient in abilities to represent the all important heuristic connections between things –the uncertain approximate or analogical links...

M. Minsky, Logical versus analogical or symbolic versus connectionist or neat versus scruffy, ***AI Magazine***, 12, 2, 1991

Knowledge integration: Two levels

Knowledge integration at the level of available data

Knowledge integration at the level of ML models

Architectures

Knowledge integration-
data level

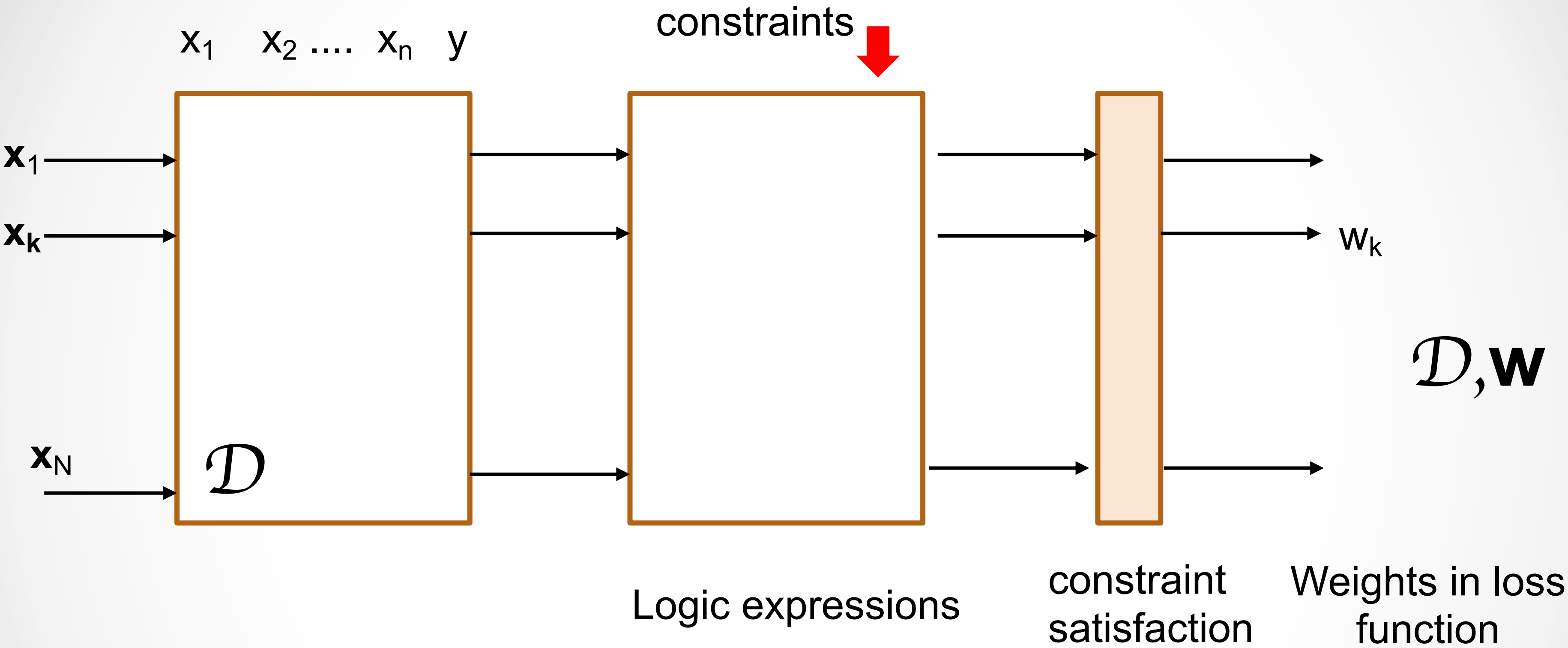
Knowledge in Machine Learning

Knowledge mechanisms in data

Moving beyond generic mechanisms of
-outlier elimination
-imputation

Accommodation of relational constraints

Knowledge-based data pre-processing



Constraint – relation (p) of ***not-acceptable*** relationships among variables*:*

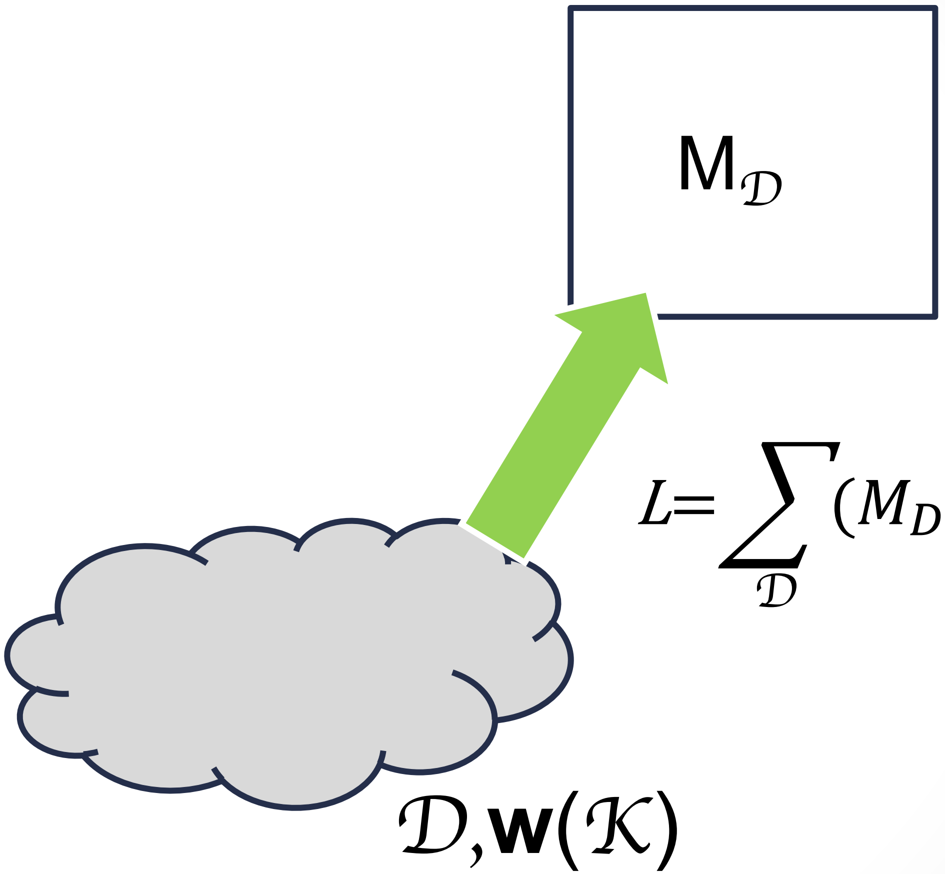
$$high(x_{ki})\ \&\ low(x_{kj}) \quad L_1$$

$$high(x_{ki})\ \&\ high(y_k) \quad L_2$$

....

$d_k = L_1(\mathbf{x}_k)\ \& L_2(\mathbf{x}_k) \& \ldots\ \&\ L_p(\mathbf{x}_k, y_k)$, $w_k = 1 - d_k$

Knowledge-based data pre-processing

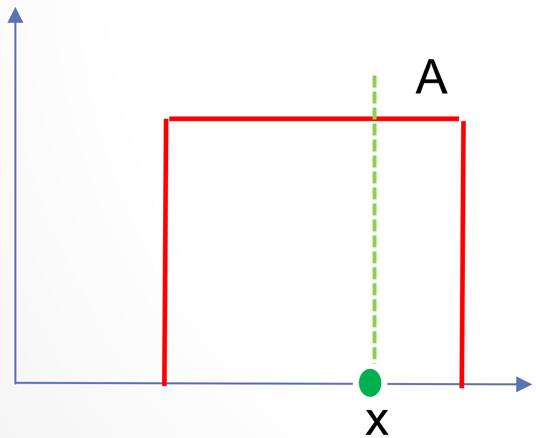


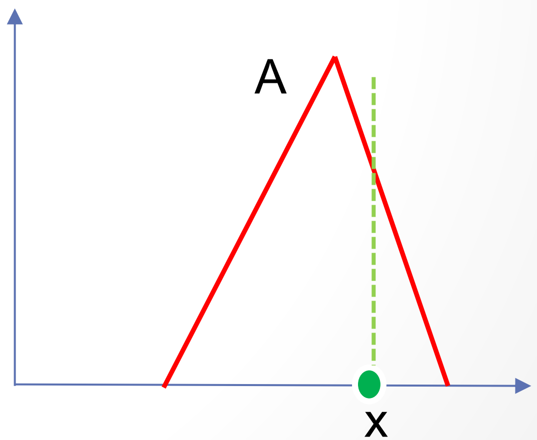
$$L = \sum_{\mathcal{D}} (M_D(\boldsymbol{x}_k) - t_k)^2 \, w_k$$

Coverage and specificity: performance criteria for matching data and information granule

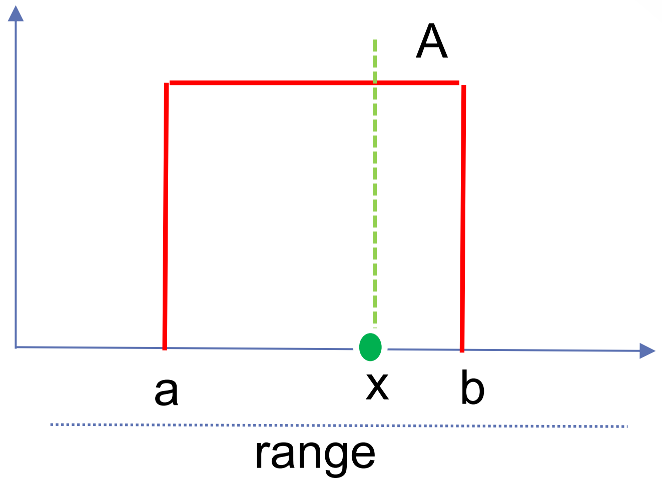
coverage
Datum **x** is included in information granule A

specificity
Expressing how detailed (specific) information granule A is





Coverage and specificity: Characterization (1)



A=[a, b]

coverage

cov(**x**, A)= A(**x**) =

1 if **x** in A,

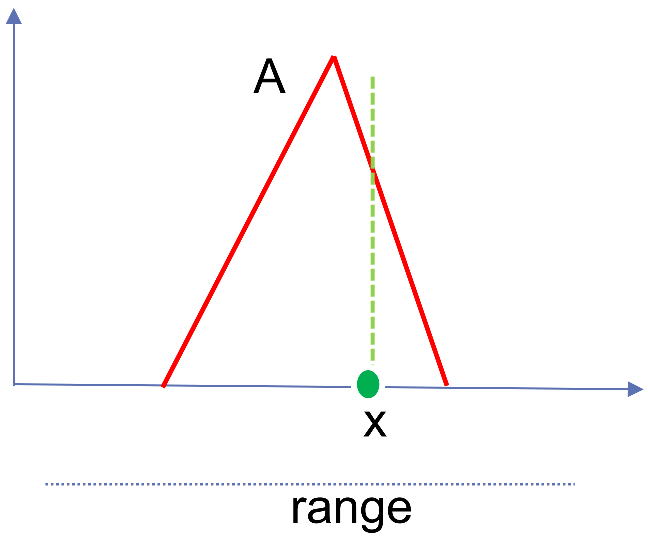
0 otherwise

specificity

sp(A)= τ(length (A)), τ- decreasing function of length of A

=1-(b-a)/range

Coverage and specificity: Characterization (2)



A-fuzzy set

coverage

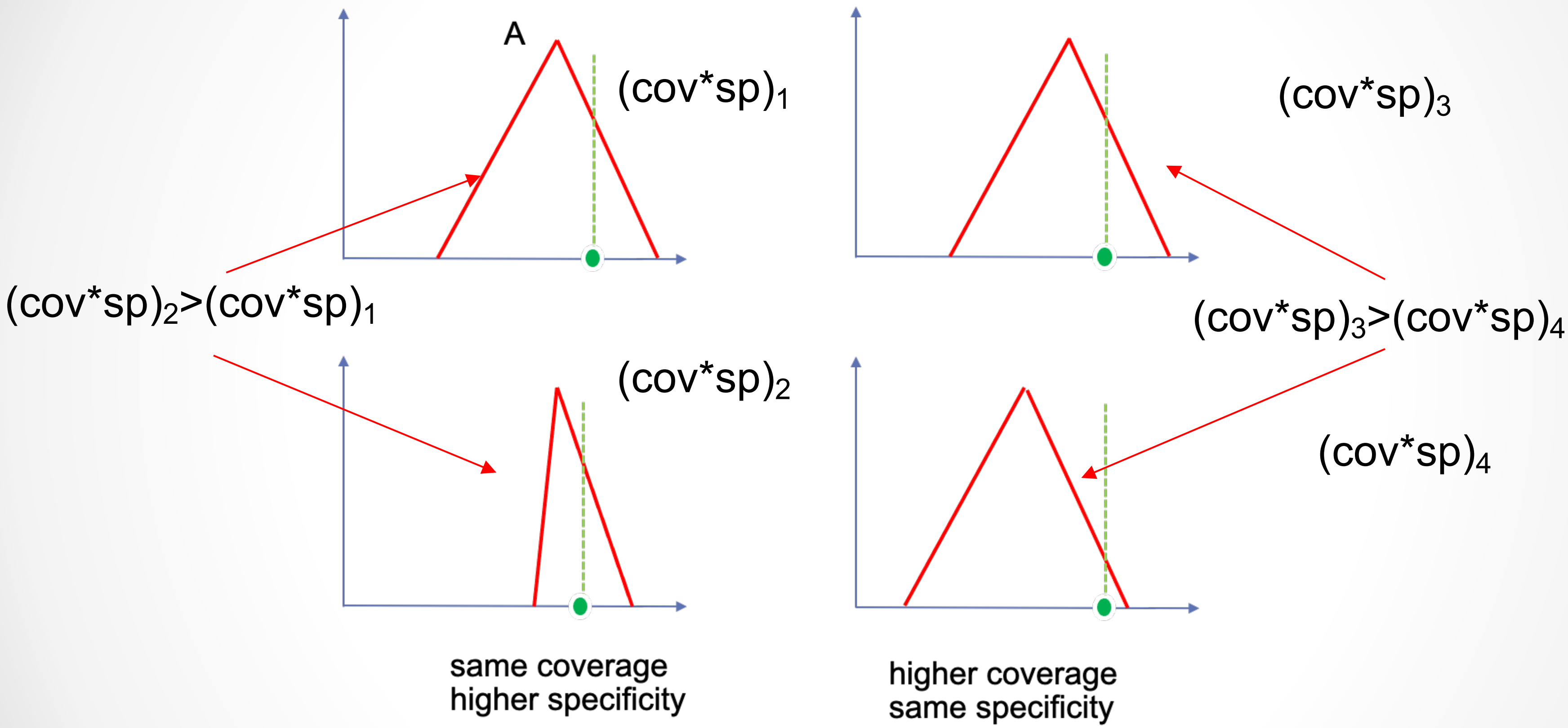
$$\text{cov}(\mathbf{x}, A) = A(\mathbf{x})$$

specificity

$$\text{sp}(A) = \int_0^1 \text{sp}(A_\alpha)\, d\alpha$$

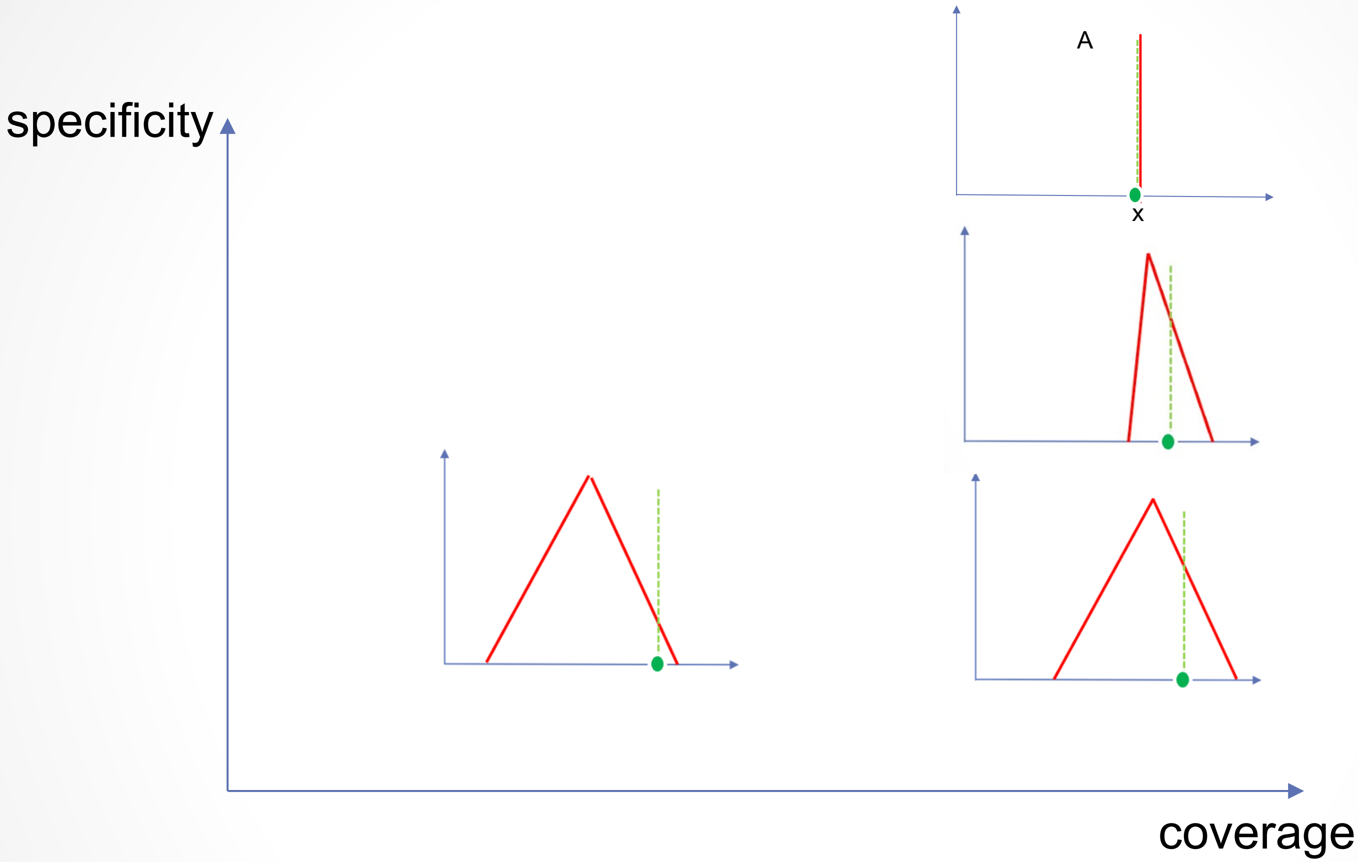
A_α - α cut of A

Coverage and specificity: matching criteria for data and information granule



coverage→max
specificity→max
Conflicting requirements
Overall performance → cov*sp

Coverage and specificity: Support of data by information granule



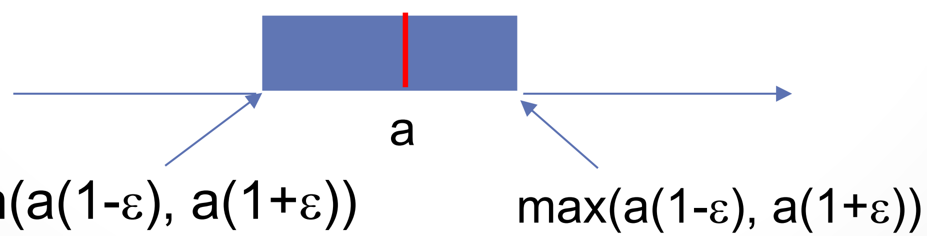
Granular embedding

Elevation of *numeric* parameters **a** of model to

granular parameters A

$$y = M(\boldsymbol{x}; a) \rightarrow Y = M(\mathbf{x}; G(a, \varepsilon)) = M(\mathbf{x}; A)$$

ε- level of information granularity (optimized)



Gaussian Process (GP) regression models

Function space view versus **parameter (weight) space** view

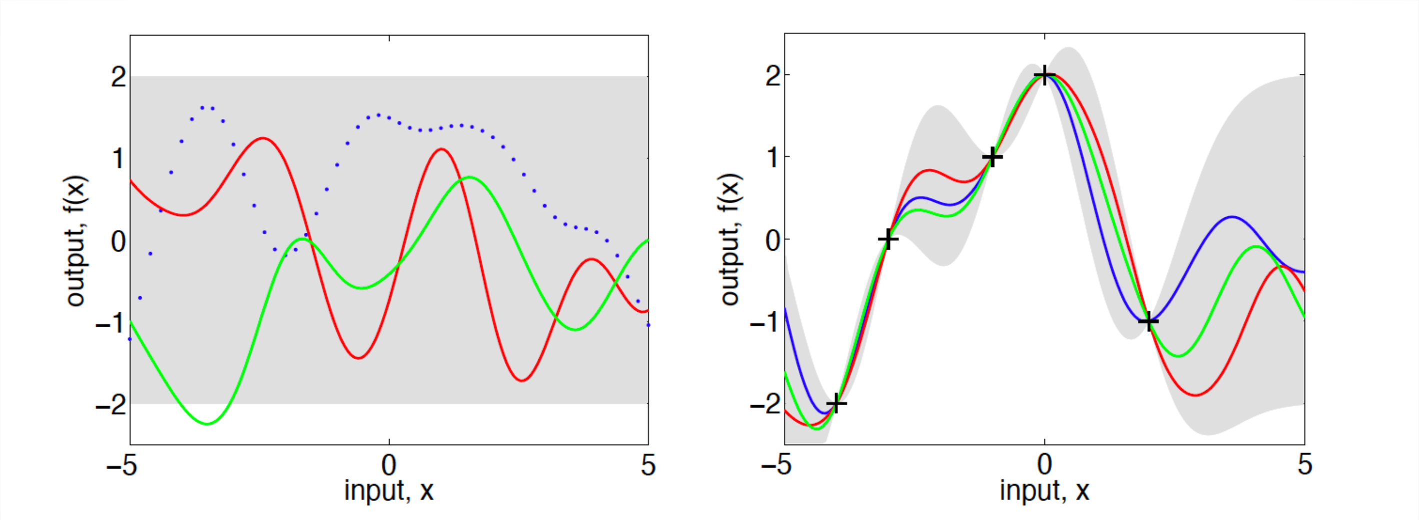
Gaussian process:
collection of random variables; any finite number have a joint Gaussian distribution

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}'))$$

mean function $m(\mathbf{x}) = E[f(\mathbf{x})]$

covariance function $k(\mathbf{x},\mathbf{x}') = E[(f(\mathbf{x})-m(\mathbf{x}))(f(\mathbf{x}')-m(\mathbf{x}'))]$

GP regression models: key ideas



prior

posterior

GP regression models: design (1)

Data $\mathcal{D} = \{(\mathbf{x}_k, target_k)\}$, k=1, 2,..., N

Given $\mathbf{x}^*$, determine output $P(f(\mathbf{x}^*)|f(X))$

$$k(X,X)=\begin{bmatrix} k(x_1,x_1) & \cdots k(x_1,x_N) \\ k(x_2,x_1) & \cdots \\ \cdots & \\ k(x_N,x_1) & \cdots k(x_N,x_N) \end{bmatrix}$$

$$k(X,\mathbf{x}^*)=\begin{bmatrix} k(x_1,x^*) \\ k(x_2,x^*) \\ \cdots \\ k(x_N,x^*) \end{bmatrix}$$

$$f(X)=\begin{bmatrix} target_1 \\ target_k \\ \cdots \\ target_N \end{bmatrix}$$

GP regression models: design (2)

Data $\mathcal{D} = \{(\mathbf{x}_k, \text{target}_k)\}$, k=1, 2,..., N

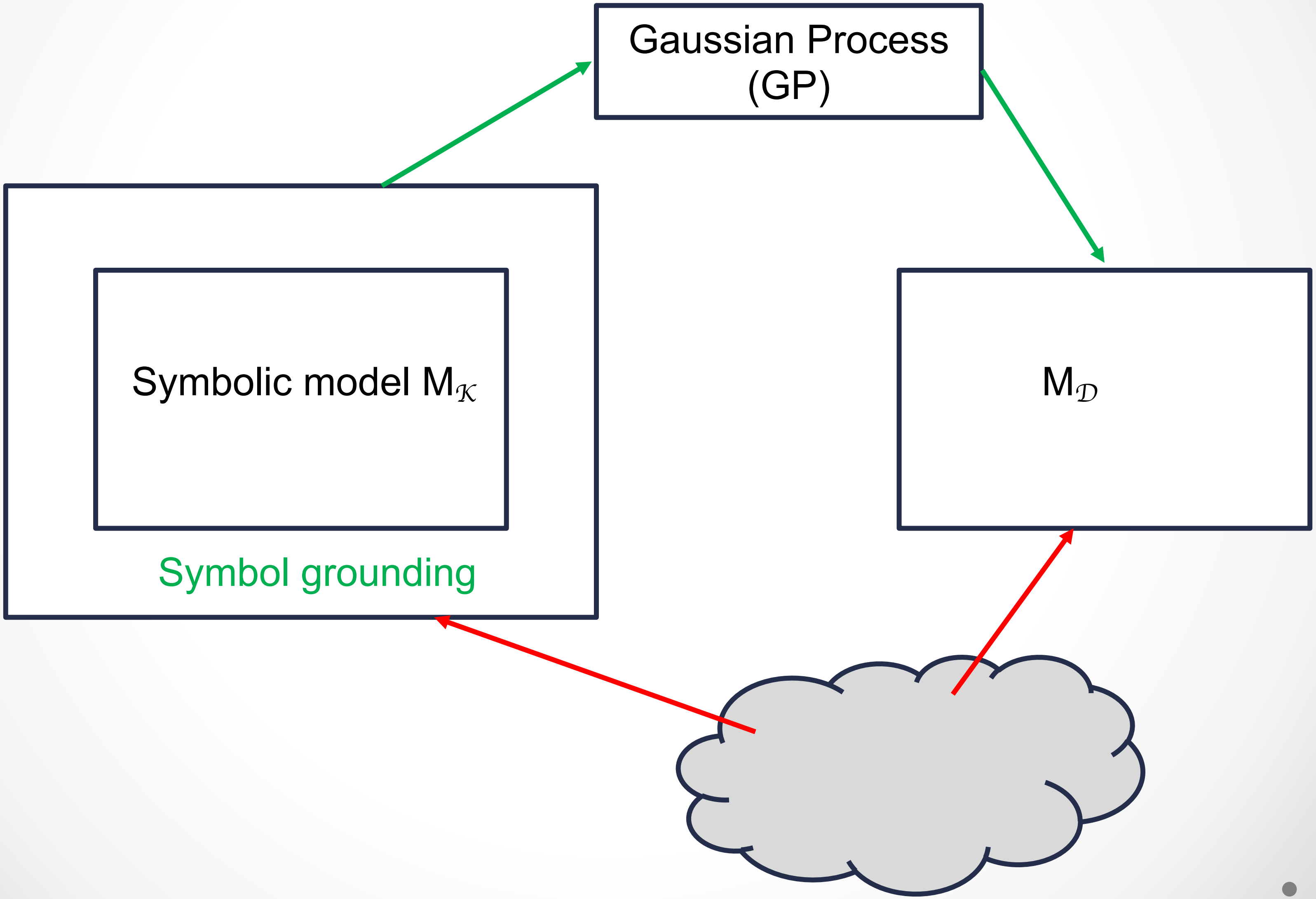
Given $\mathbf{x}^*$, determine output $P(f(\mathbf{x}^*)|f(X))$

$P(f(\mathbf{x}^*)|f(X)) = N(m(\mathbf{x}^*), \sigma(x^*))$

$m(\mathbf{x}^*) = k(X, x^*)^T k(X,X)^{-1} f(X)$

$\sigma(\mathbf{x}^*) = k(\mathbf{x}^*, \mathbf{x}^*) - k(X, \mathbf{x}^*)^T k(X,X)^{-1} k(X, \mathbf{x}^*)$

Data-knowledge ML architecture



Symbolic model $M_{\mathcal{K}}$

Symbolic model

Relationships elicited among symbols expressed over input and output variables

Domain knowledge expressed through symbols

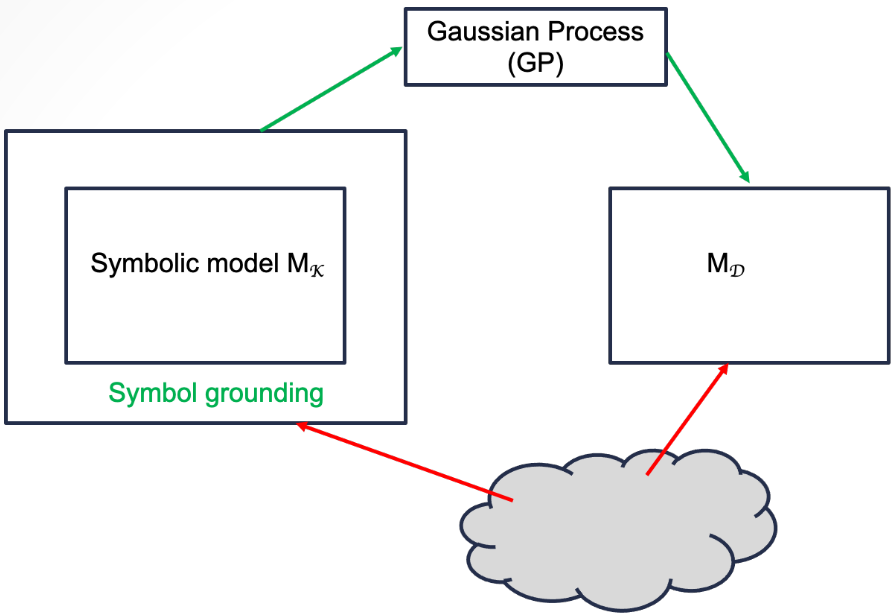
Symbols: logic framework ={linguistic terms, logic connectives, rules...}

-if x is small & y is negative large then output is medium
- larger x entails smaller z

Assumption: symbol ordering (small $<$ medium $<$ large...etc)

Symbol grounding: connecting symbols to their actual meaning

Design (1)



-numeric representatives of symbols $D_{\mathcal{K}} = \{(a_i, b_i, c_i, d_i)\}$

-symbolic model described by rules, viz. tuples

Gaussian Process (GP) built on $D_{\mathcal{K}}$

Optimization of DK with the aid population-based optimization (e.g., PSO)

Design (2)

$\mathcal{D}$={($\mathbf{x}_k$,t_k)},k=1,2,..,N

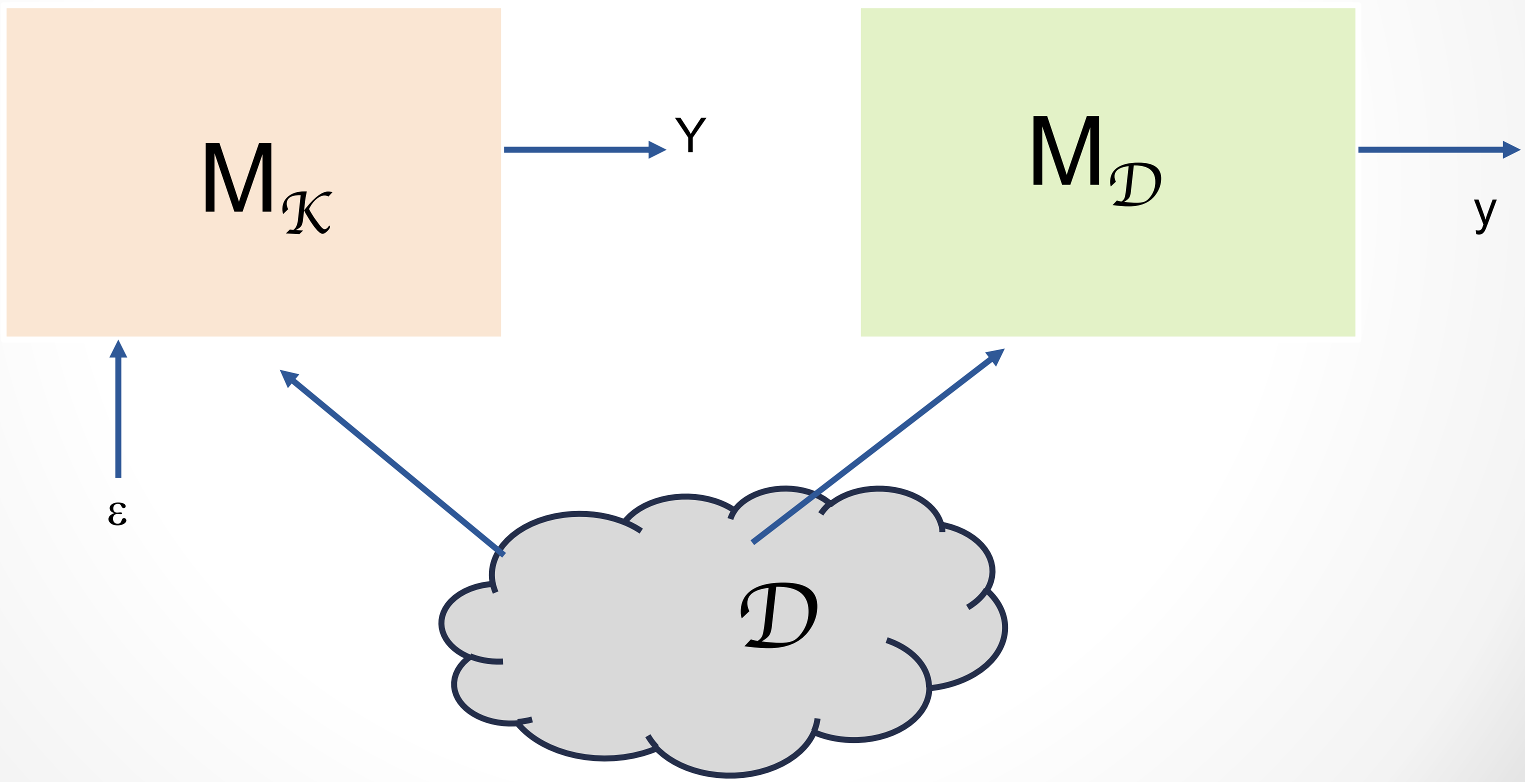
Y_k=Gran[GP(y_k|$\mathbf{x}_k$, $\mathcal{D}_{\mathcal{K}}$)]

Loss function *L* as a fitness function of PSO optimized with regard to a_i, b_i,c_i, d_i

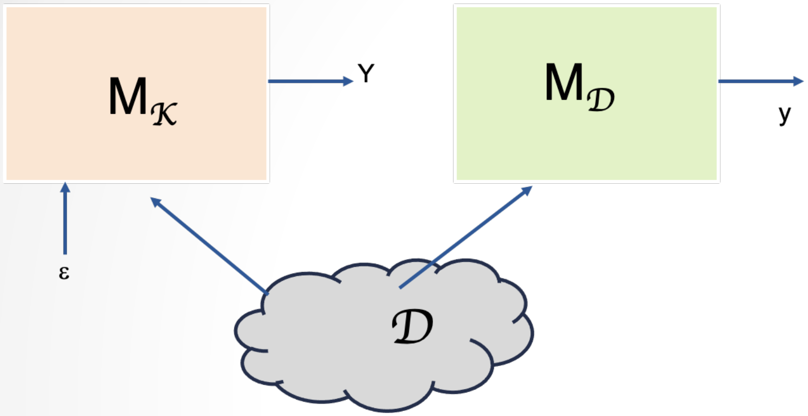
$$L=\lambda \sum_{D} (M_D(\boldsymbol{x}_k) - t_k)^2 + (1-\lambda) \sum_{D} (1 - cov(M_D(x_k), Y_k) sp(Y_k))$$

Min $_{M\mathcal{D},\, \mathbf{w},\lambda}$ L

Parameterized knowledge-based model(1)



Parameterized knowledge-based model(2)



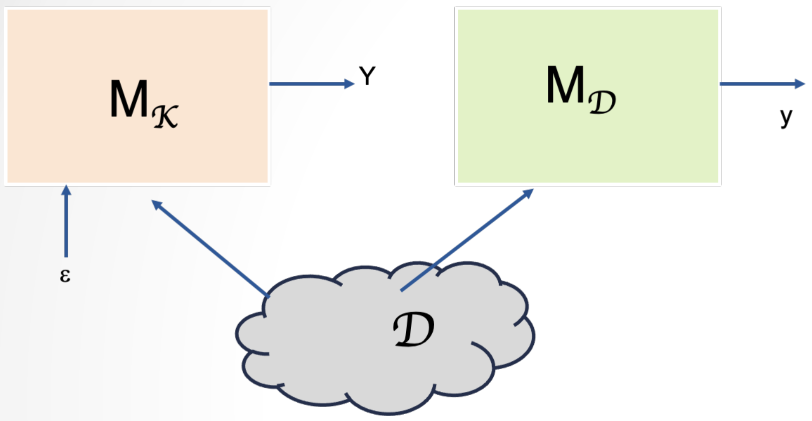
$\mathcal{D}=\{(\mathbf{x}_k,t_k)\}$, k=1,2,...,N

values of parameters of $M_{\mathcal{K}}$ estimated on a basis of $\mathcal{D}$

Granular embedding $\boldsymbol{a} \rightarrow A(\varepsilon)$

$M_{\mathcal{K}}$ $Y_k=M_{\mathcal{K}}(\mathbf{x}_k, A(\varepsilon))$

Parameterized knowledge-based model(2)



Granular embedding a→ A(ε)
a-nominal values of parameters

$M_{\mathcal{K}}$ $Y_k = M_{\mathcal{K}}(\mathbf{x}_k, A(\varepsilon))$

$M_{\mathcal{D}}$ $y_k = M_{\mathcal{D}}(\mathbf{x}_k, w)$

$\mathcal{D} = \{(\mathbf{x}_k, t_k)\}$, k=1,2,...,N

$$L = \lambda \sum_{D} (M_D(\mathbf{x}_k) - t_k)^2 + (1-\lambda) \sum_{D} (1 - cov(M_D(x_k), Y_k(\varepsilon)) sp(Y_k(\varepsilon)))$$

$\mathrm{Min}_{M,\lambda,\varepsilon} L$

Conclusions

Data and knowledge as an essential unified Machine Learning design framework

Key challenges and opportunities:

- Knowledge representation
- Integration of knowledge in the learning environment

The role of Granular Computing