Fuzzy Logic in Machine Learning

Eyke Hüllermeier
Knowledge Engineering & Bioinformatics Lab
Department of Mathematics and Computer Science
Marburg University, Germany

Philipps Universität Marburg

PROLOGUE

Use of concepts, tools, and techniques
from fuzzy sets and fuzzy logic to
enhance machine learning methods

Fuzzy Logic for Machine Learning

things in-between, e.g.,
neuro-fuzzy sytems

Machine Learning for Fuzzy Logic

Use of machine learning methods to support the
data-driven design of (rule-based) fuzzy systems
(help to overcome the knowledge acquisition bottleneck)

AGENDA

1. Introduction and Background
   a. Machine Learning
   b. Fuzzy Logic
2. Fuzzy Modeling in Machine Learning
3. Fuzzy Pattern Trees
4. Summary & Conclusions

This is a purely non-technical talk!

BACKGROUND ON MACHINE LEARNING

- The major goal of machine learning (ML) is to create computer systems that are able to automatically **improve their performance on the basis of experience**, and to adapt their behavior to changing environmental conditions → „creating machines that can learn"

Computer Poker

BACKGROUND ON MACHINE LEARNING

- The major goal of machine learning (ML) is to create computer systems that are able to automatically **improve their performance on the basis of experience**, and to adapt their behavior to changing environmental conditions → „creating machines that can learn"

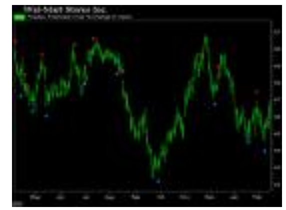
- During 30-40 years of research, ML has developed
  - **sound theoretical frameworks** (such as statistical learning theory, algorithmic learning theory, MDL, PAC learning, …) , as well as
  - a **large variety of methods** (including kernel methods, neural networks, statistical models, Bayesian methods, graphical models, rules and decision trees, instance-based learning, inductive logic programming, relational learning, …)

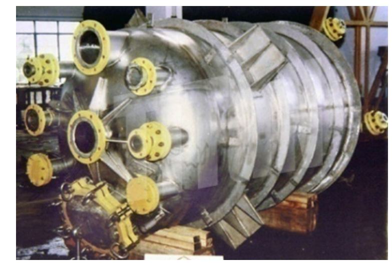
MACHINE LEARNING APPLICATIONS

business (CRM, response prediction, …)

smart environments

banking and finance (stock prediction, fraud detection, …)





technical systems (diagnosis, control, monitoring, …)

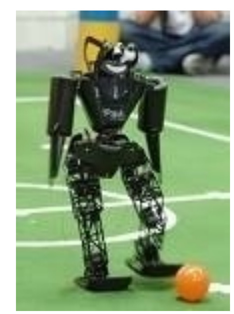
Internet (information retrieval, email classification, personalization, …)

biometrics (person identification, …)

medicine (diagnosis, prosthetics, ...)

media (speech/image recognition, video mining, …)

bioinformatics, genomic data analysis



games (e.g. poker, soccer, …)

TOPICS OF ONGOING RESEARCH

Active learning
Bagging/boosting
Constructive induction
Cooperative learning
Cost-sensitive learning
Co-training
Hierarchical classification
Deep learning
Dimensionality reduction

Ensemble learning
Graph-based learning
Incremental induction
Kernel methods
Learning on data streams
Ordered classification
Manifold learning
Multi-instance learning
Multi-strategy learning

Multi-task learning
Multi-view learning
On-line learning
Preference learning
Ranking
Relational learning
Semi-supervised learning
Transfer learning
…

… and a lot more!

MODEL INDUCTION

The core of machine learning is **model induction**, i.e., learning (predictive) models from observed training data.

A simple setting of supervised learning: Given (i.i.d.) training data

$$\mathcal{D} = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \} \subset (\mathcal{X} \times \mathcal{Y})^n$$

and a hypothesis space $\mathcal{H} \subset \{h : \mathcal{X} \to \mathcal{Y}\}$, induce a model

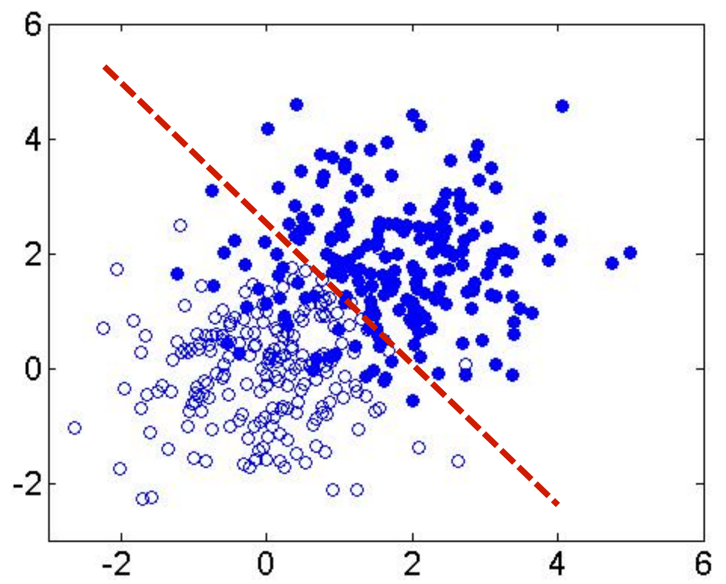
$$h^* \in \arg\min_{h \in \mathcal{H}} \int_{\mathcal{X} \times \mathcal{Y}} \underbrace{\ell\,(h(x), y)}_{\text{loss function}} \; d\,\underbrace{\mathrm{P}(X, Y)}_{\text{unknown data-generating process}}$$

MODEL INDUCTION

Example (classifier learning):

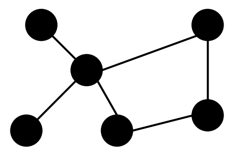
$$\mathcal{X} \subseteq \mathbb{R}^d, \quad \mathcal{Y} = \{y_1, y_2, \ldots, y_k\},$$

$$\ell(\hat{y}, y) = \begin{cases} 0 & \hat{y} = y \\ 1 & \hat{y} \neq y \end{cases}$$



LEARNING WITH/FROM STRUCTURED DATA

$$h : \text{structured input space} \xrightarrow{\mathcal{P}} \mathcal{Y} = \{y_1, y_2, \ldots, y_k\}$$



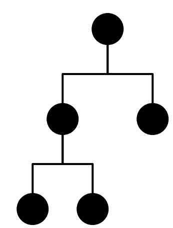
$$\longrightarrow \quad \mathcal{Y} = \{-1, +1\}$$

(e.g. a molecule)

$$h : \text{structured input space} \xrightarrow{\mathcal{P}} \text{structured output space}$$

$$\{A,\, B,\, C,\, \ldots\}^* \quad \longrightarrow$$

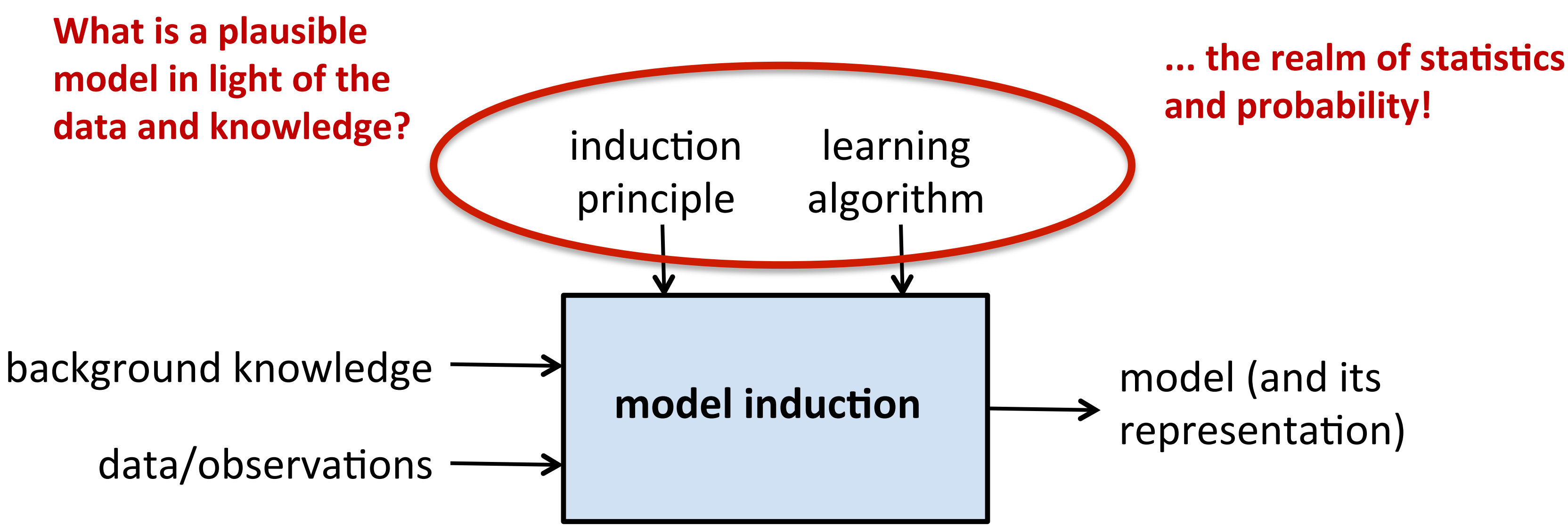
(e.g. a sentence)



(e.g. a parse tree)

INGREDIENTS OF MODEL INDUCTION

- Model induction = **searching** for an optimal model among a set of **candidate models** as specified by the **hypothesis (model) space**

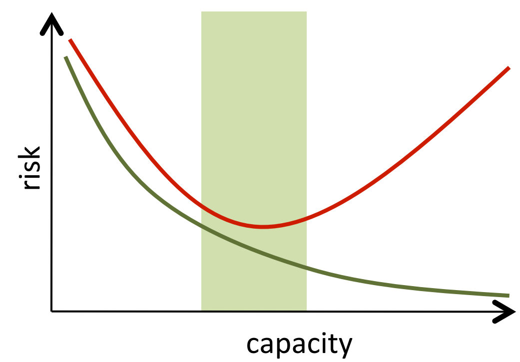


CAPACITY CONTROL

The key to successful learning/generalization is a proper **capacity control**: The model class must be flexible enough (to allow approximation of the pointwise loss-minimizer) but not too flexible (to prevent overfitting the data)

Typical bound on the true risk: With probability $1-\delta$

$$\underbrace{R(h)}_{\text{true risk}} \;\le\; \underbrace{R_{emp}(h)}_{\text{empirical risk}} + \underbrace{\sqrt{2\frac{\mathrm{VC}(\mathcal{H})\log\left(\frac{2e|\mathcal{D}|}{\mathrm{VC}(\mathcal{H})}\right)+\log\left(\frac{2}{\delta}\right)}{|\mathcal{D}|}}}_{\text{correction depending on capacity and sample size}}$$



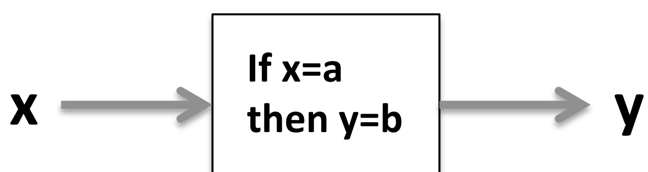
→ the more **prior knowledge** is available (about the **right type of model** and the **right level of flexibility**), the easier learning becomes !

BEYOND ACCURACY

- Apart from accuracy, a model can be judged on the basis of many other properties, notably **interpretability**.

- In many application domains (e.g., medical diagnosis), **„black box“ models** like neural networks are refused.

- Instead, more comprehensible and transparent **„white box“ models** are sought.



AGENDA

1. Introduction and Background
   a. Machine Learning
   b. Fuzzy Logic
2. Fuzzy Modeling in Machine Learning
3. Fuzzy Pattern Trees
4. Summary & Conclusions

What is Fuzzy Logic?

- Fuzzy Logic in the **narrow sense** is a special type of multi-valued logic (branch of mathematical logic, see e.g. works by P. Hajek).

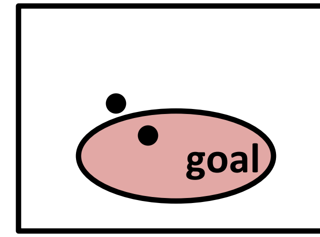
- Fuzzy Logic in the **wide sense** is considered as a collection of methods, tools, and techniques for modeling and reasoning about **vagueness** and **vague concepts.**

  „Fuzzy logic as a mathematical framework for formalizing (human-like) approximate reasoning.“

IS TRUTH A BIVALENT CONCEPT?

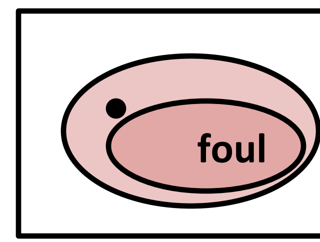
P: „The Wembley goal was indeed a goal.“

… uncertain but indeed either true of false!

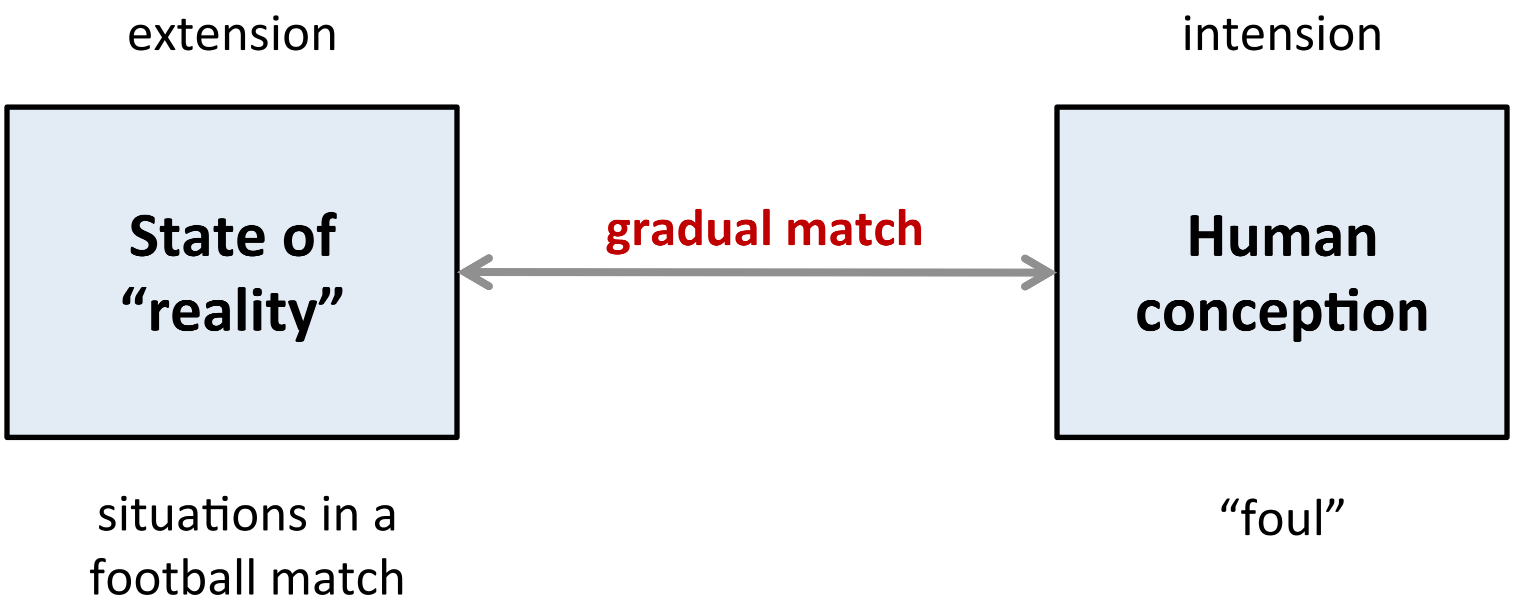


P: „This was a foul by Adriano.“

Truth degree might be unclear, since „foul“ is a vague concept …



IS TRUTH A BIVALENT CONCEPT?



An intelligent system expected to **mimic the learning or decision making behavior of a human being** should have a proper model of human conception.

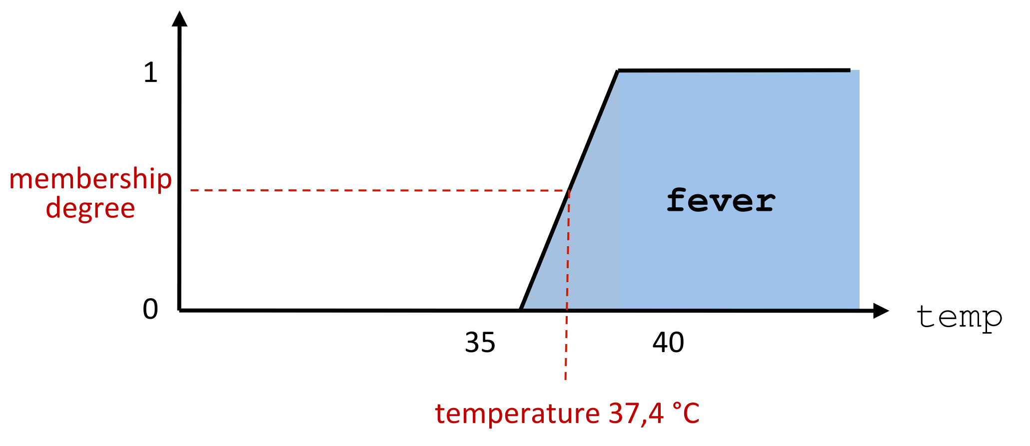
MEMBERSHIP FUNCTIONS

A fuzzy subset of a referential set X is characterized by its

$$\textbf{membership function } A : X \longrightarrow \mathcal{L} \ ,$$

where $\mathcal{L}$ is a complete lattice. For each $\boldsymbol{x} \in X$, $A(\boldsymbol{x})$ is called the degree of membership of $\boldsymbol{x}$ in A.

$\mathcal{L}$ is usually the unit interval $[0, 1]$ (equipped with the natural order relation on the reals). However, it can also be an ordinal or partially ordered scale.



FUZZY LOGIC TOOLS

Fuzzy Logic offers sound theoretical foundations as well as tools and techniques for **modeling and (approximate) reasoning with vague concepts**:

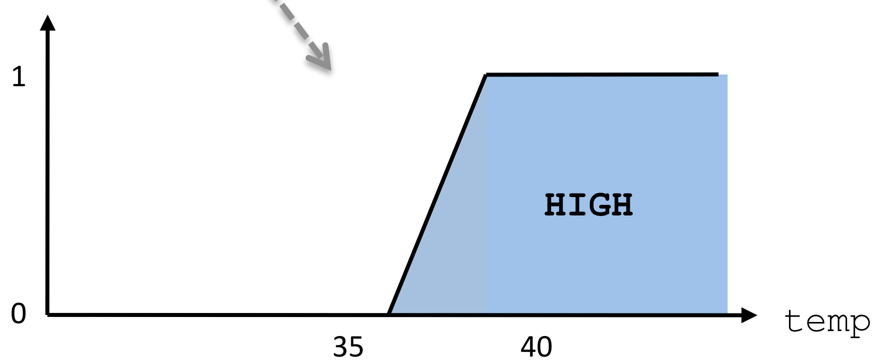
- Fuzzy sets and fuzzy relations
- Generalized logical connectives
  - conjunction,
  - disjunction,
  - …
- Fuzzy rules
- Generalized aggregation operators
- Fuzzy quantifiers („for most“, „for some“, …)
- Generalized (non-additive) measures
- …

The calculus of fuzzy logic is **truth-functional** (in contrast to probability) !

FUZZY RULES

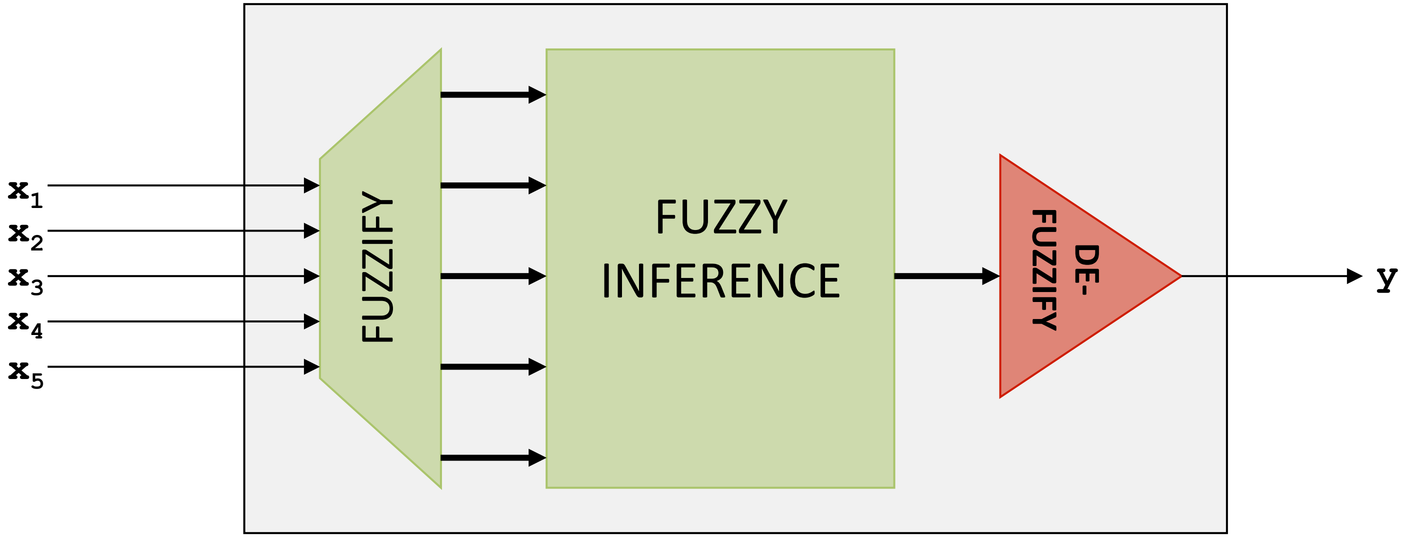
IF (x_1 is A_1) AND (x_2 is A_2) THEN (y is B)

temp is HIGH



FUZZY RULES

IF $(x_1$ is $A_{11})$ AND $(x_2$ is $A_{21})$ THEN $(y$ is $B_1)$
IF $(x_1$ is $A_{12})$ AND $(x_2$ is $A_{22})$ THEN $(y$ is $B_2)$
IF $(x_1$ is $A_{13})$ AND $(x_2$ is $A_{23})$ THEN $(y$ is $B_3)$
IF $(x_1$ is $A_{14})$ AND $(x_2$ is $A_{24})$ THEN $(y$ is $B_4)$
IF $(x_1$ is $A_{15})$ AND $(x_2$ is $A_{25})$ THEN $(y$ is $B_5)$



AGENDA

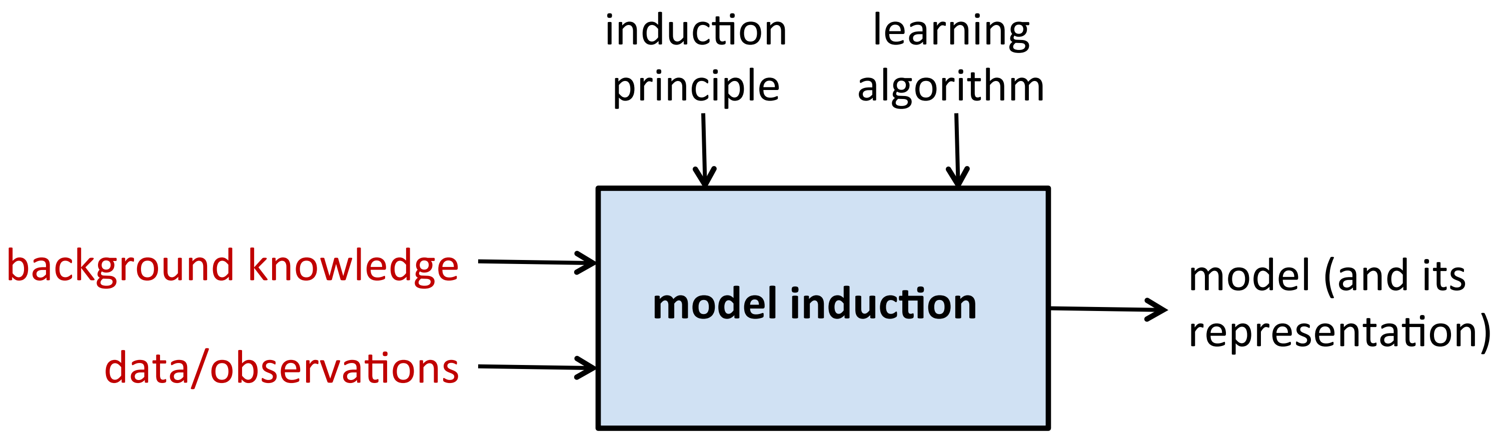
1. Introduction and Background
   a. Machine Learning
   b. Fuzzy Logic
2. **Fuzzy Modeling in Machine Learning**
3. Fuzzy Pattern Trees
4. Summary & Conclusions

POTENTIAL CONTRIBUTIONS OF FL TO ML

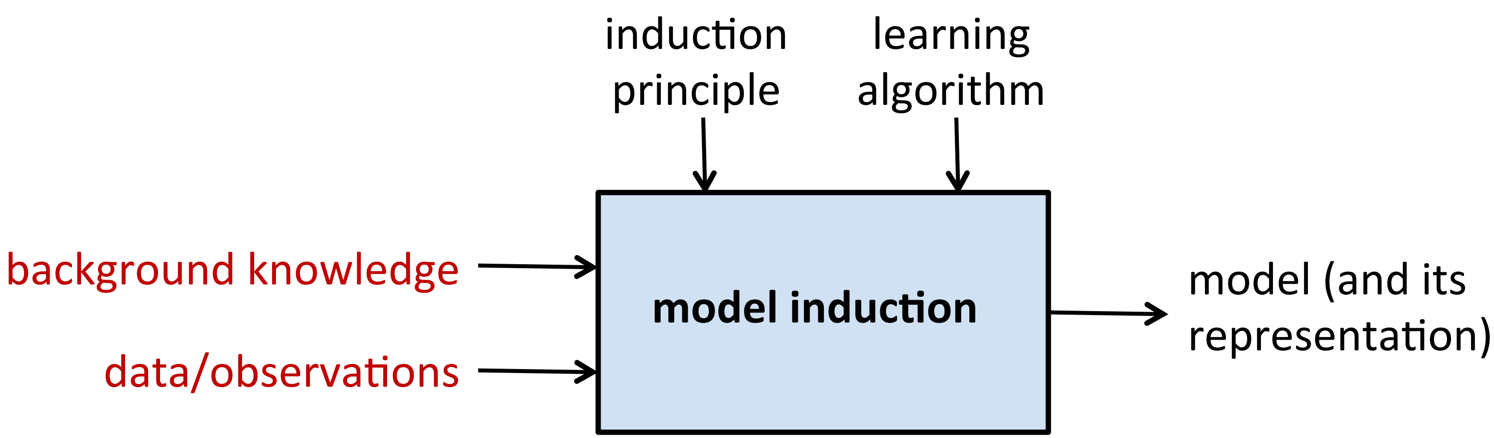
- Graduality
- Granularity
- Robustness
- Representation of Uncertainty
- Incorporation of Background Knowledge
- Aggregation, Combination and Information Fusion

E.H. Fuzzy Machine Learning and Data Mining. WIREs Data Mining and Knowledge Discovery, 2011.

FUZZY MODELING IN MACHINE LEARNING



What you get out strongly depends on what you put in!



There is a need for modeling !

$$y = h(x_1, x_2, \ldots, x_m)$$

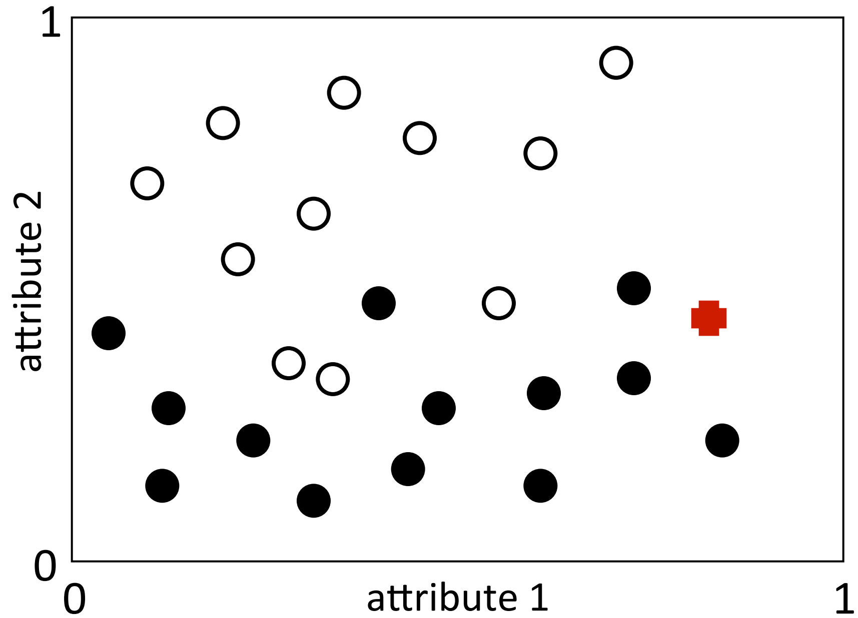
output modeling

specification of the model space (with the right capacity)

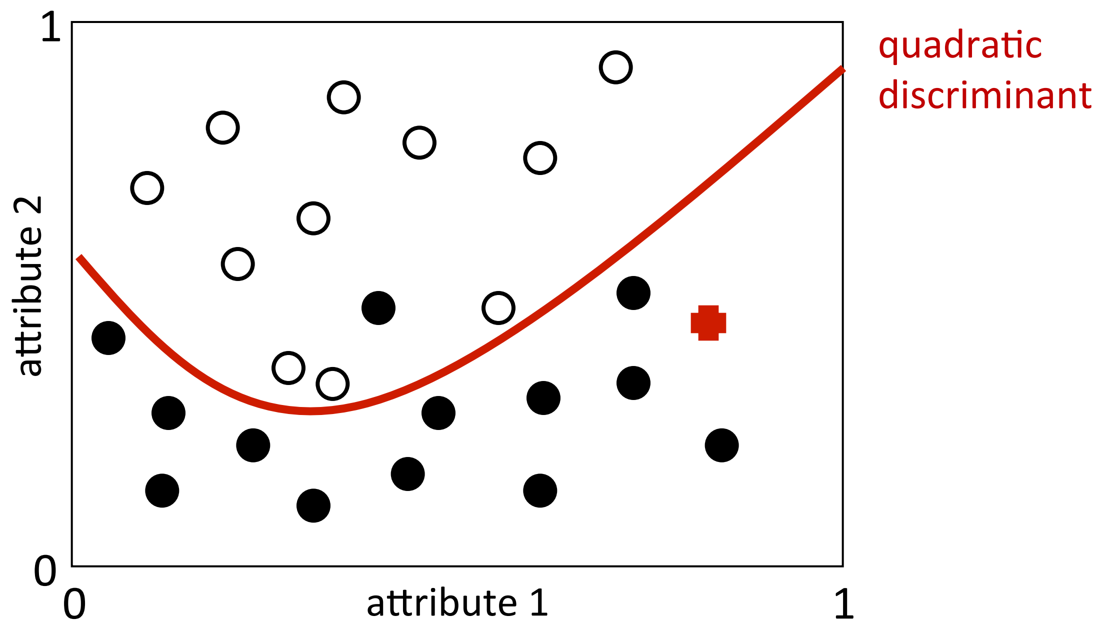
feature modeling and selection

FUZZY MODELING IN MACHINE LEARNING

Learning/generalization without a bias (prior knowledge about the model to be learned) is impossible!

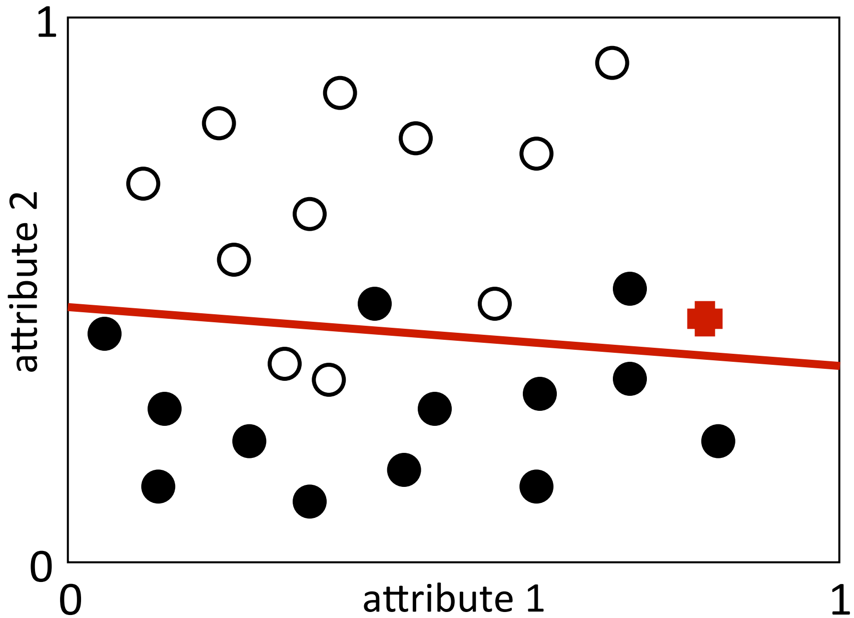


Learning/generalization without a bias (prior knowledge about the model to be learned) is impossible!



FUZZY MODELING IN MACHINE LEARNING

Learning/generalization without a bias (prior knowledge about the model to be learned) is impossible!



best “explanation” of the data in terms of a linear model

→ significantly different results for the same data!
→ learning method should support the easy incorporation of an expert‘s background knowledge

FUZZY MODELING IN MACHINE LEARNING

Fuzzy logic supports a seamless transition between „knowledge-driven" and „data-driven" model specification!

IF (x_1 is A_{11}) AND (x_2 is A_{21}) THEN (y is B_1)
IF (x_1 is A_{12}) AND (x_2 is A_{22}) THEN (y is B_2)
IF (x_1 is A_{13}) AND (x_2 is A_{23}) THEN (y is B_3)
IF (x_1 is A_{14}) AND (x_2 is A_{24}) THEN (y is B_4)
IF (x_1 is A_{15}) AND (x_2 is A_{25}) THEN (y is B_5)

Complete model specified by hand, using linguistic modeling techniques.

FUZZY MODELING IN MACHINE LEARNING

Fuzzy logic supports a seamless transition between „knowledge-driven“ and „data-driven“ model specification!

$$
\begin{aligned}
&\mathbf{IF}\ (x_1 \text{ is } A_{11})\ \mathbf{AND}\ (x_2 \text{ is } A_{21})\ \mathbf{THEN}\ (y \text{ is } B_1)\\
&\mathbf{IF}\ (x_1 \text{ is } A_{12})\ \mathbf{AND}\ (x_2 \text{ is } A_{22})\ \mathbf{THEN}\ (y \text{ is } B_2)\\
&\mathbf{IF}\ (x_1 \text{ is } A_{13})\ \mathbf{AND}\ (x_2 \text{ is } A_{23})\ \mathbf{THEN}\ (y \text{ is } B_3)\\
&\mathbf{IF}\ (x_1 \text{ is } A_{14})\ \mathbf{AND}\ (x_2 \text{ is } A_{24})\ \mathbf{THEN}\ (y \text{ is } B_4)\\
&\mathbf{IF}\ (x_1 \text{ is } A_{15})\ \mathbf{AND}\ (x_2 \text{ is } A_{25})\ \mathbf{THEN}\ (y \text{ is } B_5)
\end{aligned}
$$

Model structure (qualitative part) specified by hand,
membership functions (quantitative part) learned from data.

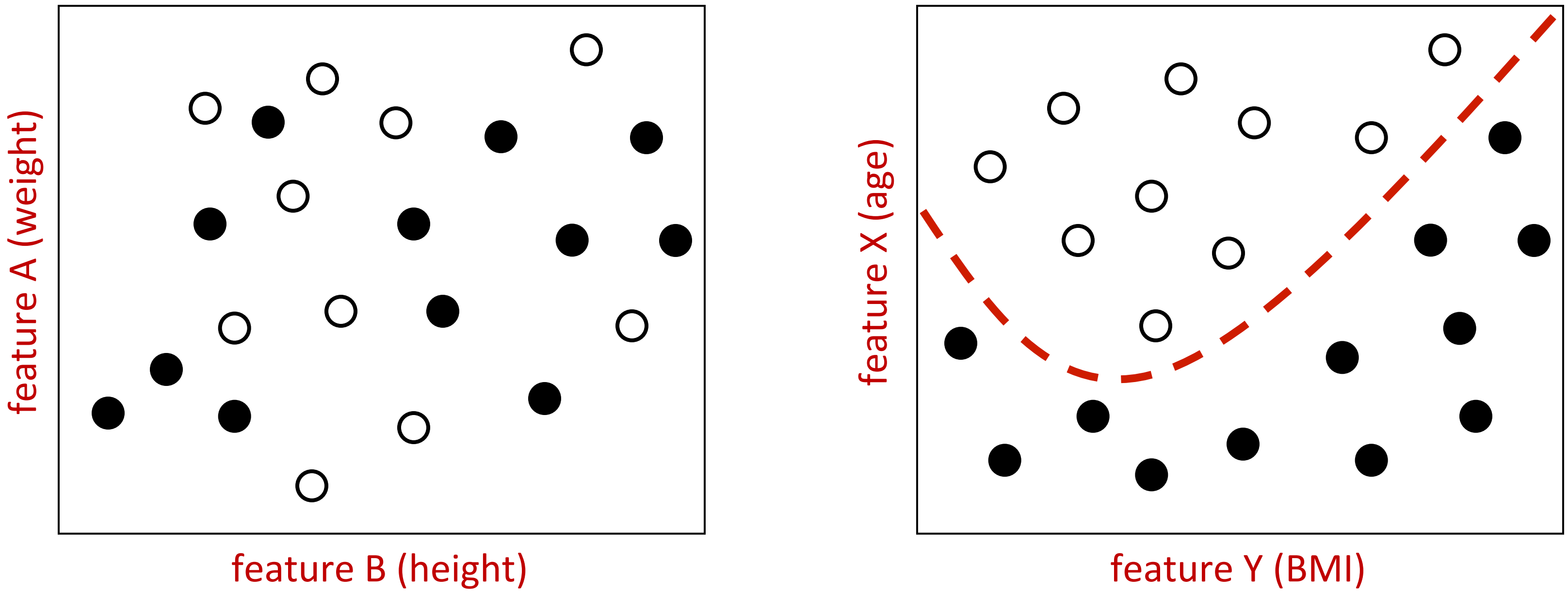
FUZZY MODELING IN MACHINE LEARNING

Fuzzy logic supports a seamless transition between „knowledge-driven“ and „data-driven“ model specification!

```
IF (x₁ is A₁₁) AND (x₂ is A₂₁) THEN (y is B₁)
IF (x₁ is A₁₂) AND (x₂ is A₂₂) THEN (y is B₂)
IF (x₁ is A₁₃) AND (x₂ is A₂₃) THEN (y is B₃)
IF (x₁ is A₁₄) AND (x₂ is A₂₄) THEN (y is B₄)
IF (x₁ is A₁₅) AND (x₂ is A₂₅) THEN (y is B₅)
```

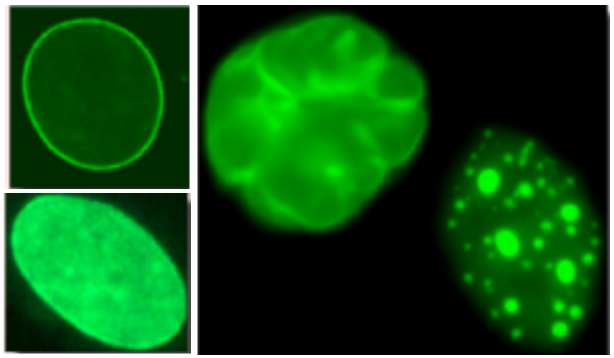
Model structure (qualitative part) partly specified by hand, partly learned from data.

FEATURE MODELING AND SELECTION



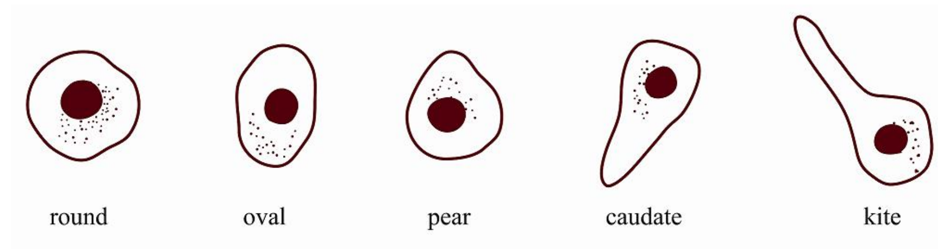
A description of instances in terms of features X, Y allows for a much better **discrimination** between the two classes !

FUZZY FEATURE MODELING

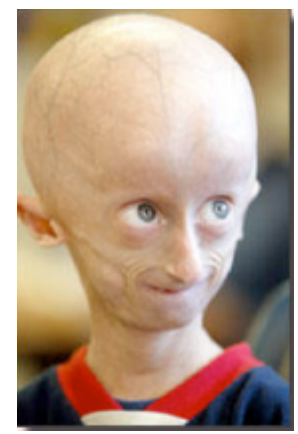


healthy pathological

shapes of cells



Hudchinson-Gilford syndrome



[Thibault et al. *Cells nuclei classification using* ***shape*** *and* ***texture*** *indexes*. WSCG 2008.]

FUZZY FEATURE MODELING

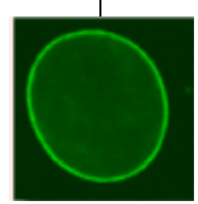
$$y = \alpha + \beta_1 x_1 + \beta_2 \boxed{x_2} + \ldots + \beta_m x_m$$

meaningful
high-level feature (e.g., roundness)

expert knowledge

LINGUISTIC MODELING
FUZZY INFERENCE

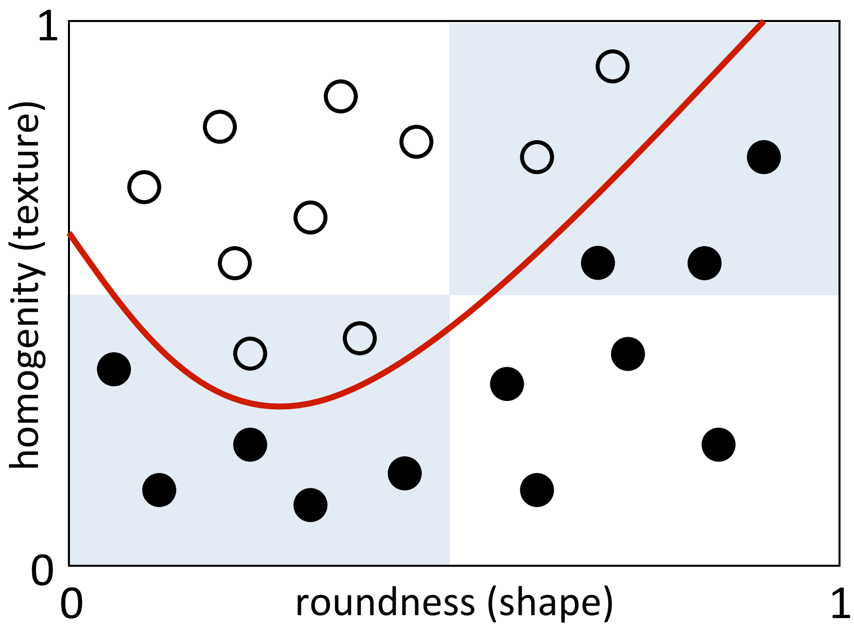
low-level feature z_1 z_2 z_3 z_4 ... z_d



Reduces dimensionality and increases interpretability!

FUZZY FEATURE MODELING

Use of gradual (instead of binary) features can increase **discriminative power**!



	homo-geneous	not homo-geneous
round	● ○	●
not round	○	● ○

separable

non-separable

MODELING DATA/OBSERVATIONS

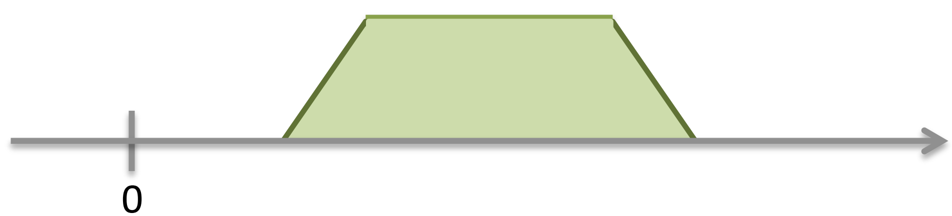
unordered, categorical
(classification)

ordered, categorical
(ordinal classification)

* < ** < *** < **** < *****

numerical
(regression)

fuzzy
(modeling imprecise data)



FUZZY MODELING IN MACHINE LEARNING

Despite being largely „data-driven“, machine learning is not a „knowledge-free“ methodology. Instead, successful learning requires a-priori **background knowledge** and a proper **modeling**

- of the **data** (which is often overlooked)
- and the underlying **hypothesis space** (type of model),

and **fuzzy logic can be very useful in this regard** (incorporating expert knowledge into the learning process) !

→ fuzzy logic is complementary to statistics and probability!

→ this aspect (like the „true strengths“ of FL in general) are still not sufficiently emphasized in the current literature

AGENDA

1. Introduction and Background
   a. Machine Learning
   b. Fuzzy Logic
2. Fuzzy Modeling in Machine Learning
3. **Fuzzy Pattern Trees**
4. Summary & Conclusions

FUZZY PATTERN TREES

- FPT is a type of fuzzy model that was independently introduced in
  - Z. Huang, TD. Gedeon, and M. Nikravesh. Pattern trees induction: A new machine learning approach. IEEE TFS 16(4), 2008.
  - Y. Yi, T. Fober and E.H. Fuzzy Operator Trees for Modeling Rating Functions. Int. J. Comp. Intell. and Appl. 8(1), 2009.
- It has recently been further developed in
  - R. Senge and E.H. Pattern Trees for Regression and Fuzzy Systems Modeling. Proc. WCCI-2010, Barcelona, Spain, 2010.
  - R. Senge and E.H. Top-Down Induction of Fuzzy Pattern Trees. IEEE TFS, 19(2), 2011.
- **Two main motivations:**
  - disadvantages of rule-based fuzzy models
  - quality assessment in production (measurement robotics)

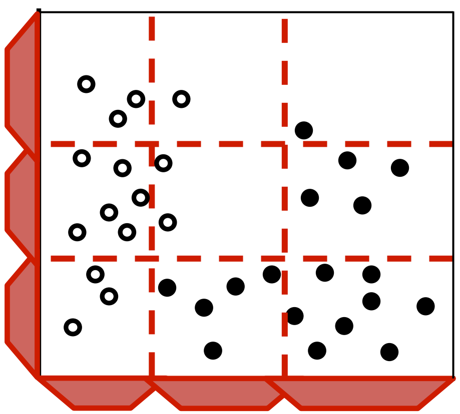
INTERPRETABILITY OF RULE-BASED MODELS

- „reasonable“ fuzzy sets are normally not guaranteed by a learning algorithm (and doing so may compromise accuracy);
- without knowing the membership functions, the true meaning remains ambiguous;
- interpretability is further compromised by
  - length and number of rules,
  - the interaction between rules,
  - inference mechanism (logical operators, rule weighing, …)

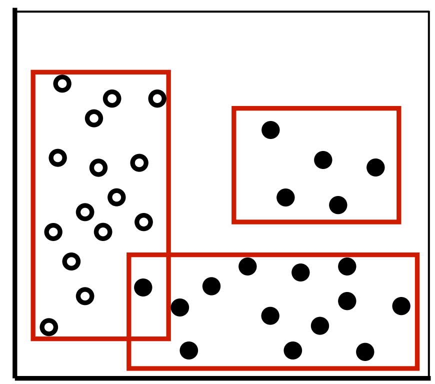
After all, fuzzy rule-based models might be less „white-box“ than many people tend to believe ...

KEY PROBLEMS OF RULE-BASED METHODS

- flat structure of the model (→ curse of dimensionality)



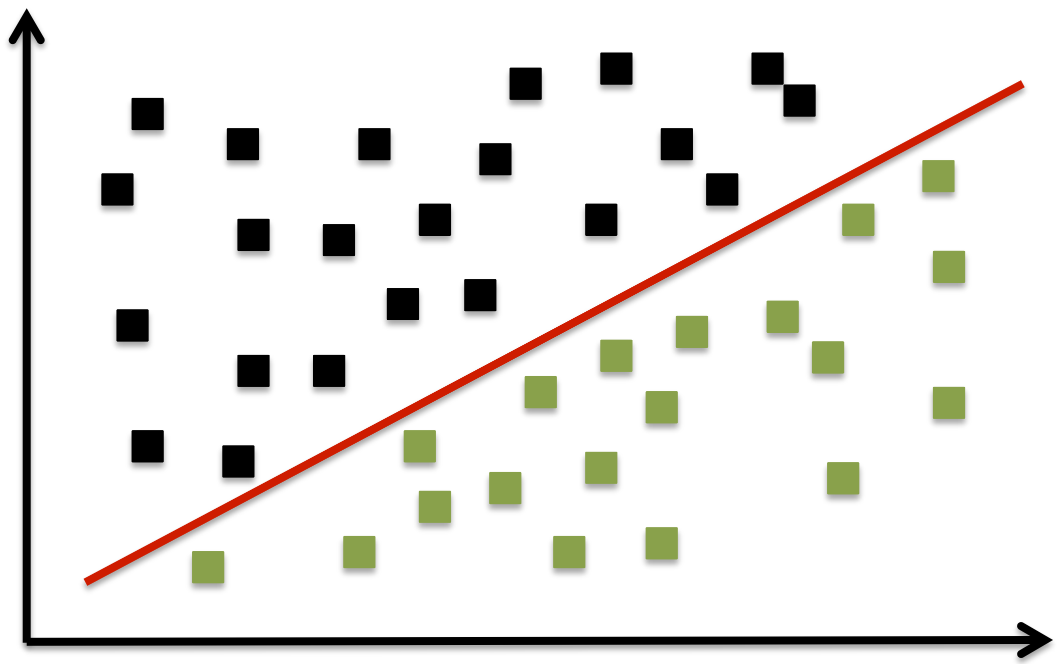
grid-based models



sequential covering

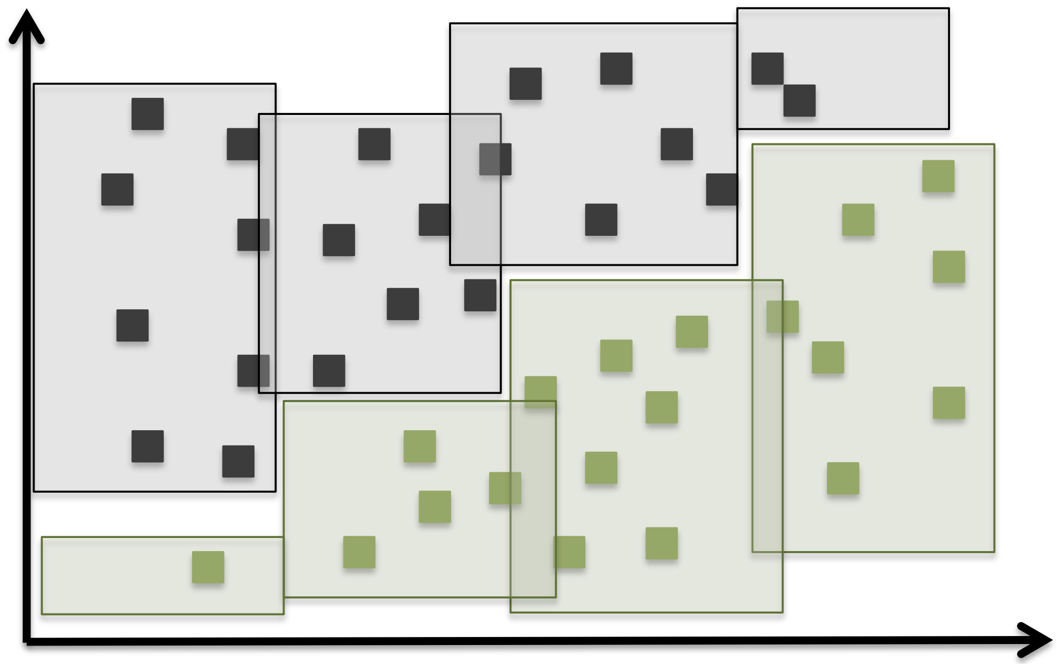
KEY PROBLEMS OF RULE-BASED METHODS

- flat structure of the model (→ curse of dimensionality)
- restriction to (quasi-)axis-parallel decision boundaries



KEY PROBLEMS OF RULE-BASED METHODS

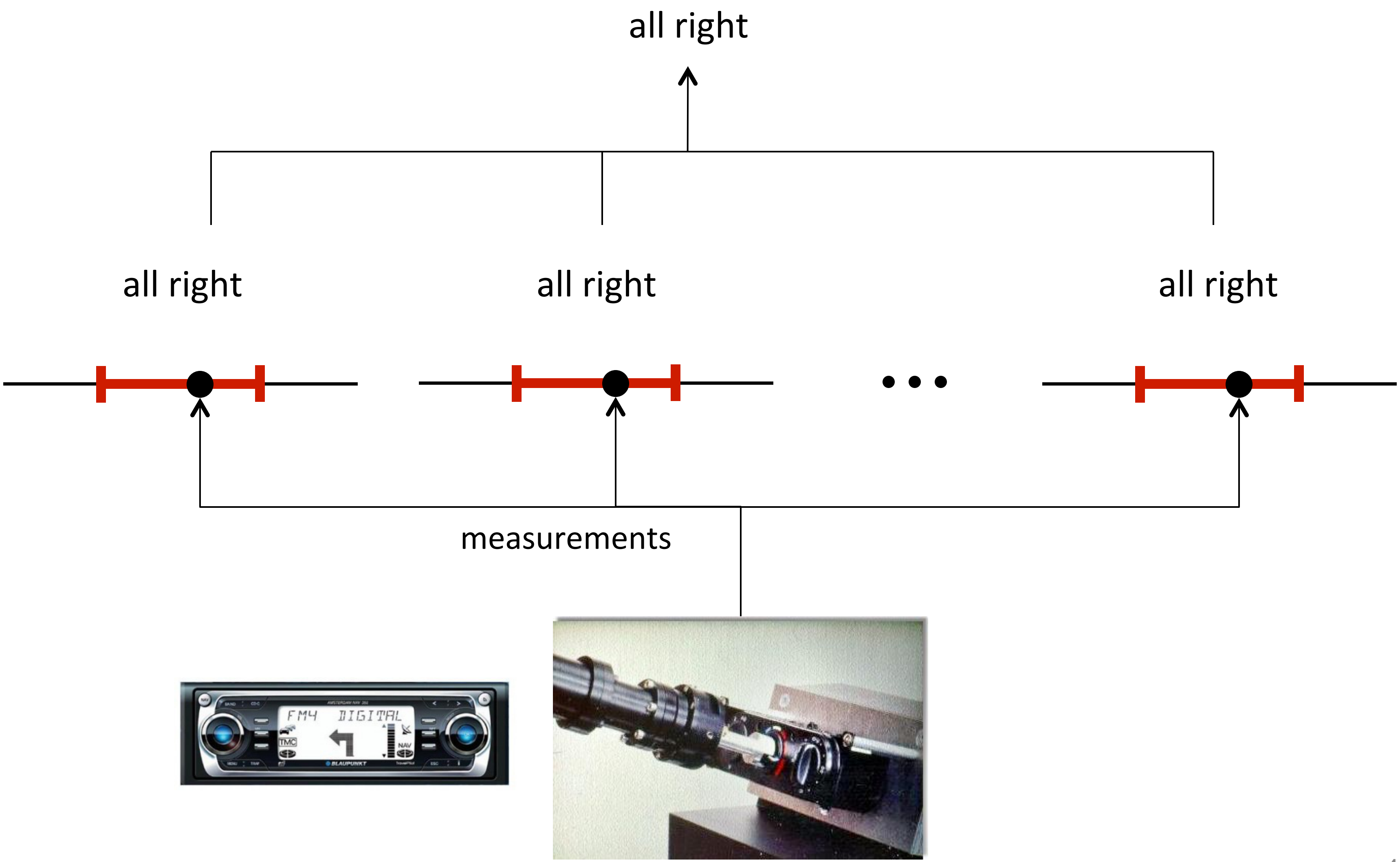
- flat structure of the model (→ curse of dimensionality)
- restriction to (quasi-)axis-parallel decision boundaries



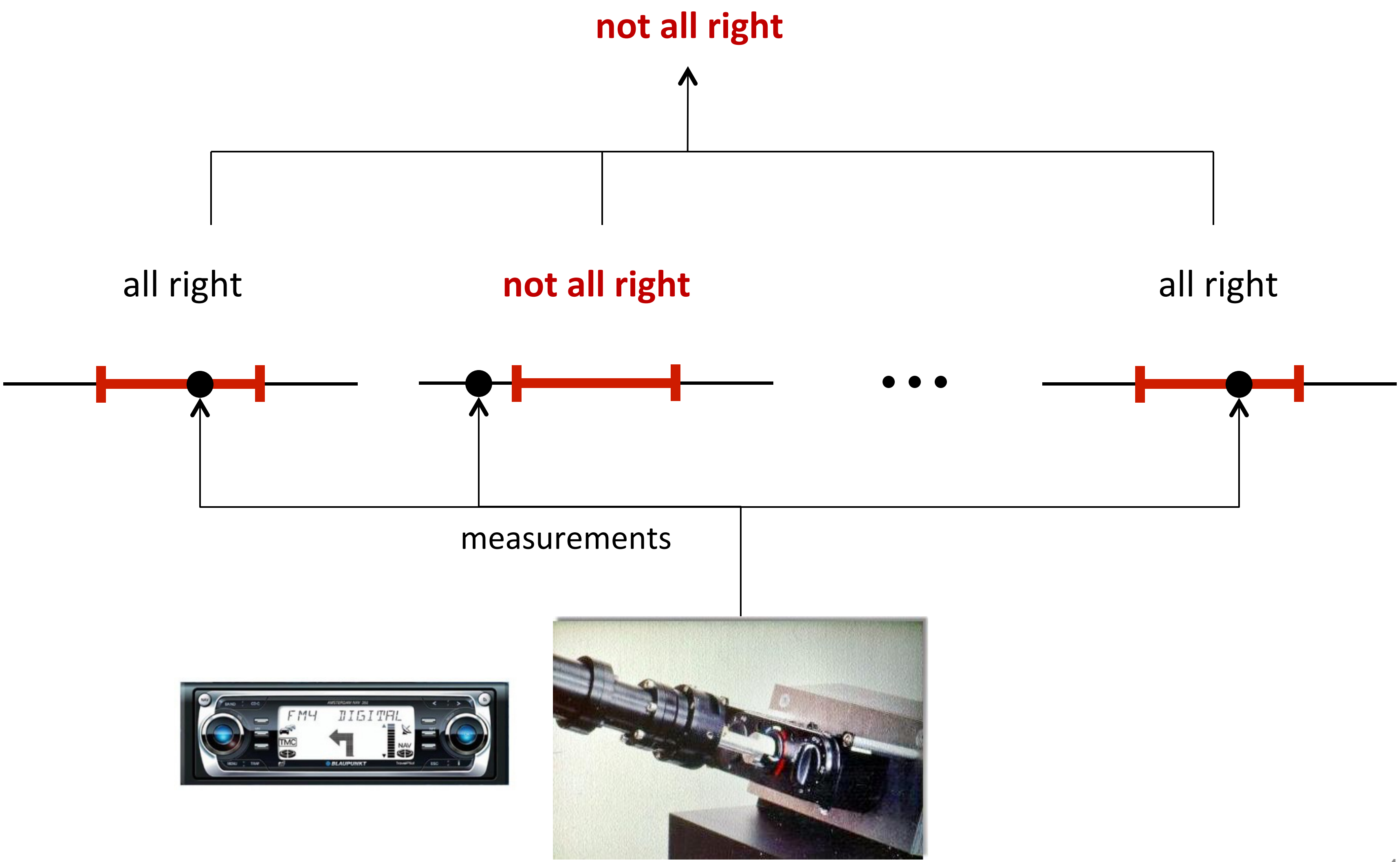
Flexibility of fuzzy models requires many rules!

Fuzzy pattern trees overcome these problems by using a **hierarchical** model structure and **generalized aggregation operators**.

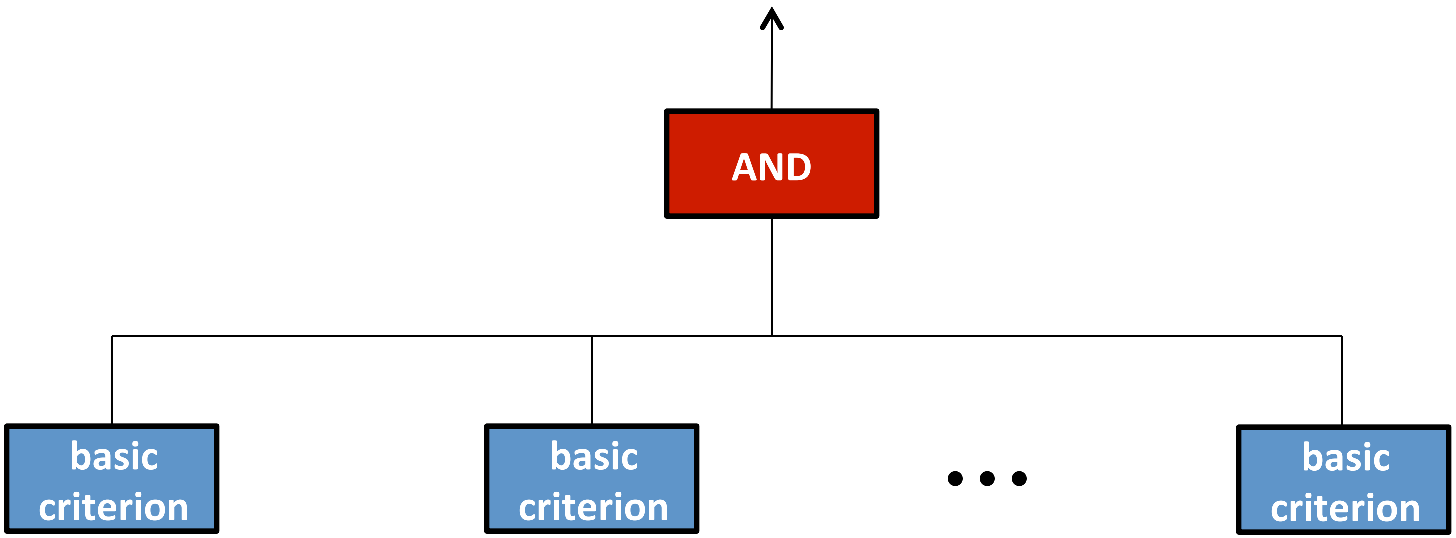
SIMPLE QUALITY CONTROL



SIMPLE QUALITY CONTROL



SIMPLE QUALITY CONTROL



$$Q(x_1, x_2, \ldots, x_n) = \bigwedge_{i=1}^{n} P_i(x_i)$$

DRAWBACKS OF THIS APPROACH

- Bivalent, **non-gradual evaluation** is not natural and does not support a proper ranking of products.
- **No compensation**: Several good properties cannot compensate for a single bad one.
- Extremely **sensitive toward noise**.
- „Flat“ structure of the evaluation scheme is **not scalable**.

single device


interior of a car

DRAWBACKS OF THIS APPROACH

- Bivalent, **non-gradual evaluation** is not natural and does not support a proper ranking of products.

fuzzy logic-based evaluation

- **No compensation**: Several good properties cannot compensate for a single bad one.
- Extremely **sensitive toward noise**.

generalized aggregation operators

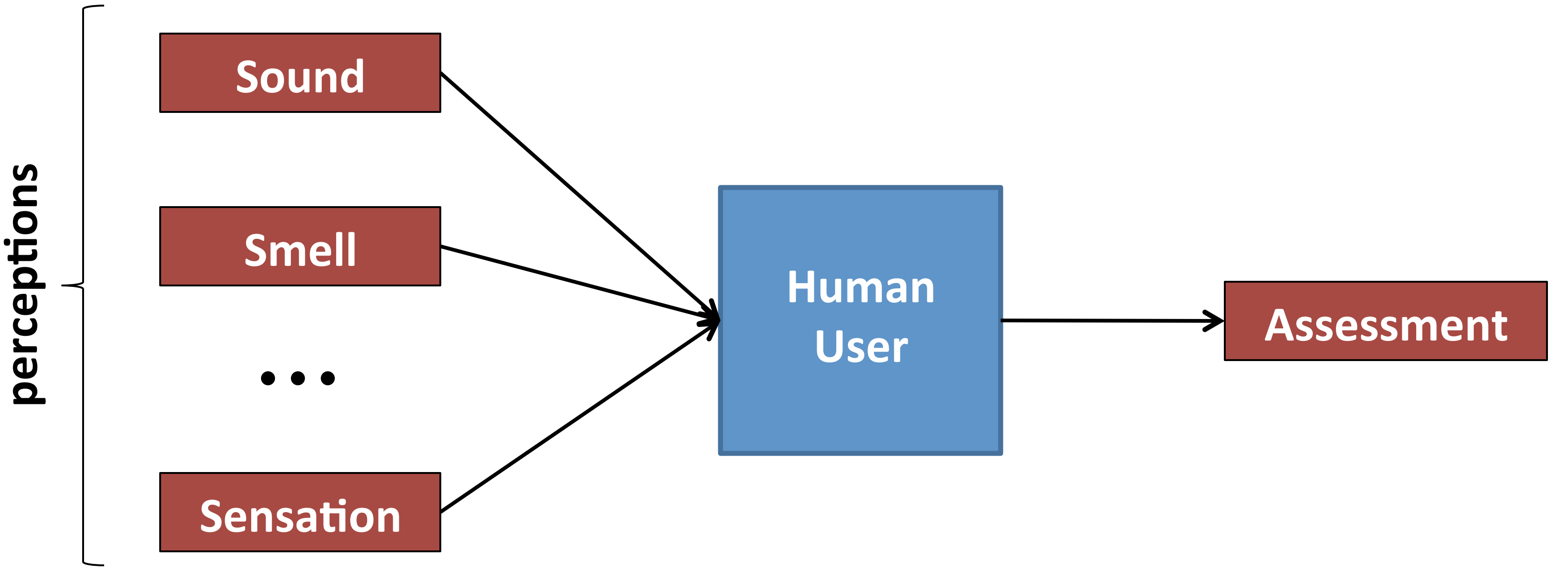
- „Flat“ structure of the evaluation scheme is **not scalable**.

hierarchical modeling

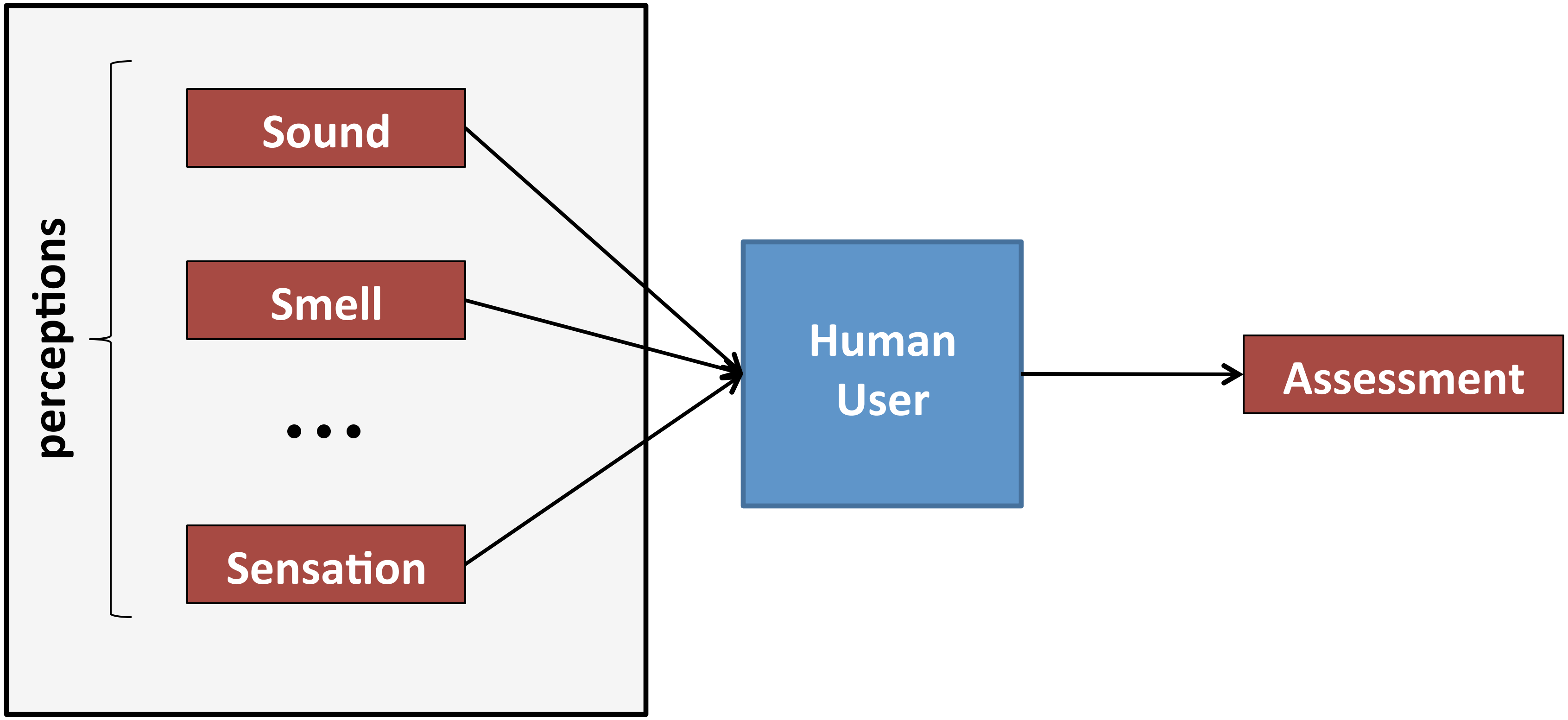
single device

interior of a car

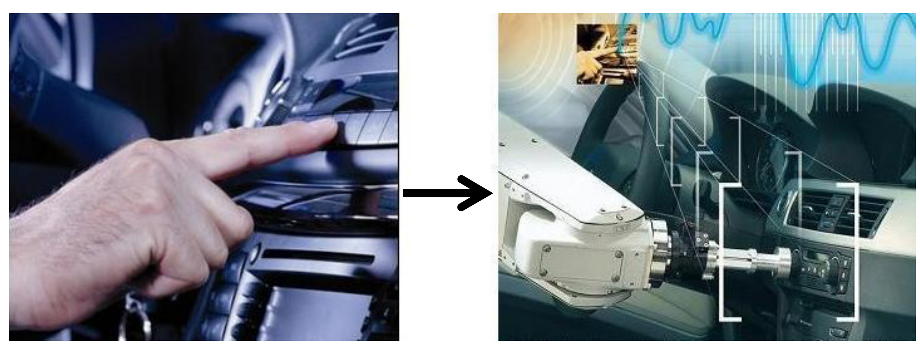
A SYSTEM IMITATING HUMAN ASSESSMENT



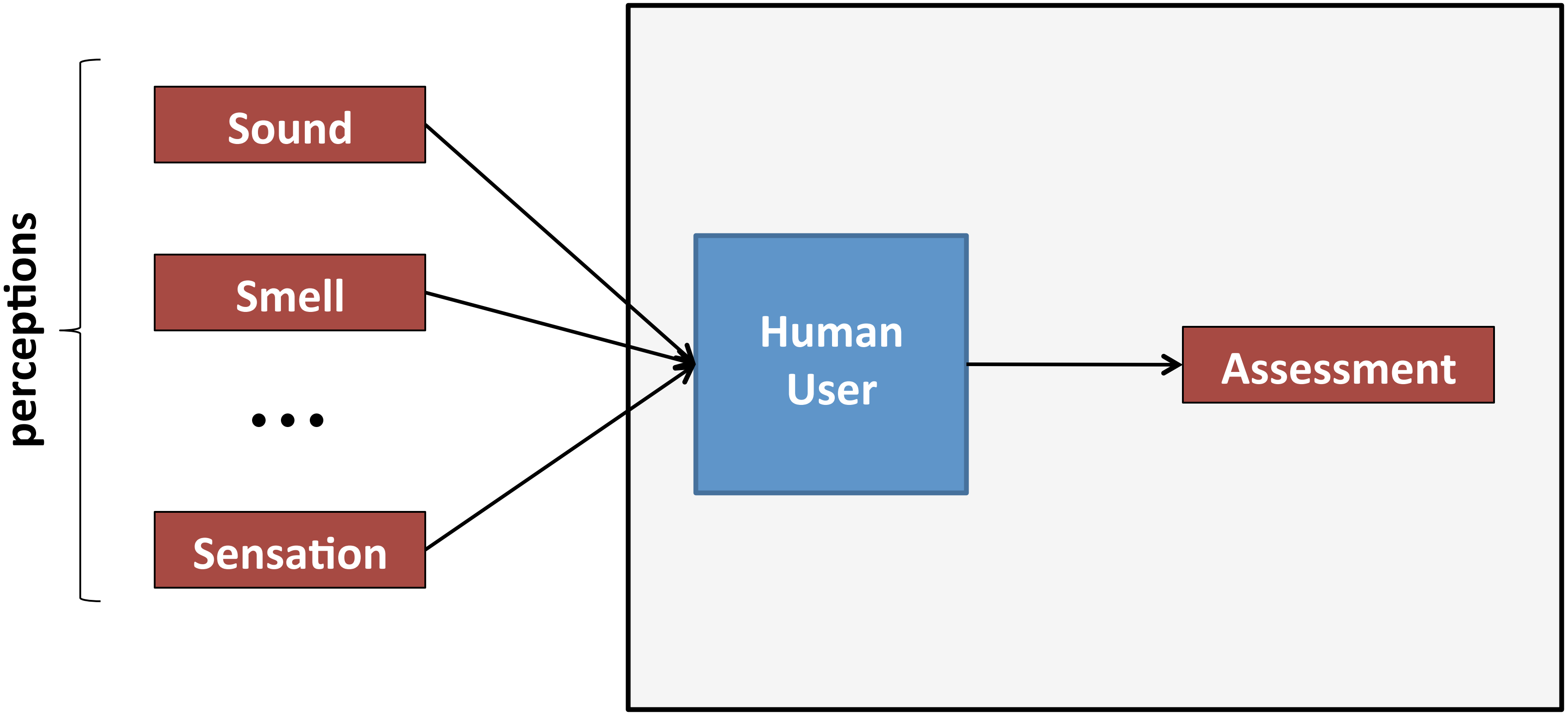
A SYSTEM IMITATING HUMAN ASSESSMENT



Quantification of individual criteria through measurement robotics

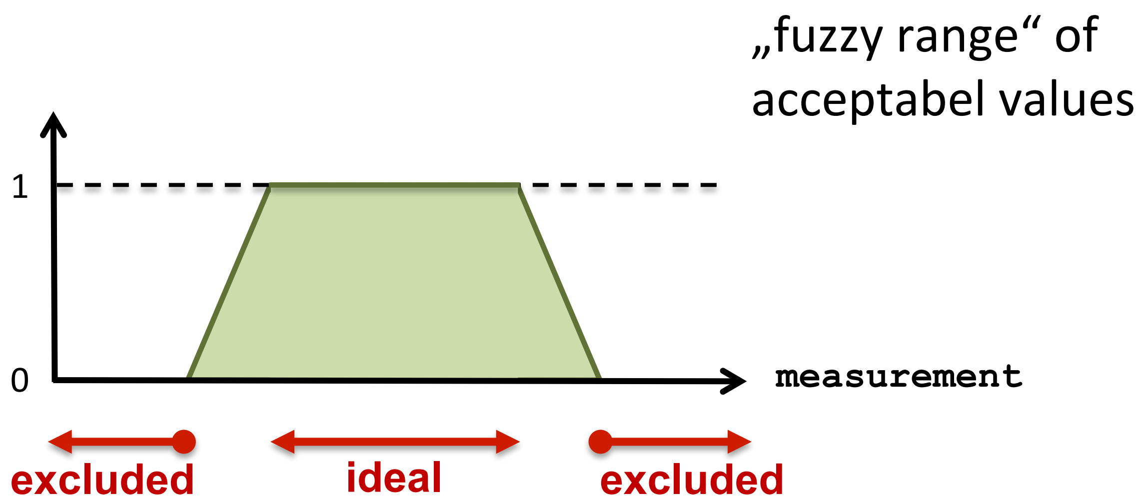


A SYSTEM IMITATING HUMAN ASSESSMENT



Evaluation of individual criteria and aggregation into an overall assessment

EVALUATION OF INDIVIDUAL CRITERIA



GENERALIZED AGGREGATION

Three modes of aggregation:

- **Conjunctive** („and“): both criteria must be fulfilled
- **Disjunctive** („or“): either of the criteria must be fulfilled
- **Averaging**

GENERALIZED AGGREGATION: CONJUNCTION

T-norms $\top : [0,1]^2 \to [0,1]$ as generalized conjunctions:

- $\top(x, 0) = 0$, $\top(1, x) = x$
- $\top(x, y) = \top(y, x)$
- $\top(x, y) \geq \top(x, z)$ for $y > z$
- $\top(x, \top(y, z)) = \top(\top(x, y), z)$

Examples:

- $\top_M(x, y) = \min(x, y)$
- $\top_P(x, y) = x \times y$
- $\top_L(x, y) = \max(x + y - 1, 0)$
- $\top_\alpha(x, y) = \frac{x \cdot y}{\max\{x, y, \alpha\}}$

Order relation: $\top_L \leq \top_P \leq \top_M$

GENERALIZED AGGREGATION: DISJUNCTION

T-conorms $\top : [0,1]^2 \to [0,1]$ as generalized conjunctions:

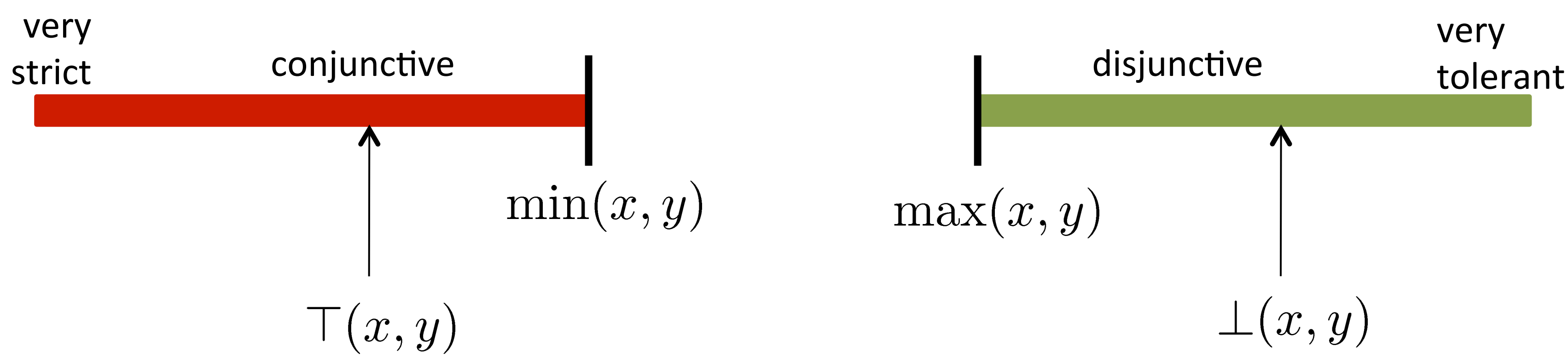
- $\perp(x,0) = \alpha$, $\perp(1,x) = 1$
- $\perp(x,y) = \perp(y,x)$
- $\perp(x,y) \geq \perp(x,z)$ for $y > z$
- $\perp(x, \top(y,z)) = \perp(\perp(x,y),z)$

Examples:

- $\perp_M(x,y) = \max(x,y)$
- $\perp_P(x,y) = x + y - x \times y$
- $\perp_L(x,y) = \min(x+y,1)$
- $\perp_\alpha(x,y) = \frac{x+y-x\cdot y-\min\{x,y,1-\alpha\}}{\max\{1-x,1-y,\alpha\}}$

Order relation: $\perp_M \leq \perp_P \leq \perp_L$

GENERALIZED AGGREGATION



GENERALIZED AGGREGATION: AVERAGING

An ordered weighted average (OWA) combination of k numbers $v_1, v_2, \ldots, v_k$ is defined by

$$\mathrm{OWA}_w(v_1, v_2, \ldots, v_k) = \sum_{i=1}^{k} w_i \cdot v_{\tau(i)},$$

where τ is a permutation of $\{1, 2, \ldots, k\}$ such that

$$v_{\tau(1)} \leq_{\tau(2)} \leq \ldots \leq v_{\tau(k)}$$

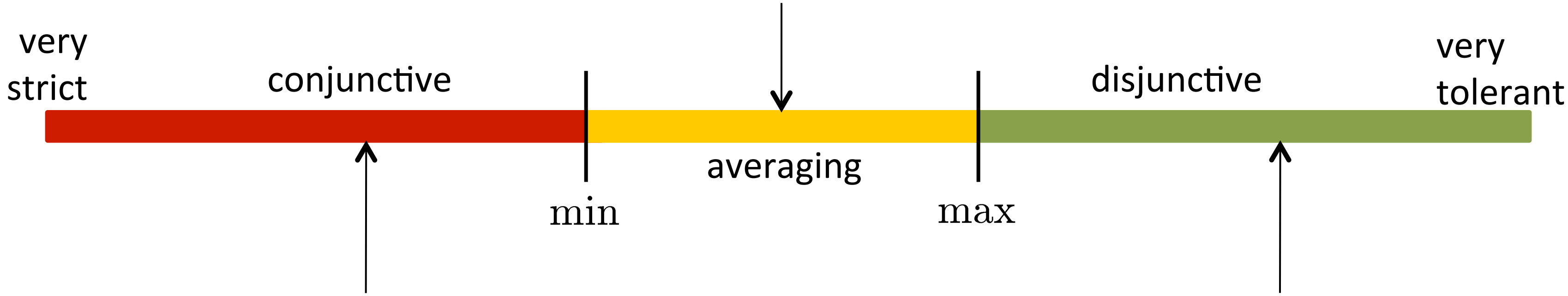
and $w = (w_1, w_2, \ldots, w_k)$ is a weight vector satisfying $w_i \geq 0$ and $\sum_{i=1}^{k} w_i = 1$.

For $k = 2$, this is simply a convex combination of the minimum and the maximum:

$$\mathrm{OWA}_w(v_1, v_2) = w_1 \cdot \min(v_1, v_2) + w_2 \cdot \max(v_1, v_2)$$

SEAMLESS TRANSITION FROM STRICT TO TOLERANT

$$\begin{aligned}\mathrm{OWA}_w(x,y) = \quad & w_1 \cdot \min\{x,y\} \\ & + w_2 \cdot \max\{x,y\}\end{aligned}$$

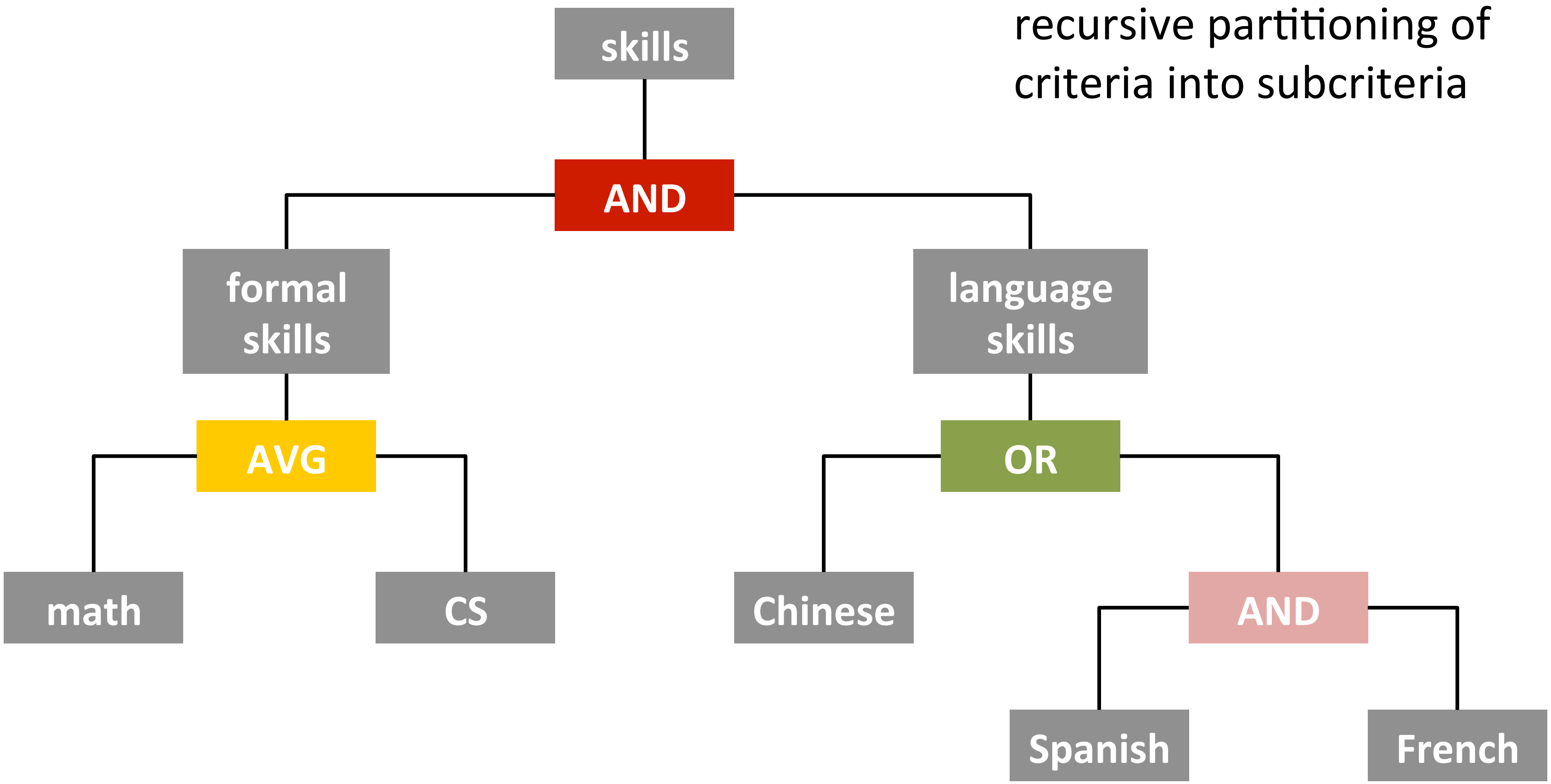


$$\top_\alpha(x,y) = \frac{x \cdot y}{\max\{x,y,\alpha\}}$$

$$\perp_\alpha(x,y) = \frac{x + y - x \cdot y - \min\{x,y,1-\alpha\}}{\max\{1-x,1-y,\alpha\}}$$

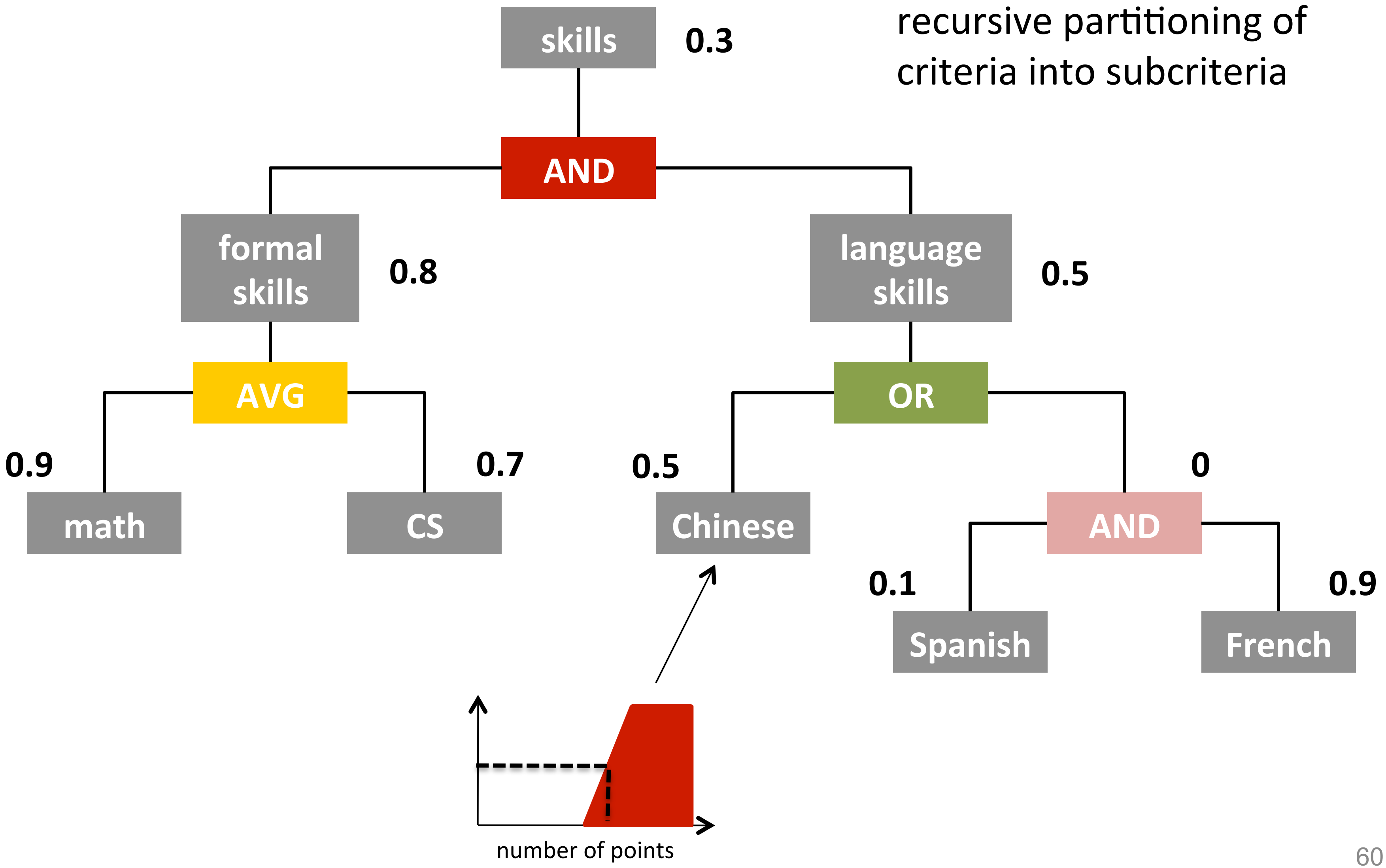
HIERARCHICAL MODELING

recursive partitioning of criteria into subcriteria



HIERARCHICAL MODELING

recursive partitioning of criteria into subcriteria



KNOWLEDGE-DRIVEN MODELING

A human expert (quality engineer) is specifying the model by hand …

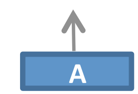
DATA-DRIVEN MODELING

- Known versus unknown structure
  - estimation of parameters (model calibration)
  - **structure + parameter learning (model induction)**

- Direct versus indirect feedback
  - **direct: product + rating (numeric or ordinal)**
  - indirect: comparison between products

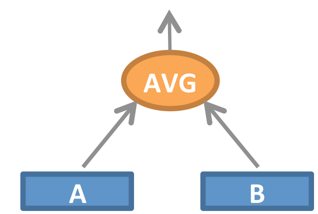
FUZZY PATTERN TREE INDUCTION

- Starting with primitive pattern trees (fuzzy subset of an attribute's domain),
- candidate trees are iteratively expanded
- and selected based on a tree performance measure (MSE on training data),
- until a stopping condition is met.



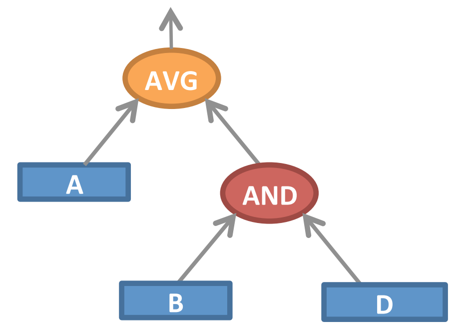
FUZZY PATTERN TREE INDUCTION

- Starting with primitive pattern trees (fuzzy subset of an attribute's domain),
- candidate trees are iteratively expanded
- and selected based on a tree performance measure (MSE on training data),
- until a stopping condition is met.



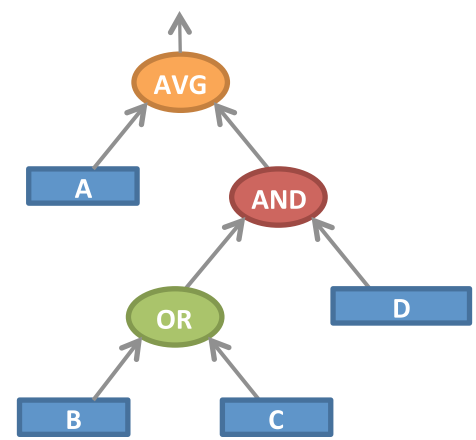
FUZZY PATTERN TREE INDUCTION

- Starting with primitive pattern trees (fuzzy subset of an attribute's domain),
- candidate trees are iteratively expanded
- and selected based on a tree performance measure (MSE on training data),
- until a stopping condition is met.



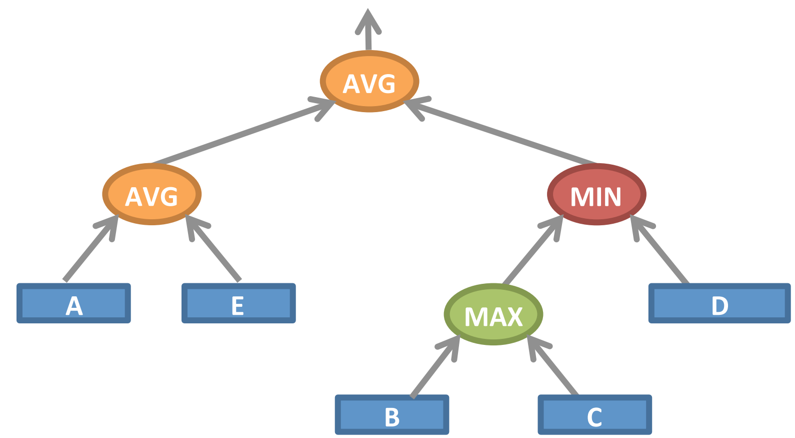
FUZZY PATTERN TREE INDUCTION

- Starting with primitive pattern trees (fuzzy subset of an attribute's domain),
- candidate trees are iteratively expanded
- and selected based on a tree performance measure (MSE on training data),
- until a stopping condition is met.



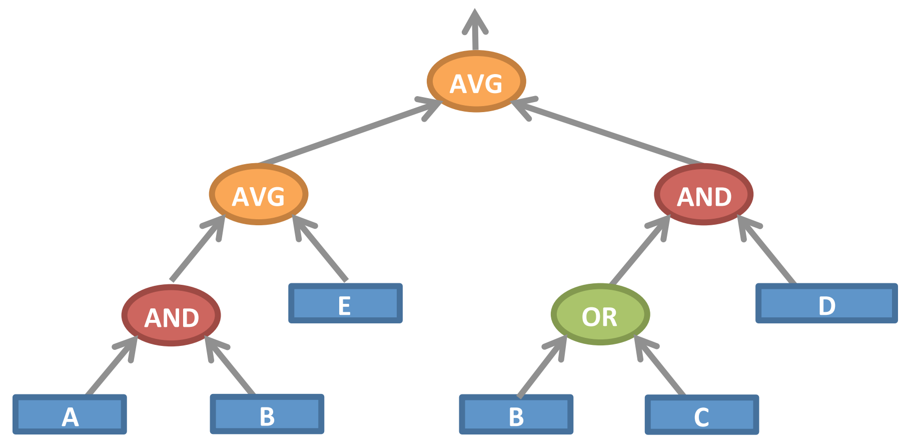
FUZZY PATTERN TREE INDUCTION

- Starting with primitive pattern trees (fuzzy subset of an attribute's domain),
- candidate trees are iteratively expanded
- and selected based on a tree performance measure (MSE on training data),
- until a stopping condition is met.



FUZZY PATTERN TREE INDUCTION

- Starting with primitive pattern trees (fuzzy subset of an attribute's domain),
- candidate trees are iteratively expanded
- and selected based on a tree performance measure (MSE on training data),
- until a stopping condition is met.



→ **greedy beam search**

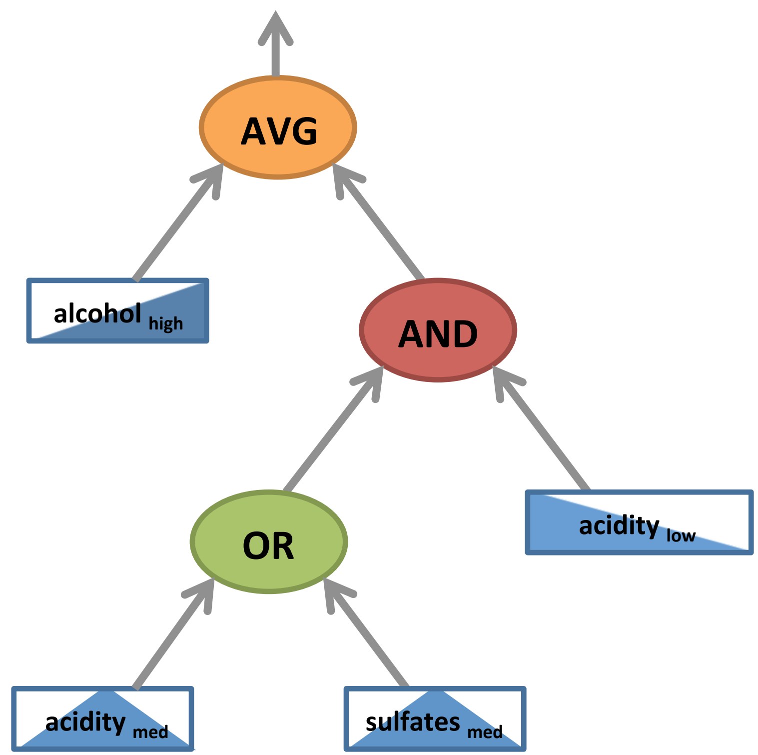
THE WINE QUALITY DATA

The task is to predict the quality of wine (on a scale from 0 to 10) based on properties like acidity, level of alcohol, etc. (UCI benchmark data).

acidity	alcohol	…	sulfates	quality
7.4	9.4	…	0.56	5
7.8	10	…	0.46	3
7.8	10.5	…	0.80	6
11.2	9.3	…	0.91	3
7.4	9.8	…	0.55	5
7.3	10.6	…	0.53	4
…	…	…	…	…

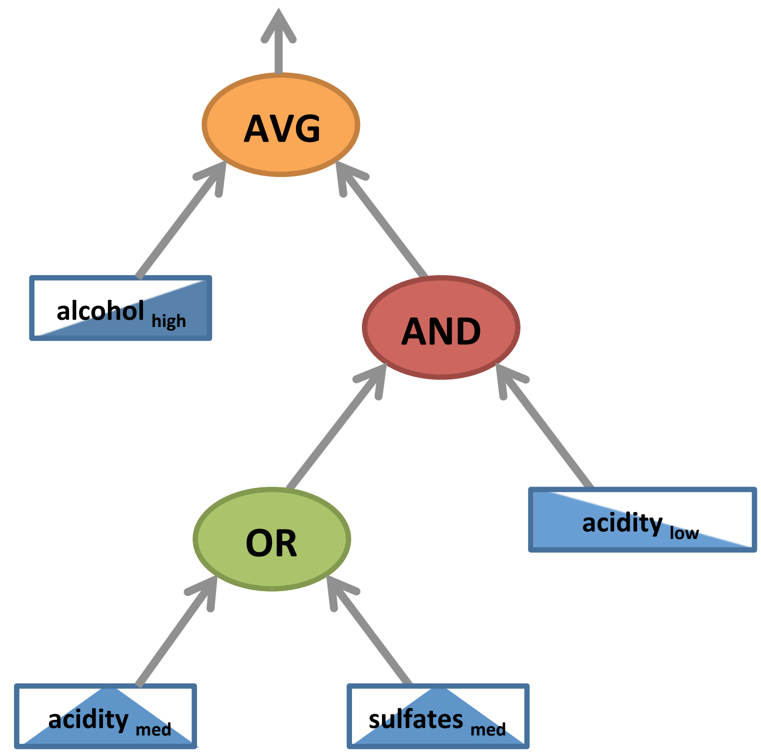
THE WINE QUALITY DATA

Pattern tree induced from a given set of data (wine properties + rating):



FEATURES OF FUZZY PATTERN TREES

- **interpretability** of the model class
- **modularity:** recursive partitioning of critria into sub-criteria
- **flexibility** without the tendency to overfit the data
- **monotonicity** in single attributes
- built-in **feature selection**
- yield **competitive performance** for **classification** and **regression**



SUMMARY & CONCLUSIONS

- ML is a **more recent research topic** in the fuzzy community (compared to control, optimization, decision making, …).
- Like in other fields, FL can **reasonably extend, generalize, and complement** existing foundations and methods in ML.
- **Modeling** (background knowledge) is one of the main strengths, and crucial for successful learning (which does not only require data).
- Besides, ML can benefit from **other fuzzy logic tools**, including generalized uncertainy formalisms, aggregation operators, …

SUMMARY & CONCLUSIONS

- **Fuzzy pattern trees** as a novel model class for machine learning.
- I**nteresting features:** interpretability , modularity, non-linearity, monotonicity, handling imprecise data, …
- Viable **alternative to rule-based models** (more compact models thanks to hierarchical instead of flat model structure)
- Competitive performance in **classification** and **regression**.
- **Evolving fuzzy pattern trees** (eFPT) for learning on data streams.

Literature

Fuzzy logic and machine learning (survey articles):

- E.H. Fuzzy Sets in Machine Learning and Data Mining: Status and Prospects. Fuzzy Sets and Systems, 156(3), 2005.
- E.H. Fuzzy Machine Learning and Data Mining. WIREs Data Mining and Knowledge Discovery, 2011.

Fuzzy pattern trees:

- Huang, TD. Gedeon, and M. Nikravesh. Pattern trees induction: A new machine learning approach. IEEE TFS 16(4), 2008.
- Y. Yi, T. Fober and E.H. Fuzzy Operator Trees for Modeling Rating Functions. Int. J. Comp. Intell. and Appl. 8(1), 2009.
- R. Senge and E.H. Pattern Trees for Regression and Fuzzy Systems Modeling. Proc. WCCI-2010, Barcelona, Spain, 2010.
- R. Senge and E.H. Top-Down Induction of Fuzzy Pattern Trees. IEEE TFS, 19(2), 2011.
- A. Shaker and E.H. Evolving Fuzzy Pattern Trees for Binary Classification on Data Streams. Information Sciences, to appear.

Java implementation of fuzzy pattern induction:

`http://www.uni-marburg.de/fb12/kebi`