# PATTERN RECOGNITION
# Delimiting the territory

Václav Hlaváč

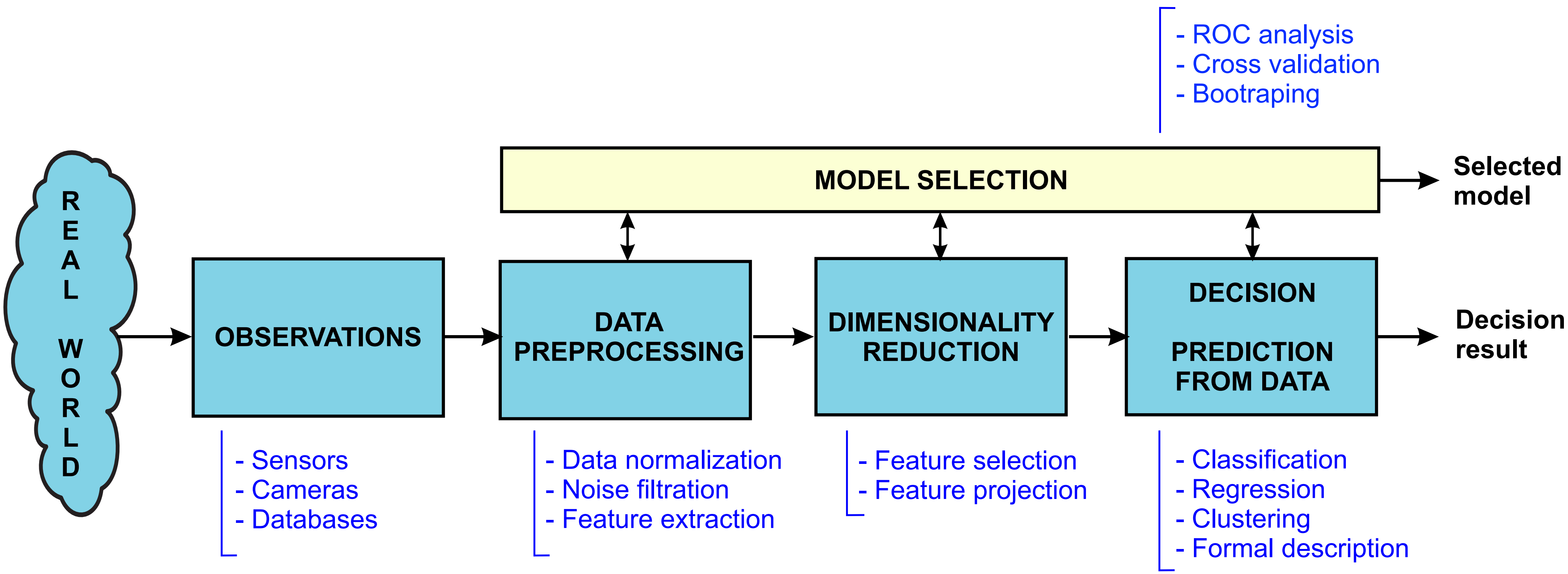
Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics
Center for Machine Perception
http://cmp.felk.cvut.cz/~hlavac, hlavac@fel.cvut.cz

**Outline of the talk:**

- Modeling and system theory.
- Pattern recognition, task formulation, Bayesian task.
- Motivating examples.
- Supervised and unsupervised classification; task formulations.
- Probability, statistics, informal introduction.
- Conditional dependence, Bayes theorem.

# The overall picture, components






- Input: Data, training (multi)-set.
- Models and its parameters learned from the training data.
- Outputs: Can be different ones, discussed at the next slide.

# An old scientific problem



- The nature of classification and decision had been a central theme in the discipline of philosophical epistemology, the study of the nature of knowledge.

- The foundations of pattern recognition can be traced back to Plato and later Aristotle, who distinguished between:

  - an *essential property* — shared by all members in a class or *natural kind*.

  - an *accidental property* — which would differ among members in the class.

# Classification/categorization (or the functional description)



# H. Bülthof's counterexample

# Types of decision / prediction problems



Classification – Assigns the observation to a class from a small set of possible classes. The output is a label, an identifier of the the class, e.g. the system grades apples as A, B, C, and a reject.

Regression – predicts a value from the observation. It is a generalization of the classification. The output could be, e.g. a real number as a company value based on its past performance and stock market indicators.

Unsupervised learning (clustering) – organizes observations into meaningful classes based on their mutual similarities. E.g. in transcriptomics, it builds groups of genes with related expression patterns (called coexpressed genes).

---

Structural relations representation – the objects is described using basic primitives, e.g. observation of a human by a surveillance camera as composition of prototypical actions, body positions. A structure comes into play.

# Other disciplines, share similar core ideas



Neural networks – one of formalisms to solve a decision problem without necessarily creating a model of a real biological system.

Machine learning – given a set of training examples, learn the decision rules automatically. No manual (subjective) definition of rules is involved. A different situation requires a different set of training examples.

Data Mining – extraction of implicit, previously unknown and potentially useful information from the data.

Statistical modelling – finds a (generative) model describing the studied object, e.g. using probability distributions and assesses its quality using statistical techniques.

Scientific visualization – A high-dimensional problem should be visualized in 2D or 3D (we humans do not see more).

# Learning: approaches in different disciplines

Engineering – signal processing, system identification, adaptive and optimal control, information theory, robotics, . . .

Computer science – artificial Intelligence, machine learning, computer vision, information retrieval, . . .

Statistics – learning theory, data mining, learning and inference from data, . . .

Cognitive science and psychology – perception, sensorimotor control, learning, mathematical psychology, computational linguistics, . . .

Computational neuroscience – neural networks, neural information processing, . . .

Economics – decision theory, game theory, operational research, . . .

---

*Pedagogy – different approach, statistical models are not used, . . .*

# Biological motivation

- A human is consider the most advanced animal also due to the ability to think about the way she/he reasons.
- There is a general interest in mimicking biological perception in machines. One of the aims is to imitate intelligent behavior in partly unknown environment.
- The ability to learn using stimuli from surrounding world is a basic attribute of intelligent behavior. Pattern recognition provides certain insight how learning can be performed.
- There is a key question – knowledge representation. Among us humans, the observable means for sharing knowledge – the natural language is the most advanced tool for expressing observations, description of phenomena, problem formulations, their solution and related learning issues.

# Complex phenomena and system approach



- A desire to understand complex phenomena, e.g., in biology, social sciences, technology requires to analyze involved phenomena in a complex way taking into account very many relations and different contexts.
- The system approach contrasts Newtonian endeavor to reduce every phenomenon to relations among basic elements and their basic properties.

# A few concepts from the system theory

- While analyzing a complex phenomenon, we restrict ourselves to the part which is of our interest. We call it the object (or sometimes the system).
- The rest (which is unimportant from the chosen point of view) is called background.
- Objects are not often analyzed in their entire complexity. Instead, only those properties are observed or measured in one study which seem to be of interest. The system theory uses term resolution for different points of view.
- The object description (often mathematical) varies both quantitatively and qualitatively when the resolution is changed. The change of resolution provides a meta-view allowing to find a qualitative change in object description.

# Two possible approaches to describe objects



The attempt to exact description of objects (complex phenomena) using mathematical tools leads (roughly speaking) to two possible approaches:

1. Understanding and expressing physical/other principles $\Rightarrow$ mathematical modeling (in a Newtonian sense).

   Called also generative approach.

2. Pattern recognition way $\Rightarrow$ classification.

   Called also discriminative approach.

---

*Note: We will talk about generative/discriminative approach to learning of a decision rule (of a classifier) later in the course too. There are similarities due to names of used concepts and also differences to the taxonomy mentioned here.*

# Mathematical modeling

- The important properties of the objects are mimicked using mathematical equations. The relation between the input and the output is often sought.
- The approach is often close to the Newtonian approach as the desire is to obtain a detailed and preferably a deterministic explanation.

---

- *Example:* A feasible mathematical model of a power house boiler used in control engineering predicts almost identical behavior as the real boiler.
- *Counterexample 1:* In many cases, we are not able to create a mathematical model of a complex system, e.g., the model describing how a human body is functioning.
- *Counterexample 2:* Computer vision. The inverse task to the physical process of the image formation is too complex.

# Pattern recognition as an alternative to modeling



- Pattern recognition assigns observations according to some decision rule to a priori known classes of objects.
- Equivalence classes (reflexivity, symmetry, transitivity).
- Objects within classes are more similar to each other than objects from different classes.
- The understanding to the object is often weaker in pattern recognition than in modeling.

# Concepts



- A pattern is an object, process or event that can be given a name.
- A pattern class (or category) is a set of patterns sharing common attributes and usually originating from the same source.
- During recognition (or classification) given objects are assigned to prescribed classes.
- A classifier is a machine which performs classification.

# The role of learning in pattern recognition

- The advantage of PR is that a human creating the recognition rule does not need to understand the complex nature of the object.
- A decision rule can be learned empirically from many observed examples (knowledge engineering paradox).

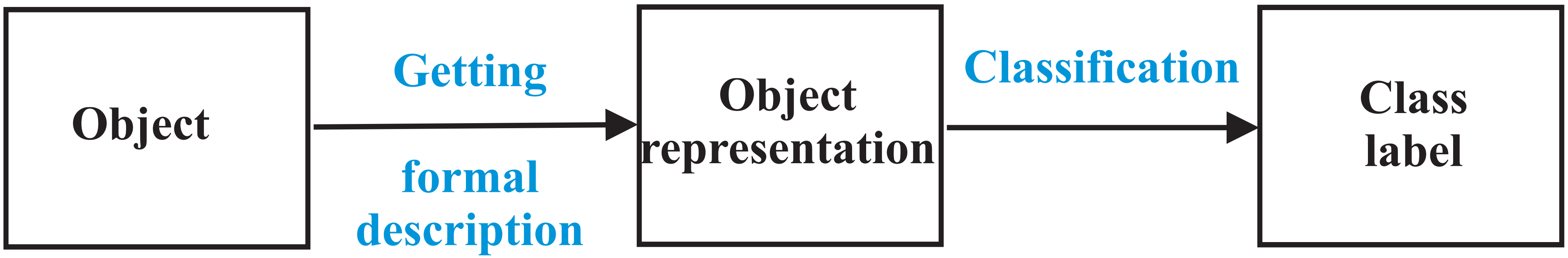
  Supervised learning based on the training set comprising of observations and corresponding decisions assigned by a teacher (an expert).

  Unsupervised learning seeks for similarities among observations without having an expert classification at hand.

  Reinforcement learning explores reward information (positive, negative) from the environment. A cumulative reward is maximized.

# Pattern recognition and applications

Pattern recognition theory and tools can be separated from applications.



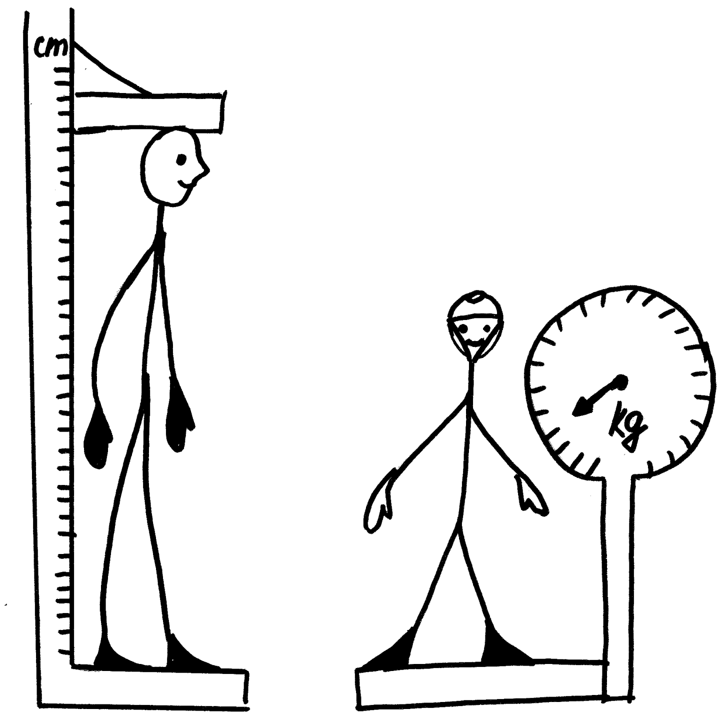
# Pattern recognition, A motivating example

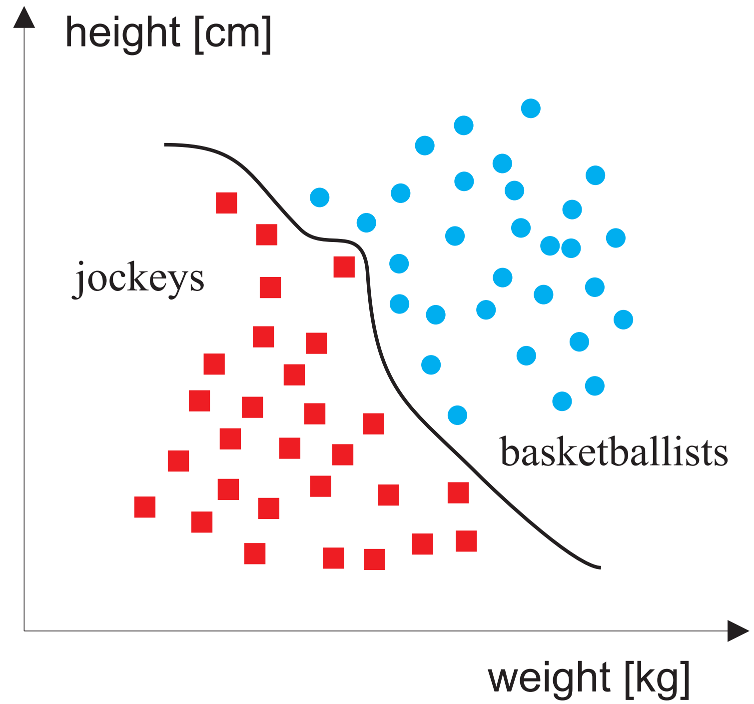


**Object (situation)** is described by two parameters:

- $x$ – observable *feature* (also observation).
- $y$ – *hidden parameter* (state, special case—a class).

**Example** statistical PR: jockeys and and basketballists.



# Main approaches to pattern recognition



1. Statistical (feature-based) pattern recognition.
   - Objects are represented as points in a vector space.
   - The coordinate axes correspond to individual observations (measurements) expressed by a numerical values.
2. Structural pattern recognition.
   - There is a structure among observations. The aim is to represent and explore it.
   - Formal grammars are the oldest and the most advanced tool to represent the structure.
3. Artificial neural networks. The classifier is represented as a network of cells modeling neurons of the human brain (connectionist approach, e.g., a feedforward model of the neural network (McCulloch, Pitts, 1943).

# Bayesian decision making



Bayesian task of statistical decision making seeks

for sets $X$ (observations), $Y$ (hidden states) and $D$ (decisions), a joint probability $p_{XY}\colon X \times Y \to \mathbb{R}$ and the penalty function $W\colon Y \times D \to \mathbb{R}$

a strategy $Q\colon X \to D$ which minimizes the Bayesian risk

$$R(Q) = \sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y)\, W(y, Q(x))\,.$$

The solution to the Bayesian task is the Bayesian strategy $Q$ minimizing the risk.

---

Notes: deterministic strategy, separation into convex subsets.

# Generality of the Bayesian formulation

Motto: *"Let set $X$ (observations) and set $K$ (hidden states) be two finite sets."*

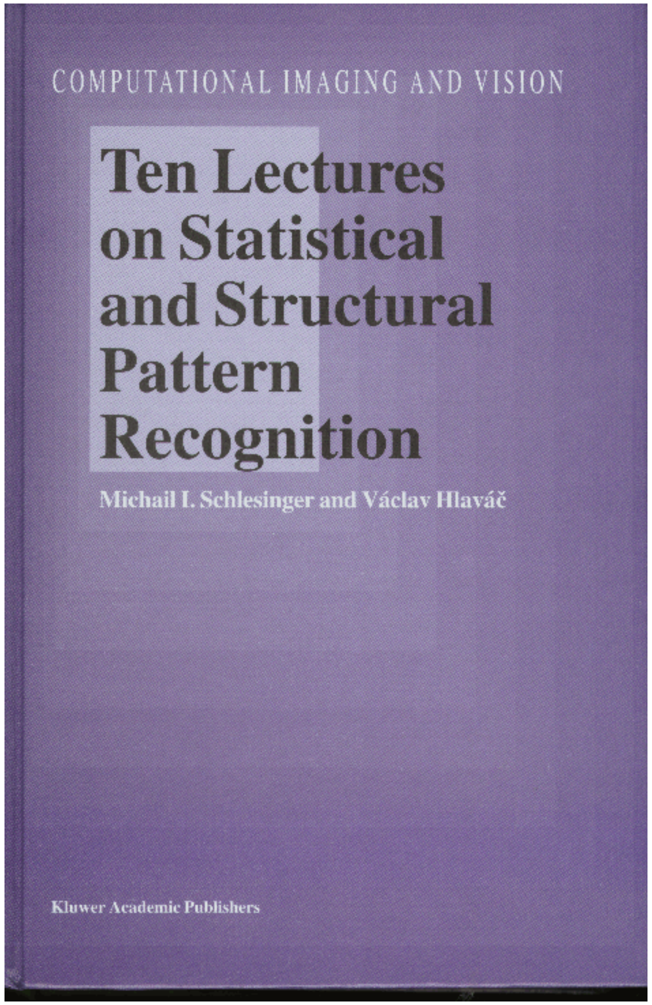
- Statistical pattern recognition results are very general. Properties of sets $X$ (observations) and $Y$ (hidden parameters) were not constrained.
- Sets $X$ and $Y$ can have formally a (mathematically) diverse structure.
- The approach can be and is used in very different applications.

---

Observation $X$ can be a number, a symbol, a function of two variables (e.g., an image), a graph, other algebraical structure.

# Recommended reading

- Duda Richard O., Hart Peter E., Stork, David G.:, Pattern Classification, John Wiley & Sons, New York, USA, 2001, 654 p.
- Schlesinger M.I., Hlaváč V.: Ten lectures on statistical and syntactic pattern recognition, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002, 521 p.
- Bishop C.: Pattern Recognition and Machine Learning, Springer-Verlag New York 2006, 758 p.



# Our tool for experiments

- Franc V. (our ex PhD student): Statistical Pattern Recognition Toolbox in MATLAB, in development since 2000, http://cmp.felk.cvut.cz/cmp/software/stprtool/

# What is known in statistical pattern recognition?



- Solution to some non-Bayesian tasks, e.g., class "I don't know" (called also reject option) can be introduced. Tasks with non-random interventions.
- Linear classifiers and their learning. E.g., a popular special case—Support Vector Machines. Learning.
- Estimate of needed length of training multi-set for prescribed precision and reliability of classification (e.g., Vapnik's theory of learning).
- Embedding of a non-linear problem to a higher dimensional vector space (feature space straightening, allowing linear classifiers), kernel methods.
- Unsupervised learning, variants of EM algorithm.

# What is known in statistical pattern recognition?



- Solution to some non-Bayesian tasks, e.g., class "I don't know" (called also reject option) can be introduced. Tasks with non-random interventions.
- Linear classifiers and their learning. E.g., a popular special case—Support Vector Machines. Learning.
- Estimate of needed length of training multi-set for prescribed precision and reliability of classification (e.g., Vapnik's theory of learning).
- Embedding of a non-linear problem to a higher dimensional vector space (feature space straightening, allowing linear classifiers), kernel methods.
- Unsupervised learning, variants of EM algorithm.

# Application of mathematical statistics



- The most developed part of statistics is the statistics of random numbers.
- Recommendations are based on concepts as: mathematical expectation, dispersion, correlation, covariance matrix, ...
- Tools of mathematical statistics can be used to solve many practical problems provided the random object can be represented by a number (or a vector of numbers).
- Substantial success in statistical pattern recognition for vectors of features.

---

- Failure for images $f(x, y)$, where $f$ is brightness or color of a pixel and $x, y$ are pixel coordinates.