

# PATTERN RECOGNITION

## Introduction; Delimiting the territory

Václav Hlaváč

Czech Technical University in Prague

Czech Institute of Informatics, Robotics and Cybernetics

160 00 Prague 6, Jugoslávských partyzánů 1580/3, Czech Republic

<http://people.ciirc.cvut.cz/hlavac>, [vaclav.hlavac@cvut.cz](mailto:vaclav.hlavac@cvut.cz)

*Courtesy: M.I. Schlesinger, V. Franc*

### Outline of the talk:

- ◆ Global picture, epistemology.
- ◆ Modeling and system theory approach.
- ◆ Pattern recognition, learning.
- ◆ Statistical × structural PR.
- ◆ Bayesian formulation.
- ◆ What has been known in PR?

## What is pattern recognition ?

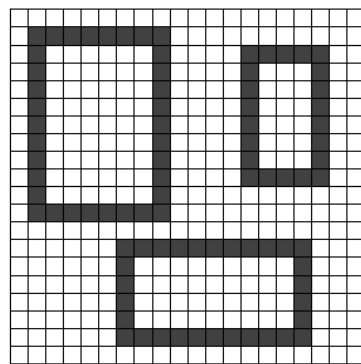
- ◆ **Epistemology** – a branch of philosophy dealing with the origin, nature, methods and scope of cognition/knowledge. Pattern recognition is one of methods.
  - ◆ **Pattern recognition** / **Machine learning** (almost synonyms) is a scientific discipline that constructs and studies algorithms that learn from data by building a statistical model, and use it for making decisions or predictions.
  - ◆ “**Pattern recognition** is the assignment of a physical object or event to one of several prespecified categories” – the book by Duda & Hart 1977, 2001.
- 
- ◆ Satoshi Watanabe, theoretical physicist (★1910 Tokyo, †1993 Tokyo, Worked in Germany, Denmark and USA): *“A pattern is the opposite of a chaos; it is an entity vaguely defined, that could be given a name.”* (1981)
  - ◆ Isaiah Berlin, philosopher (★1909 Riga, †1997 Oxford): *“To understand is to perceive patterns. To make intelligible is to reveal the basic pattern.”*

## Concepts

- ◆ A **pattern** is an object, process or event that can be given a name.
- ◆ A **pattern class** (or category) is a set  $M \subset X$  of elements (patterns) sharing common attributes, i.e. finite recognizable characteristics (features).
- ◆ **Classification** (or recognition) assigns given objects to prescribed classes.
- ◆ A **classifier** is a machine (program) which performs classification (decision making, prediction).
- ◆ A **sample** (example, object, situation) is a particular instance of the pattern we want to classify.
- ◆ A **training multi-set** (training data) is a collection of samples used to train the statistical model (classifier).
- ◆ A **test multi-set** (test data) is a collection of samples distinct from the training multi-set.
- ◆ **Features** are attributes characterizing properties of samples relevant to considered classification.

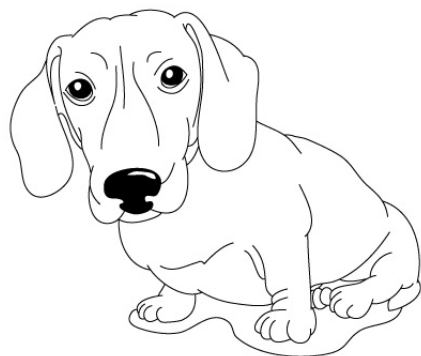
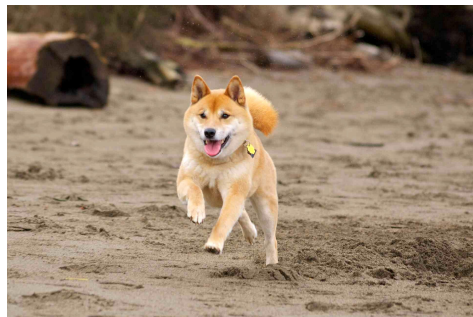
## A pattern class, examples (1)

- ◆ Set of all syntactically correct arithmetic expression like, e.g.,  $2x(a + 3b) - 6y + (x - y)/7$ .
- ◆  $M$  is a subset of the set  $X$  of all finite strings over an alphabet.  $M$  can be described by a context-free grammar.
- ◆ Set of all binary valued images containing non-overlapping and non-touching one pixel wide rectangular frames.  $M$  is a subset of the set  $X$  of all rectangular binary valued images.



## A pattern class, examples (2)

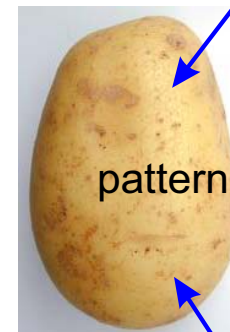
- ◆ Set of all dogs in images.



## Basic concepts, an illustration

- ◆ A **pattern** is studied (a potato in our example, see the illustration).
- ◆ A **feature vector**  $\mathbf{x} \in X$  is a vector of observations (measurements). Vector  $\mathbf{x}$  constitutes a single point in the feature (vector) space  $X$ .
- ◆ A **hidden state** (class label in a special case)  $y \in Y$  cannot be measured directly. Pattern with equal hidden states belong to the same class.
- ◆ The **task is to design a classifier** (a decision rule)  $q: X \rightarrow Y$  assigning a pattern instance into a hidden state.

hidden state  
(or class label)  $y$



feature vector  $\mathbf{x}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

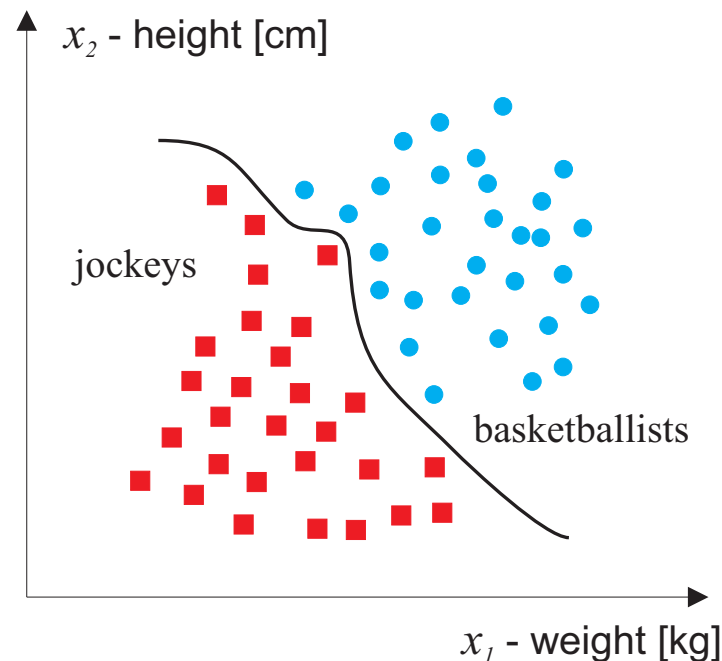
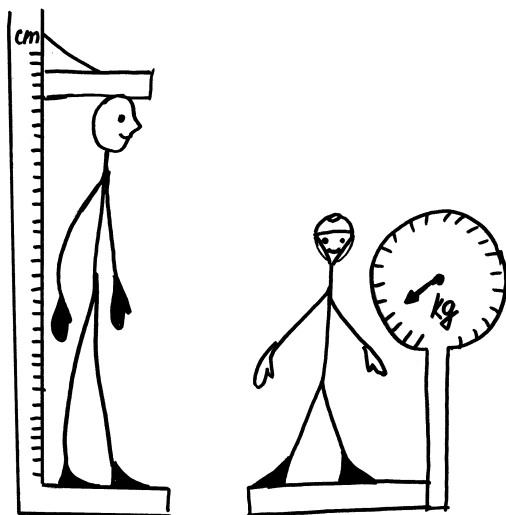
# Pattern recognition, A motivating example

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

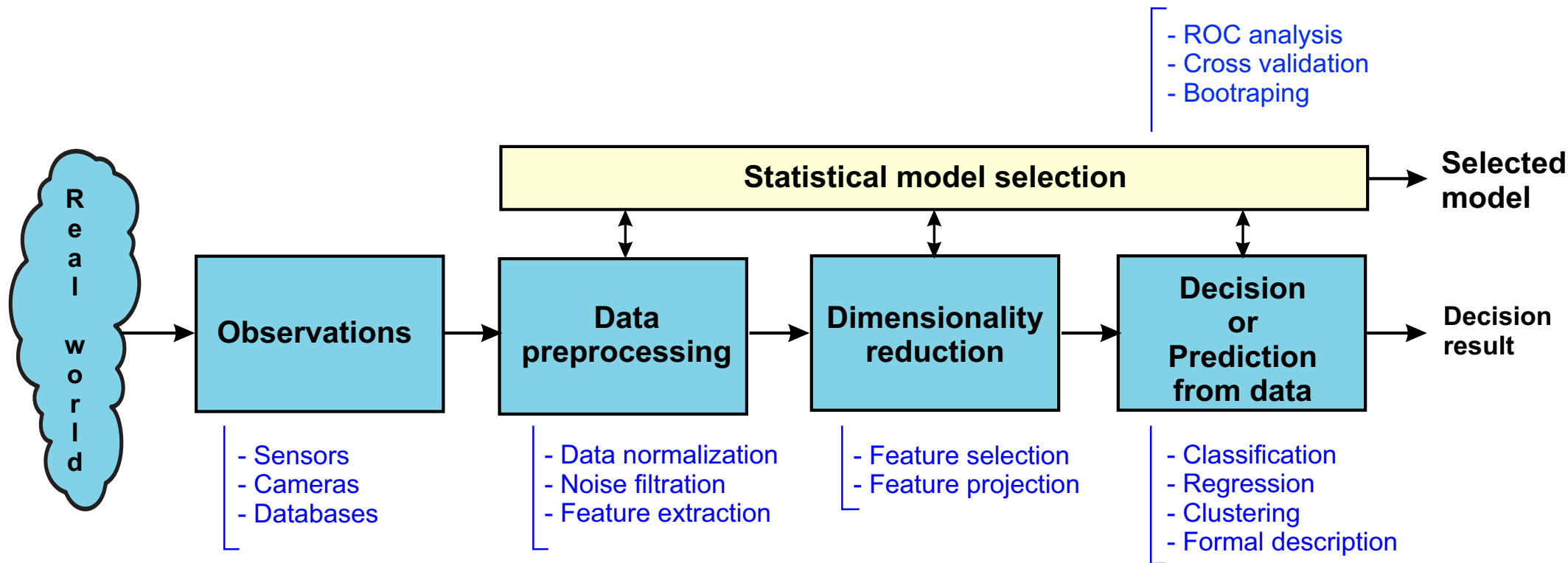
$x$  – observable *feature* (also observation).

$y$  – *hidden parameter* (state, special case—a class).

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



## The overall picture, components



- ◆ Input: Data, training (multi)-set.
- ◆ Statistical models and their parameters are learned empirically from the training data.
- ◆ Outputs: diverse decision; see the diagram.



# Classification is 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](#) (Πλάτων, 428 BC - 348 BC) and his student [Aristotle](#) (Ἀριστοτέλης, 384 BC – 322 BC),
- ◆ They both 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)



## Types of decision / prediction problems

**Classification** – Assigns the observation to a class from a small (discrete) 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 sharing similar core ideas

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

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

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

**Scientific visualization** – A high-dimensional problem should be visualized as a 2D image or a 3D scene. We humans do not see more dimensions.

**Neural networks** – one of mathematical formalisms aimed at solving a decision problem without necessarily creating a model of a real biological system.

## Biological motivation

- ◆ A human is considered 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 the 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.

# Generative × discriminative object representation

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

1. **Generative**  $\Rightarrow$  modeling.

- ◆ Attempts to understand physical / other principles and express them by models.
- ◆ This model is able to generate data similar to those observed empirically.
- ◆ The example be the mathematical modeling of a physical / technological phenomenon (in the Newtonian sense).

2. **Discriminative**  $\Rightarrow$  classification.

- ◆ Attempts to understand the outer behaviour without knowing detailed principles (what is unknown for complex objects / phenomena).
- ◆ The output are decisions / prediction in the regression sense.
- ◆ The example is recognition (classification), e.g. determining the diagnosis of the disease by a physician / computer program.



## 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 and thus it is not useful in practice.

## Pattern recognition as an alternative to modeling

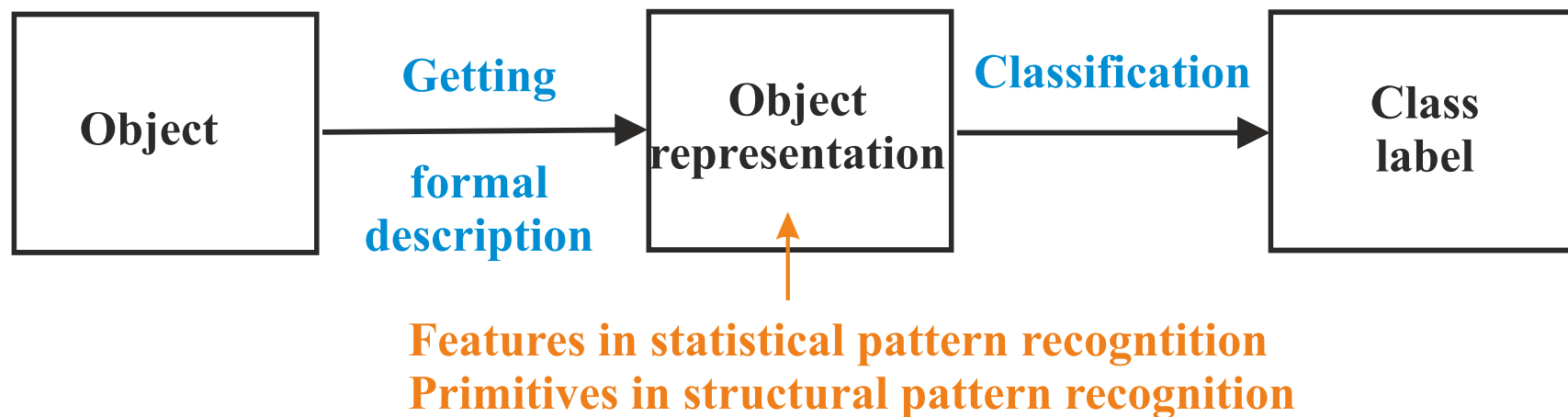
- ◆ Pattern recognition **assigns** observations according to some decision rule **to** a priori known **classes of objects**.
- ◆ Equivalence classes (reflexivity, symmetry, transitivity).
- ◆ Objects within classes are more similar to each other than objects from different classes.
- ◆ The understanding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

# The role of learning in pattern recognition

- ◆ The advantage of pattern recognition 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: It is easier for humans to give examples of a correct classification than to express an explicit classification rule.
- ◆ Three main approaches to learning:
  - Supervised learning** based on the training multiset 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.



# Main approaches to pattern recognition

## 1. Statistical (feature-based) pattern recognition.

- ◆ Statistical model of patterns and pattern classes is assumed.
- ◆ The coordinate axes correspond to individual observations (features, measurements) expressed by a numerical values.
- ◆ Objects are represented as points in a vector space.

## 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  $p_{XY}: X \times Y \rightarrow \mathbb{R}$  and the penalty function  $W: Y \times D \rightarrow \mathbb{R}$

a strategy  $q: X \rightarrow D$  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.
  - ◆ **Classification** is a special case of the decision-making problem where the set of decisions  $D$  and hidden states  $Y$  coincide.

## Generality of the Bayesian formulation (1)

Motto: *“Let set  $X$  (observations) and set  $Y$  (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.

# Generality of the Bayesian formulation (2)

Observation  $x$  can be a number, symbol, function of two variables (e.g., an image), graph, algebraic structure, etc.

Application	Observation	Decisions
value of a coin in a slot machine	$\mathbf{x} \in \mathbb{R}^n$	value
optical character recognition	2D bitmap, gray-level image	characters, words
license plate recognition	2D bitmap, gray-level image	characters, numbers
fingerprint recognition	2D bitmap, gray-level image	personal identity
speech recognition	signal from a microphone $x(t)$	words
EEG, ECG analysis	$\mathbf{x}(t)$	diagnosis
forfeit detection	various	{yes, no}
speaker identification	signal from a microphone $x(t)$	personal identity
speaker verification	signal from a microphone $x(t)$	{yes, no}



## Generative × discriminative classifier

- ◆ Cf. a more general distinction between generative and discriminative models, slide 16 of this lecture.
- ◆ We wish to learn either a decision strategy  $q: X \rightarrow Y$  or the posterior probability  $P(Y|X)$ .

**Generative classifiers**, e.g. naïve Bayes classifier, model-based as Gaussian mixture model, ...

- ◆ Assume  $P(X|Y)$ ,  $P(Y)$  be functions.
- ◆ Estimate  $P(X|Y)$ ,  $P(y)$  from training data directly.
- ◆ Use Bayes rule to calculate  $P(Y|X = x_i)$
- ◆ ‘Generative’ means that a model produces data subject to the probability distribution via sampling.

**Discriminative classifiers**, e.g. perceptron, SVM,  $k$ -NN, ...

- ◆ Assume posterior  $P(Y|X)$  be a function.
- ◆ Estimate  $P(Y|X)$  from training data.
- ◆ ‘Discriminative’ means that the model enables classification of  $x$  and cannot generate  $x$  complying the probability model.

# What has been known in statistical pattern recognition?

- ◆ Bayesian formulation based on a known statistical model.
- ◆ Solution to some special non-Bayesian tasks, e.g., with the class “I do not know.” (called also reject option), minimax classifier, tasks with non-random interventions.
- ◆ Linear classifiers and their learning. E.g., a popular special case—Support Vector Machines.
- ◆ Embedding of a non-linear problem to a higher dimensional vector space, mainly locally acting kernel methods.
- ◆ Estimate of needed length of the training set for prescribed precision and reliability of classification (e.g., Vapnik-Chervonenkis’ theory of learning).
- ◆ Unsupervised learning, variants of EM algorithm.

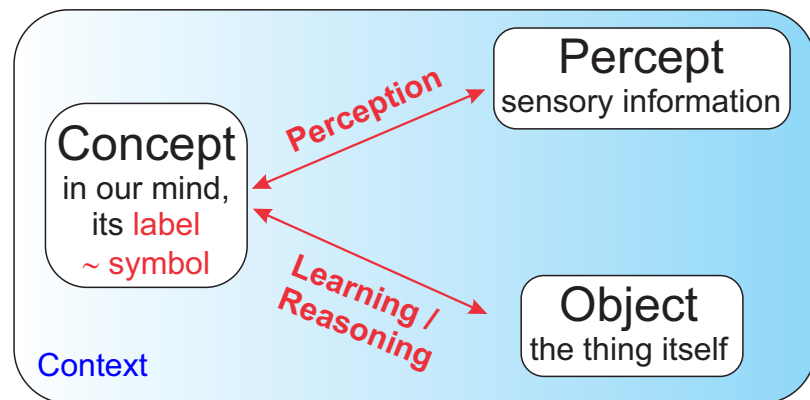
## Application of mathematical statistics

- ◆ 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**. See the next slide.

## Image analysis & objects

- ◆ Failure for images  $f(x, y)$ , where  $f$  is brightness or color of a pixel and  $x, y$  are pixel coordinates.
- ◆ Inverting image formation process leads to an ill-posed task and thus useless practically.
- ◆ We need to anchor to the concept **objects** and explore its semantics..
- ◆ The object detection, its segmentation in images is a chicken and egg problem. The link between semantics and the object appearance is needed.

- 
- ◆ Knowledge  $\sim$  Observations + Context + Experience
  - ◆ A problematic symbol grounding issue.



## 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.

