UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

CONTEXT & VOCABULARY

WHAT REPRESENTS ARTIFICIAL INTELLIGENCE?

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

What is Artificial intelligence?

- The term ***Artificial Intelligence***, as a research field, was coined at the conference on the campus of Dartmouth College in the summer of **1956**, even though the idea was around since antiquity.
- For instance in the first manifesto of Artificial Intelligence, "*Intelligent Machinery*", in **1948** Alan Turing distinguished two different approaches to AI, which may be termed ***"top-down"*** or ***knowledge-driven AI*** and ***"bottom-up"*** or ***data-driven AI***

(sources: Wikipedia & http://www.alanturing.net & Stanford Encyclopedia of Philosophy)

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

What is Artificial intelligence?

- The two different approaches to AI can be detailed:
  - ***"top-down"*** or ***knowledge-driven AI***
    - cognition = high-level phenomenon, independent of low-level details of implementation mechanism, first neuron (1943), first neural network machine (1950), neucognitron (1975)
    - Evolutionary Algorithms (1954,1957, 1960), Reasoning (1959,1970), Expert Systems (1970), Logic, Intelligent Agent Systems (1990)…
  - ***"bottom-up" or data-driven AI***
    - opposite approach, start from data to build incrementally and mathematically mechanisms taking decisions
    - Machine learning algorithms, Decision Trees (1983), Backpropagation (1984-1986), Random Forest (1995), Support Vector Machine (1995), Boosting (1995), Deep Learning (1998/2006)…

(sources: Wikipedia & http://www.alanturing.net & Stanford Encyclopedia of Philosophy)

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

What is Artificial intelligence?

- *AI is originally defined, by Marvin Lee Minsky, as "the construction of computer programs doing tasks, that are, for the moment, accomplished more satisfyingly by human beings because they require high level mental processes such as: learning. perceptual organization of memory and critical reasoning".*

- There are so the "artificial" side with the usage of computers or sophisticated electronic processes and the side "intelligence" associated with its goal to imitate the (human) behavior.

(sources: Wikipedia & http://www.alanturing.net & Stanford Encyclopedia of Philosophy)

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

What is Artificial intelligence?

- The concept of ***strong artificial intelligence*** makes reference to a machine capable not only of producing intelligent behavior, but also to experience a feeling of a real sense of itself, “real feelings” (whatever may be put behind these words), and "an understanding of its own arguments”.

- The notion of ***weak artificial intelligence*** is a pragmatic approach of engineers: targeting to build more autonomous systems (to reduce the cost of their supervision), algorithms capable of solving problems of a certain class, etc. But this time, the machine *simulates* the intelligence, it seems to act as if it was smart.

(sources: Wikipedia & http://www.alanturing.net & Stanford Encyclopedia of Philosophy)

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

Why Artificial Intelligence is so difficult to grasp?

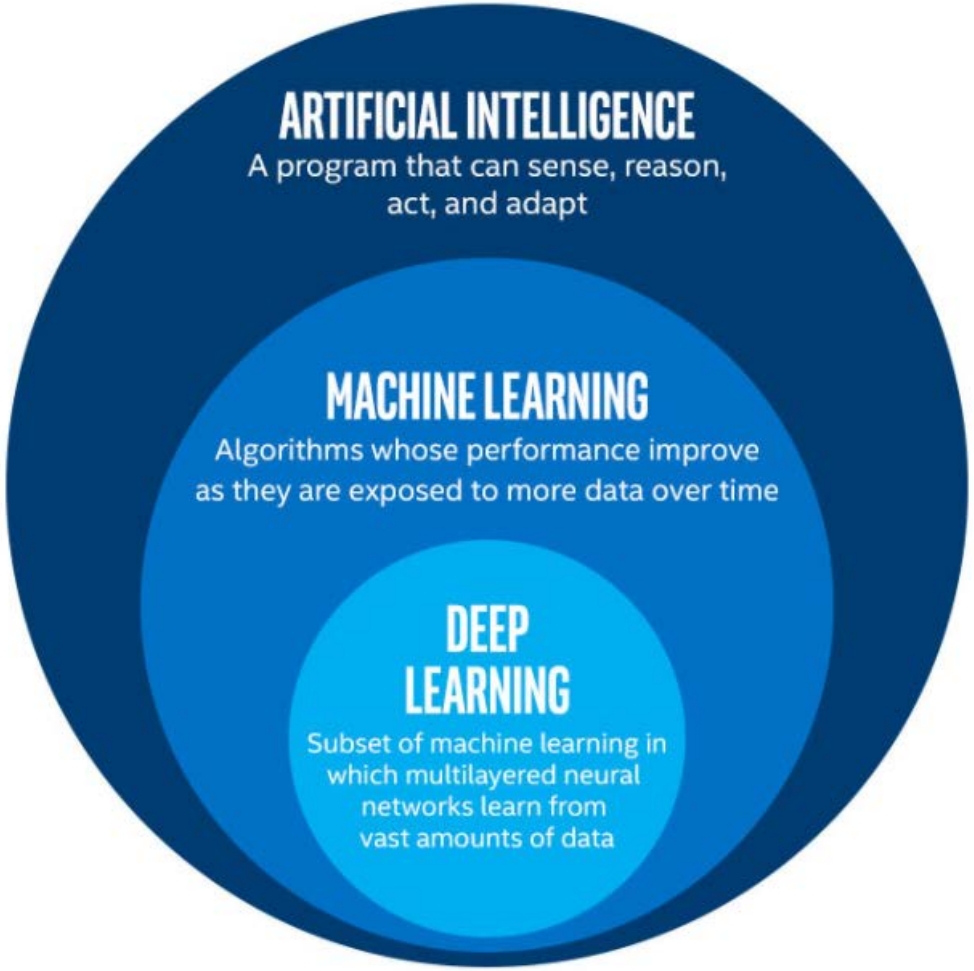
- Frequently, when a technique reaches mainstream use, it is no longer considered as artificial intelligence; this phenomenon is described as the AI effect: "AI is whatever hasn't been done yet.“ (*Larry Tesler's Theorem*) -> e.g. Path Finding (GPS), Checkers game, Chess electronic game, Alpha Go…

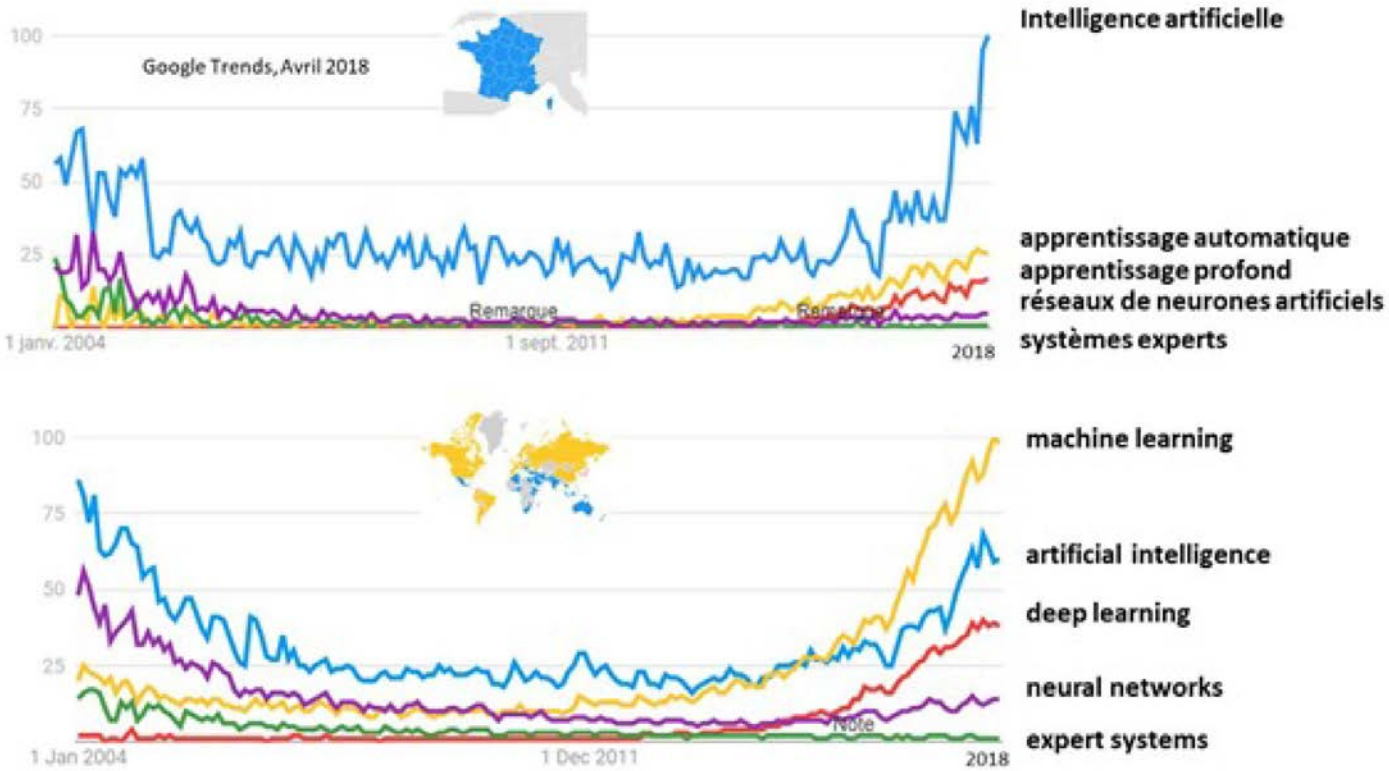
- Consequently, AI domain is continuously evolving and so very difficult to grasp.

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

AI vs Machine Learning vs Deep Learning



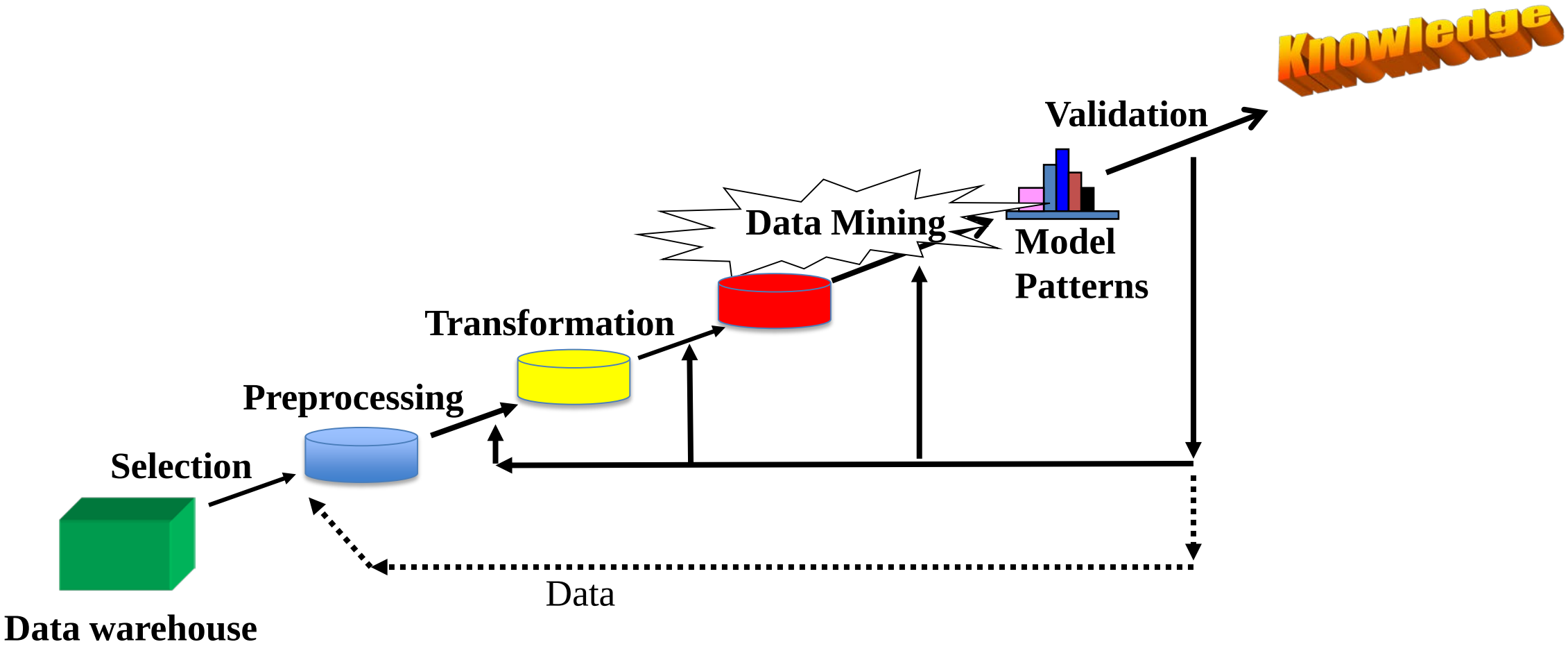


MACHINE LEARNING VS DATA MINING?

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

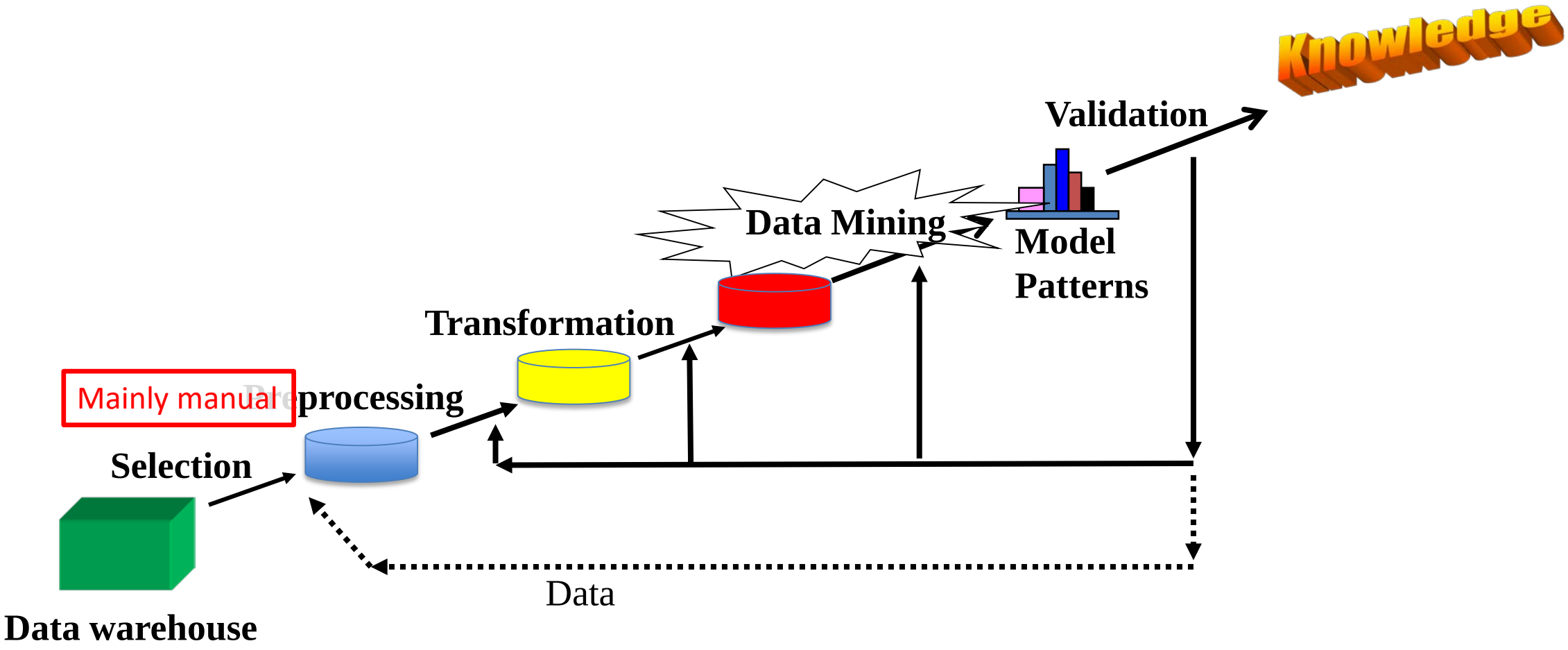
Data Mining Workflow



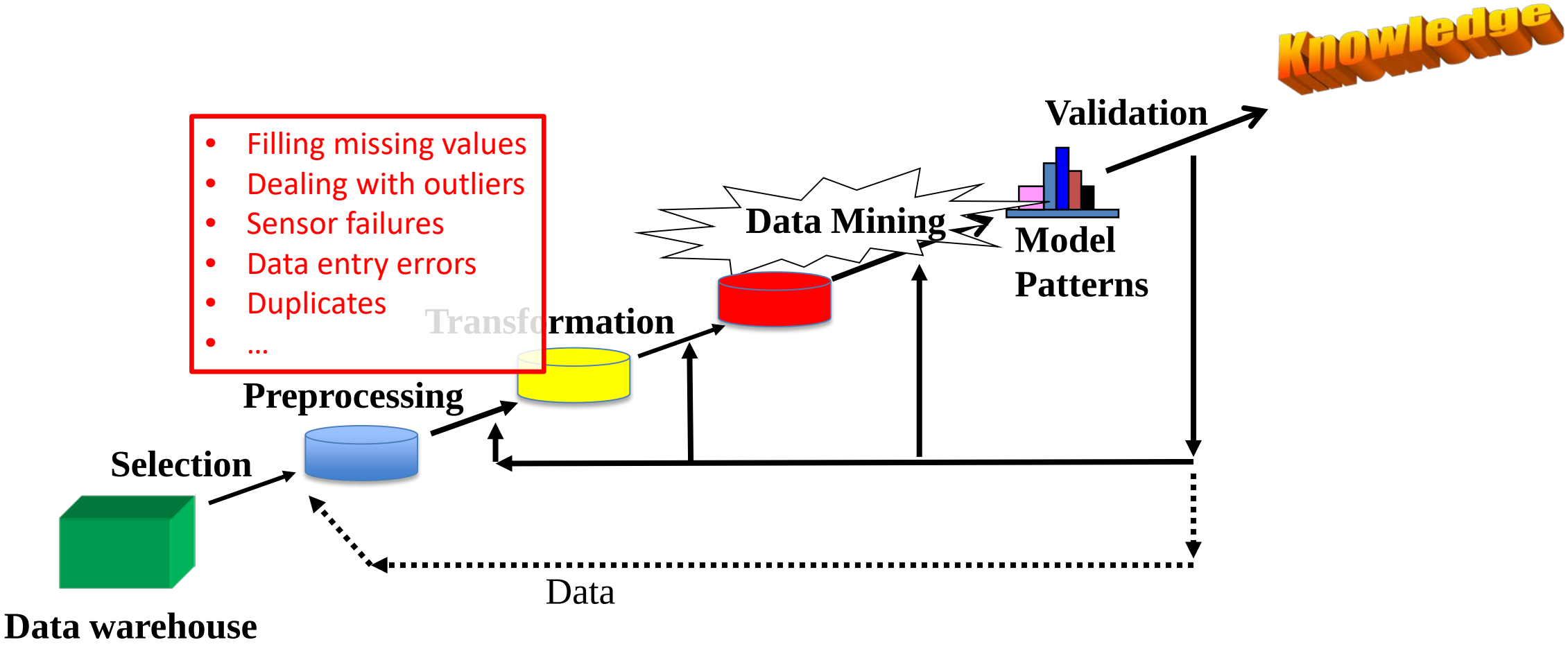
UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

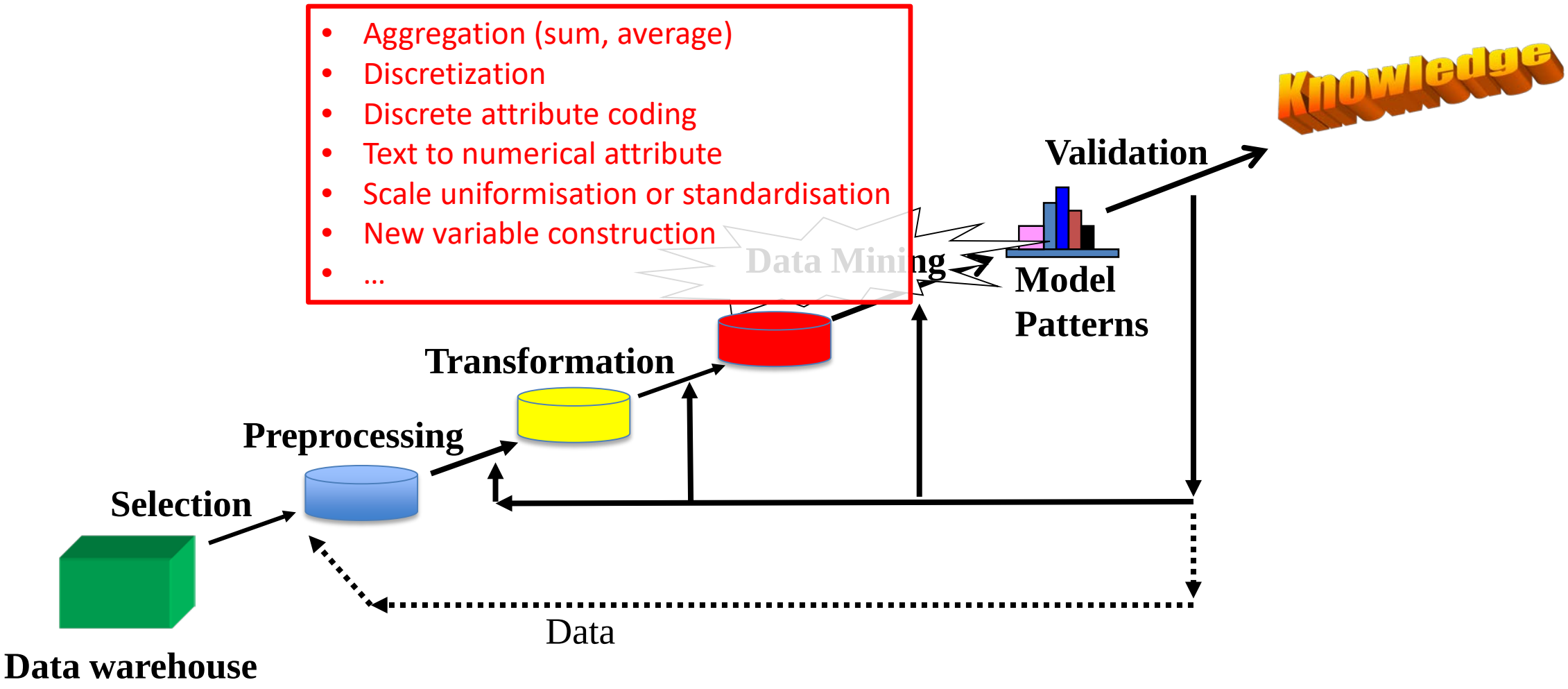
Data Mining Workflow



Data Mining Workflow



Data Mining Workflow

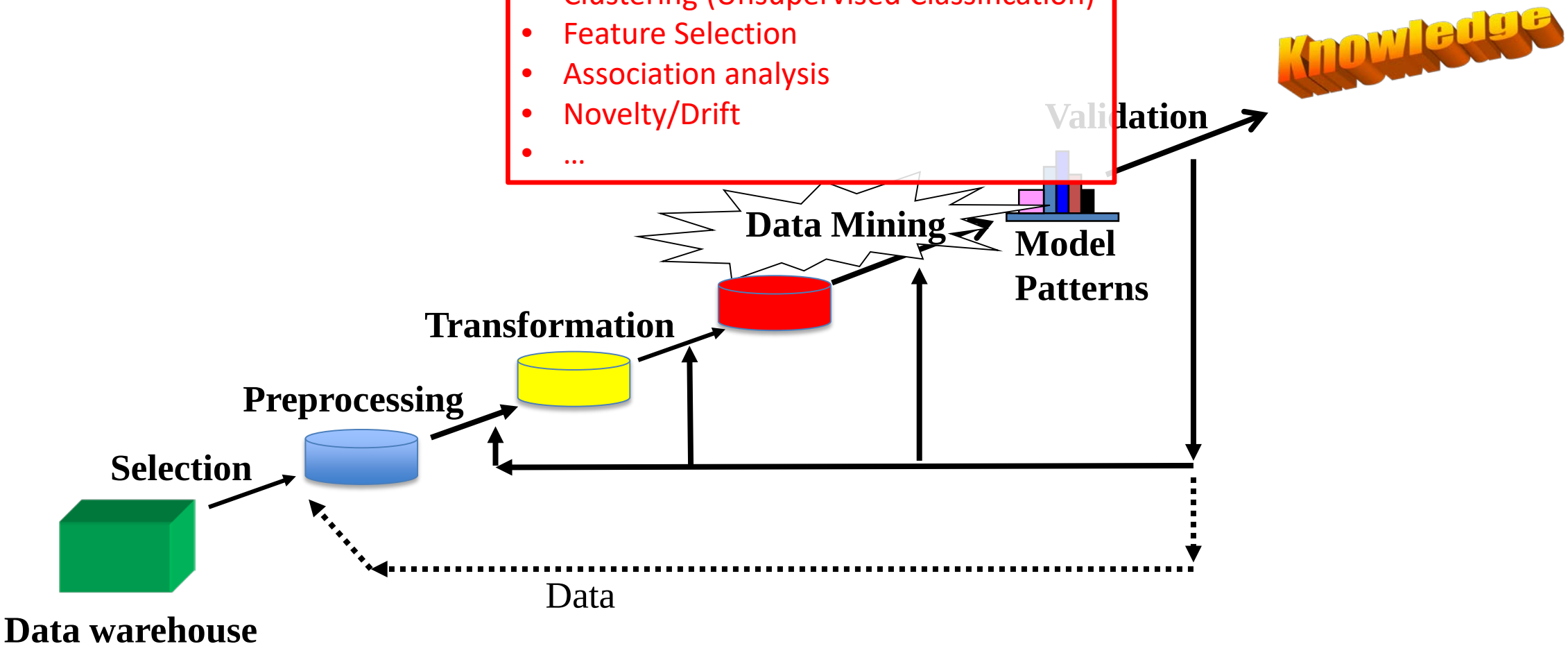


UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Data Mining Workflow

- Regression
- (Supervised) Classification
- Clustering (Unsupervised Classification)
- Feature Selection
- Association analysis
- Novelty/Drift
- …

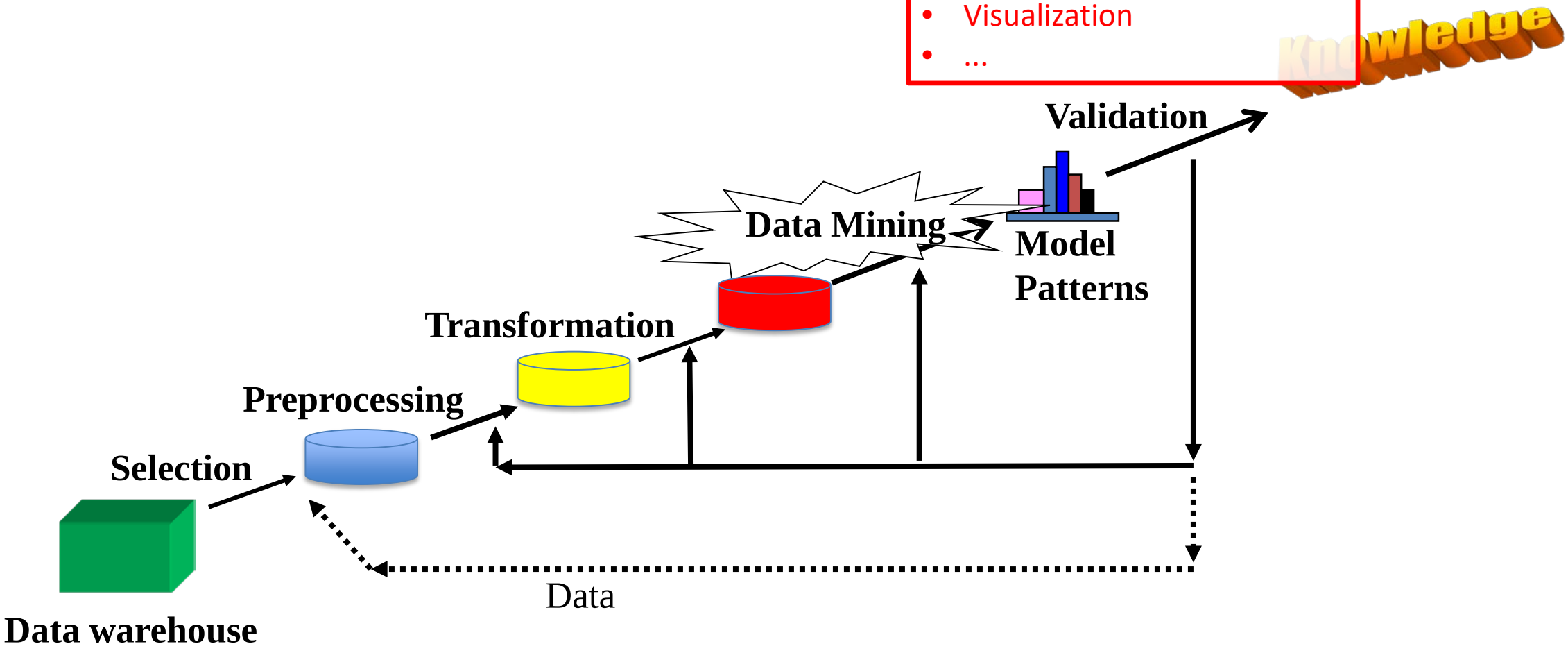


UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Data Mining Workflow

- Evaluation on Validation Set
- Evaluation Measures
- Visualization
- ...

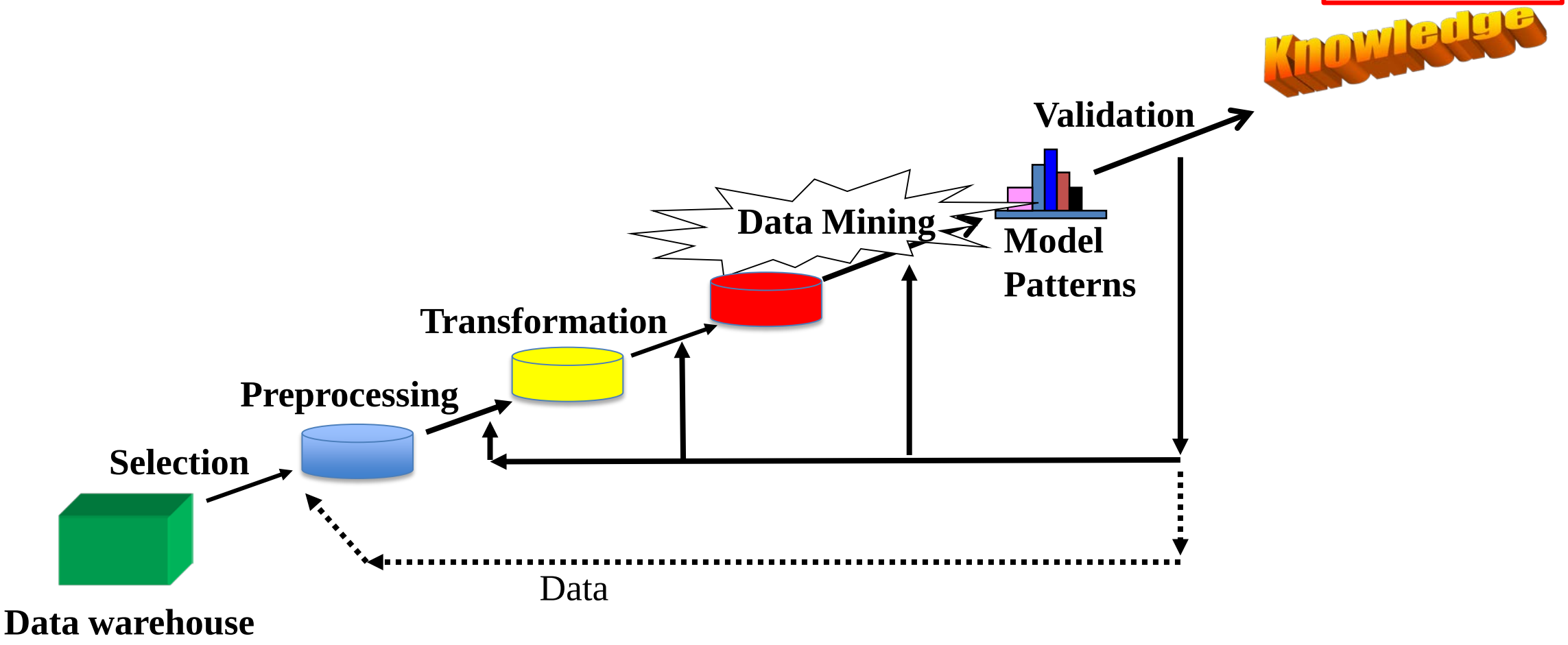


UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Data Mining Workflow

- Visualization
- Reporting
- Knowledge
- ...



UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Data Mining Workflow

Problems

- Regression
- (Supervised) Classification
- Density Estimation / Clustering (Unsupervised Classification)
- Feature Selection
- Association analysis
- Anomality/Novelty/Drift
- …

Possible Solutions

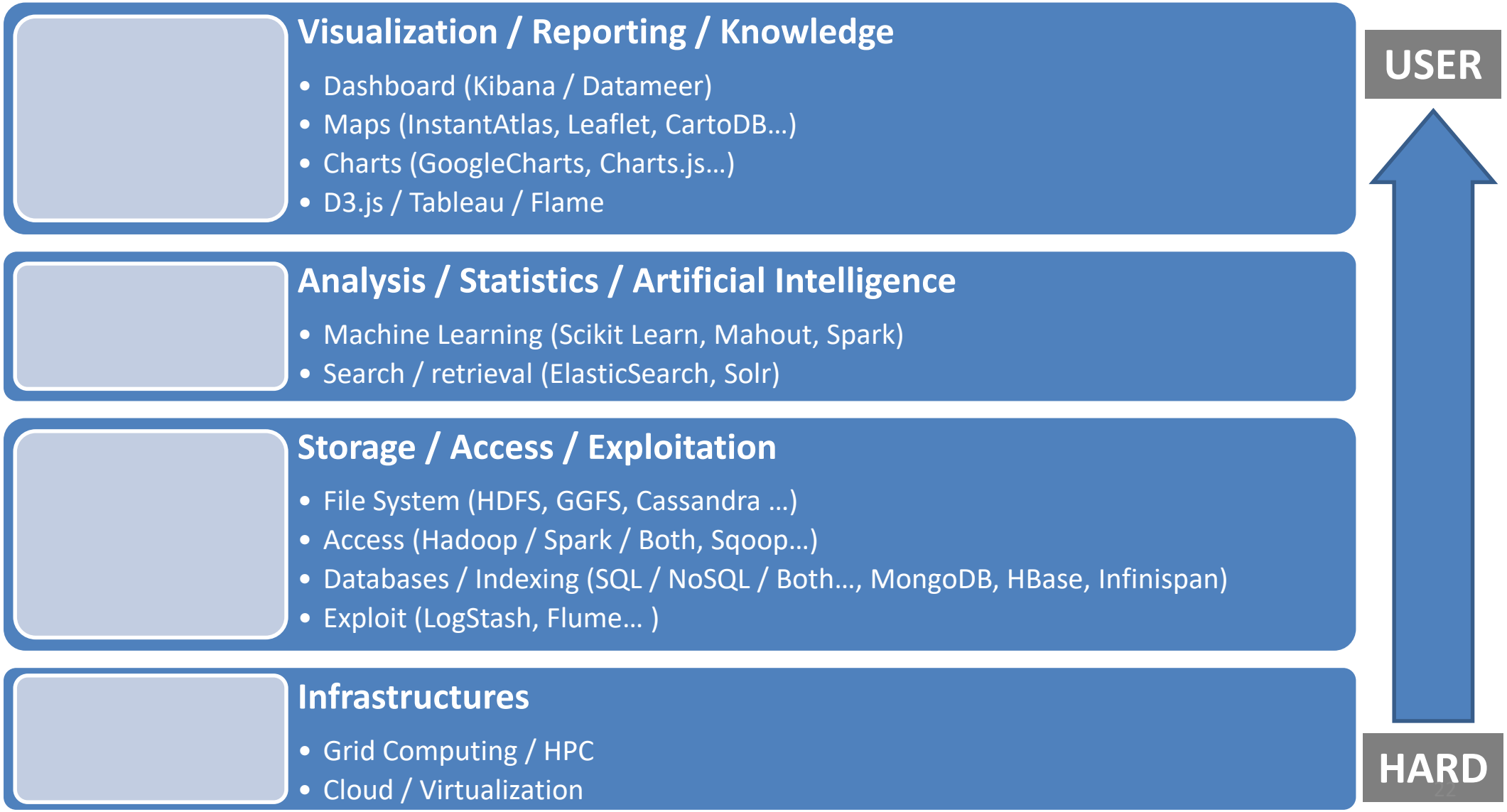
- Machine Learning
  - Support Vector Machine
  - Artificial Neural Network
  - Boosting
  - Decision Tree
  - Random Forest
  - …
- Statistical Learning
  - Gaussian Models (GMM)
  - Naïve Bayes
  - Gaussian processes
  - …
- Other techniques
  - Galois Lattice
  - …

MACHINE LEARNING VS DATA SCIENCE?

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Data Science Stack

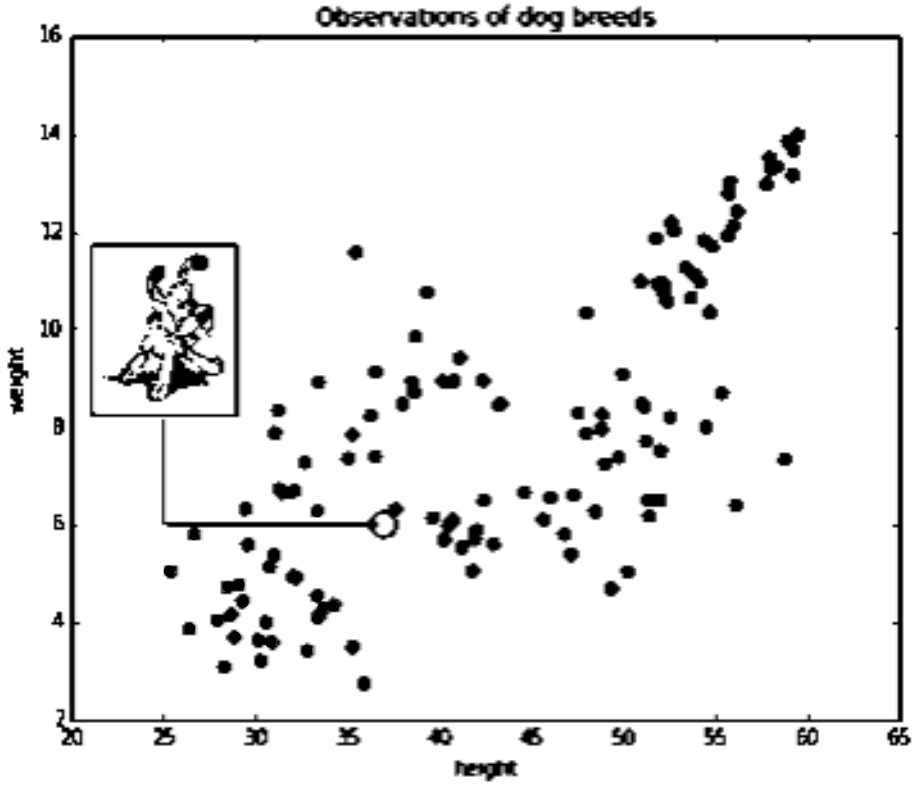


MACHINE LEARNING VS STATISTICS?

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

What breed is that Dogmatix (Idéfix) ?

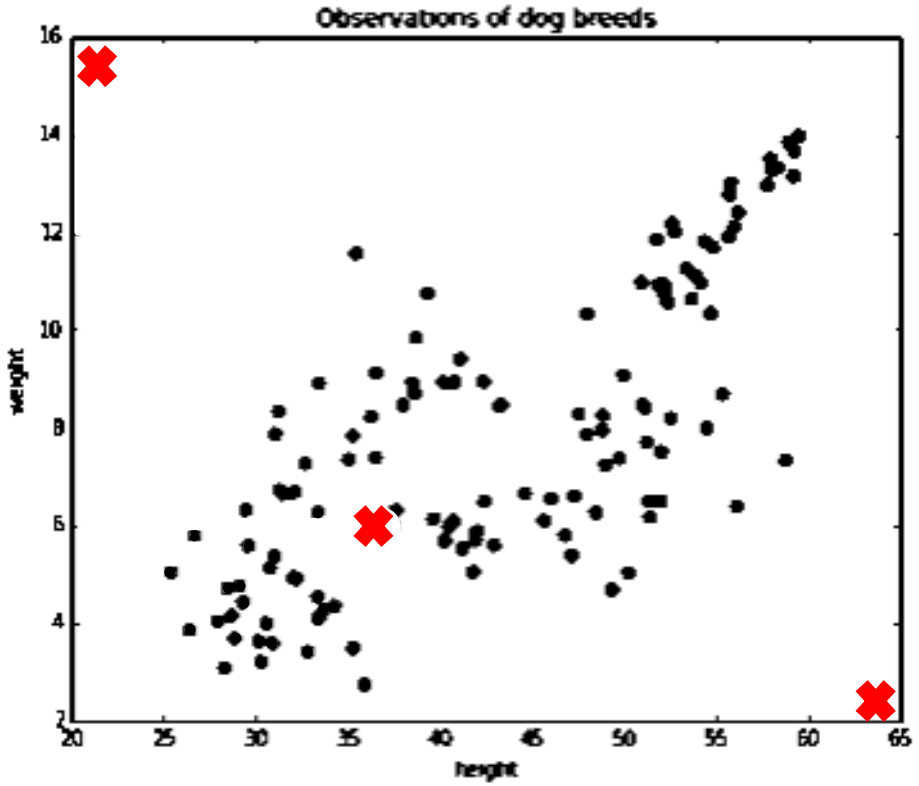


The illustrations of the slides in this section come from the blog "Bayesian Vitalstatistix: What Breed of Dog was Dogmatix?"

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

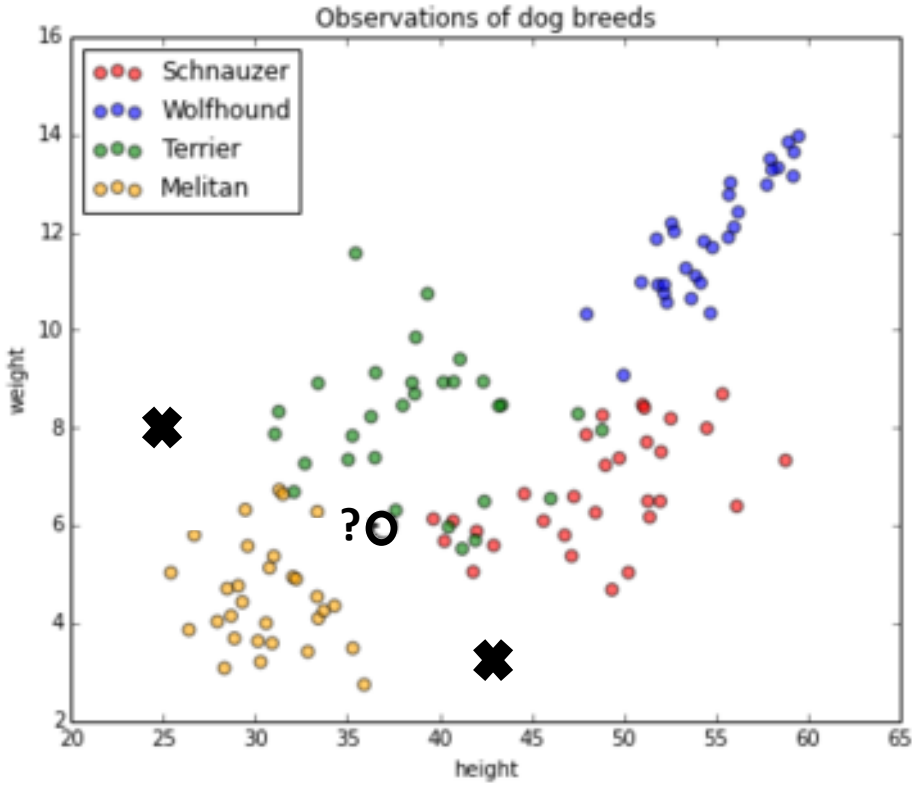
Does any real dog get this height and weight?



- Let us consider ***x***, vectors independently generated in $\boldsymbol{R}^d$ (here $\boldsymbol{R}^2$), following a probability distribution fixed but ***unknown*** $P(\boldsymbol{x})$.

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

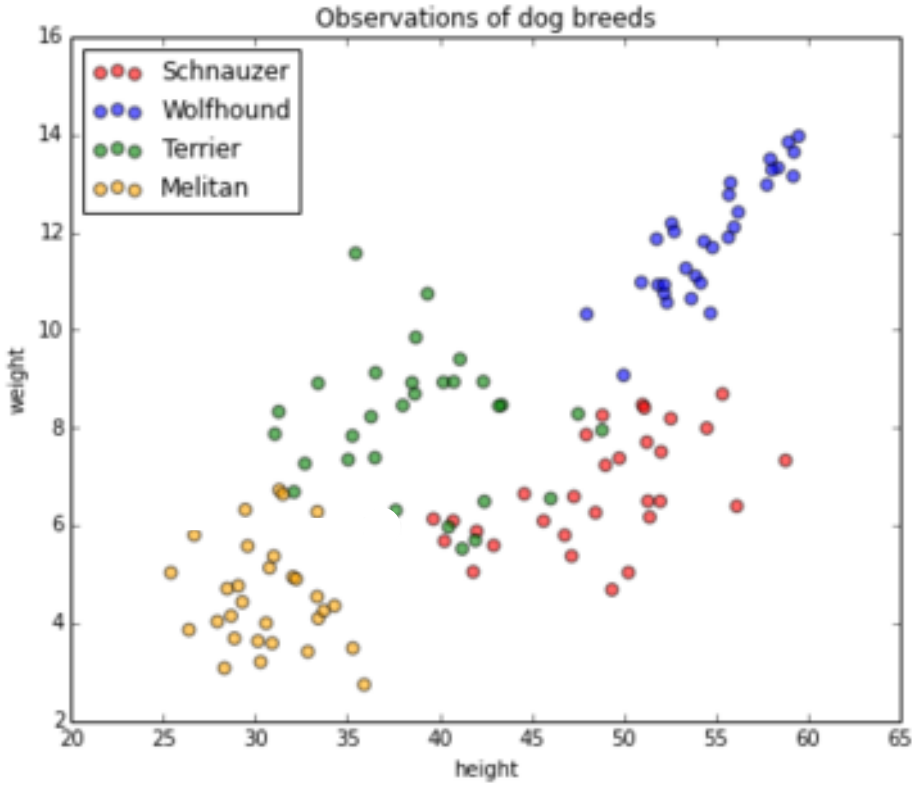
What should be the breed of these dogs?



- An Oracle assignes a value *y* to each vector ***x*** following a probability distribution *P(y|**x**)* also fixed but ***unknown***.

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

An oracle provides me with examples?

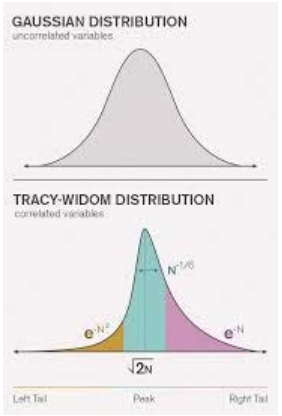


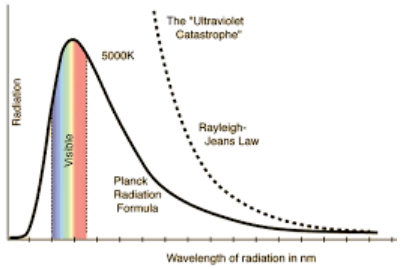
- Let S be a training set
  $S = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, \boldsymbol{y}_m)\}$,
  with m **training samples i.i.d.** which follow the **joint probability**
  $P(\boldsymbol{x}, y) = P(\boldsymbol{x})P(y|\boldsymbol{x})$.

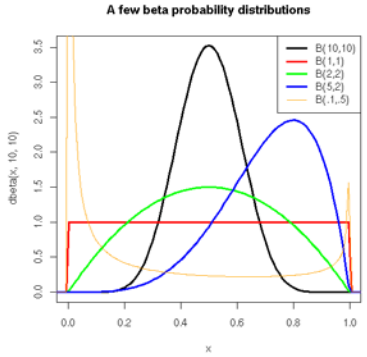
UNIVERSITÉ CÔTE D'AZUR

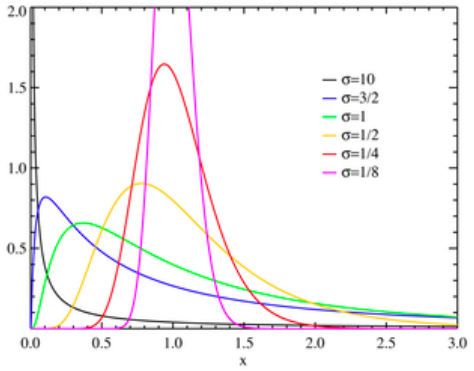
univ-cotedazur.fr

Statistical solution: Models, Hypotheses…





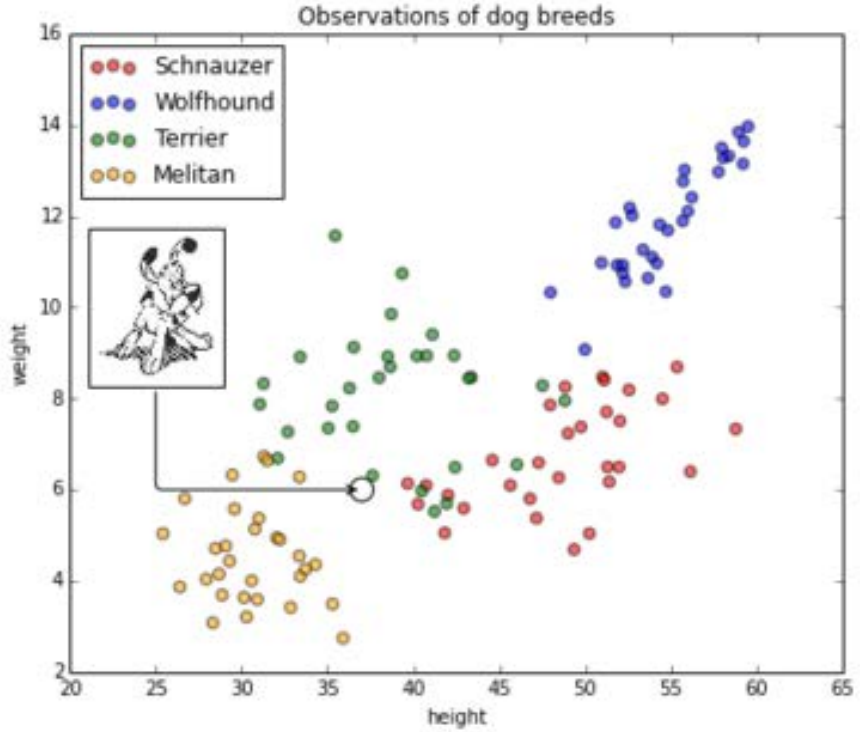


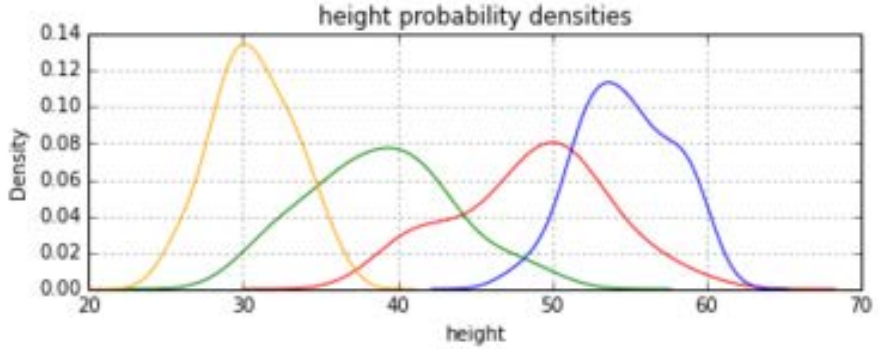


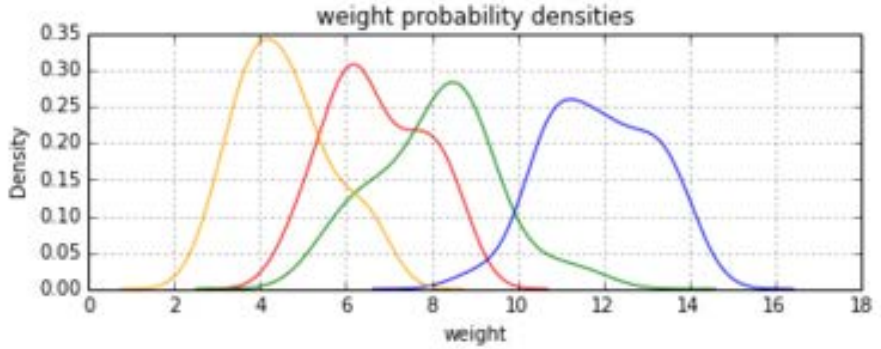
UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Statistical solution P(height, weight|breed)…



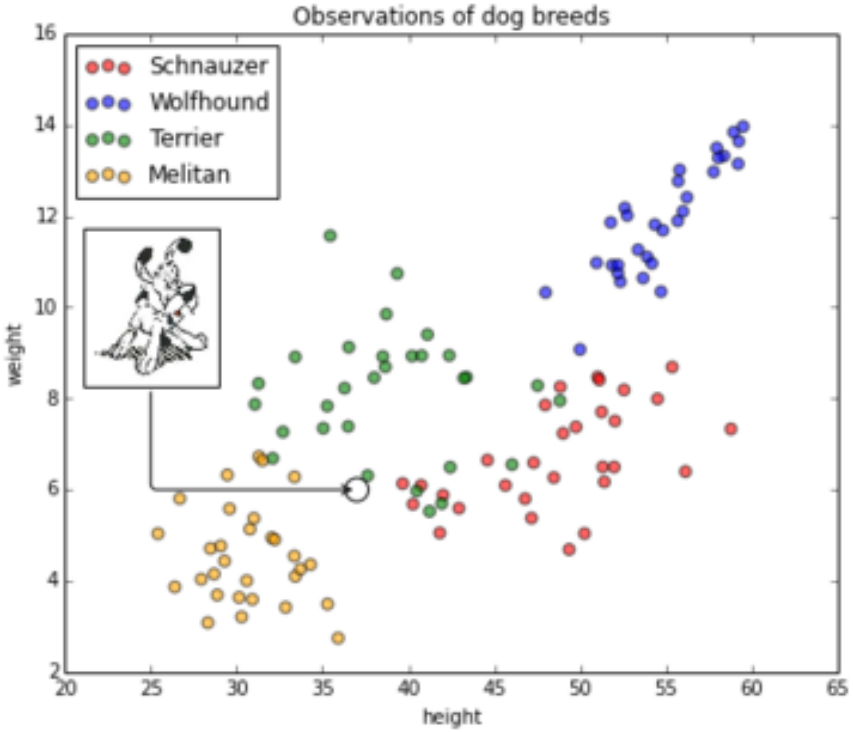


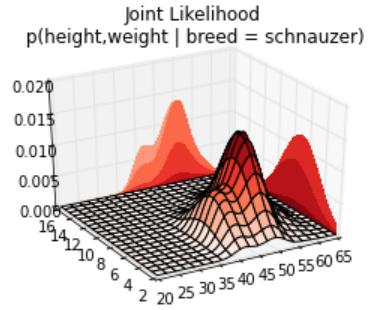


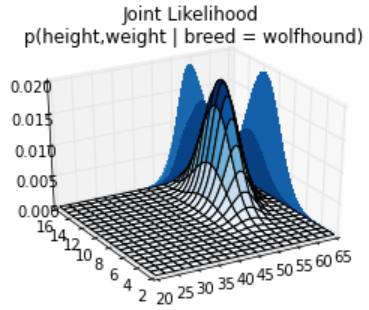
UNIVERSITÉ CÔTE D'AZUR

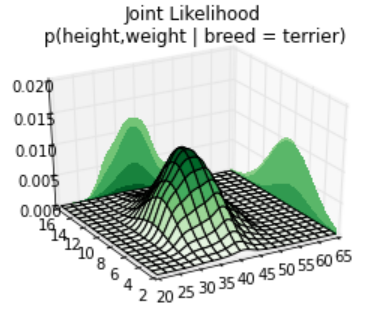
univ-cotedazur.fr

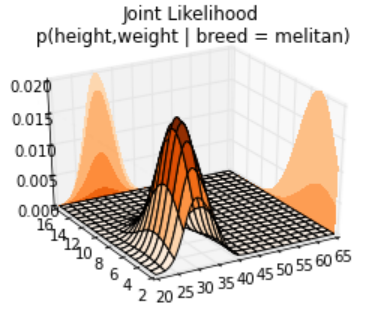
Statistical solution P(height, weight|breed)…







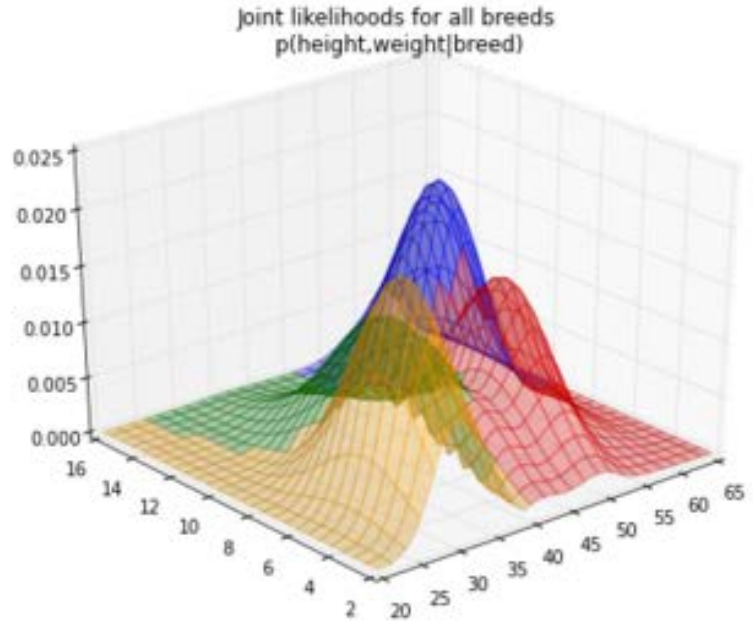


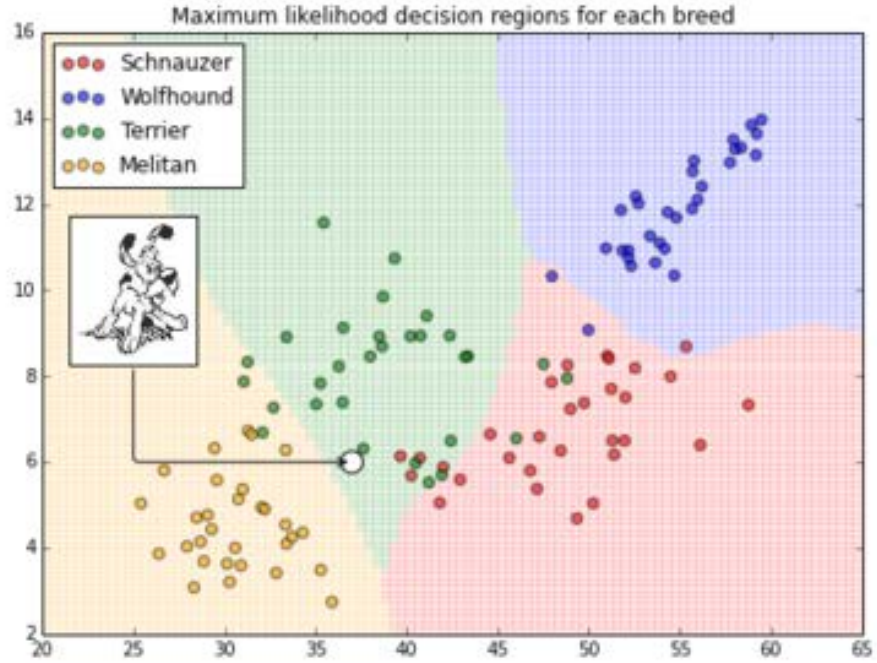


UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Statistical solution P(height, weight|breed)…

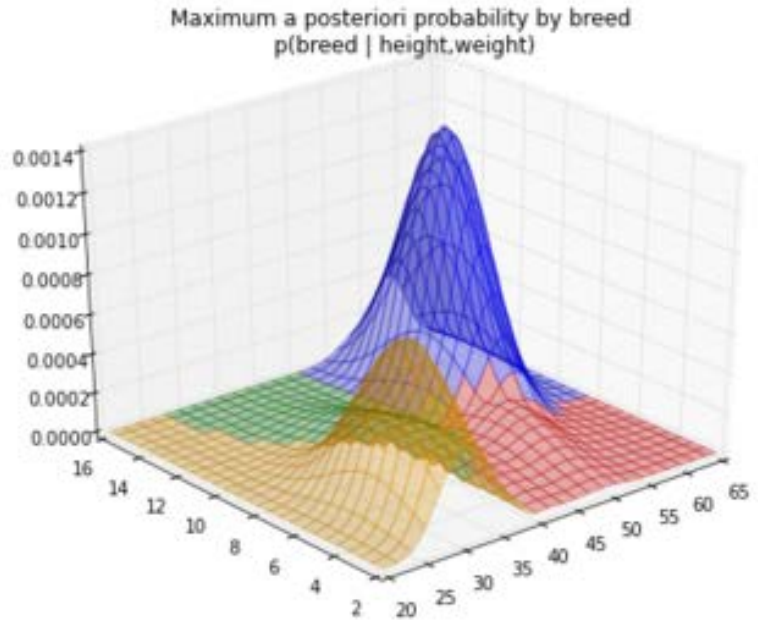


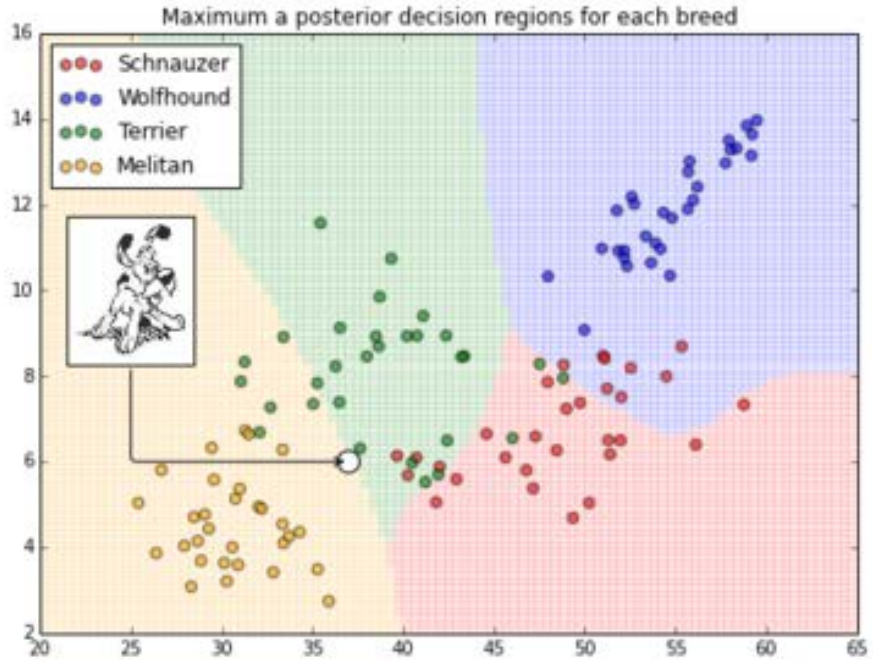


UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Statistical solution: Bayes, P(breed|height, weight)…

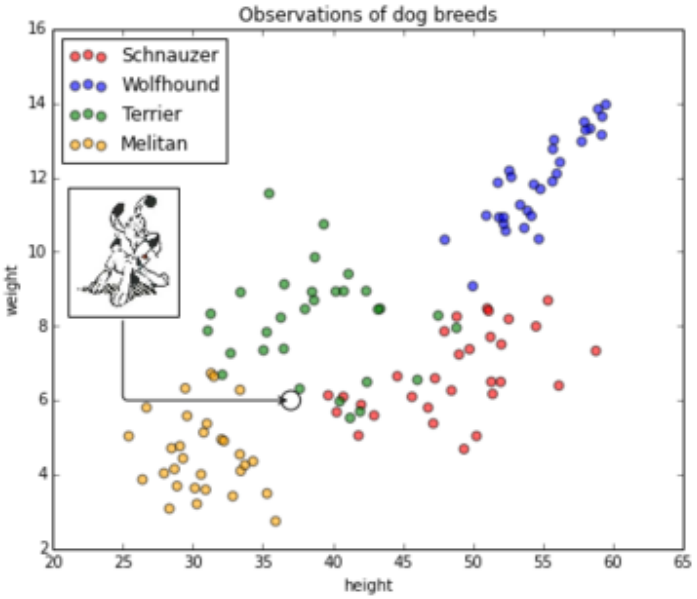




UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Machine Learning



- we have a learning machine which can provide a family of functions *{f(**x**;α)}*, where α is a set of parameters.

$$\left(\mathbf{x}\right) \xrightarrow{f(\mathbf{X},\boldsymbol{\alpha})\ ?} y$$

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

The problem in Machine Learning

$$\left(\mathbf{x}\right) \xrightarrow{f(\mathbf{X},\boldsymbol{\alpha})\ ?} y$$

- The problem of learning consists in finding the function (among the $\{f(\boldsymbol{x};\alpha)\}$) which provides the best approximation $\hat{y}$ of the true label y given by the Oracle.
- best is defined in terms of minimizing a specific (error) cost ***related to your problem/objectives***

  $Q((\boldsymbol{x}, y), \alpha) \in [a; b]$.
- Examples of cost functions Q:
  - ***Hinge Loss***: error 0/1 cost, 0 if predicted and expected labels match, 1 otherwise (used in classification)
  - ***Quadratic Loss***: $(f(\boldsymbol{x})-y)^2$ (used in regression)
  - ***Cross-Entropy Loss, Logistic Loss…***

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

The problem in Machine Learning

For Clarity sake, let us note $z = (\boldsymbol{x}, y)$.

- Thus, the objective is to minimize the ***Risk***, i.e. the expectation of the error cost:

$$R(\alpha) = \int Q(z, \alpha) dP(z)$$

where $P(z)$ is unknown.

The training set $S = \{z_i\}_{i=1,\ldots,m}$ is built through an *i.i.d.* sampling according to $P(z)$. Since we cannot compute $R(\alpha)$, we look for minimizing the ***Empirical Risk*** instead:

$$R_{emp}(\alpha) = \frac{1}{m} \sum Q(z_i, \alpha)$$

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

Machine Learning fundamental Hypothesis

For Clarity sake, let us note $z = (x, y)$.

$S = \{z_i\}_{i=1,\dots,m}$ is built through an *i.i.d.* sampling according to P(z).

Machine Learning ⇐ ***Statistics***

Train through Cross-Validation

Machine Learning ⇐ ***Statistics***

Training set & Test set have to be distributed according to the same law (i.e. P(z)).

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Vapnik learning theory (1995)

Vapnik had proven the following equation $\forall m$ with a probability at least equal to $1 - \eta$:

$$R(\alpha_m) \leq R_{emp}(\alpha_m) + (b - a)\sqrt{\frac{d_{VC}(ln(2m/d_{VC}) + 1) - ln(\eta/4)}{m}}$$

Training Error ***Generalization Error***

Thus minimizing the ***Risk*** depends on minimizing the ***Empirical Risk*** and the ***confidence interval*** which is linked to the term d_{VC} corresponding to the complexity of the model family chosen, i.e. the Vapnik-Chervonenkis dimension

UNIVERSITÉ CÔTE D'AZUR
univ-cotedazur.fr

Vapnik learning theory (1995)

In his learning theory [Vapnik, 1995], Vapnik defines 4 fundamental steps:

- Study the theory of consistence of learning processes
- Define bounds on convergence speed of learning processes
- Handle the generalization power of learning processes
- Design a theory to build learning algorithms in order to find a tradeoff between minimizing the ***Empirical Risk*** and the ***confidence interval*** $\Rightarrow$ minimization of the ***Structural Risk***.

UNIVERSITÉ CÔTE D'AZUR

univ-cotedazur.fr

Machine Learning vs Statistics