



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**



AI and Predictive Analytics in Data-Center Environments

Data Science and Engineering

Introduction

“Before doing experiments, we have to know which question we want so solve”

Data Science and Engineering

Which kind of question are we asking to/about our data?

Descriptive Questions

How is that data?

- Which features?
- Which ranges?
- Descriptive, ordinal or quantity variables?
- Features follow a distribution?
- ... ?

Exploratory Questions

Are there Relations?

- Related or correlated features?
- Features result of combination of other features?
- Features derived from others?
- ... ?

Inductive Questions

Does data Repeats?

- Repetitive values?
- Repetitive patterns?
- Recurring data?
- ... ?

Predictive Questions

Can we predict data?

- Does data follow a model?
- Can be forecasted?
- Can data be classified/regressed?
- ... ?

Causal Questions

What causes that data?

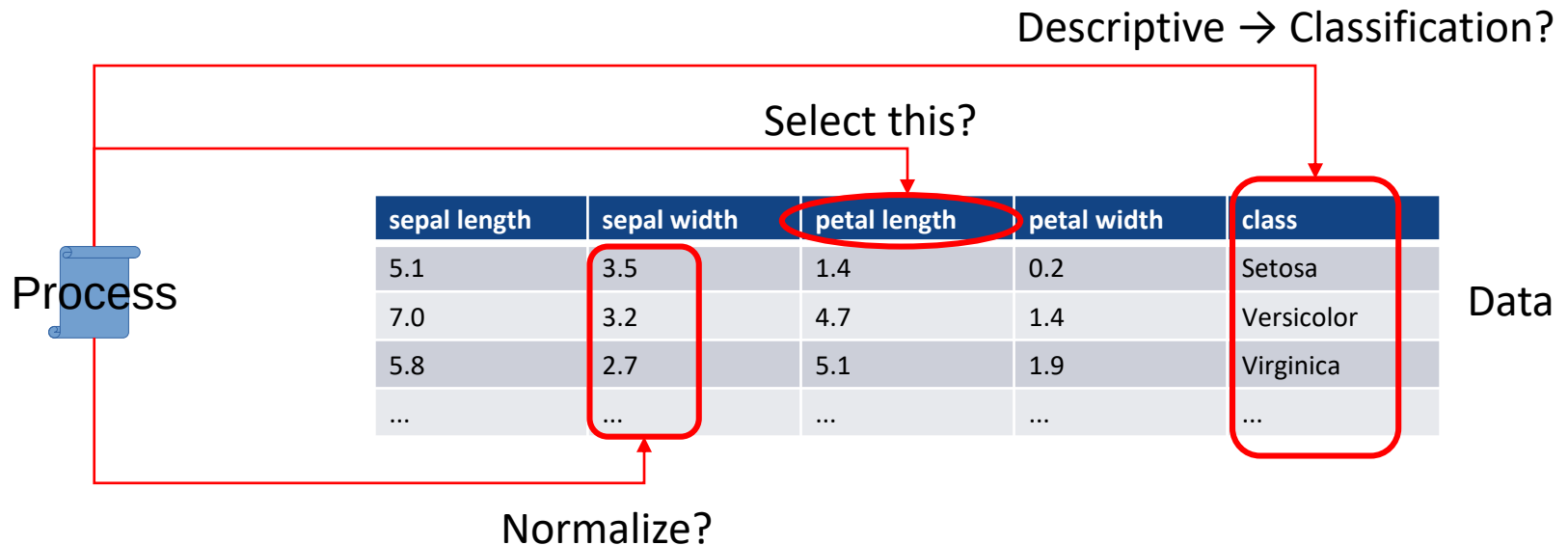
- Which mechanism is behind data?
- ... ?

Algorithms

- Statistical analysis can help solving those questions...
- ... but sometimes “algorithms” and “heuristics” are required

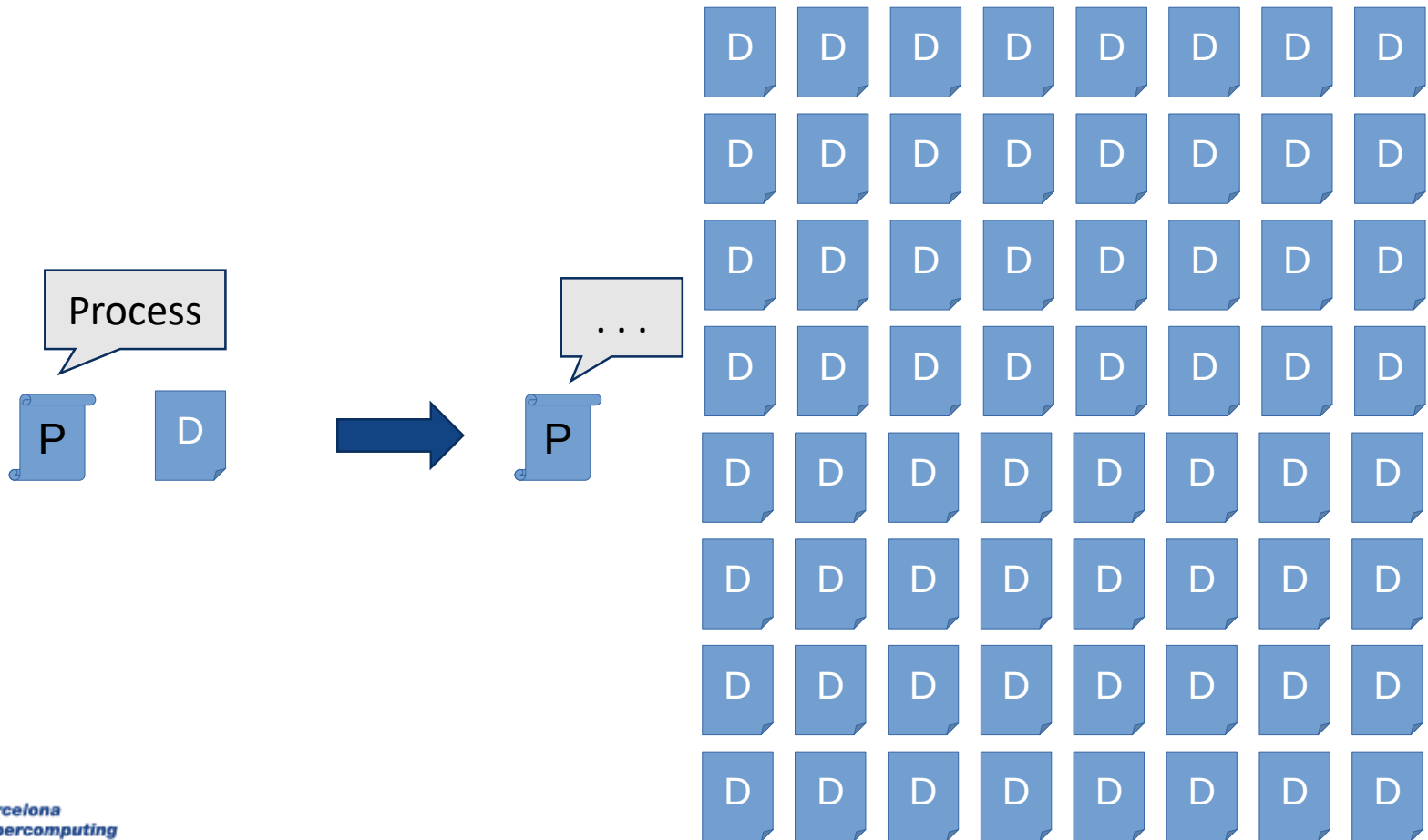
Automatic statistical analysis

- Complexity on analyzing that data:
 - More than simplistic statistical analysis
 - Decisions to be made when selecting features, format data, ...



Automatic statistical analysis

- Complexity on analyzing that data:
 - Repetitive or exhaustive analysis to be done



Automatic statistical analysis

- “Let the machine do the analysis”
 - and select which features to examine
 - and use heuristics to decide which analyses to do
 - and decide/rank when something is relevant
 - and return a reusable data model
- “Machine Learning” → “Automatic Learning” (or “Modeling”)

Machine Learning

- Automatic methods to create models
 - Engineering algorithms to perform Statistical Analyses
 - Designing heuristics to decide which analyses to be done
 - Solving problems using those methods
- An application of Statistics and Artificial Intelligence

Methodology

Example

Examples

- “Our data seems to show no patterns”
- No clear or strong correlations
 - Try to find relations on data combinations?
 - Try to find relations on data transformations?
- No clear “device” generating that data
 - Try to find similar groups on data?

Examples

- “Certain features seem to divide data better than others?”
- Divide data by those features
 - “Divide and Conquer” strategy on data!
- Repeat analyses on each group
 - Statistical Modeling, check a mechanism in similarities
- Use a heuristic to decide first-analyzed features
 - “Gain of Information” techniques to decide where to start and where to follow

Examples

- “A subset shows something like patterns”
- Repeat analysis in subset
 - That’s “Tree-Search” strategy
- ...“Now all subsets can be directly modeled”
 - FIN 😊
 - (That’s actually called the “RecPart-Tree” algorithm)