Data Mining in a complex world

Hugues Bersini
IRIDIA/CODE

Modelling the data: WHY ??

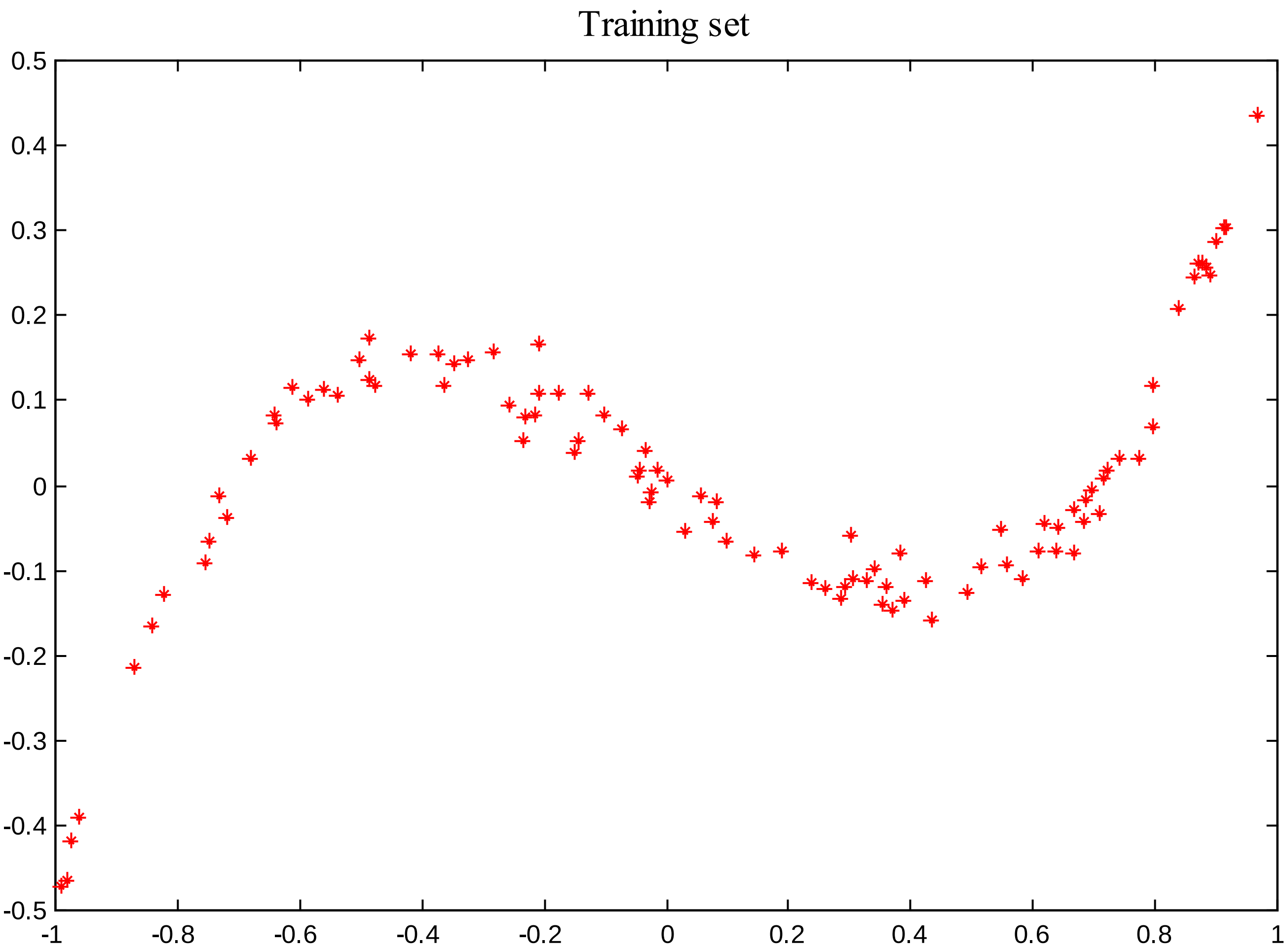
only if structure and regularities in the data
data contains the needed information in a hidden form !!

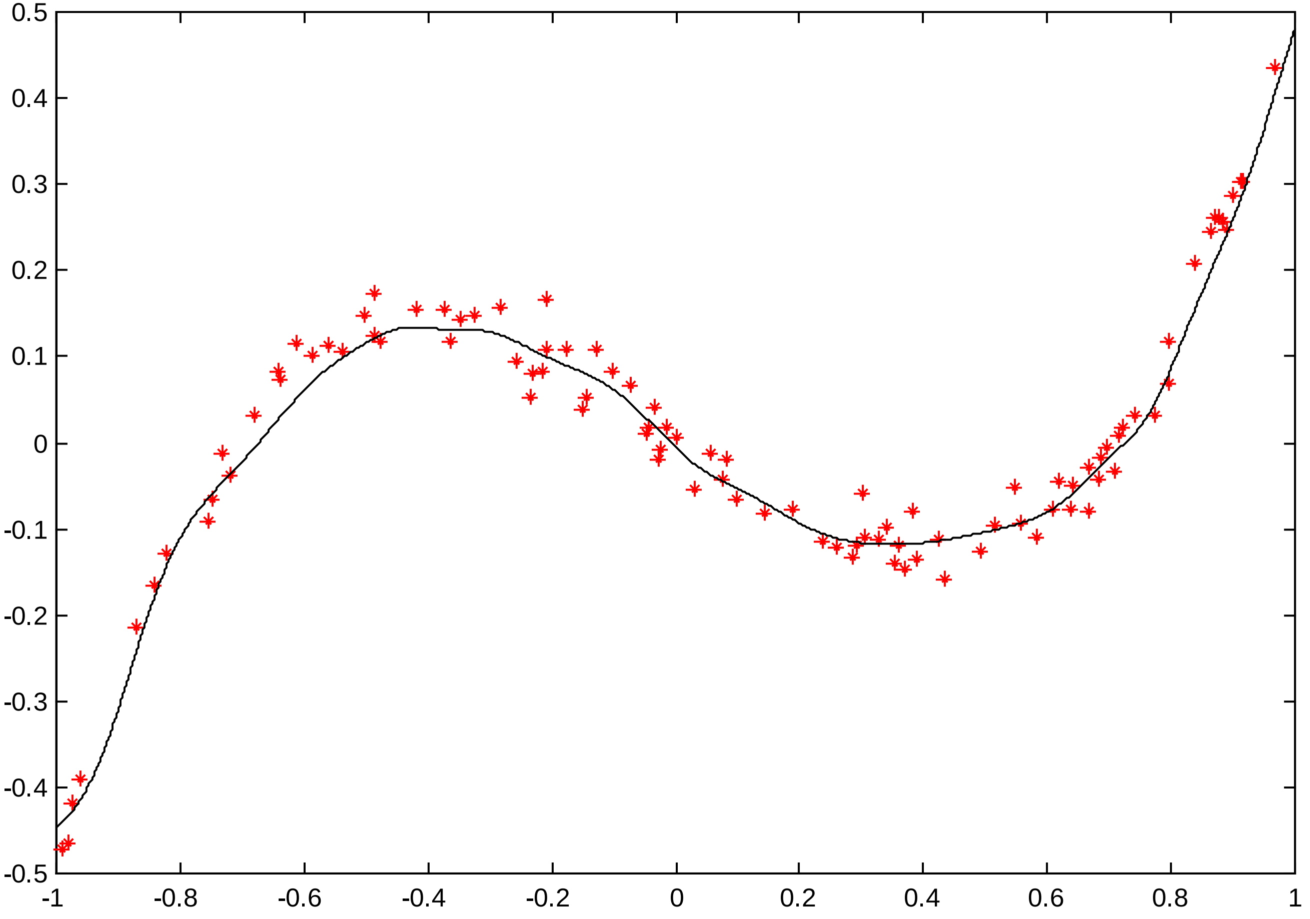
To compress the data

To understand the data

To predict new data

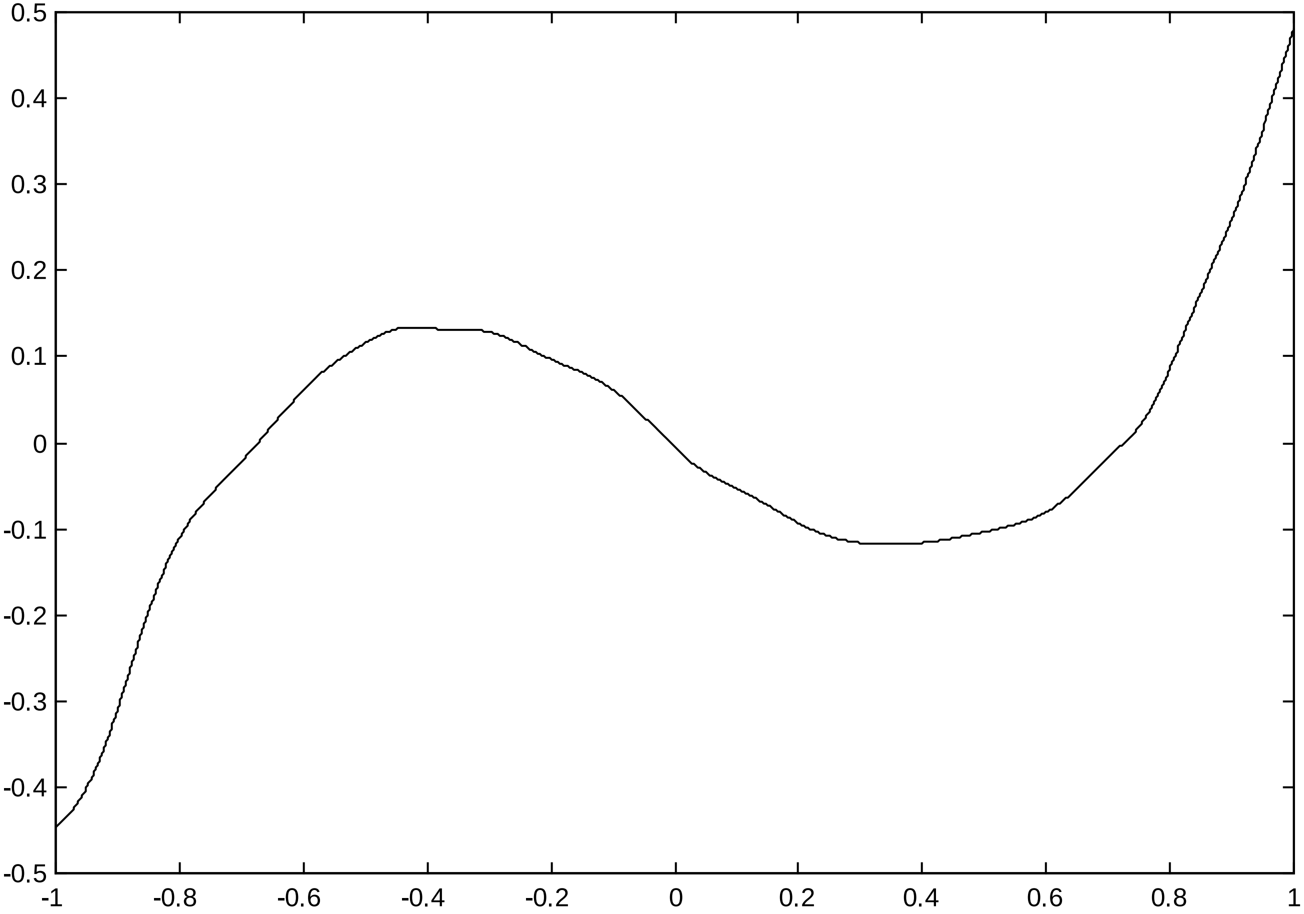
They might be antagonistic objectives


Training set
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

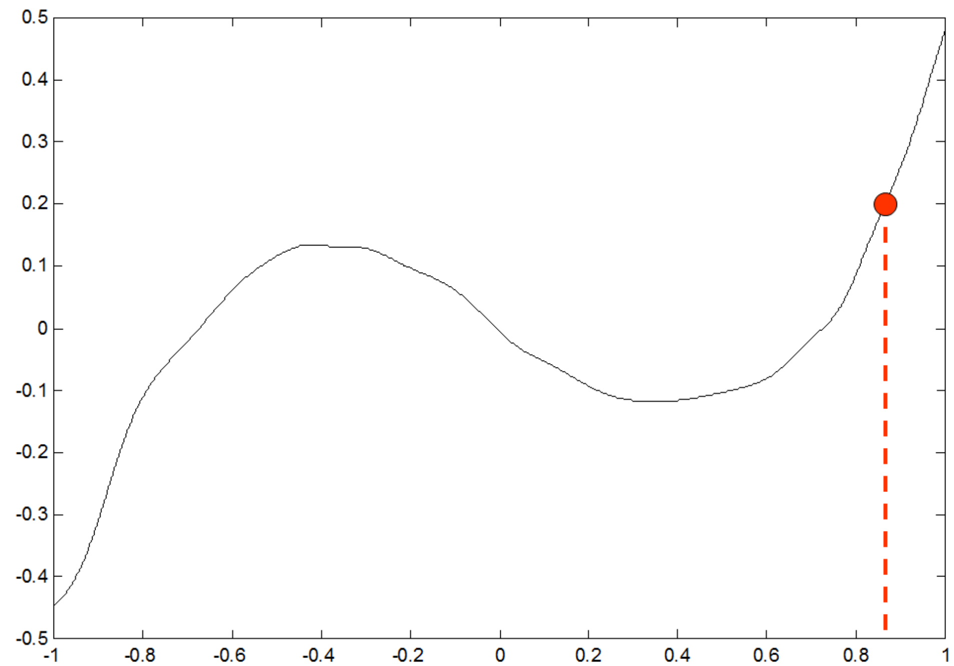


0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1

A compressed model with predictive power



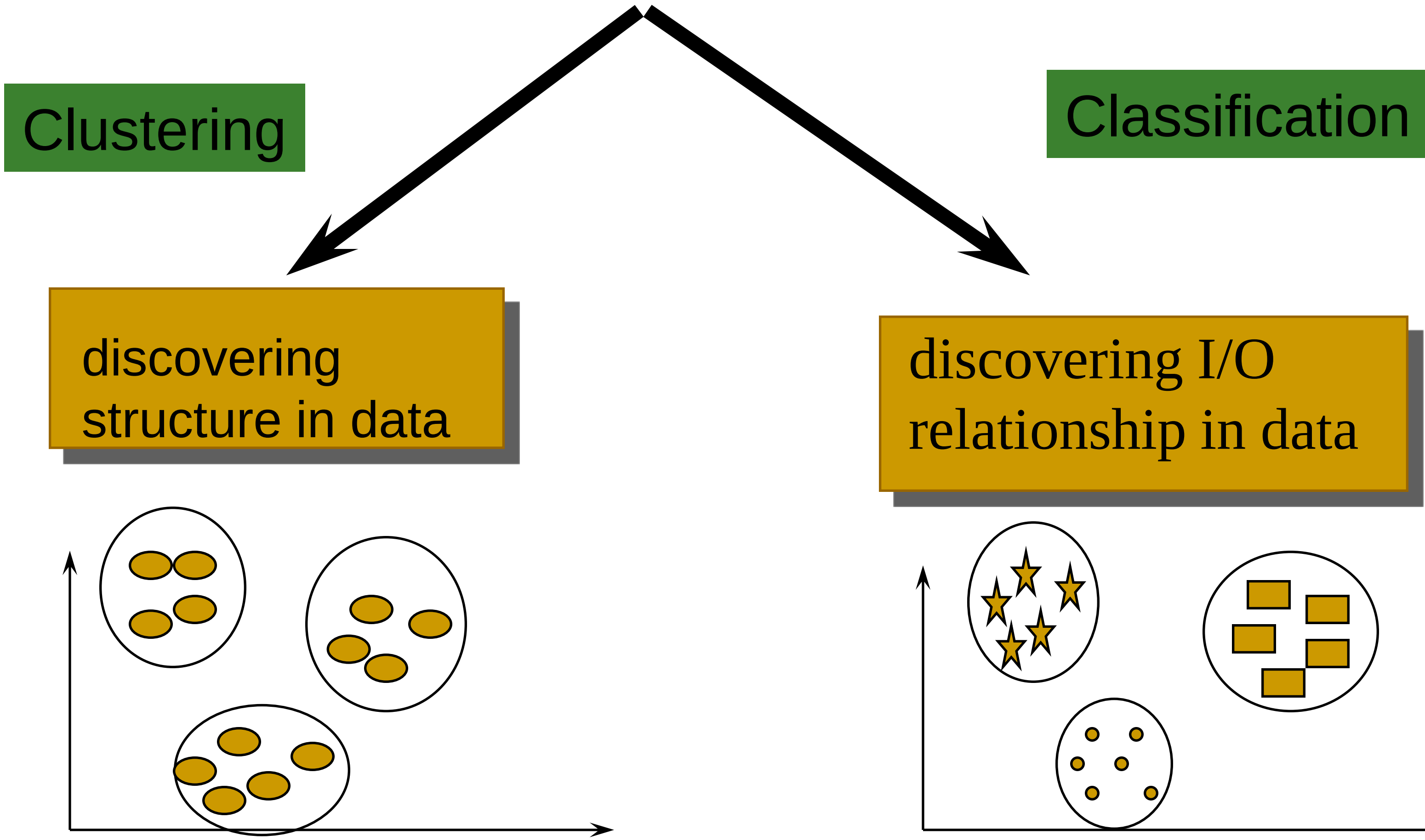
Prediction with global models



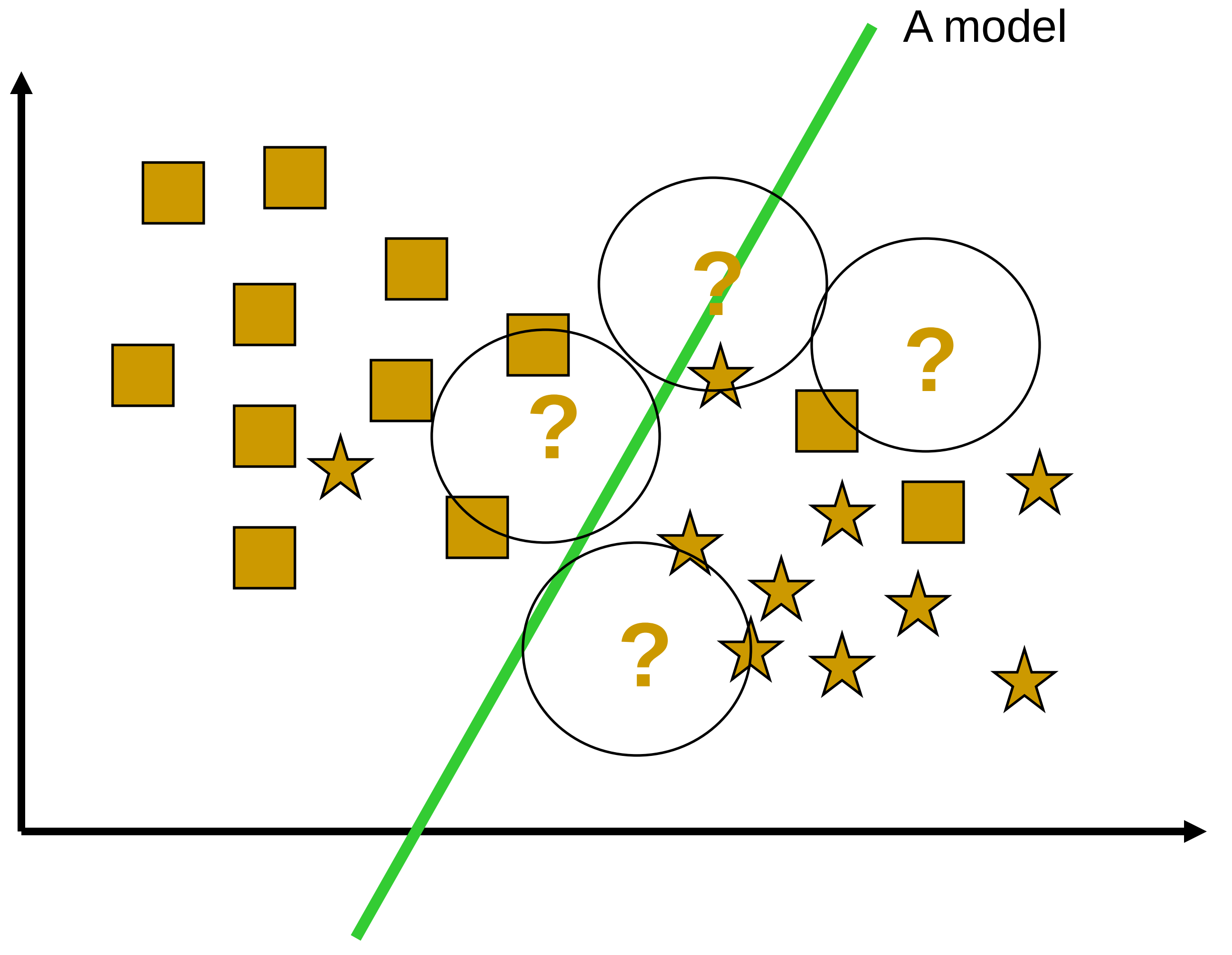
The main techniques of data-mining

- Clustering
- Classification
- Outlier detection
- Association analysis
- Regression
- Forecasting
- Why in business: personalized business, improved prediction, targeted marketing

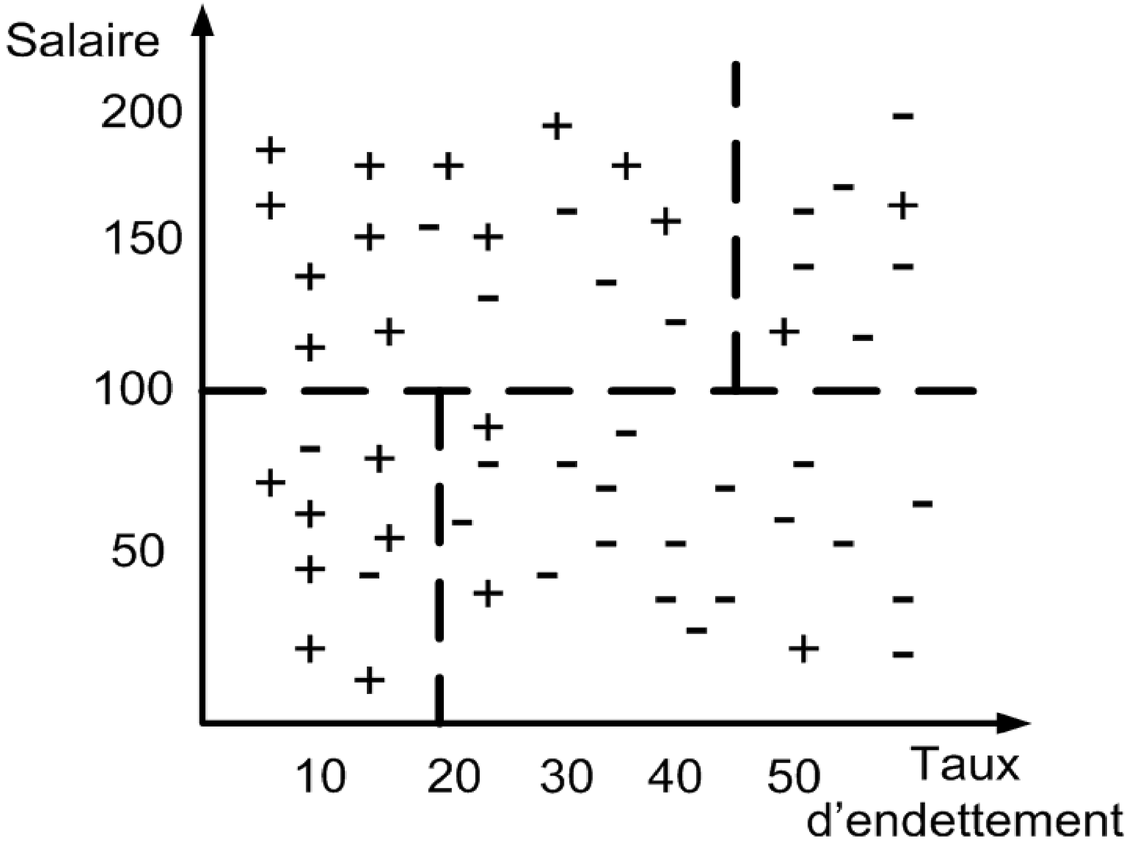
Data Classification: to understand and/or to predict

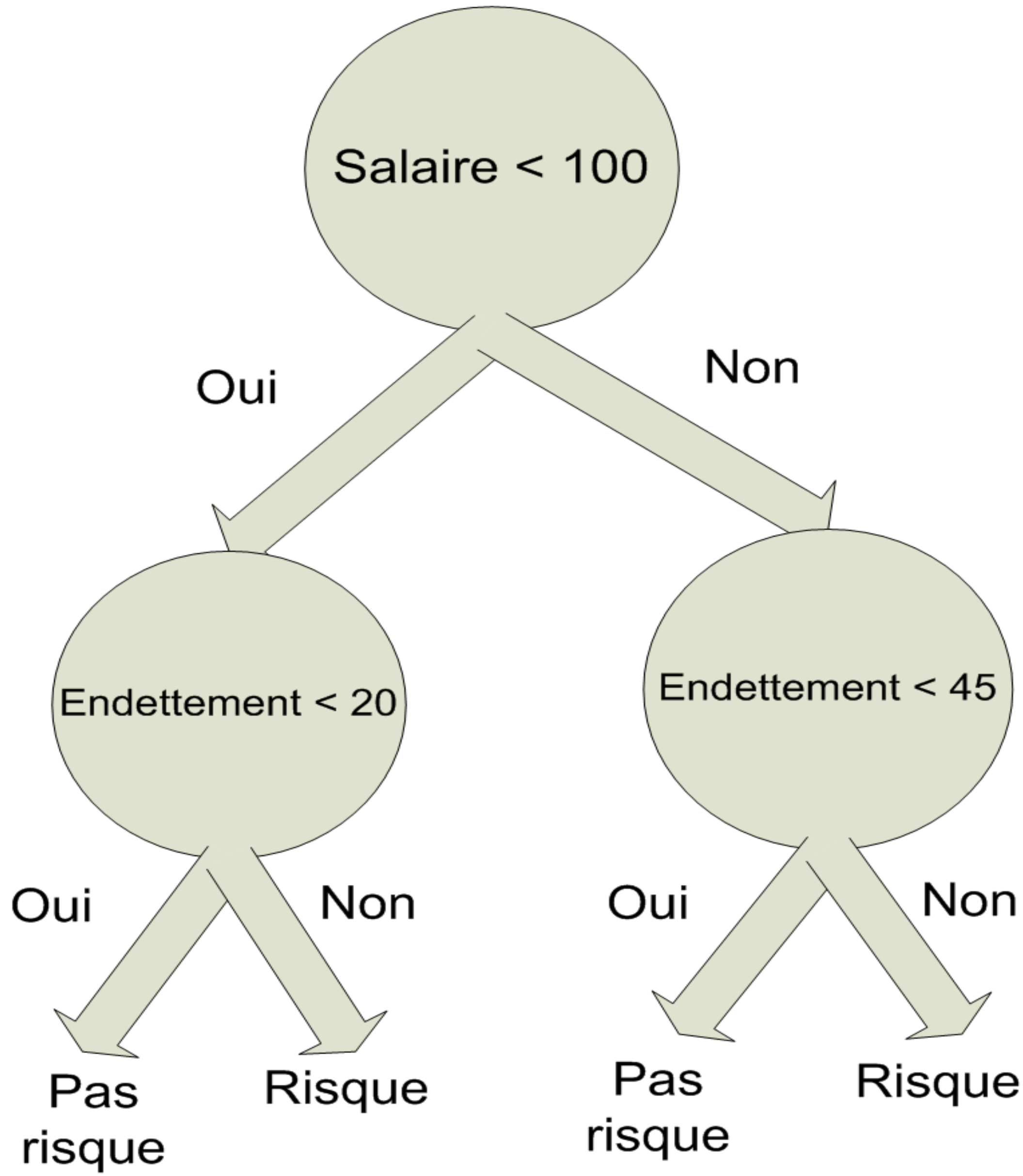


CLASSIFICATION



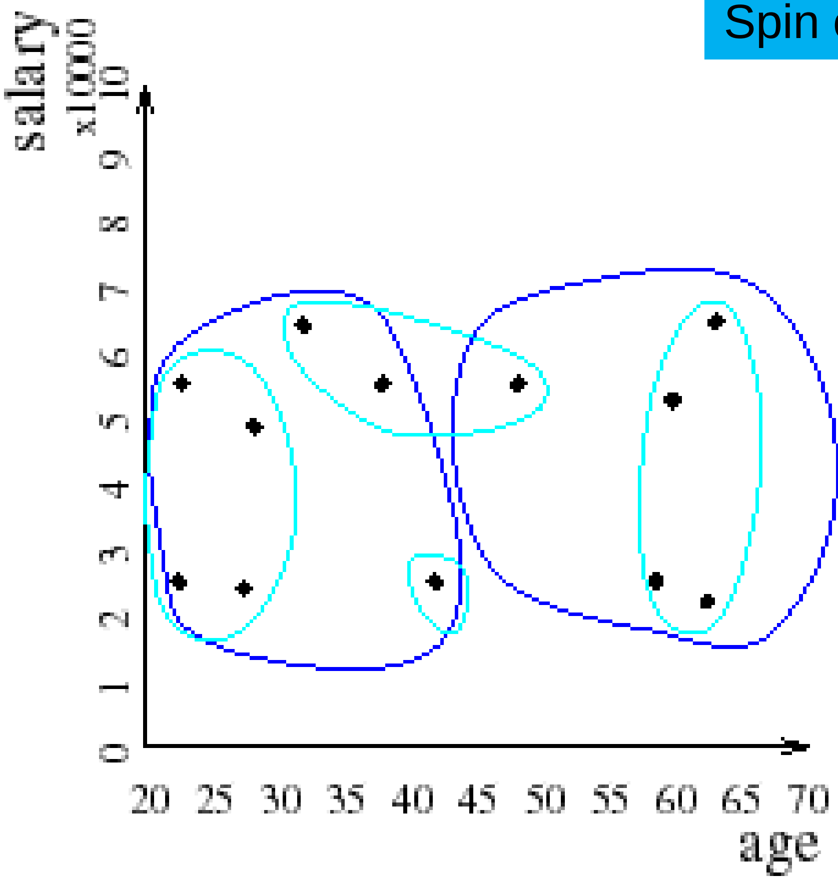
Exemple of classification: Decision tree




Salaire < 100
Oui
Non
Endettement < 20
Endettement < 45
Oui
Non
Oui
Non
Pas risque
Risque
Pas risque
Risque

Clustering and outlier

Spin off : VADIS



Intéressant petit coco

Market Basket Analysis: Association analysis

Quantity bought

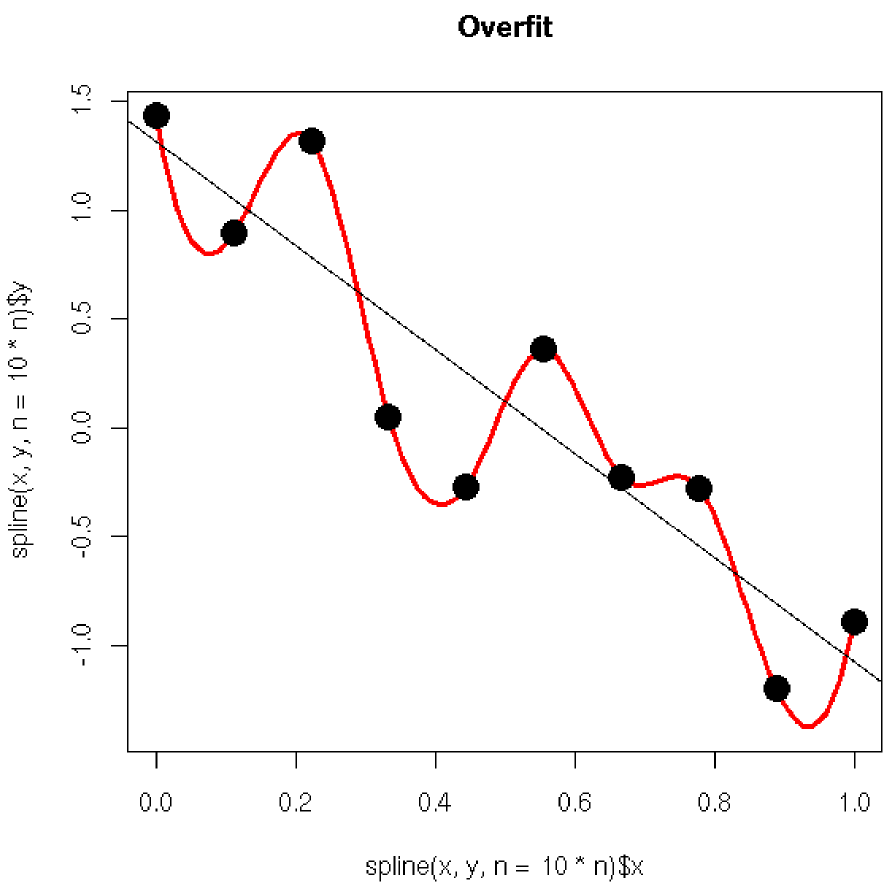
Transn.	Juice	Tea	Coffee	Milk	Sugar	Pop
1	0	0	0	0	0	0
2	0	2	2	4	3	0
3	1	0	0	0	0	0
4	0	1	0	0	0	0
5	1	2	1	1	0	0
6	0	2	1	3	2	0
7	0	0	0	0	0	6
8	0	0	0	0	0	0
9	4	0	0	0	0	0
10	0	0	1	1	0	0
11	0	0	0	0	0	6
12	0	0	1	1	0	0
13	0	0	0	0	0	5
14	0	0	0	0	0	0
15	1	2	0	2	0	0
16	0	1	1	1	2	1
17	1	0	1	0	0	0
18	2	0	0	0	0	0
19	0	0	0	0	0	2
20	3	0	0	0	0	3

Calcul of Improvement

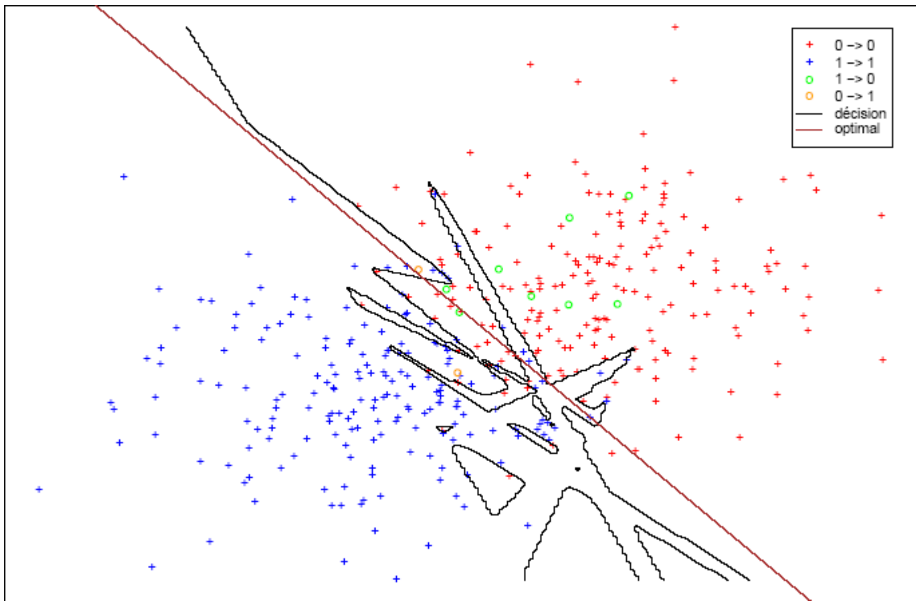
IMPROVEMENT = (N * xij) / (ni * nj)

Improvement	Juice	Tea	Coffee	Milk	Sugar	Pop
Juice	0	0,95	0,82	0,82	0	0,17
Tea	0.95	0	1,9	2.38	3,33	0.56
Coffee	0.82	1,9				
Milk	0,82					
Sugar	0	3,33				
Pop	0,17					

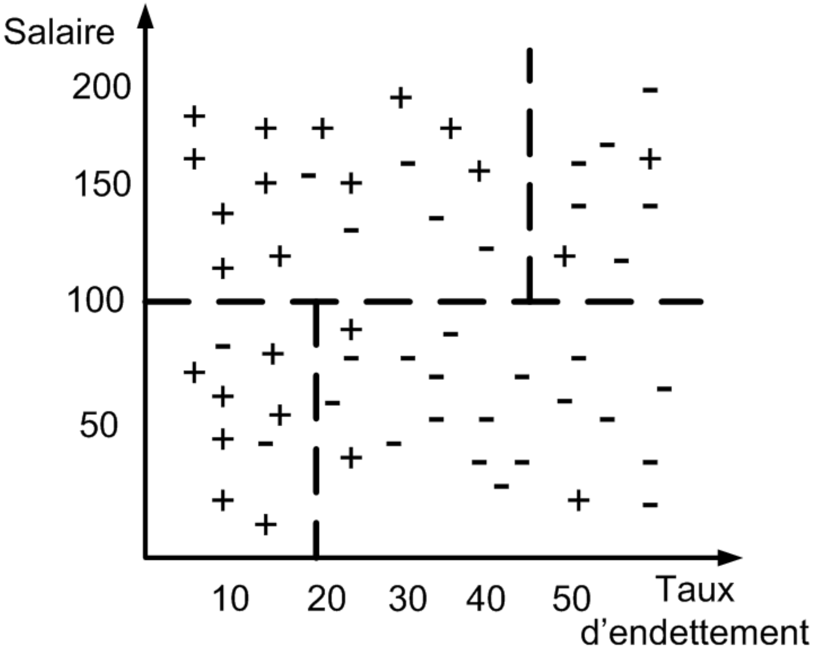
Data Regression and Prediction



Understand or predict



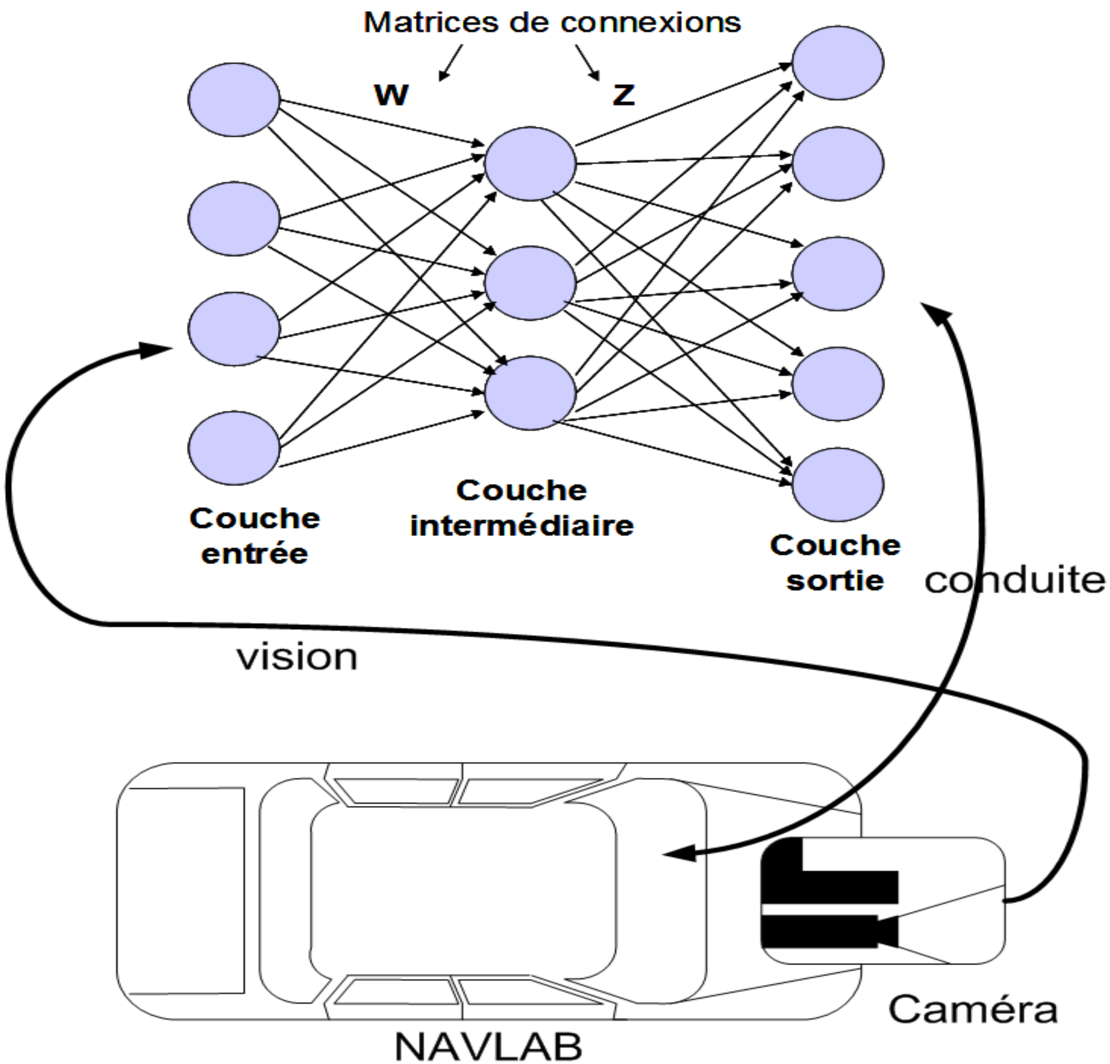
Neural networks



Decision tree

Important emblematic achievements

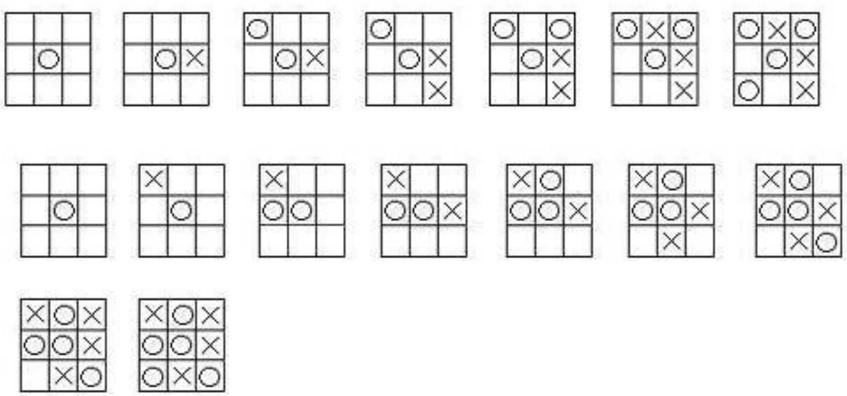
1) A new engineering approach



The Darpa Challenge

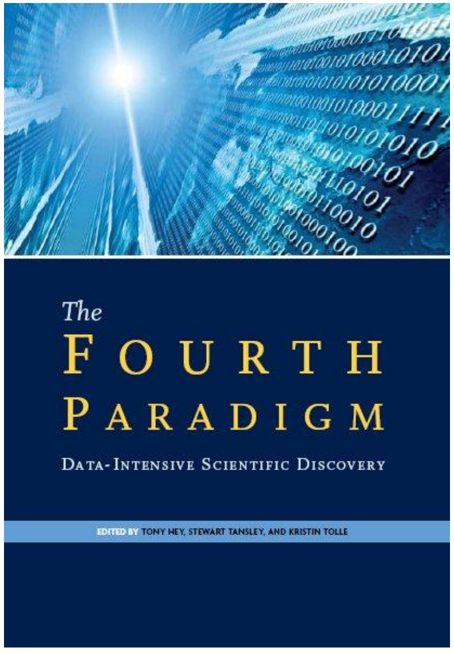
Games

Min-max



Data mining

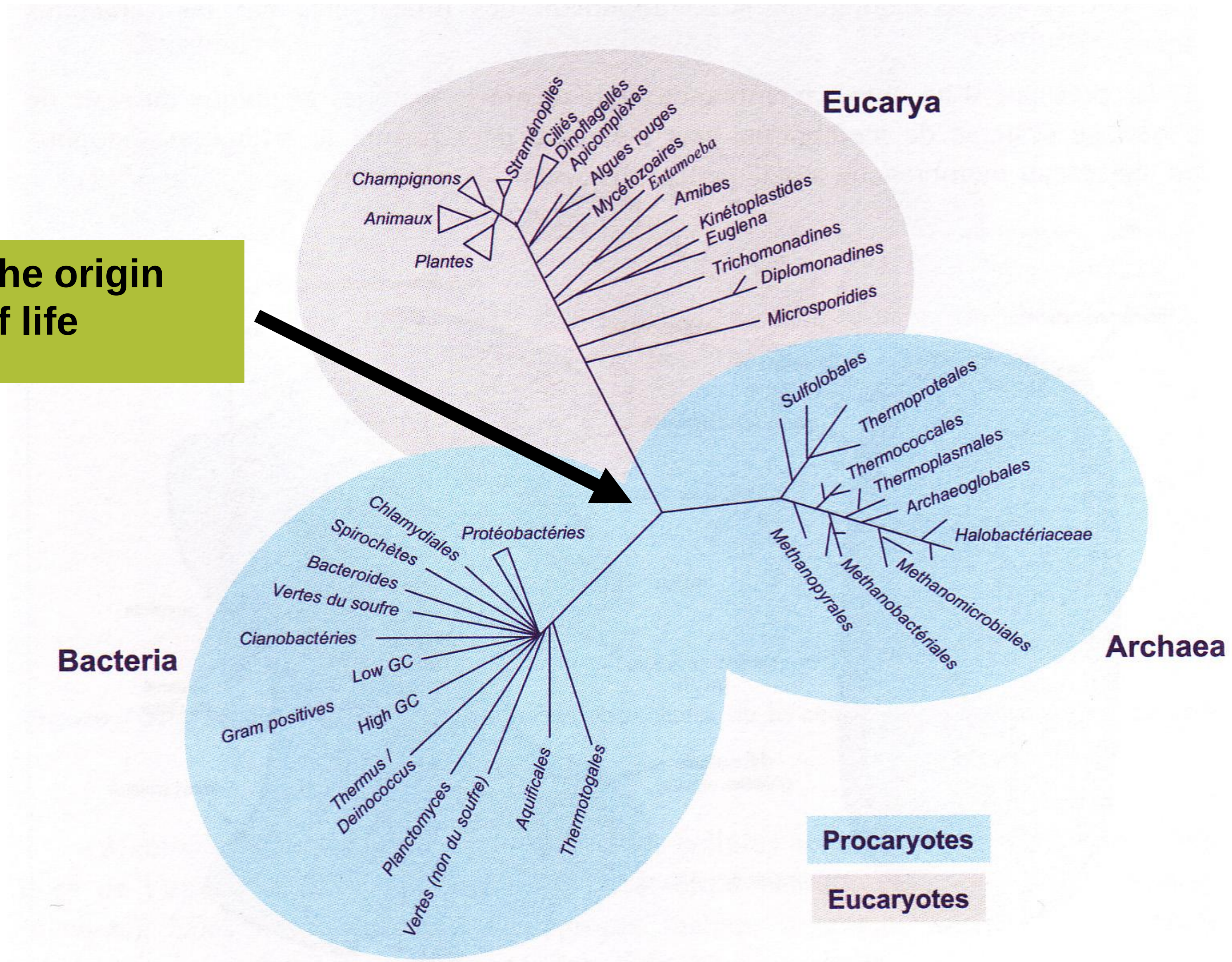
2) A new scientific Paradigm: The fourth : Microsoft



Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies.

.

The origin of life



CLIMATE FORECASTING

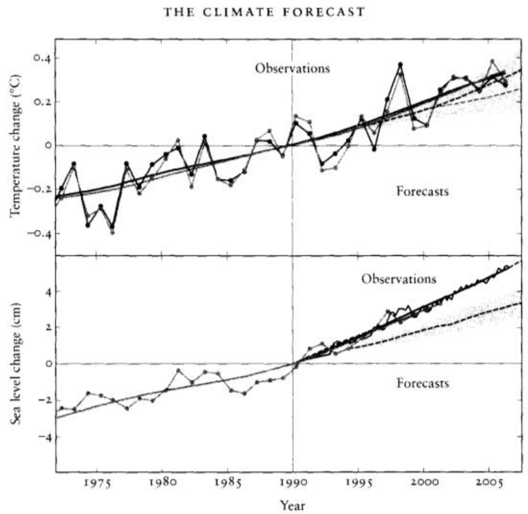


Figure 1. *Upper panel: comparison of observations of global mean temperature (joined points) with model forecasts (grey zone and dotted lines). Lower panel: comparison of observed sea level (joined points) and model forecasts (grey zone and dotted lines). Both panels cover the years 1970 to 2007.*

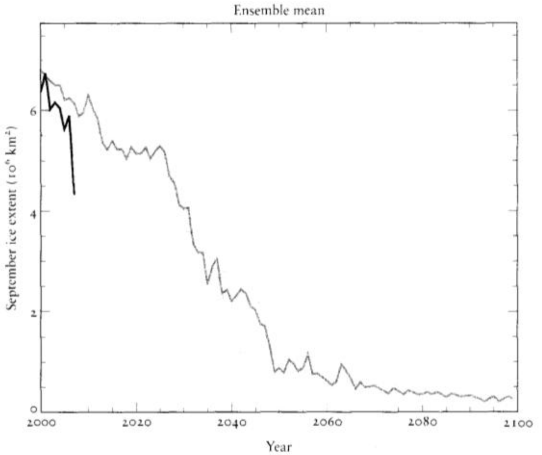
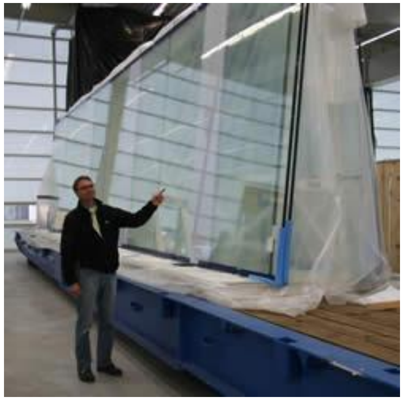


Figure 2. *The IPCC model predictions of the extent of ice covering the summer Arctic Ocean (grey zone with a solid line representing the average in its centre) and the observed ice cover (solid line to the left of the figure).*

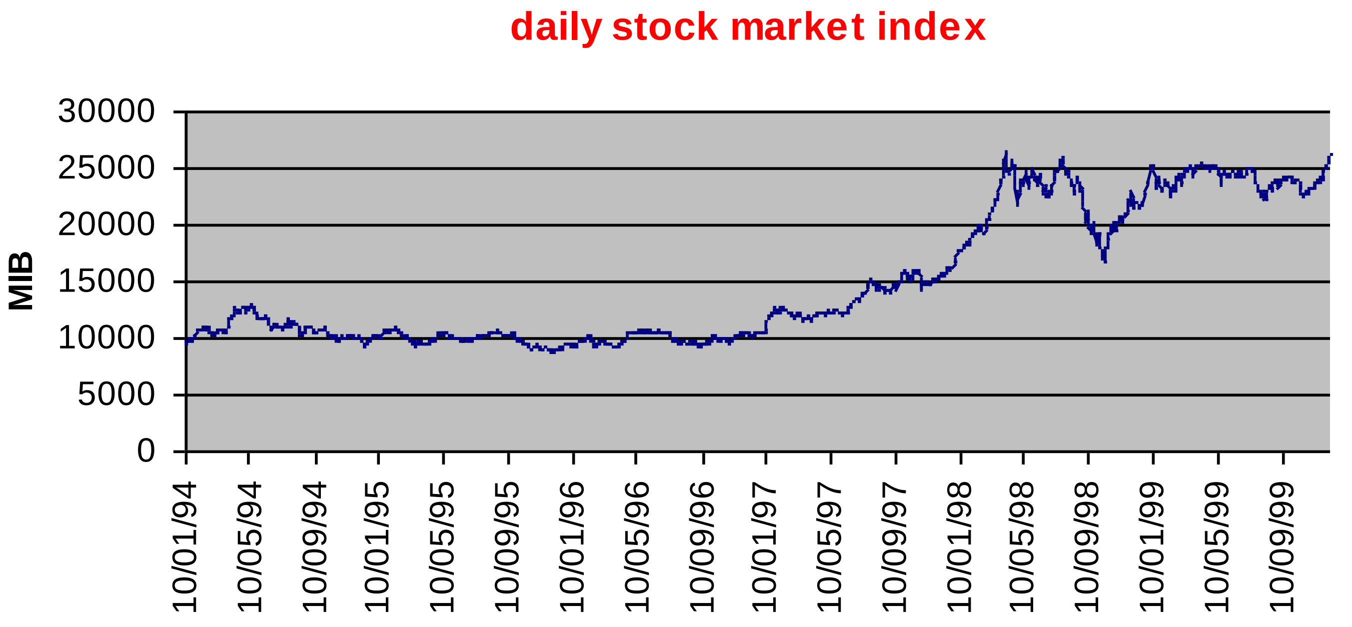
James Lovelock

3) A huge market of business opportunities: IRIDIA's CV

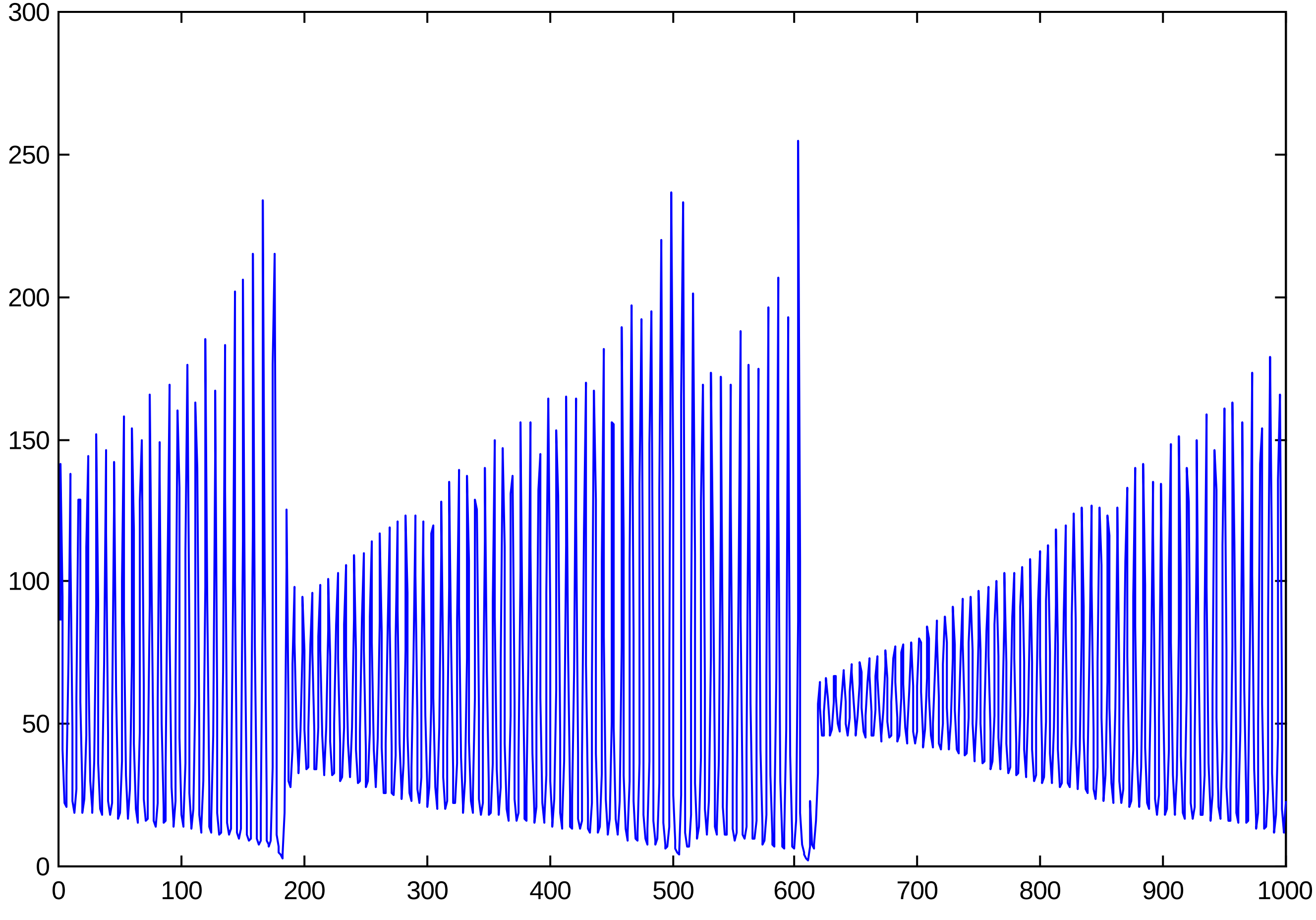
Automatic glass default recognition



Financial prediction

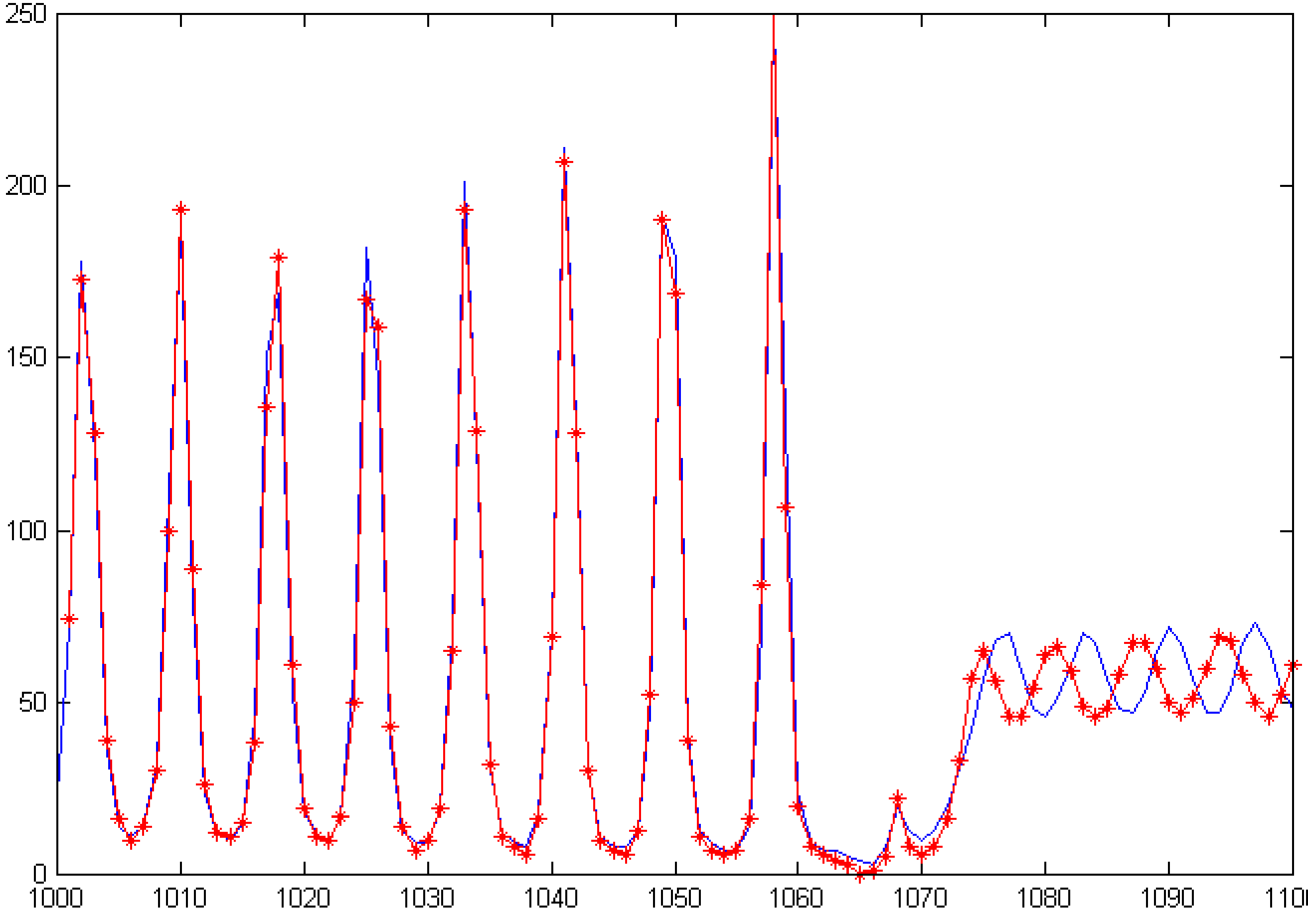


Santa Fe time series



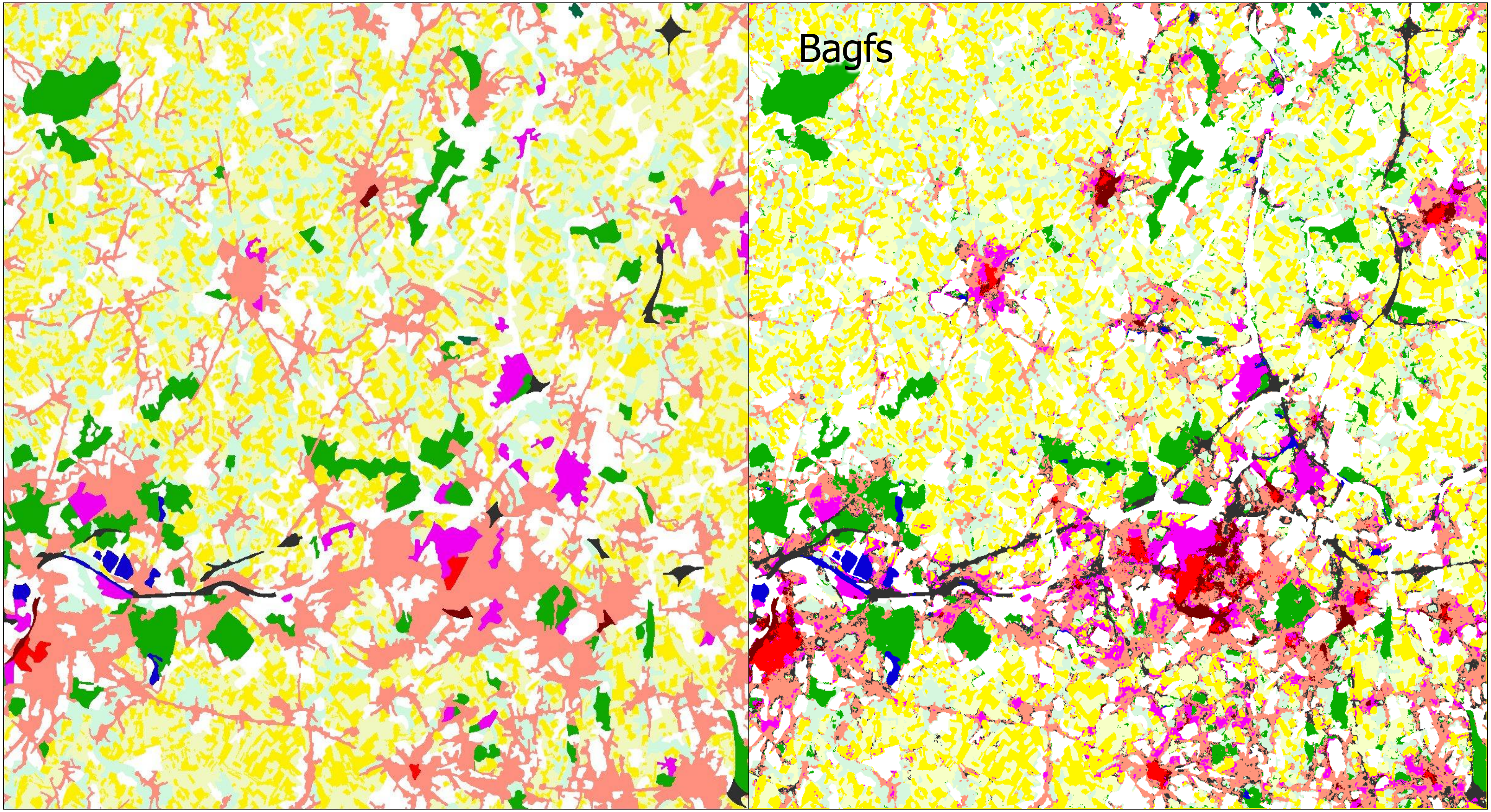
Task: predict the continuation of the series for the next 100 steps.

Lazy Learning prediction

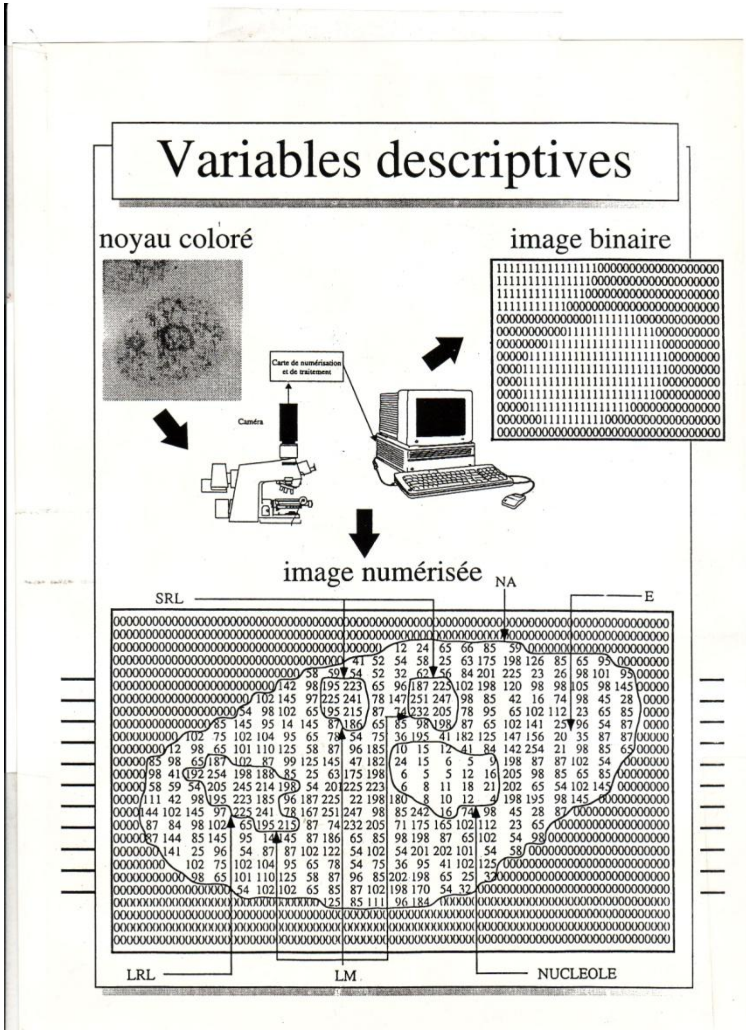


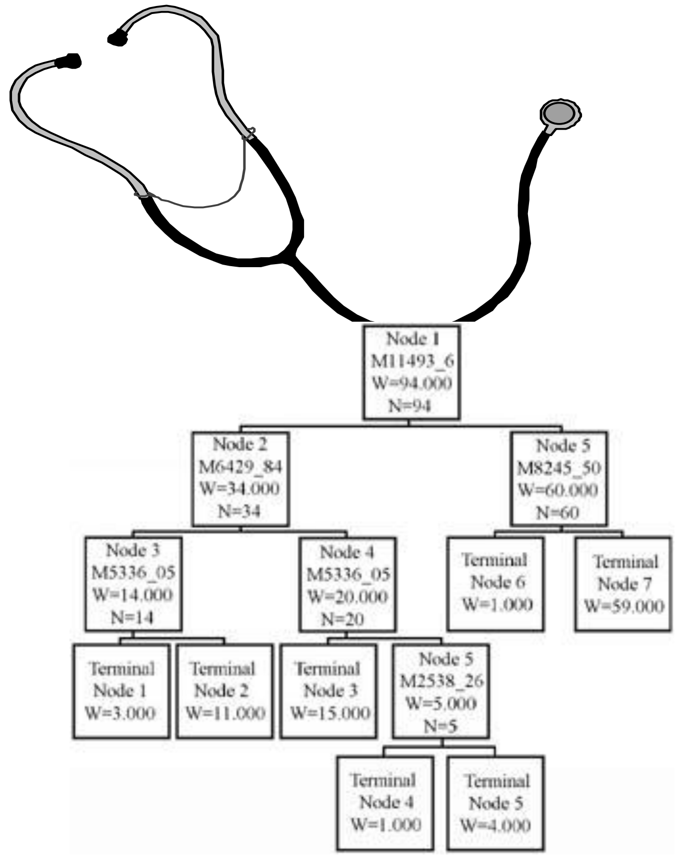
LL is able to predict the abrupt change around t =1060 !

Automatic image labelling

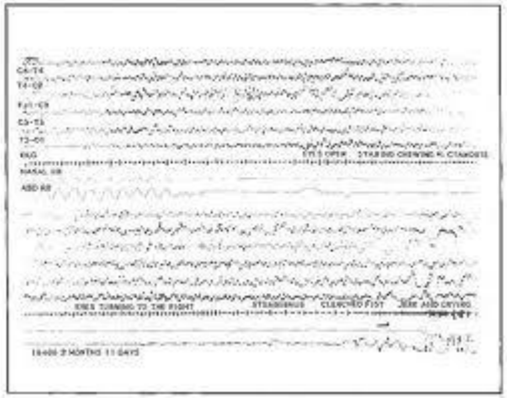


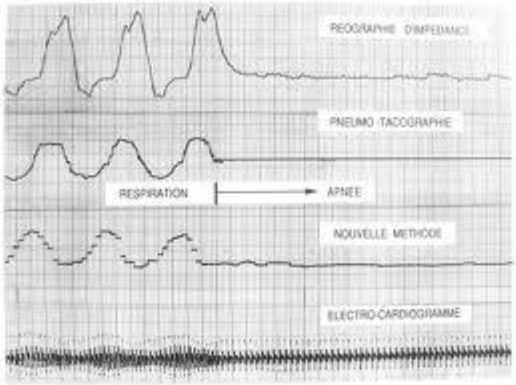
Cancer diagnosis



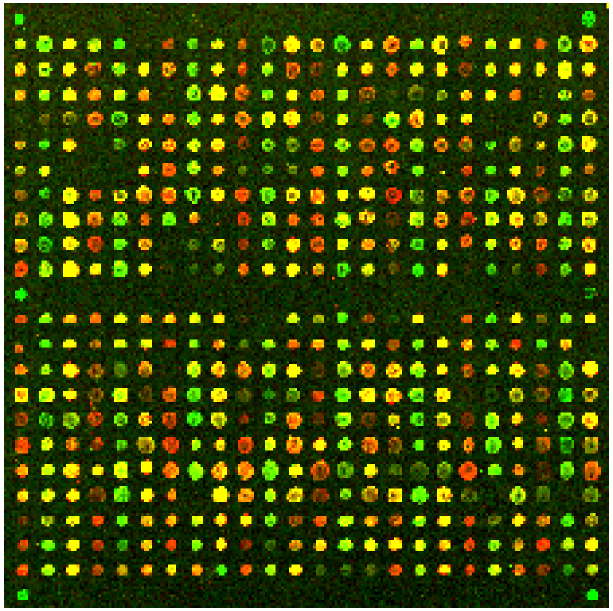


Sudden infant death syndrome

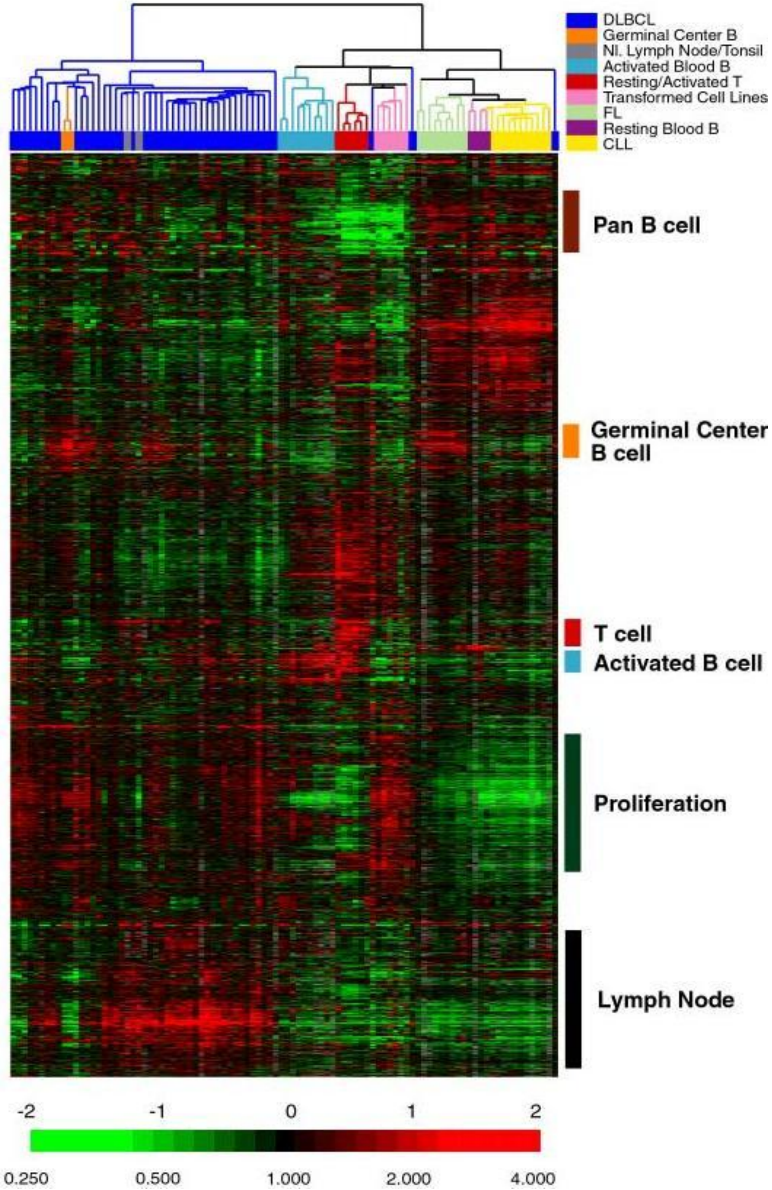




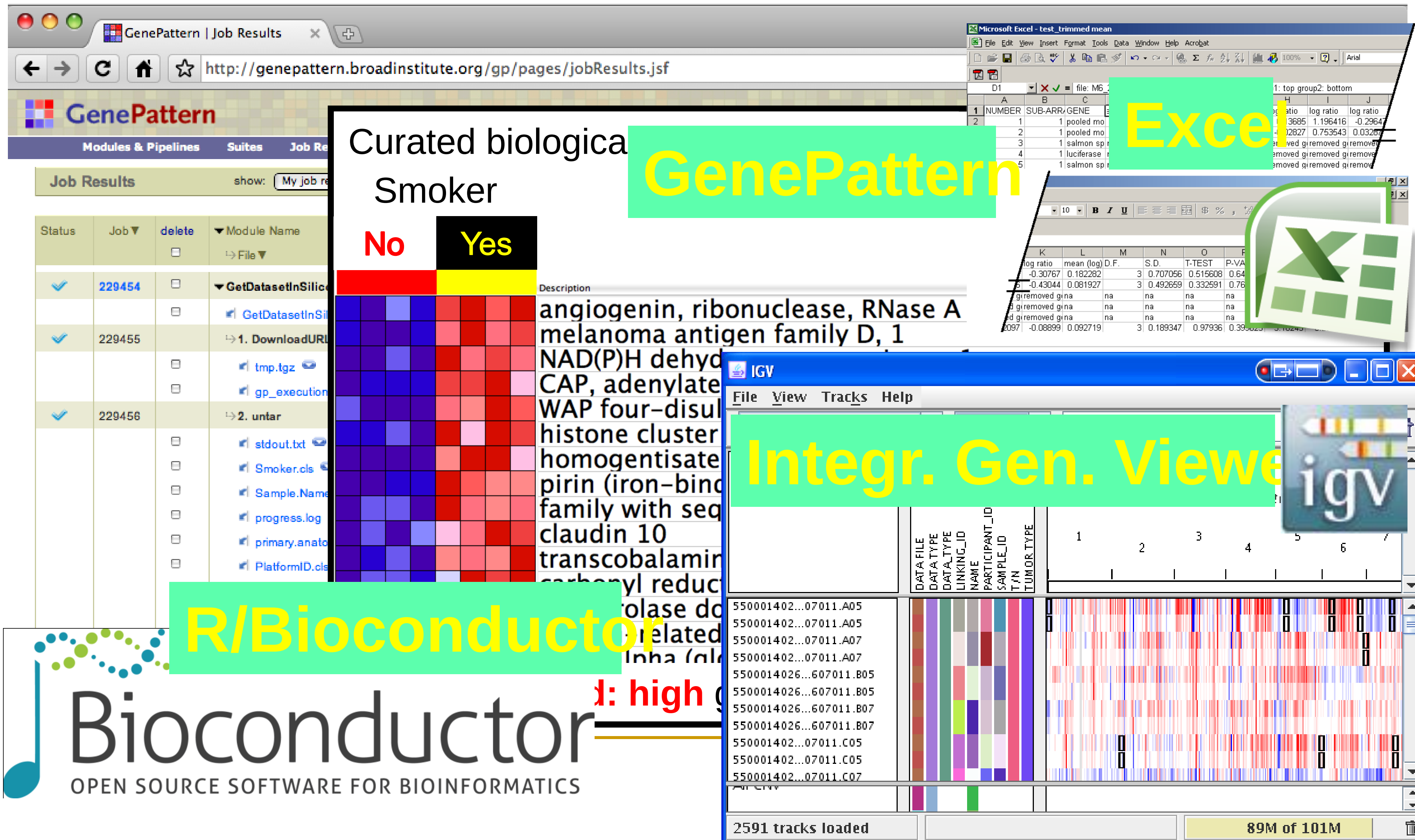
Microarrays



Microarray chip

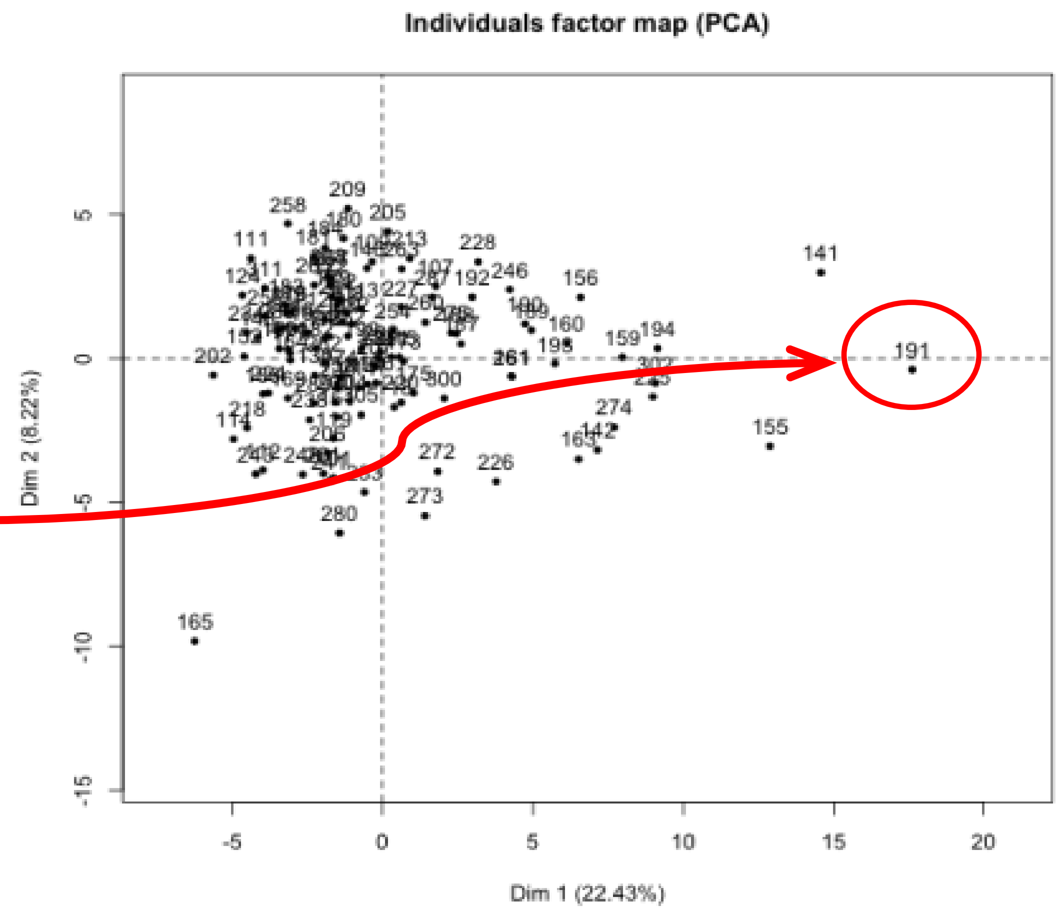


In Silico project: Integration with visualisation and analysis tools



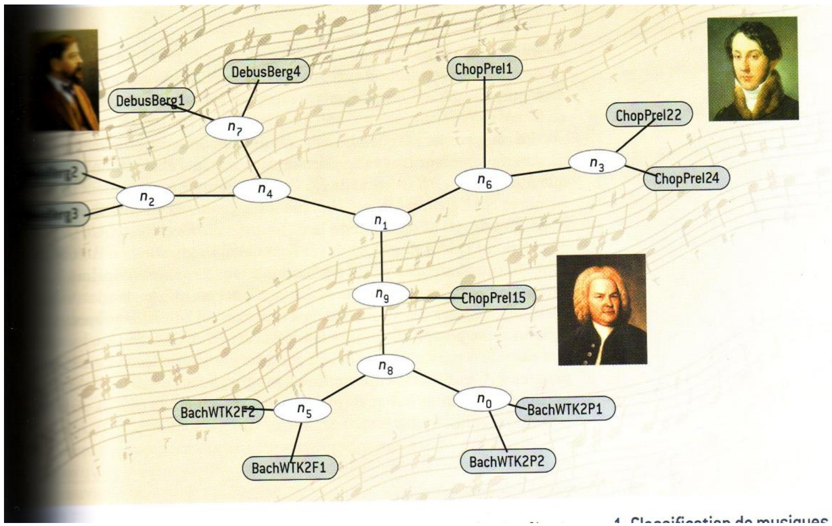
SMART : detection of outlier clinical site

- Real example
  - Known fraud in center 191
- SMART analysis
  - 191 is an outlier
- Other centers?
  - 141, 155, 165?
  - Most frauds are undetected by current methods



Summary through PCA of a SMART analysis

The future of it: More and more free documents with various contents and various own structuration



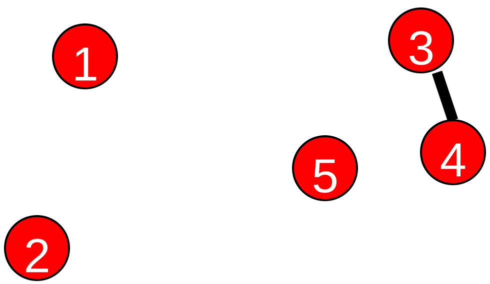
Art Mining:

- images
- musics
- movies

Exemple of clustering: hierarchical clustering

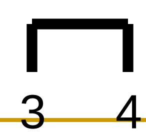
Algoritm

- *Join the two closest elements.*
- Update the distance matrix.



Closest : 3 et 4

	1	**2**	**3**	**4**	**5**
1	0	10	15	18	12
2		0	23	22	13
3			0	4	6
4				0	5

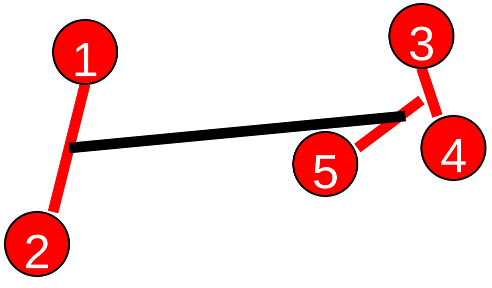


Distance matrix

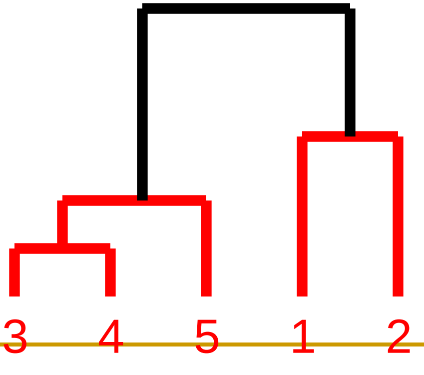
Hierarchical clustering

Algoritm

- *Join the two closest elements.*
- Update the distance matrix.



Closest : (1,2) et (3,4,5)

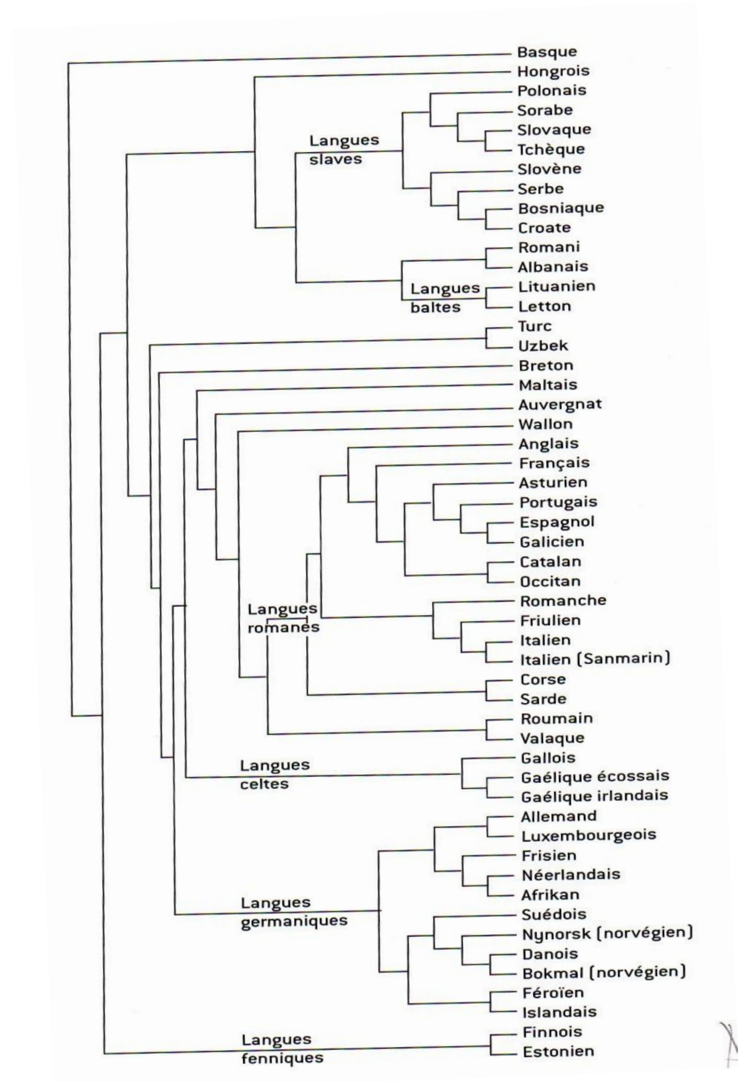


Similarity based on compression algorithm

- Suppose two documents A and B
- Compute length of compressing A: C(A)
- Compute length of compressing B: C(B)
- Compute length of compressing AB: C(AB)
- Similarity (A,B) = 1-[C(A)+C(B)-C(AB)]/C(A)

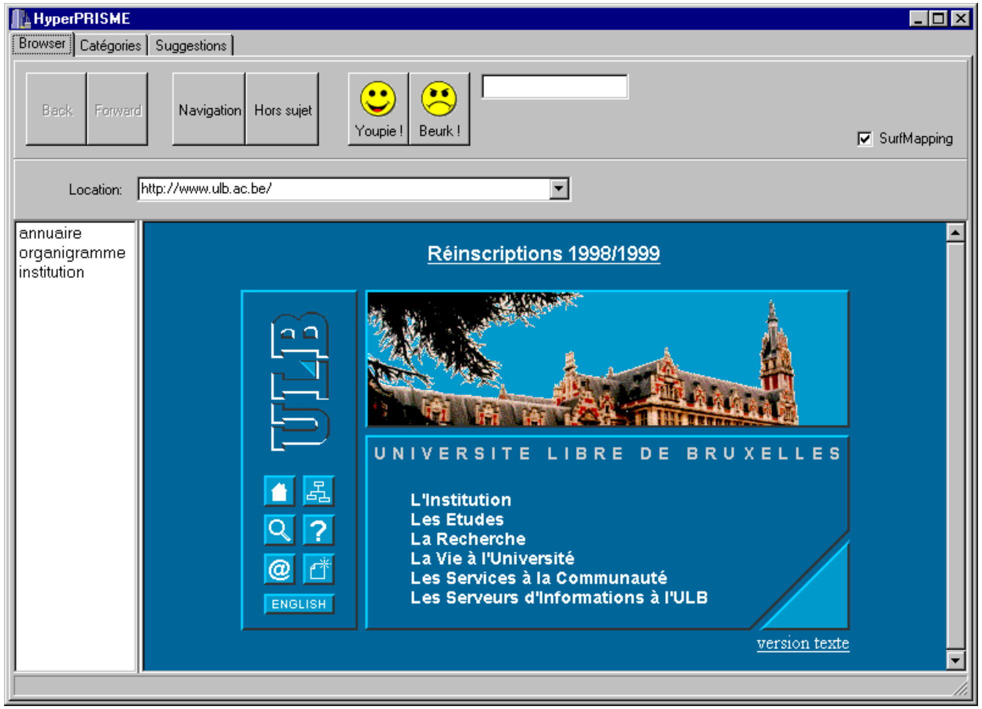
  if C(A) >= C(B)

Simalirity between natural languages



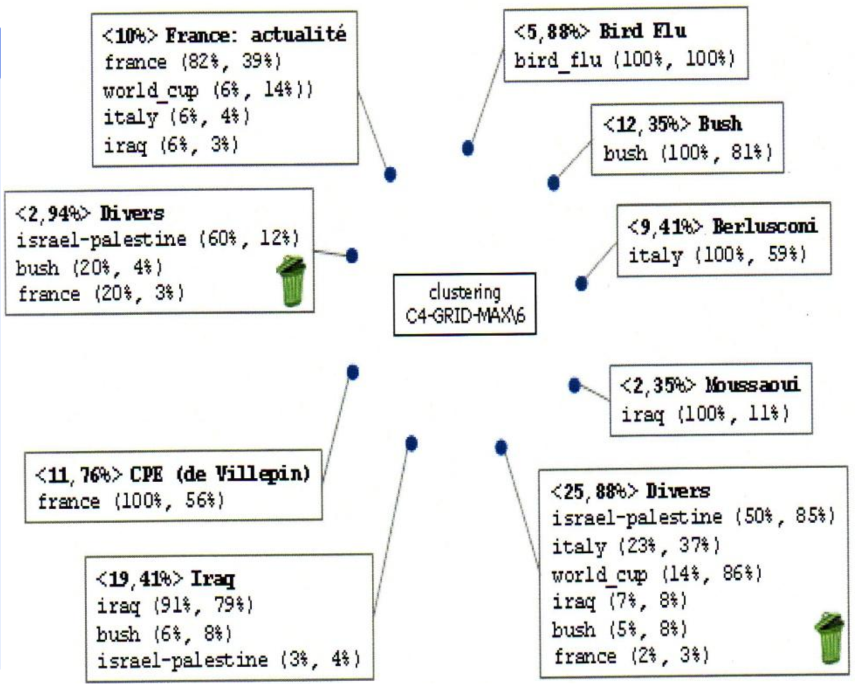
Web Mining

- The Hyperprisme project
- Spy the user and mine his clickstream
- Automatic profiling of users
    - Key words: positif, negatif,…
- Automatic grouping of users on the basis of their profiles



Text Mining: still a lot of possible improvements

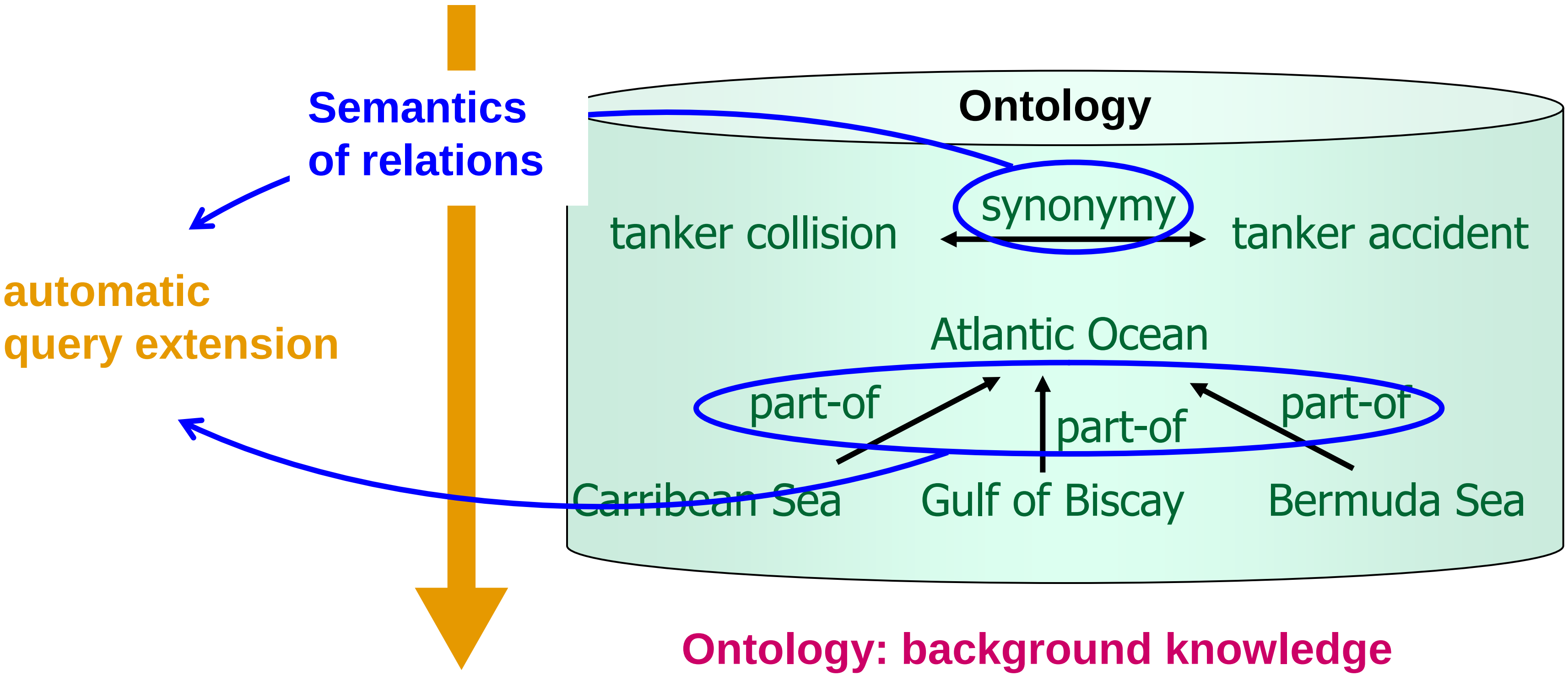
Table III Term-document matrix						
Term	*Doc1*	*Doc2*	*Doc3*	*Doc4*	*Doc5*	*Doc6*
Passenger traffic volume	1	1	0	5	2	0
Decrease	1	2	1	0	0	0
Increase	0	2	0	0	0	0
Passengers carried	5	1	0	0	0	0
Personal traffic tools	1	0	0	0	0	0
Grow up	4	1	6	0	0	0
Million	4	1	0	0	0	0
Hundred	0	0	0	0	1	0
FAST rapid transit system	0	2	0	0	0	0
Finished	0	1	0	0	0	0
A1 station	0	0	0	5	4	4
B1 station	0	0	0	1	5	0
C1 station	0	0	0	1	0	0
D1 station	0	0	0	1	0	1
E1 station	0	0	0	1	0	2
Passenger-Kilometers	0	1	7	0	0	0
Columniation	0	0	0	0	2	0
Check the number	0	0	0	0	2	0
Ticket Revenues	0	0	0	0	0	7



Semantic enrichment

Using background knowledge to extend query

"tanker accident" atlantic | Search



Ontology: background knowledge

(tanker collision OR tanker accident) AND
(Atlantic Ocean OR Carribean Sea OR Bermuda Sea OR ...)

Exploit the structure of the documents

Like for XML for instance

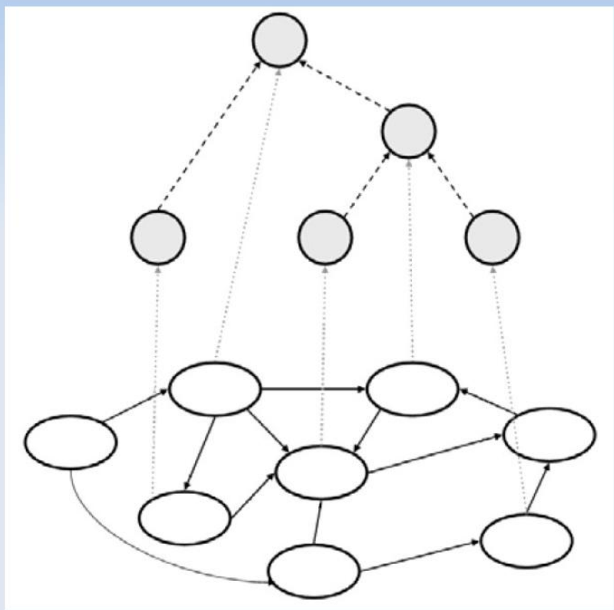
```
<Course>
     <title> Software technologies </title>
     <teacher> Bersini </teacher>
    <themes>
               <name> programming technique </name>
               <name> data representation </name>
               <name> data mining </name>
    </themes>
 </Course>
```

Exploit the graph structure of XML + the content between the tags

We are working on Wikipedia

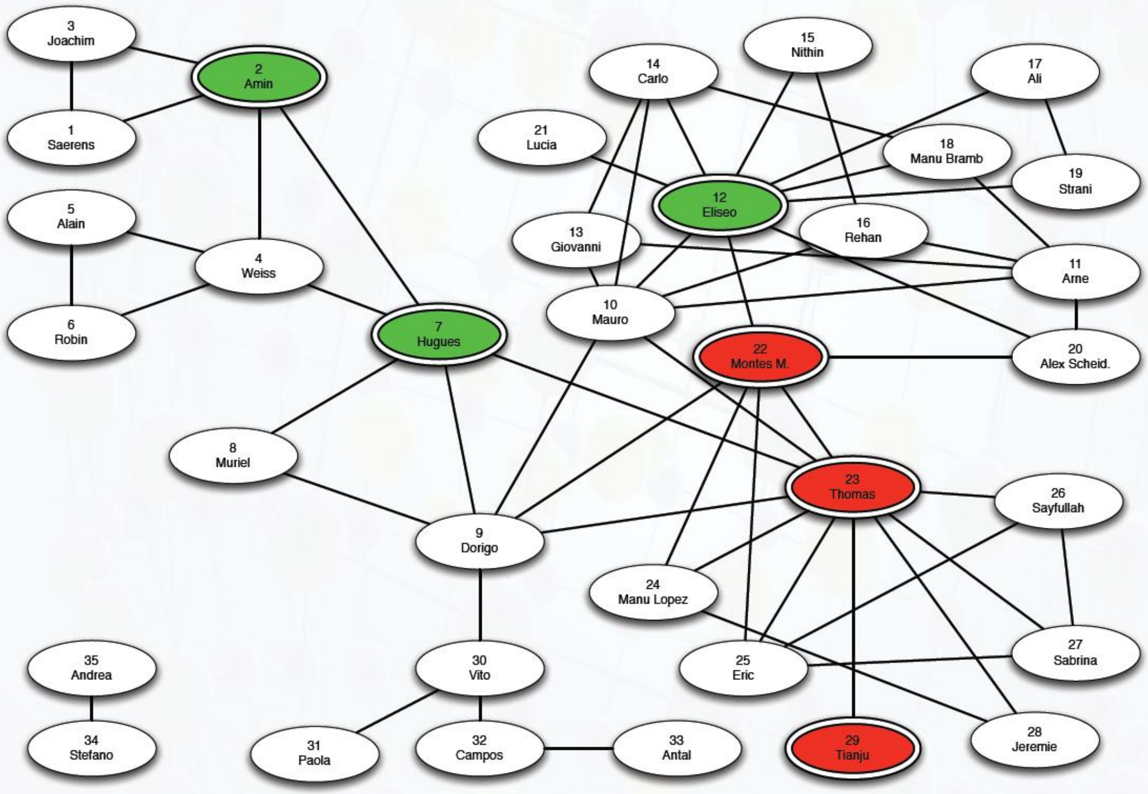
The Nature of Wikipedia

- Wikipedia is a combination of two interconnected graphs
  - A directed graph with the regular pages as nodes and the links between pages as edges
  - An acyclic directed graph with the category pages as nodes and their connections as edges
- The main regular page graph consists of ~ 3 650 000 nodes and the category graph of ~ 700 000 nodes (last count)



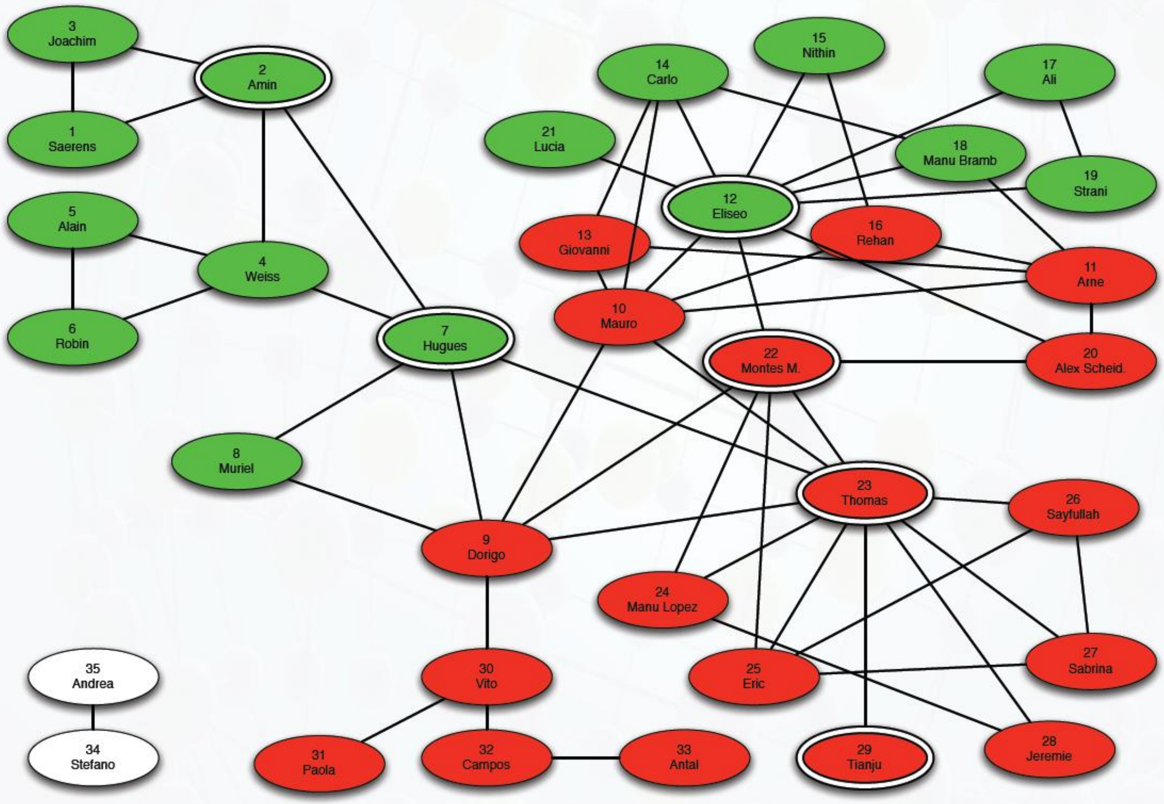
Graph Mining

Application to Classification



Let us classify all the nodes.

Application to Classification



Let us classify all the nodes.

Combine different types of information: graph and text

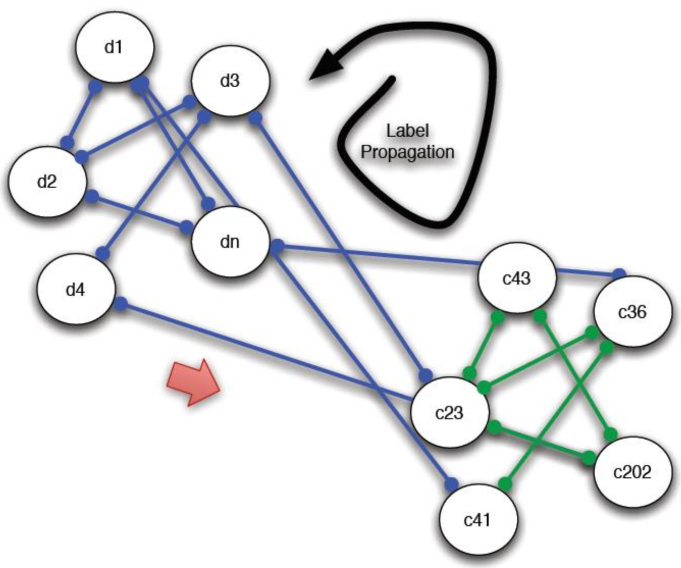
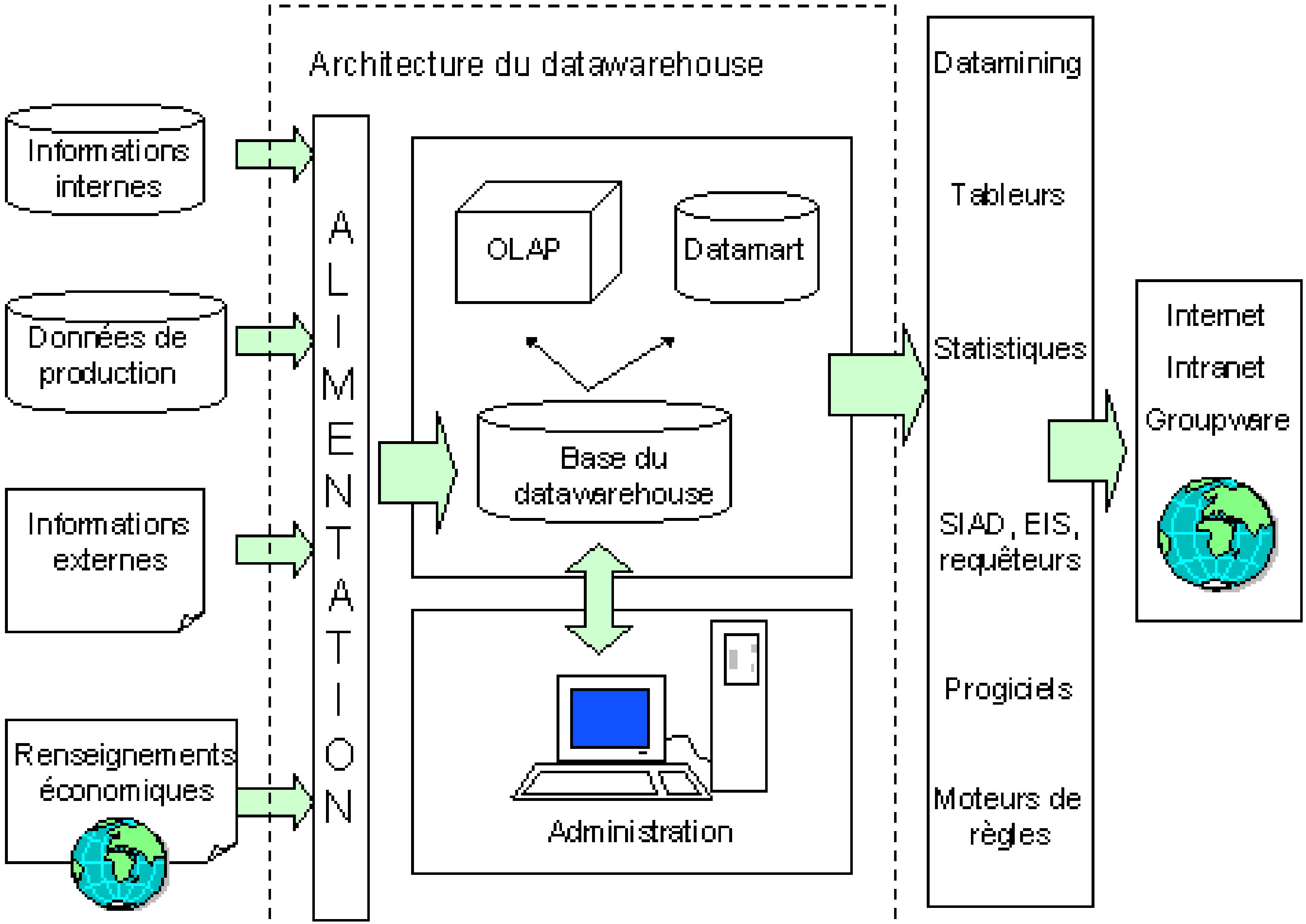
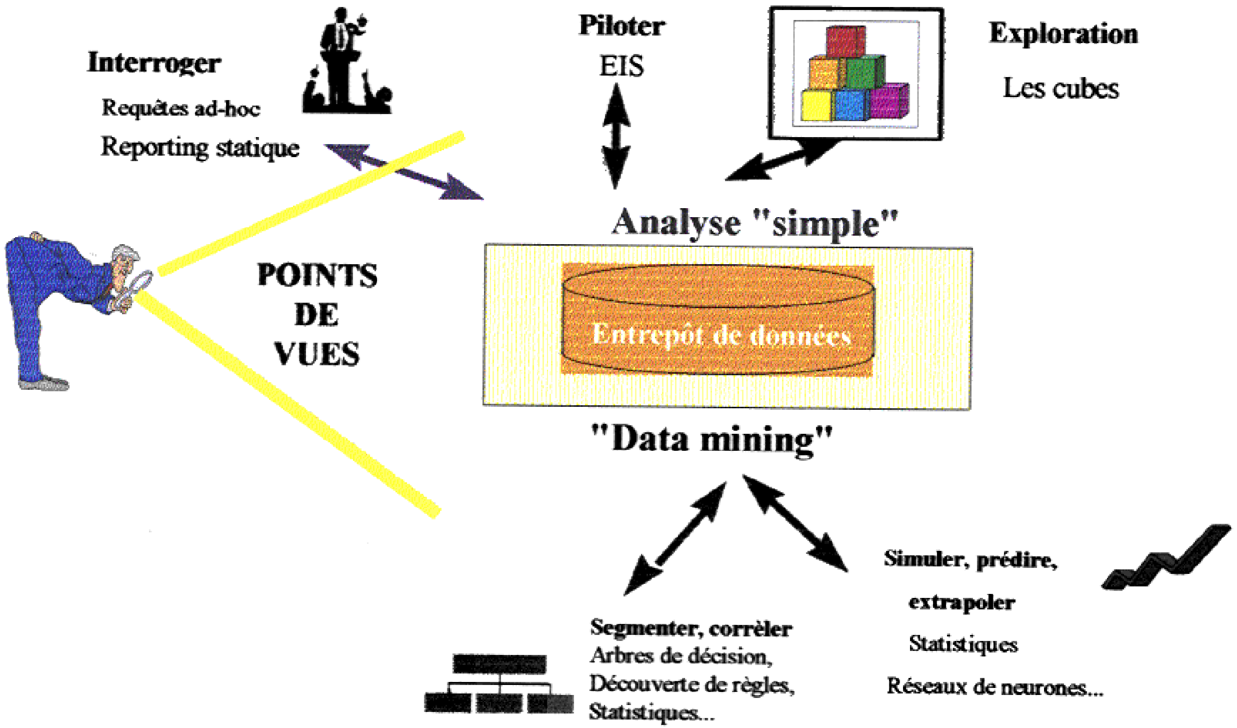


Figure 10: **The document nodes has been connected to an external preexisting citation network through inferred k nearest neighbors links (i.e. in blue). The goal is to propagate labels from the citation graph to the just connected documents.**

Data Warehousing


Architecture du datawarehouse
Informations internes
Données de production
Informations externes
Renseignements économiques
ALIMENTATION
OLAP
Datamart
Base du datawarehouse
Administration
Datamining
Tableurs
Statistiques
SIAD, EIS, requêteurs
Progiciels
Moteurs de règles
Internet
Intranet
Groupware

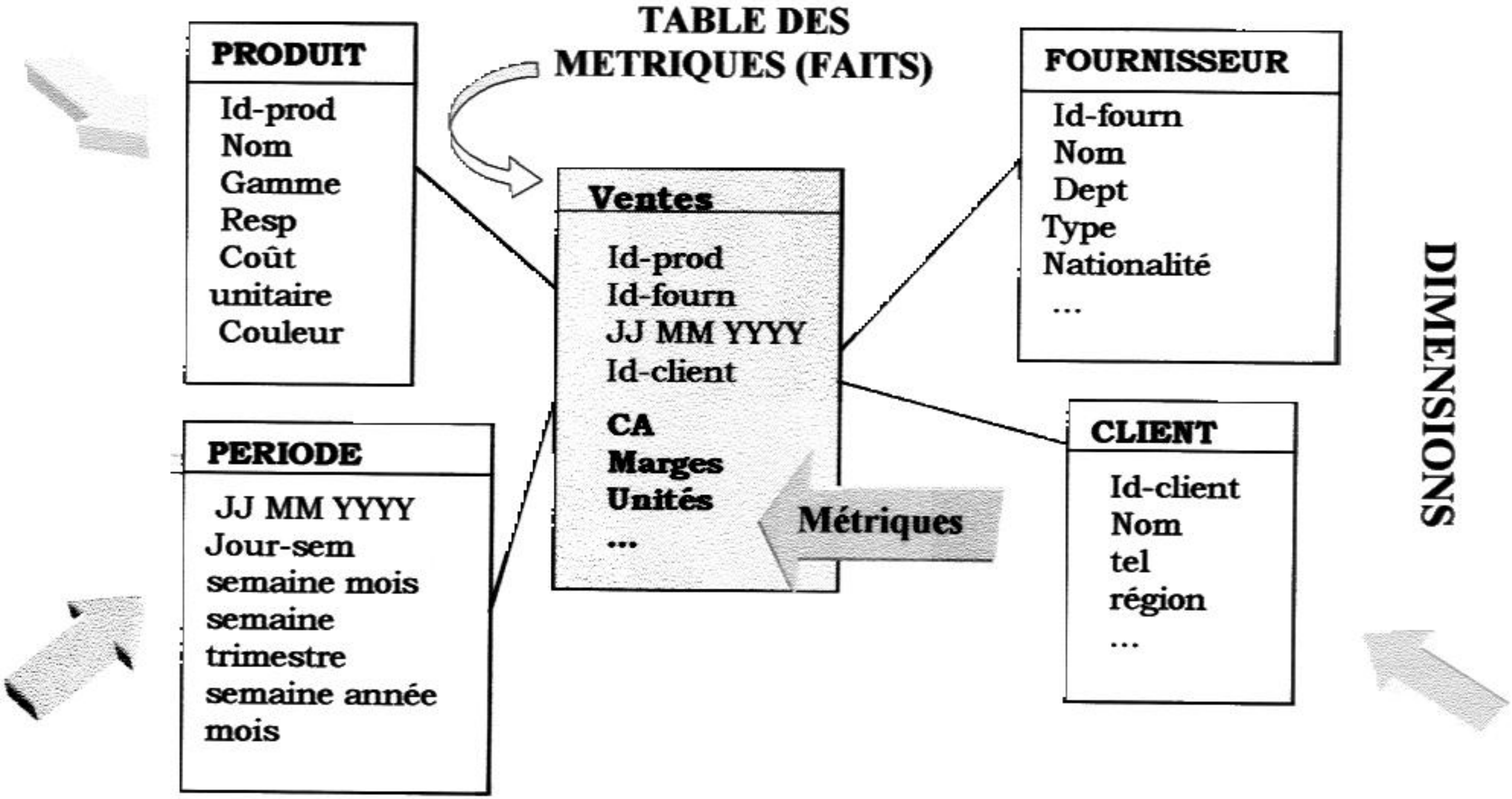


Source : *Le Data Warehouse –Le Data Mining*, Eyrolles, Paris, p. 40

Réorganisation des données

- Orientées sujet
- intégrées
- transversales
- historisées
- non volatiles
- Des données productions ---> données décision

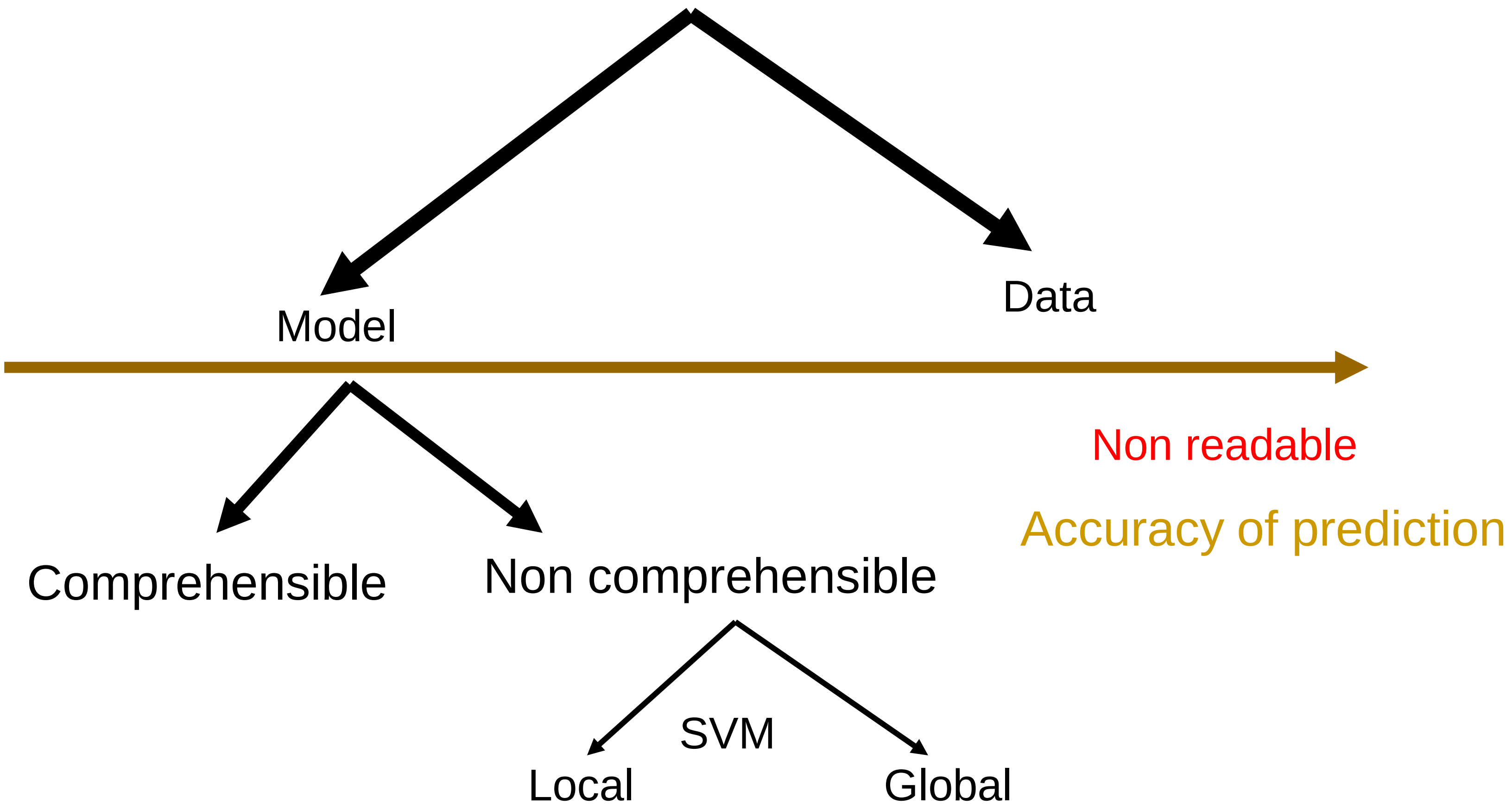
	Environnement transactionnel	Data Warehouse
Type d'utilisateurs	Font tourner les roues de l'entreprise	Vérifient si les roues de l'entreprise tournent bien
Définition de système performant	Système performant = système rapide	Notion de performance est liée au degré de prévisabilité d'une requête
Volumes manipulés	Faible	Elevé
Type d'accès	Lecture/écriture : la donnée est modifiée en ligne	Chargement par batch, mises à jour interdites car les données sont des clichés issus des systèmes de production
Types de données stockées	Dynamiques : mises à jour fréquente	Statique, évolution par chargement
Gestion des redondances	Est évitée car elle pose des problèmes d'incohérence de données	Redondance peut être nécessaire pour optimiser les performances ➔ pas de problème de cohérence car la donnée de base est déjà une copie
Domaine couvert	Modèle le plus souvent propre à une application	Rôle transversal dans l'entreprise et organisé par sujet
Mode d'accès et conséquence sur le modèle de données	Par l'intermédiaire d'application ; le modèle de données n'est visible que par l'utilisateur qui ne voit le système qu'au travers des applications qu'il utilise ➔ le modèle de données peut être complexe	Directe ou légèrement masquée par un outil d'aide à la décision➔ le modèle doit être simple
Type de requête	Simples car prévisibles ➔ le modèle de données est conçu pour éviter les requêtes trop complexes. La plupart des requêtes s'appuient sur un index, d'où des temps de réponses proportionnels au volume stocké. Les performances sont stables car toutes les requêtes sont prédéfinies	Complexe, surtout si l'utilisateur est autonome. Il est quasiment impossible de garantir que tous les accès passeront par les index : le temps de réponse peut dépendre du volume stocké et pas seulement du volume associé au résultat de la requête
Horizon temporel	Court	Long
Nombre et type d'accès	Réguliers et prévisibles	Très irréguliers et imprévisibles
Volume	Rarement supérieure à la dizaine de gigas	Supérieur car historisation



Source :Franco J.M (1997), *le Data Warehouse et le Data Mining*, Eyrolles, Paris, p. 100

Model-based vs Data-based

Different approaches



Understanding and Predicting

Building Models

A model needs data to exist but,
once it exists, it can exist without the data

Model

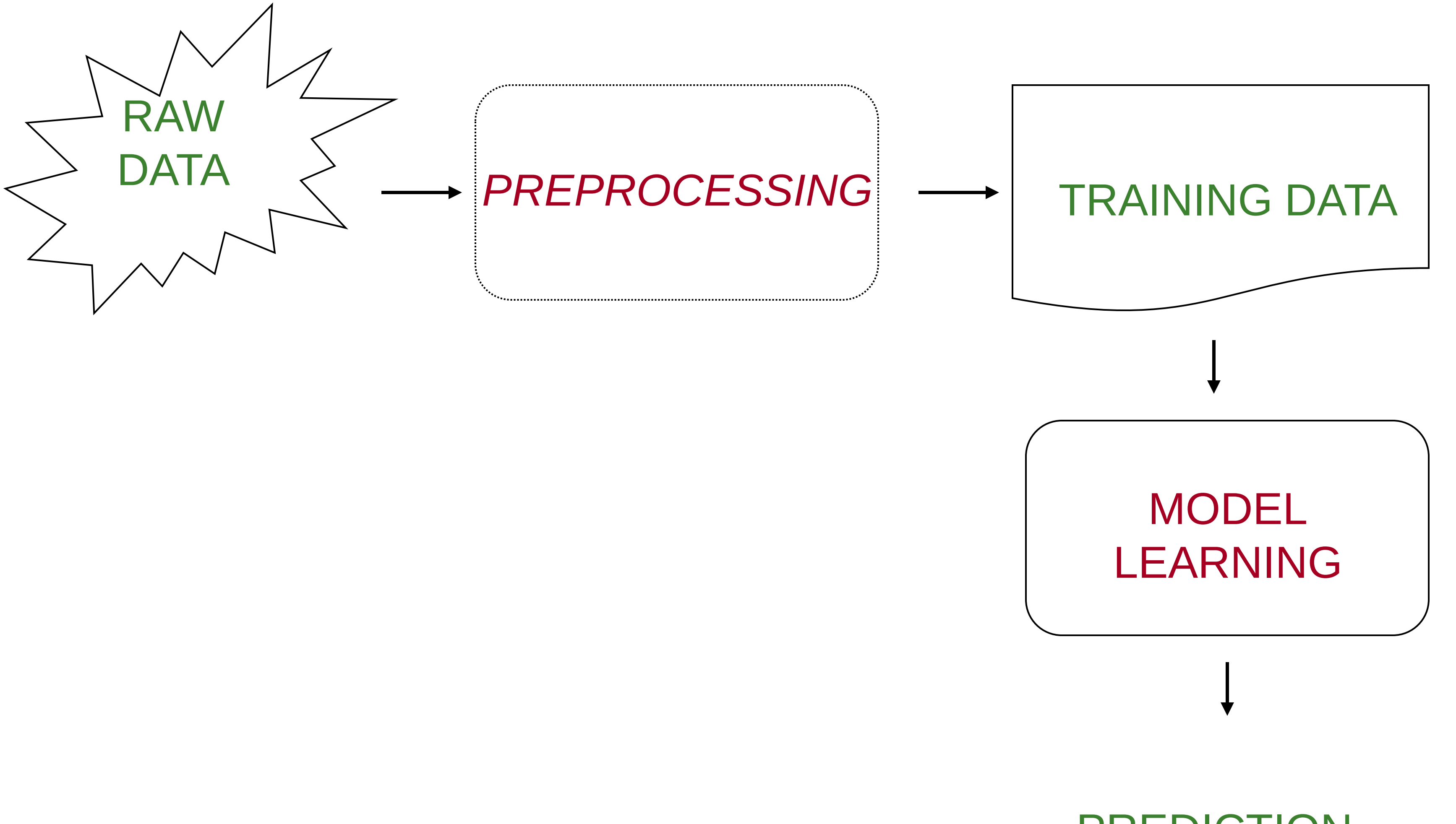
Structure

Parameters

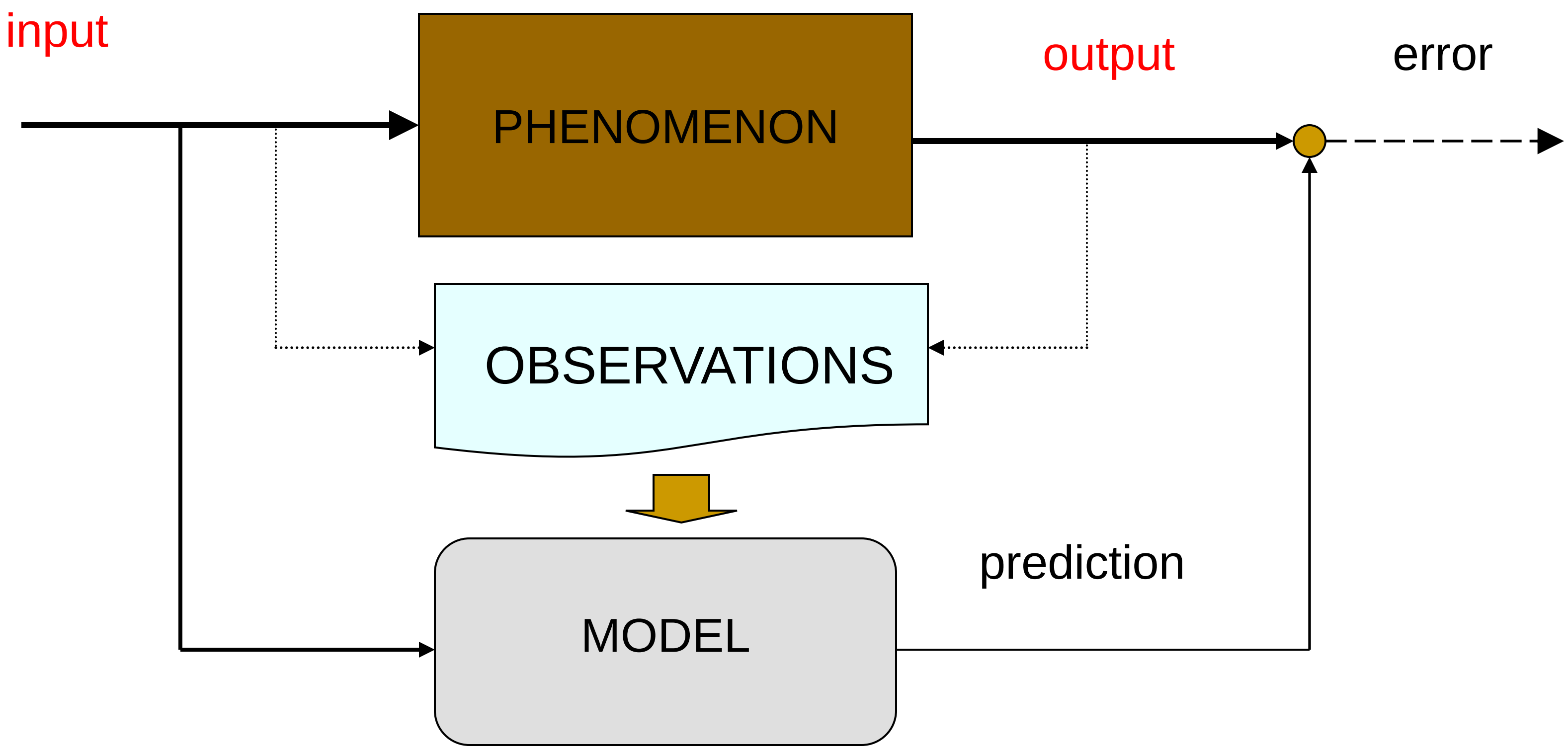
To fit the data

Linear, NN, Fuzzy, ID3, Wavelet, Fourier, Polynomes,...

From data to prediction

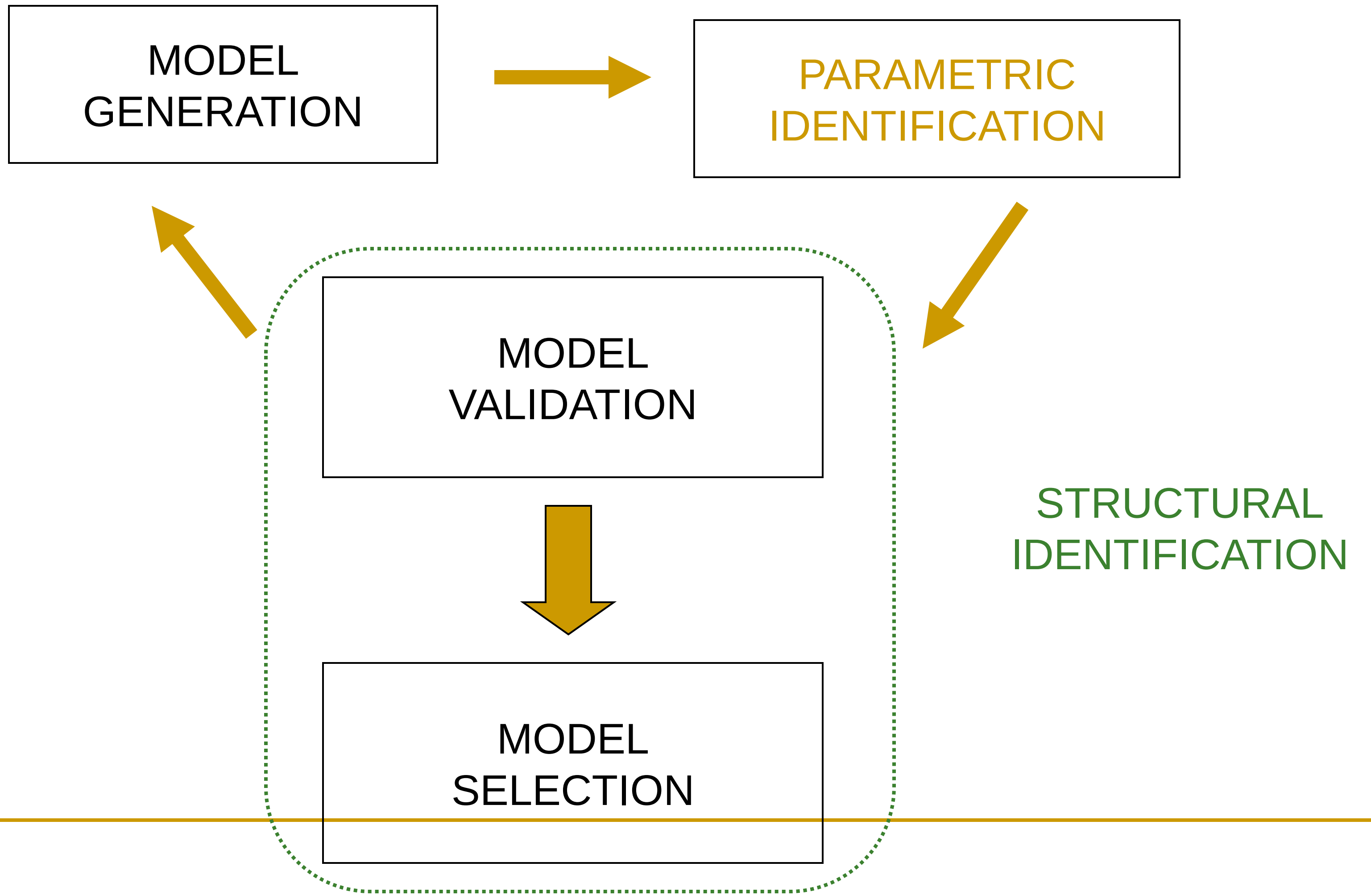


Supervised learning

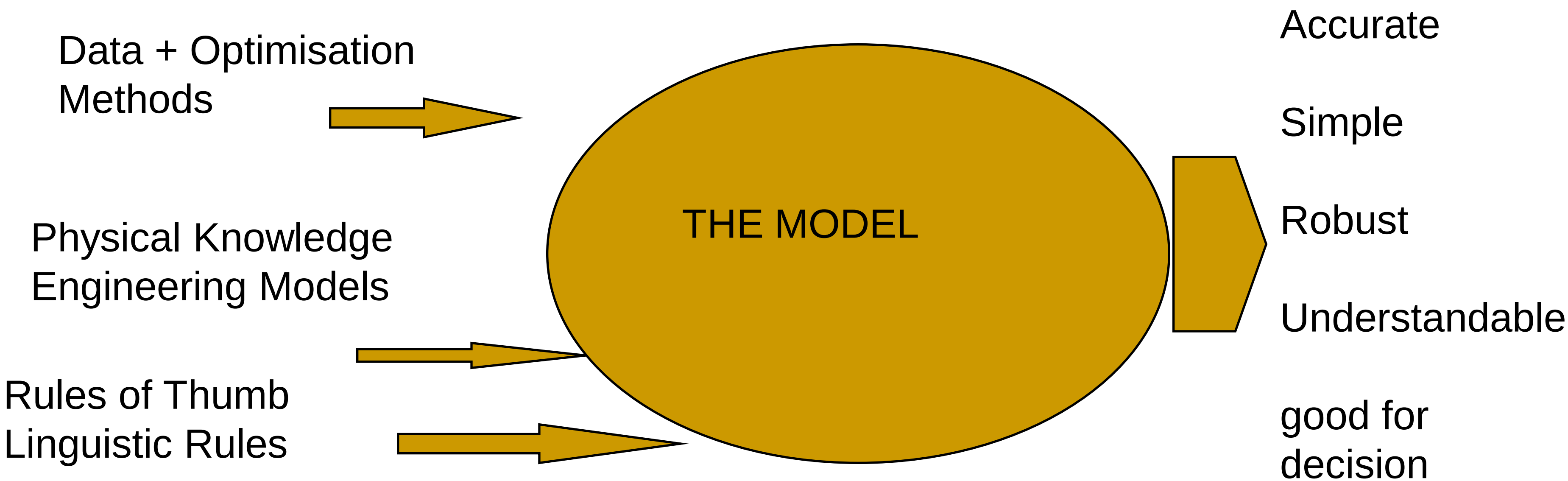


- Finite amount of noisy observations.
- No a priori knowledge of the phenomenon.

Model learning

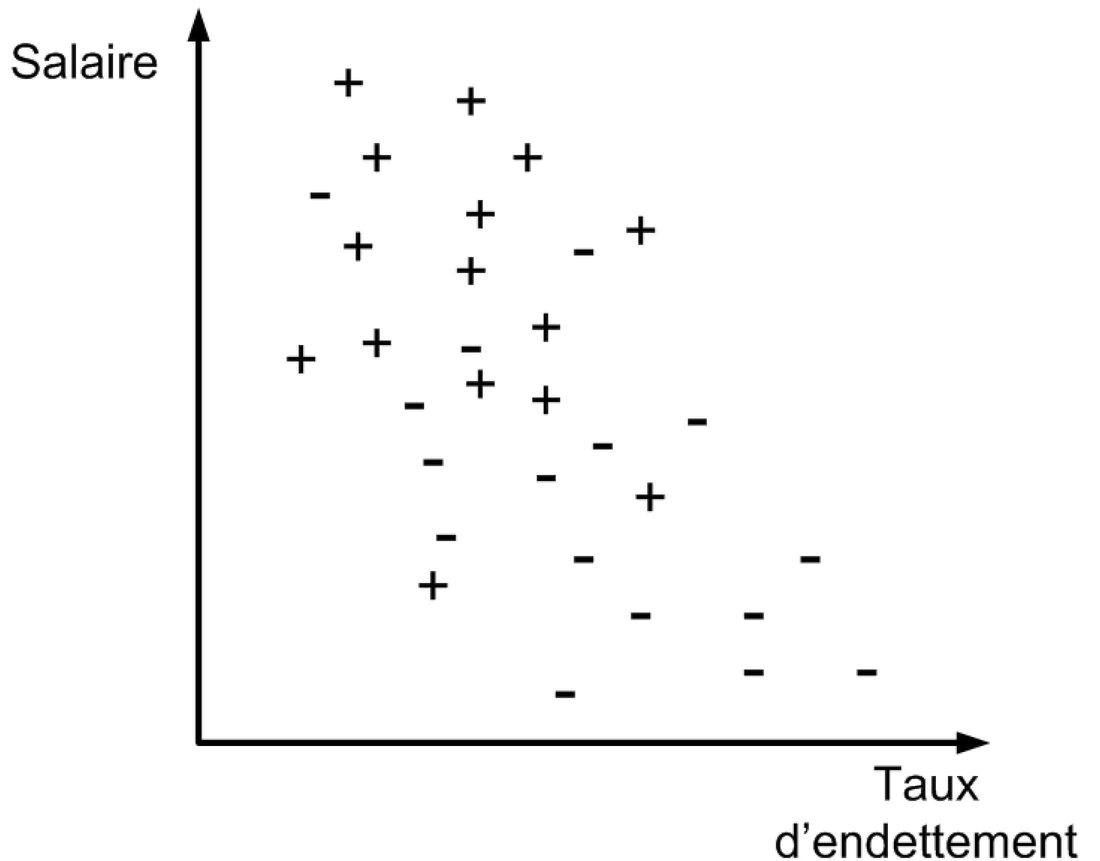


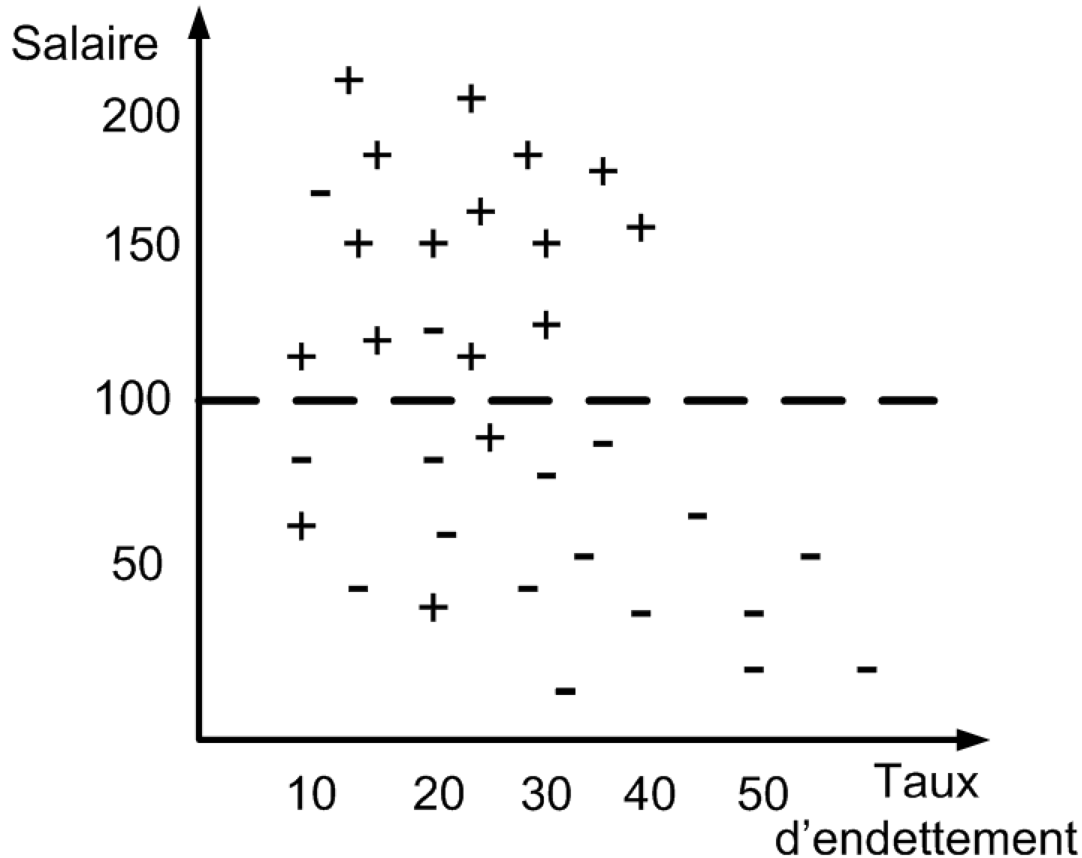
The Practice of Modelling

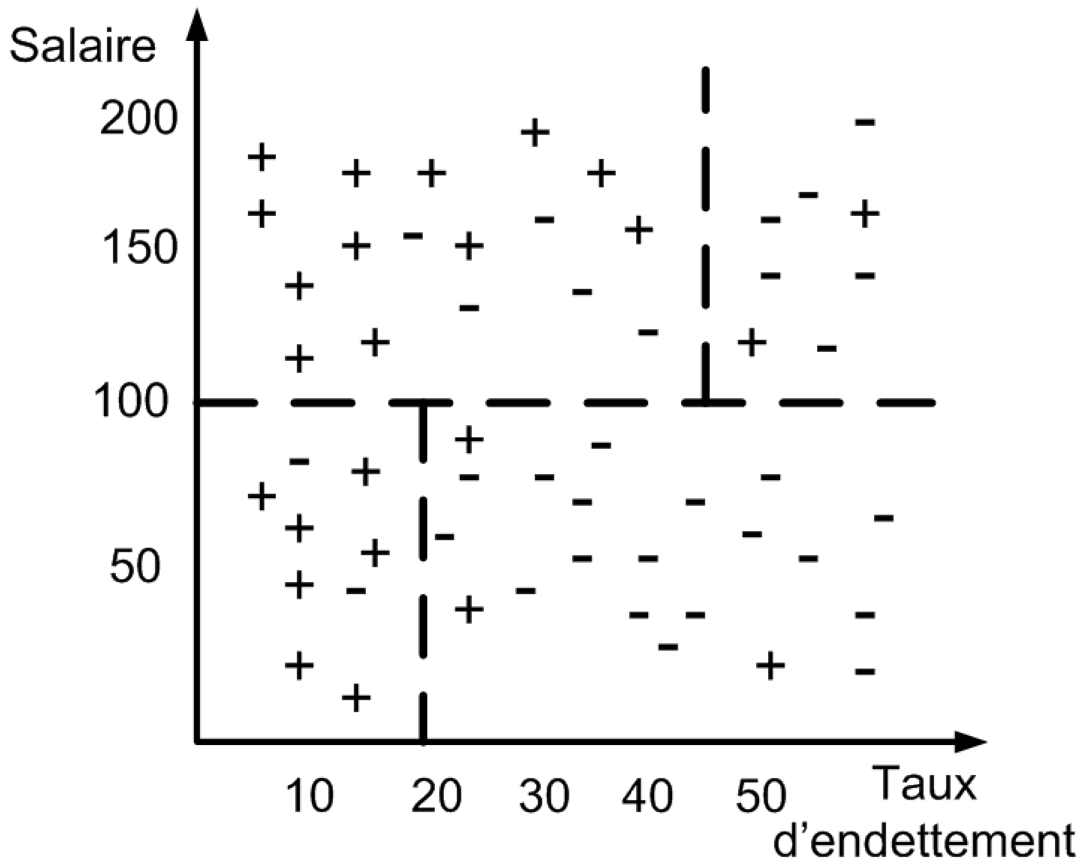


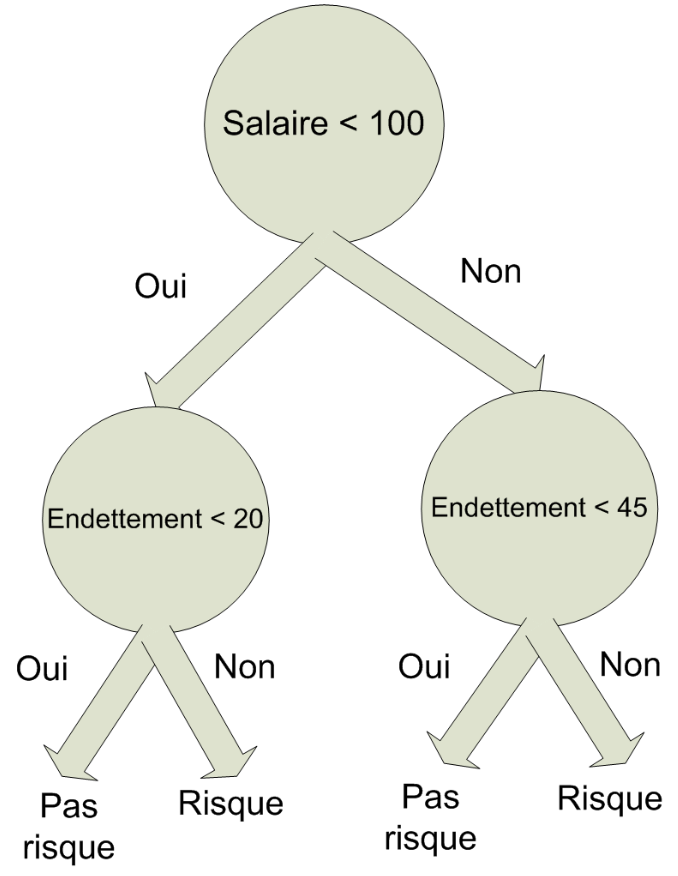
Comprehensible models

- Decision trees
- Qualitative attributes
- Force the attributes to be treated separately
- classification surfaces parallel to the axes
- good for comprehension because they select and separate the variables


Salaire
Taux
d'endettement


Salaire
200
150
100
50
10
20
30
40
50
Taux
d'endettement


Salaire
200
150
100
50
10
20
30
40
50
Taux
d'endettement


Salaire < 100
Oui
Non
Endettement < 20
Endettement < 45
Oui
Non
Oui
Non
Pas risque
Risque
Pas risque
Risque

Decision trees

- Very used in practice. One of the favorite data mining methods
- Work with noisy data (statistical approaches) can learn logical model out of data expressed by and/or rules
- ID3, C4.5 ---> Quinlan
- Favoring little trees --> simple models

- At every stage the most discriminant attribute
- The tree is being constructed top-down adding a new attribute at each level
- The choice of the attribute is based on a statistical criteria called : “the information gain”
- Entropie = $-p_{oui}\log_2 p_{oui} - p_{non}\log_2 p_{non}$
- Entropie = 0 if $P_{oui/non} = 1$
- Entropie = 1 if $P_{oui/non} = 1/2$

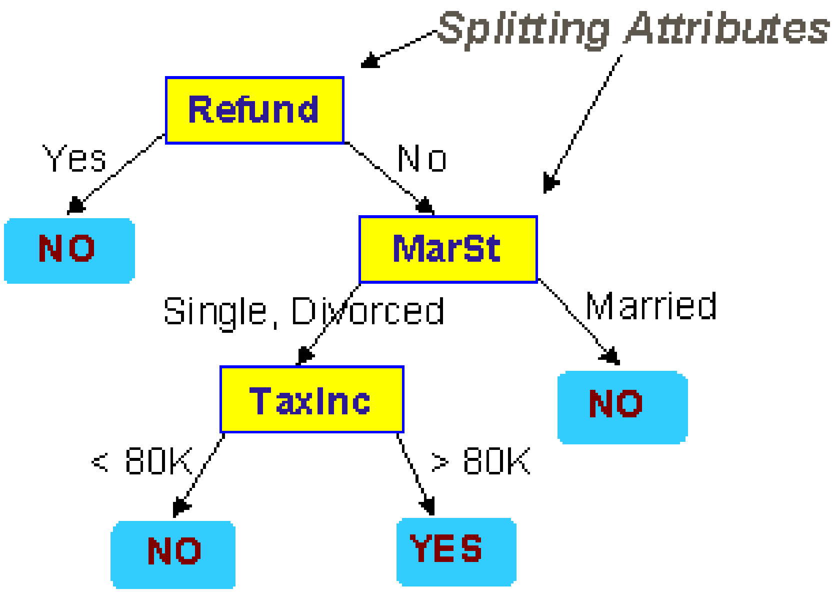
Information gain

- S = set of instances, A set of attributes and v set of values of attributes A
- Gain (S,A) = Entropie(S)-Σ_v|S_v|/|S|*Entropie(S_v)
- the best A is the one that maximises the Gain
- The algorithm runs in a recursive way
- The same mechanism is reapplied at each level

Example Decision Tree

categorical categorical continuous class

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

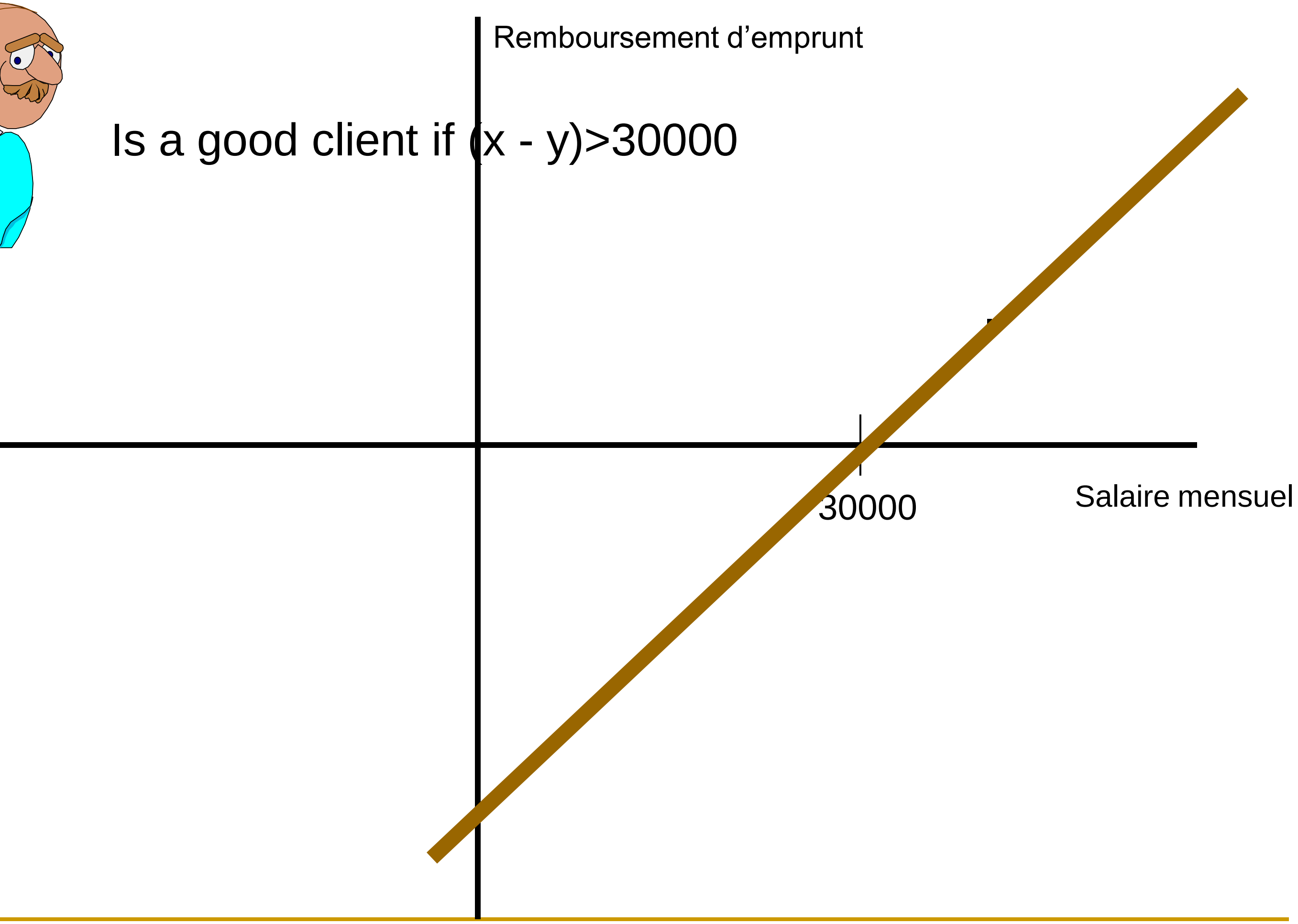


The splitting attribute at a node is determined based on the Gini index.

(C) 1999, Univ. of Minnesota

BUT !!!!

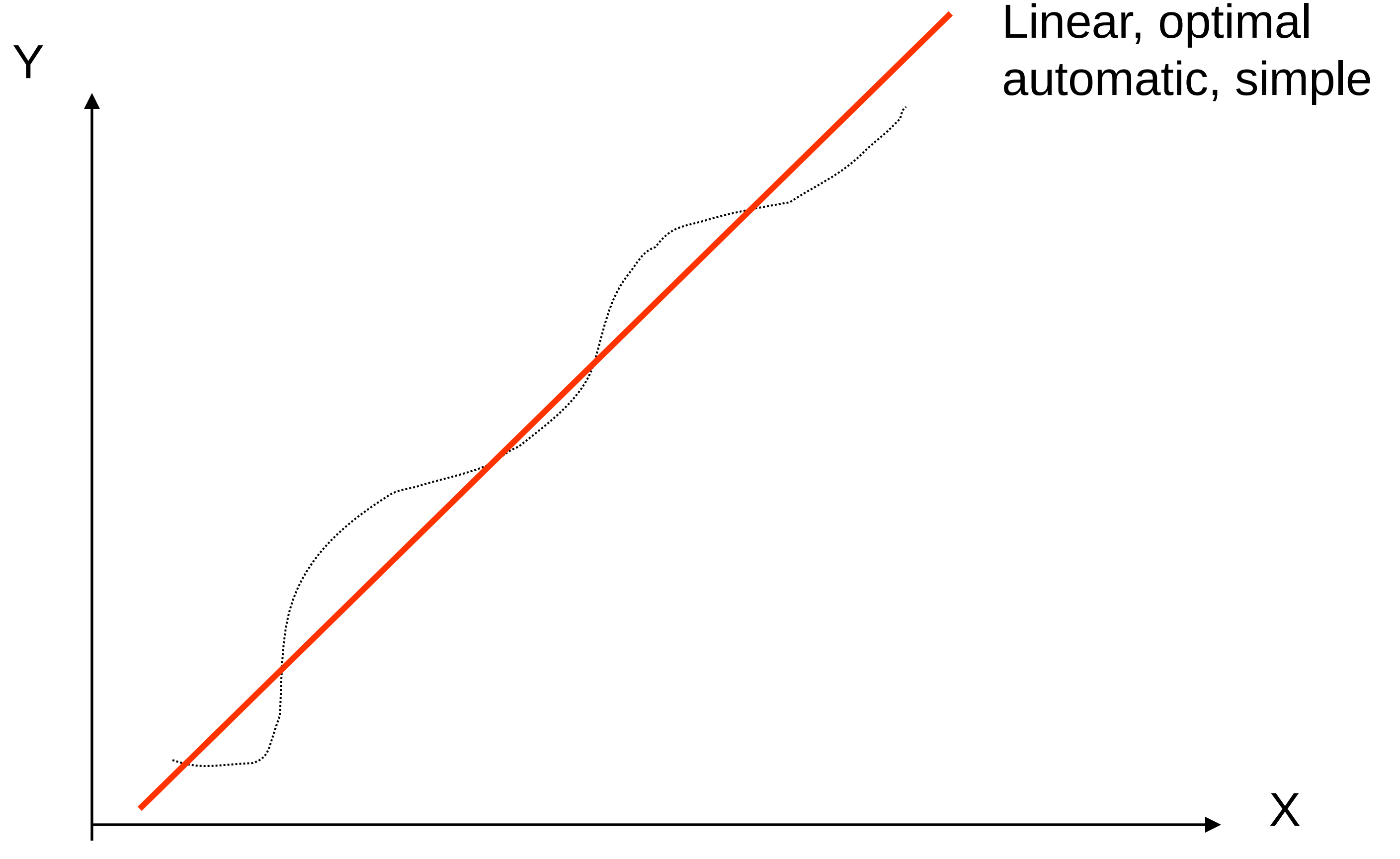
Is a good client if (x - y)>30000



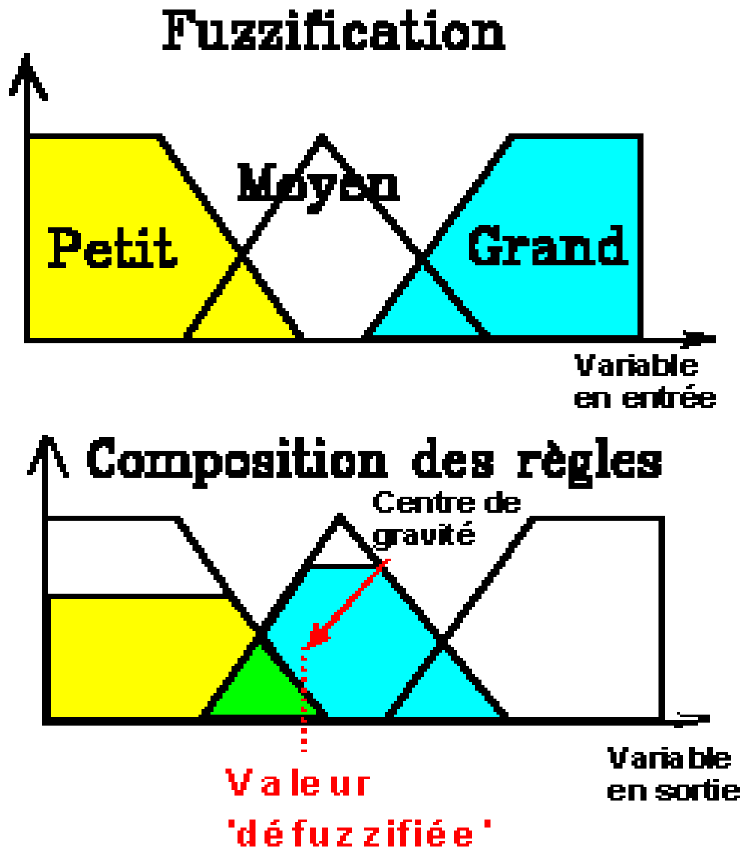
Other comprehensible models

- Fuzzy logic
- Realize an I/O mapping with linguistic rules
- If I eat “a lot” then I take weight “a lot”

Trivial example



The fuzzy


Fuzzification
Petit
Moyen
Grand
Variable
en entrée
Composition des règles
Centre de
gravité
Valeur
'défuzzifiée'
Variable
en sortie

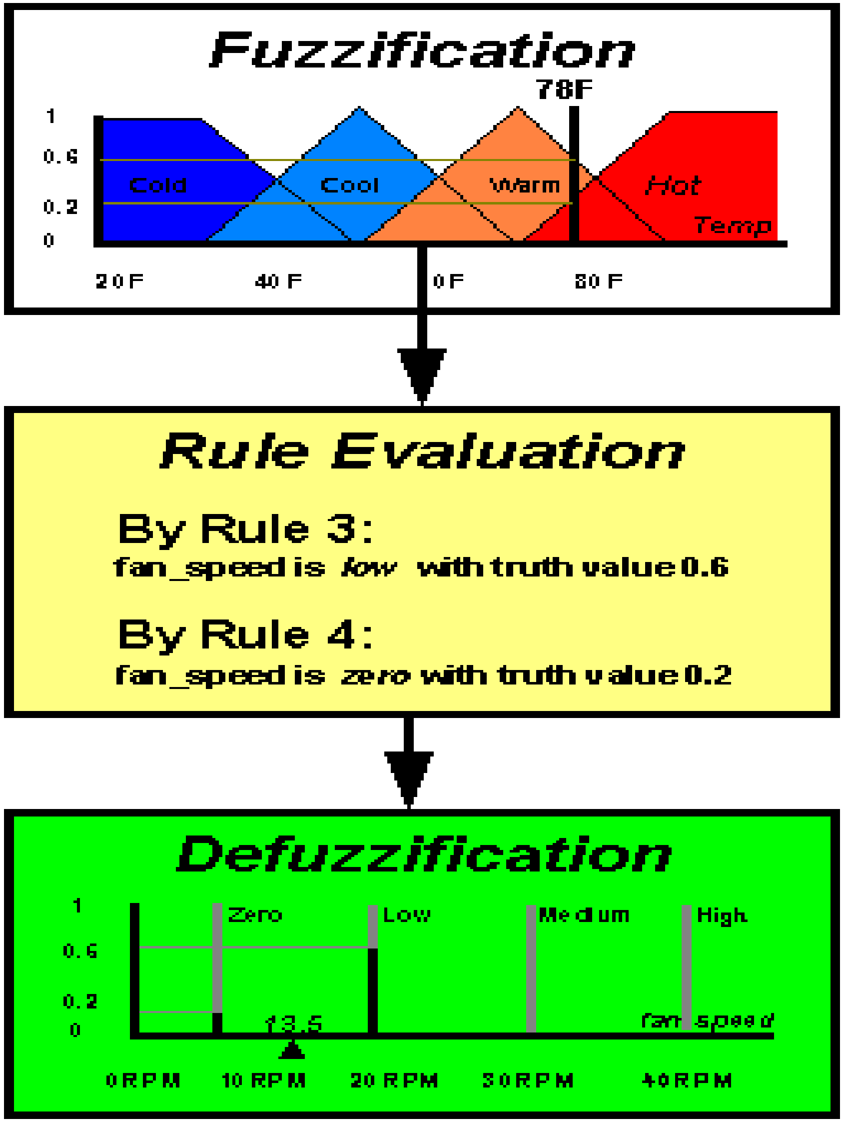
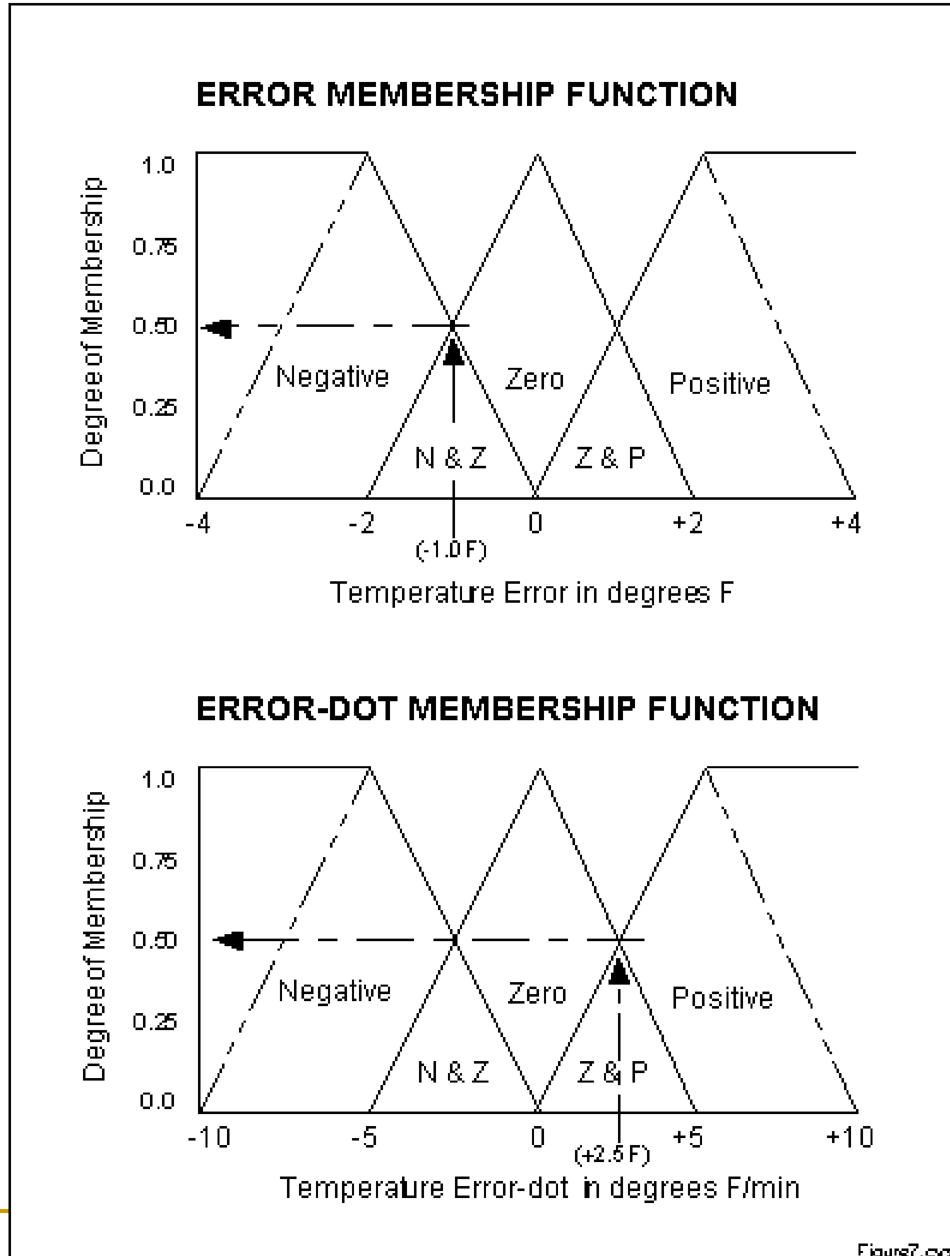
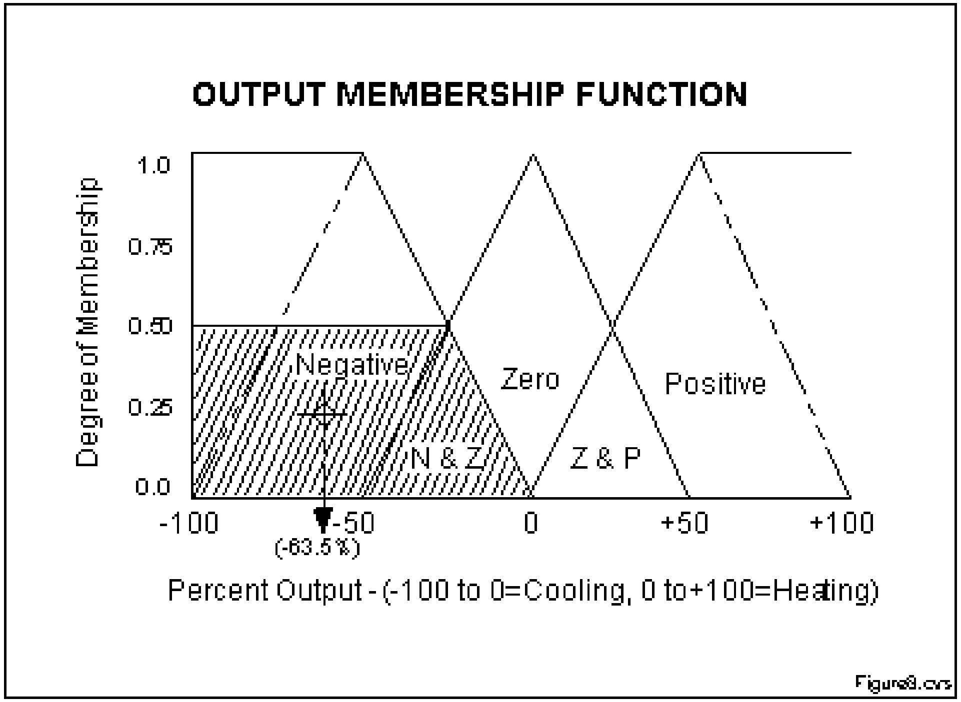
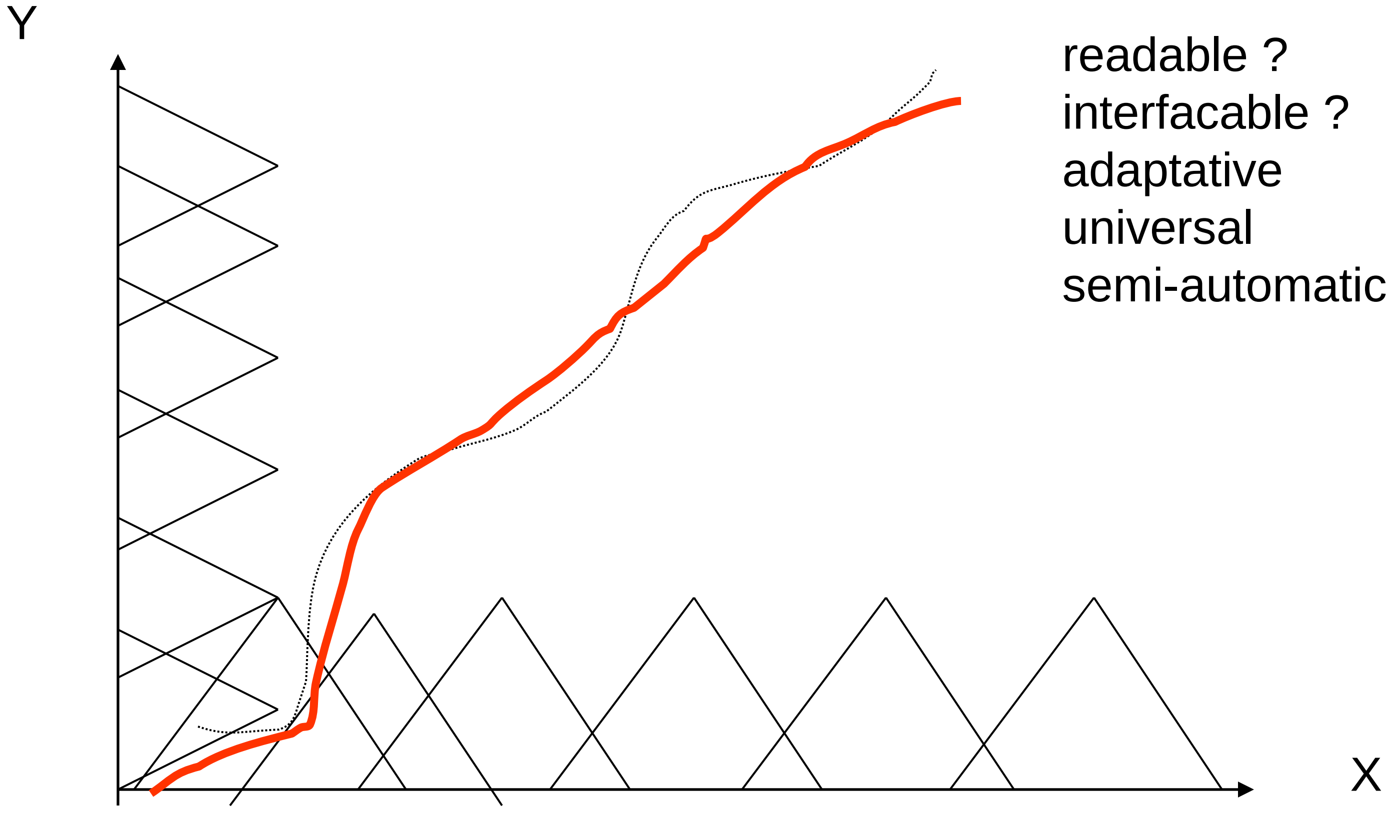


Figure 3 The Fuzzy Inference Process


ERROR MEMBERSHIP FUNCTION
1.0
0.75
0.50
0.25
0.0
Degree of Membership
Negative
Zero
Positive
N & Z
Z & P
-4
-2
0
+2
+4
(-1.0 F)
Temperature Error in degrees F
ERROR-DOT MEMBERSHIP FUNCTION
1.0
0.75
0.50
0.25
0.0
Degree of Membership
Negative
Zero
Positive
N & Z
Z & P
-10
-5
0
+5
+10
(+2.5 F)
Temperature Error-dot in degrees F/min
Figure7.cvs


OUTPUT MEMBERSHIP FUNCTION
Degree of Membership
1.0
0.75
0.50
0.25
0.0
Negative
Zero
Positive
N & Z
Z & P
-100
-50
0
+50
+100
(-63.5%)
Percent Output - (-100 to 0=Cooling, 0 to+100=Heating)
Figure8.cvs

IF x is very small THEN y is small
IF x is small THEN y is medium
IF x is medium THEN y is medium

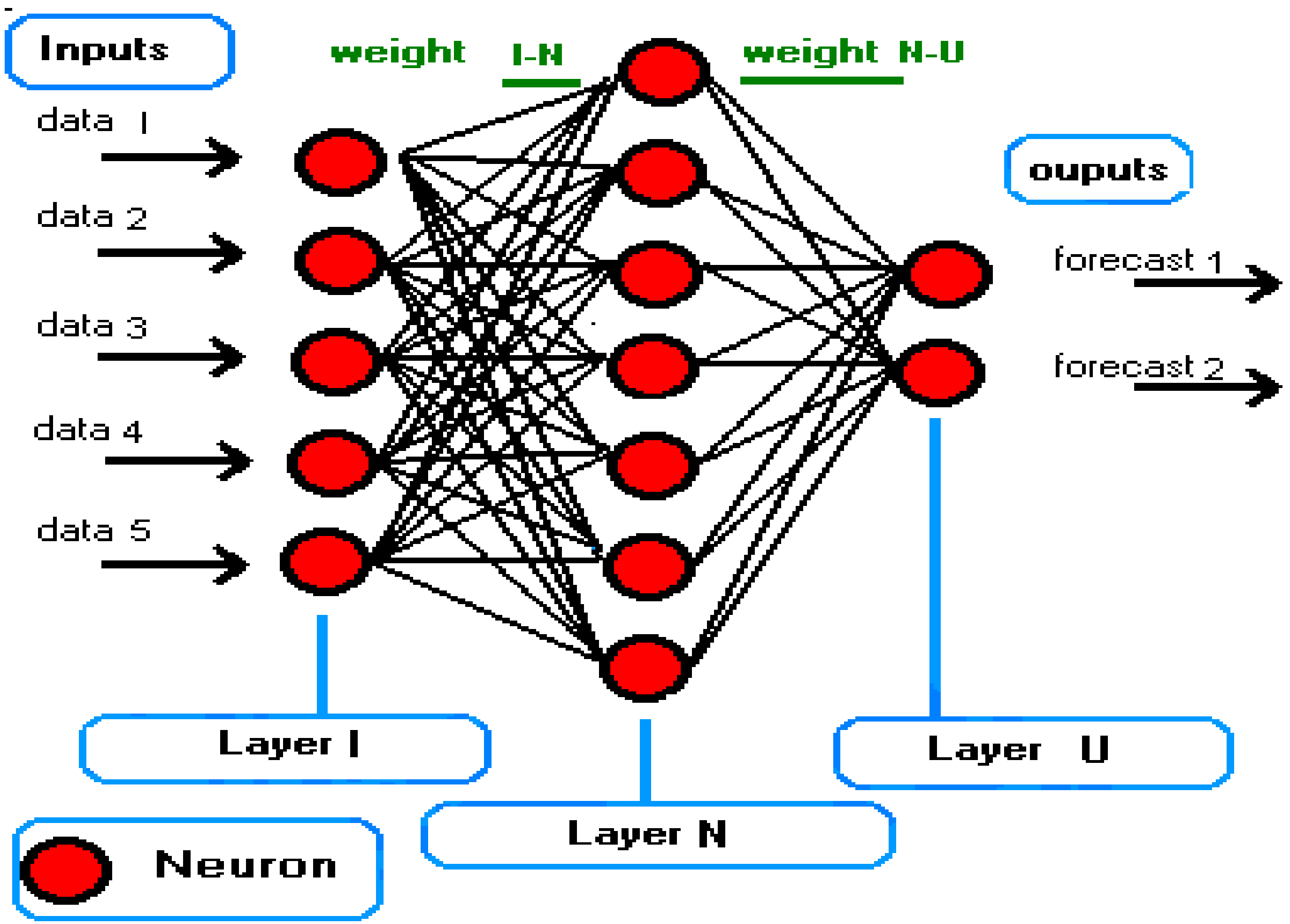


readable ?
interfacable ?
adaptative
universal
semi-automatic

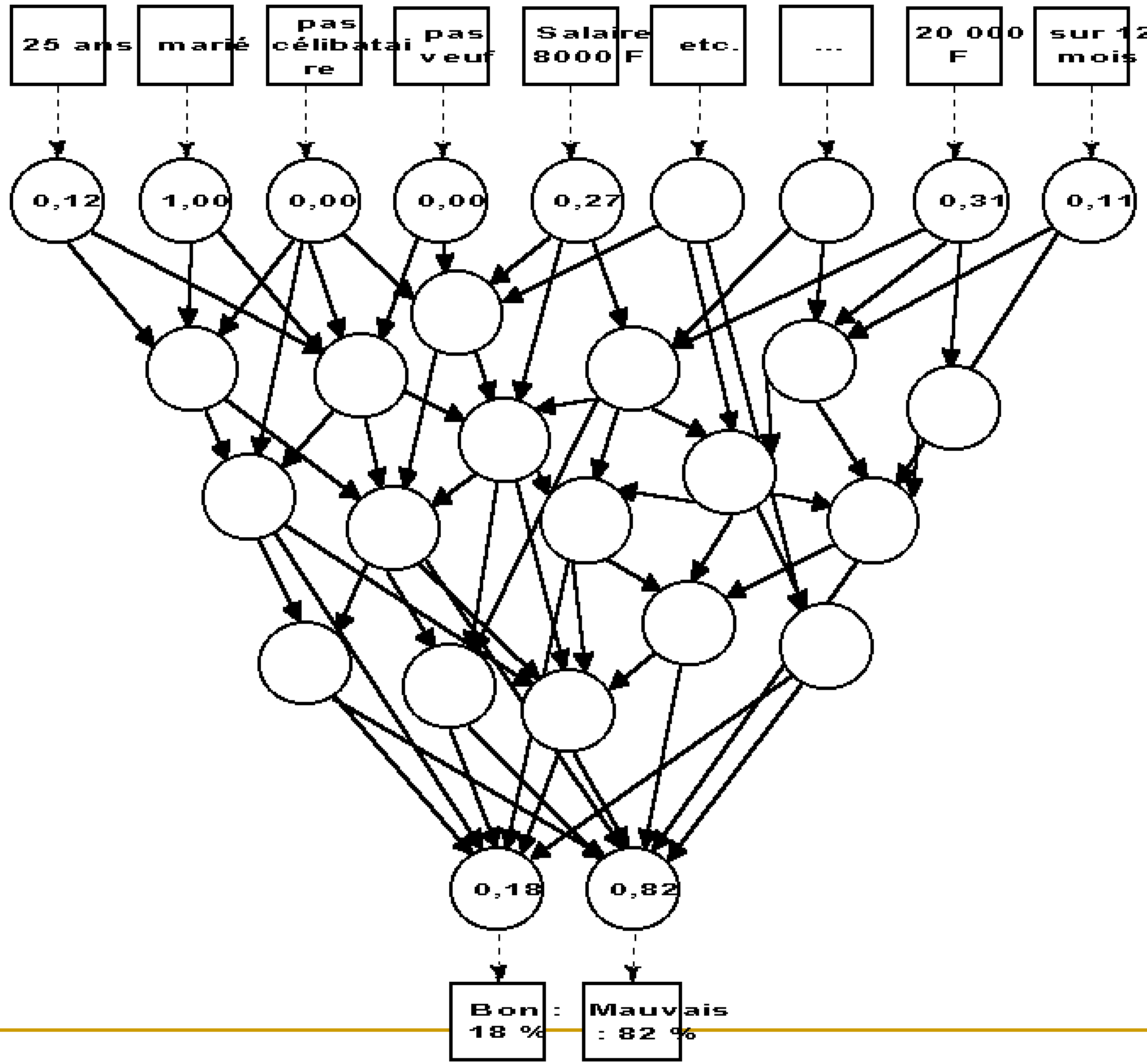
Non comprehensible models

- From more to less
  - linear discriminant
  - local approaches
    - fuzzy rules
    - Support Vector Machine
    - RBF
  - global approaches
    - NN
    - polynômes, wavelet,…
    - Support Vector Machine

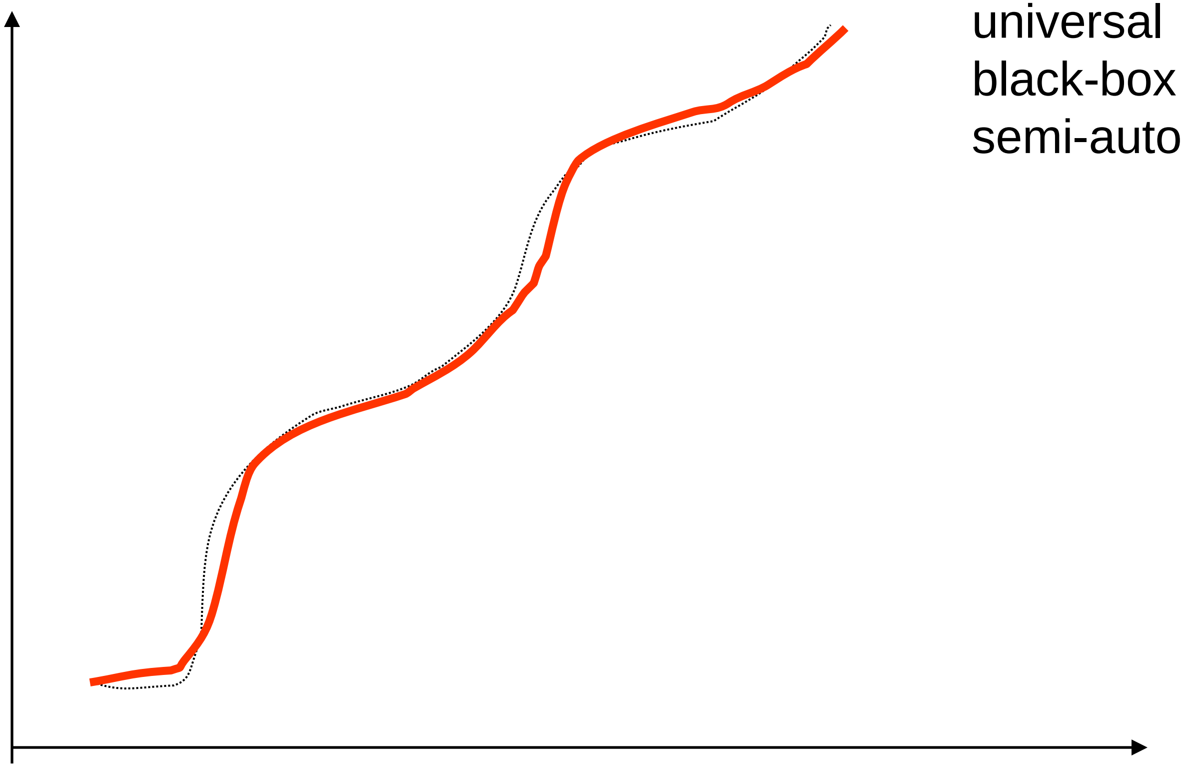
The neural network



description de dossier de prêt

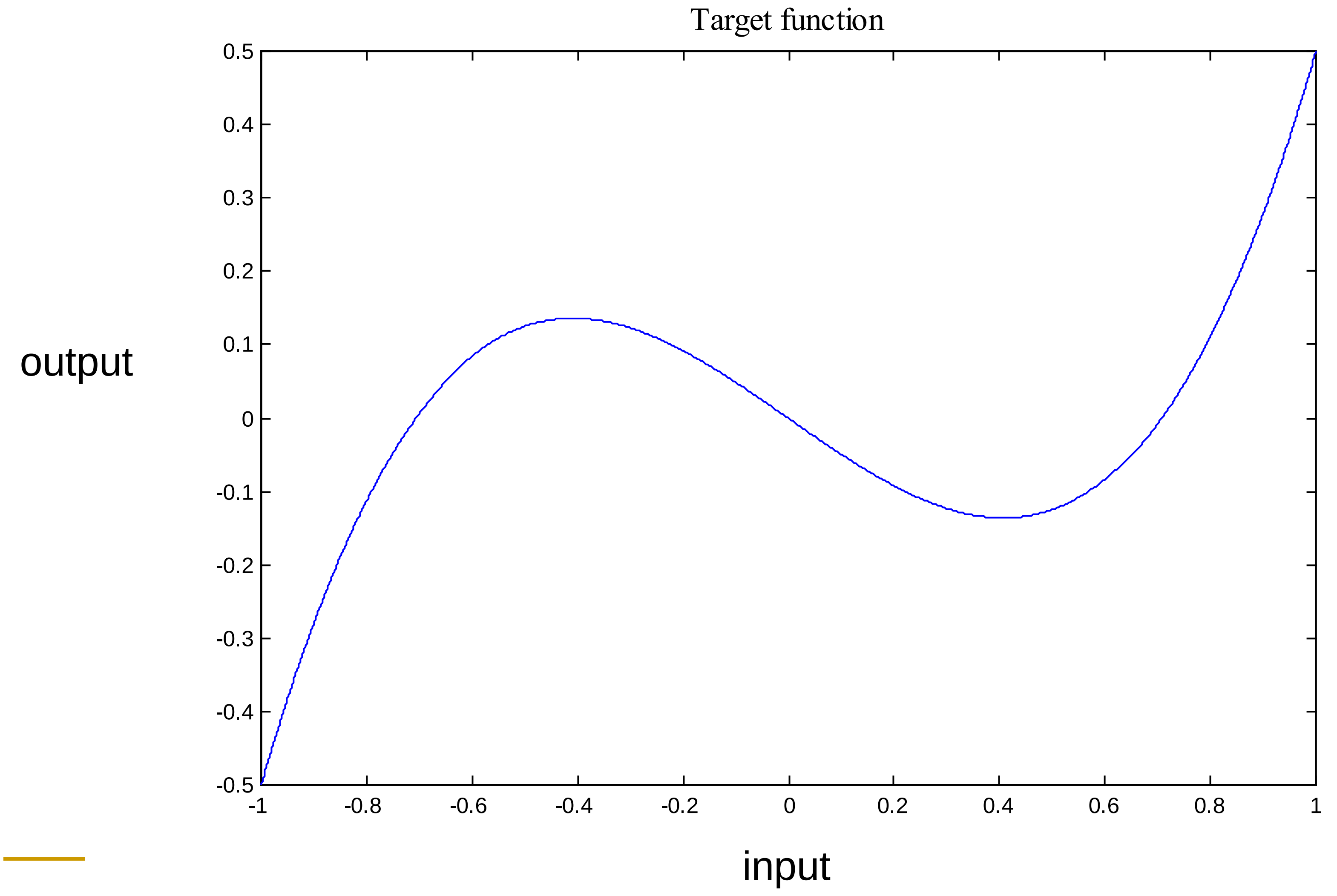


suggestion de décision

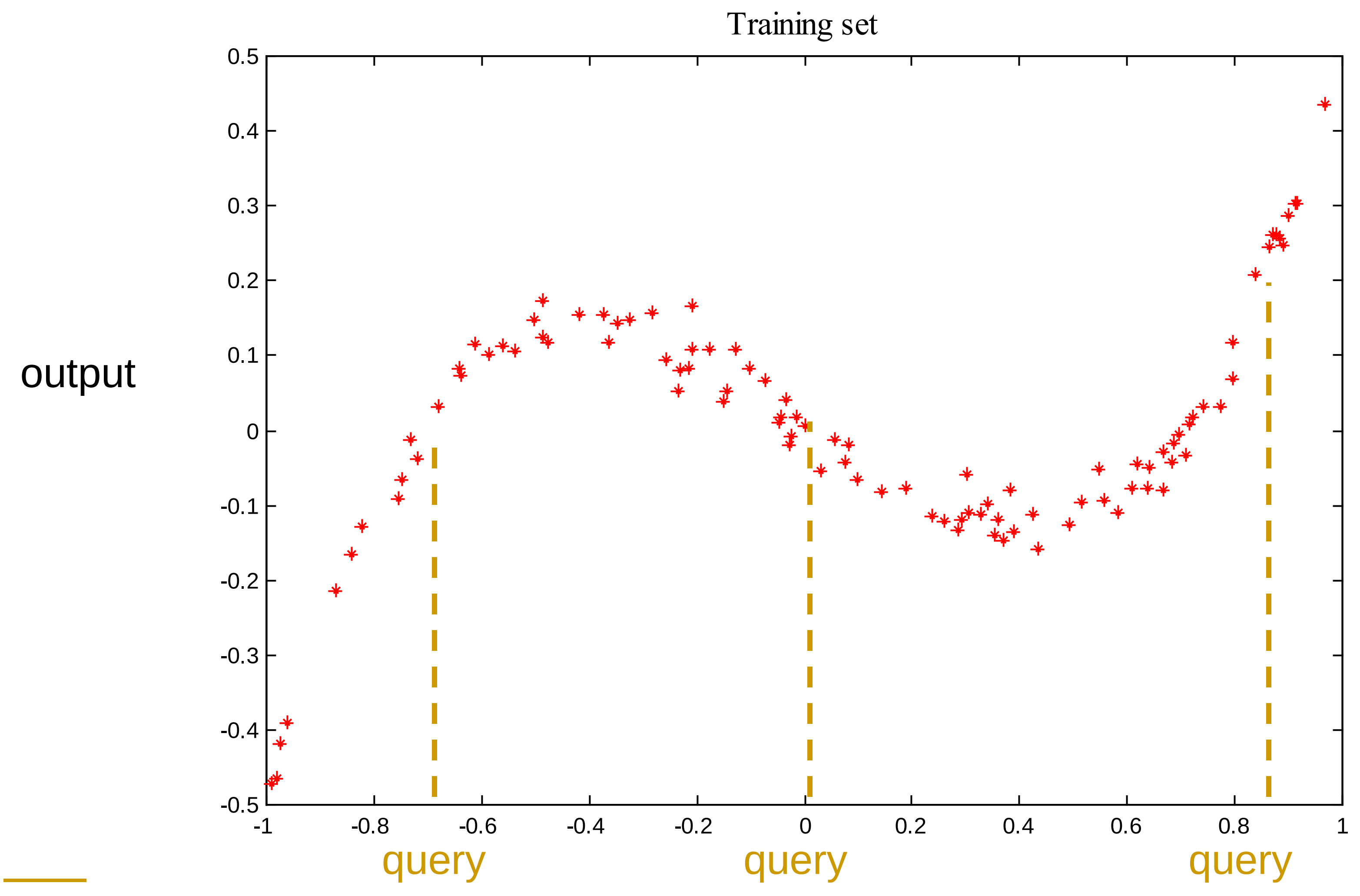


precise
universal
black-box
semi-automatic

Nonlinear relationship

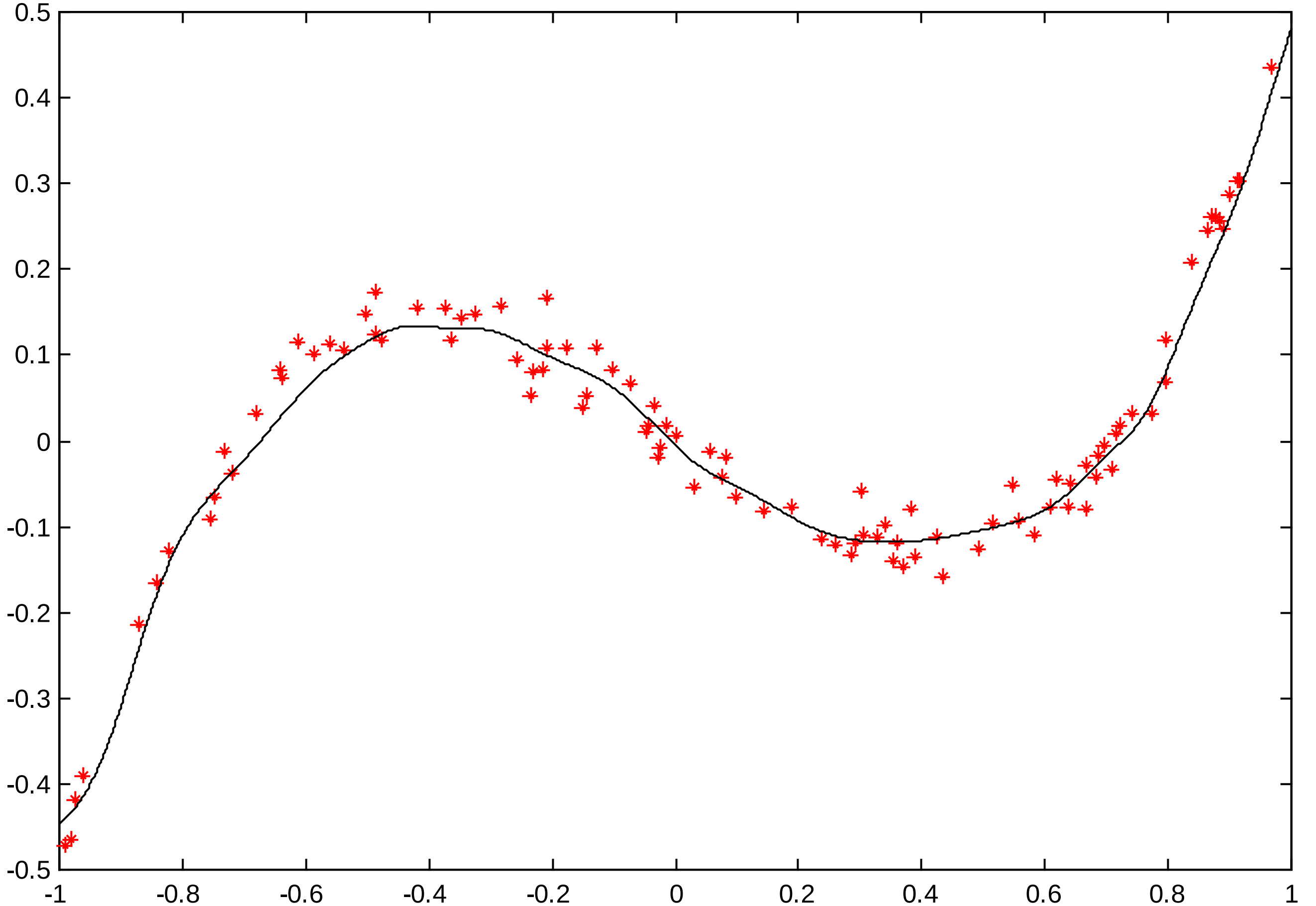


Observations



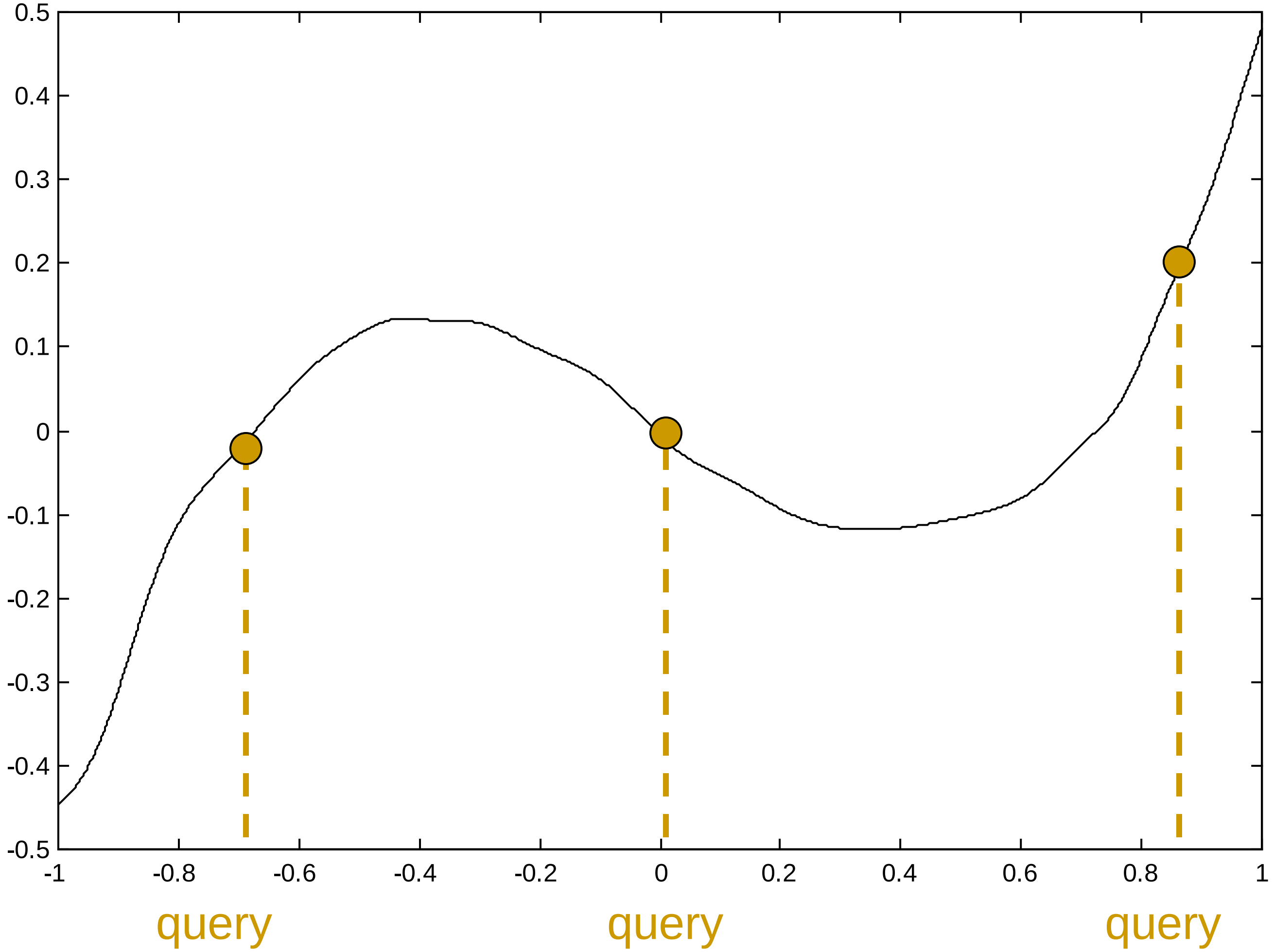
Global modeling

output



input

Prediction with global models



Advantages

- Exist without data
- Information compression
  - Mainly SVM: mathématiques, pratiques, logique et génériques.
- Detect a global structure in the data
- Allow to test the sensitivity of the variables
- Can easily incorporate prior knowledge

Drawbacks

- Make assumption of uniformity
- Have the bias of their structure
- Are hardly adapting
- Which one to choose.

BAGFS: ensemble method

`Weak classifiers´ ensembles

- Classifier capacity reduced in 2 ways :
  - simplified internal architecture
  - NOT all the available information
- Better **generalisation**, reducing **overfitting**
- Improving **accuracy**

☞ by **decorrelating** classifiers errors

☞ by increasing the **variability** in the learning space.

`Bagging´ : resampling the learning set

- Bootstraps aggregating (*Leo Breiman*)
  - random and independant perturbation of the learning set.
  - vital element : **instability** of the inducer*.
    - e.g. **C4.5**, **neural network** but not **kNN** !
  - increase ***accuracy*** by reducing ***variance***

* inducer = base learning algorithm : c4.5, kNN, ...

Learning set resampling : `Arcing´

- **Adaptive** resampling or reweighting of the learning set (*Leo Breiman* terminology).

☞ Boosting (*Freund & Schapire*)

  - sequential reweighting based on the description accuracy.
    - ☞ e.g. **AdaBoost.M1** for multi-class problems.
  - needs unstability so as bagging
  - better variability than bagging.
  - sensible to noisy databases.
  - better than *bagging* on non-noisy databases

Mutliple Feature Subsets : Stephen D. Bay (1/2)

- problem ?
  - **kNN is stable vertically** so Bagging doesn't work.

☞ **horizontally** : **MFS** - combining random selections of features with or without replacement.

- question ?
  - what about other inducers such C4.5 ??

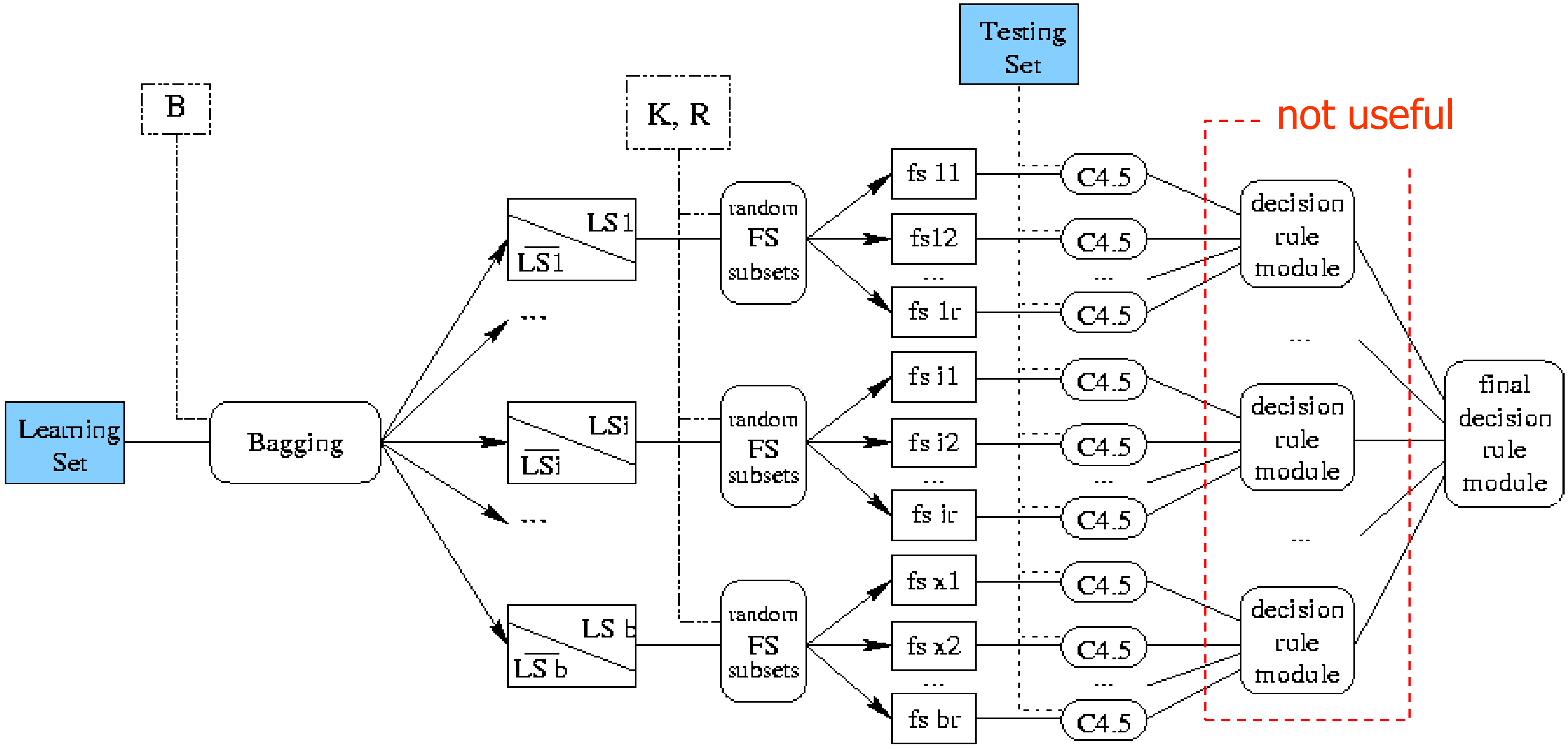
Multiple Feature Subsets : Stephen D. Bay (2/2)

- **Hypo** : kNN uses its ‘ horizontal ’ instability.
- Two parameters :
  - K=n/N, proportion of features in subsets.
  - R, number of subsets to combine.
- ☺ MFS is better than single kNN with FSS and BSS, feature selections techniques.
- ☺ MFS is **more stable** than kNN on added irrelevant features.
- ☺ MFS **decreases** variance <u>**and**</u> bias through randomness.

BAGFS : a multiple classifier system

- BAGFS = MFS inside each Bagging.
- BAGMFS = MFS & Bagging together.
- 3 parameters
  - **B**, number of bootstraps
  - **K**=n/N, proportion of features in subsets
  - **R**, number of feature subsets
- decision rule : majority vote

BAGFS architecture around **C4.5**



Experiments

- Testing parametrization
  - optimizing K between 0.1 and 1 by means of a nested 10-fold cross-validation
  - R= 7, B= 7 for two-level method : Bagfs 7x7
  - set of 50 classifiers otherwize : Bag 50, BagMfs 50, MFS 50, Boosting 50

Experimental Results

	c45	bagmfs 50	bagfs 7x7	boosting 50	bag 50	mfs 50
hepatitis	77.6	82.7	**84.1**	82.1	81.0	83.2
glass	64.8	**77.3**	76.6	74.4	74.8	75.2
iris	92.7	**93.4**	93.2	92.4	92.3	**93.5**
ionosphere	90.9	93.7	93.5	93.2	92.8	93.6
liver disorders	64.1	**73.5**	70.5	72.3	72.8	65.6
new-thyroid	92.0	**94.9**	94.5	93.5	93.8	92.7
ringnorm	91.9	**97.9**	97.7	95.3	95.6	97.6
twonorm	85.4	**96.9**	96.7	96.4	96.6	96.6
satimage	86.8	**91.4**	**91.3**	90.0	90.8	92.1
waveform	76.2	**84.6**	83.9	84.0	83.2	83.9
breast-cancer-w	94.7	**96.9**	**96.8**	95.5	95.3	**96.8**
wine	85.7	**92.3**	90.8	91.3	91.3	89.6
segmentation	93.4	**98.2**	**98.4**	95.1	96.6	**98.7**
Image	96.5	97.3	**97.8**	96.7	**97.6**	**97.6**
car	92.1	**93.2**	92.5	92.1	**93.2**	92.2
diabetes	72.4	75.7	75.7	**76.2**	75.7	74.0
	84.8	90.0	89.6	88.8	89.0	88.9

- McNemar test of significance (95%) : Bagfs performs never signif. worse and even sign. better on at least 4 databases (see red databases).

BAGFS : discussions

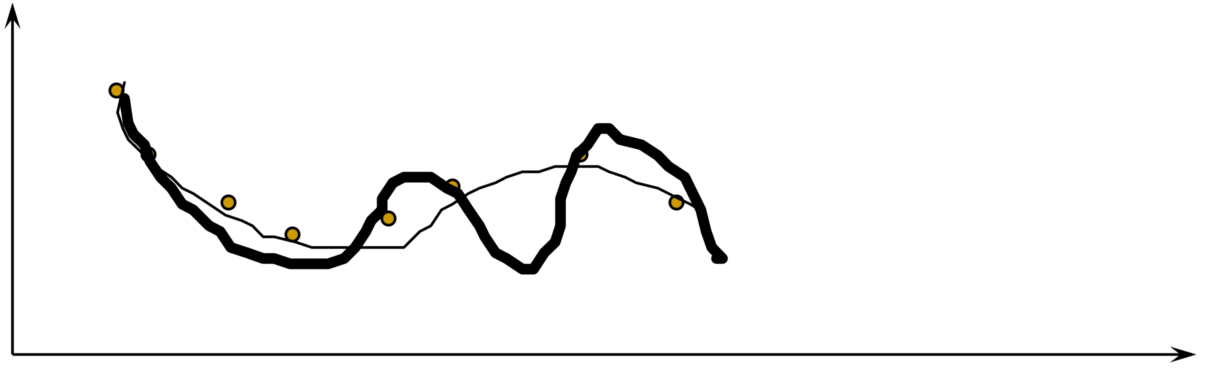
- **How adjusting the parameters B, K, R**
  - internal cross validation ?
  - dimensionality and variability measures *hypothesis*
- **Interest of a second level ?**
  - About irrelevant and (un)informative features ?
  - Does bagging + feature selections work better ?
  - How proving the interest of MFS randomness ?
- **How using bootstraps complementary ?**
  - Can we ?
  - What to do ?
- **How proving horizontal unstability of C4.5 ?**
- **Comparison with 1-level bagging and MFS**
  - Same number of classifiers ?
  - Advantage of tuning parameters ?

Which best model ??

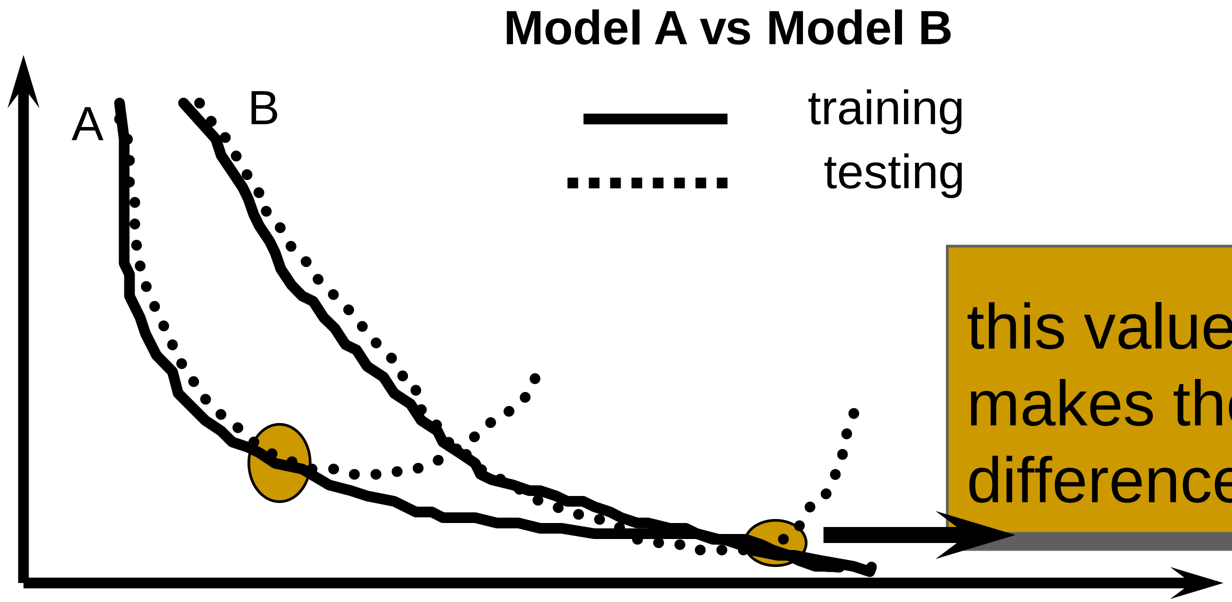
when they all can perfectly fit the data

They all can perfectly fit the data but

they don't approach the data
in the same way. This approach
depends on their structure



This explains the importance of Cross-validation



this value makes the difference

Which one to choose

- Capital role of crossvalidation.
- Hard to run
- One possible response

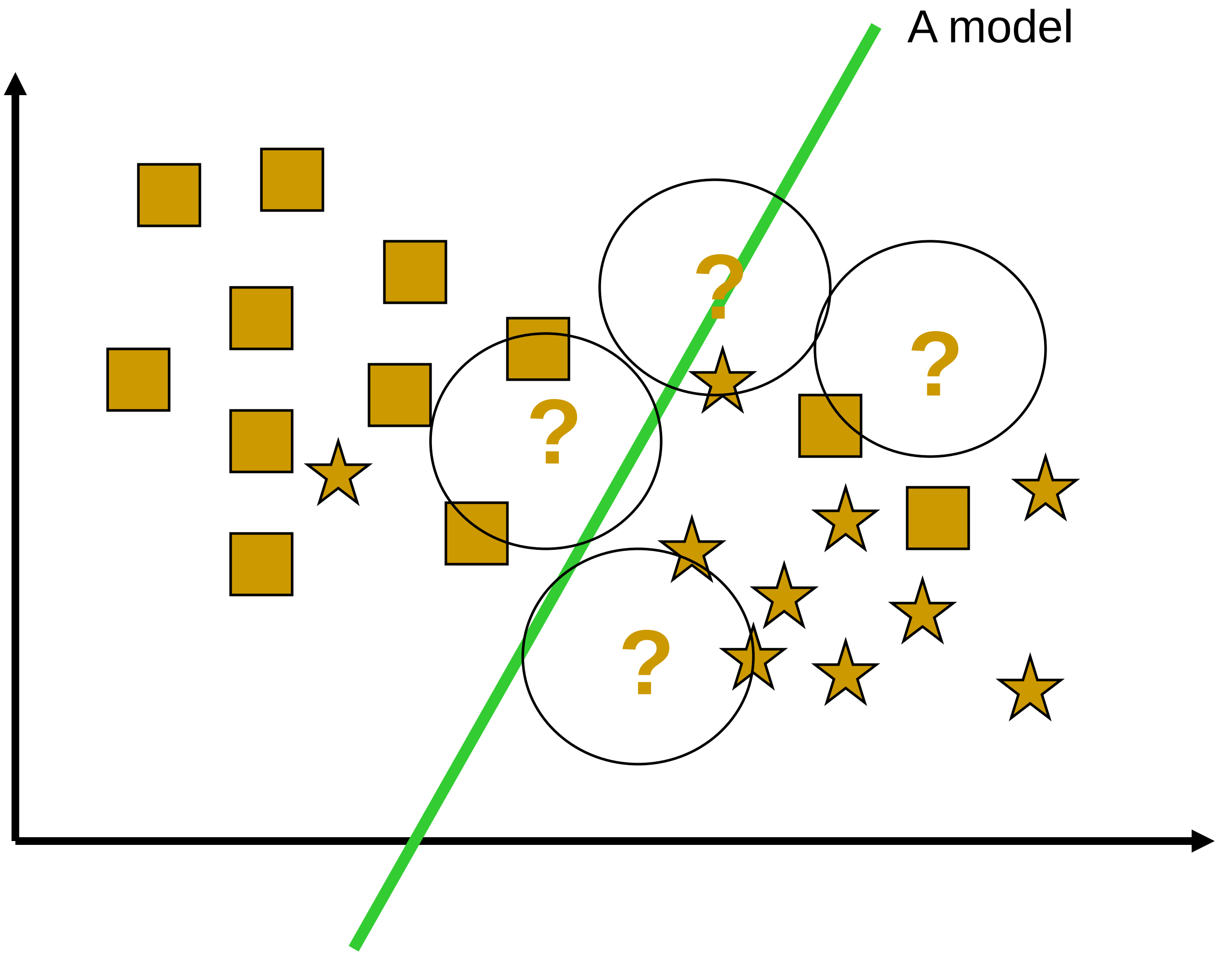
- Lazy methods
- Coming from fuzzy

Model or Examples ??

Build a Model

Prediction
based on the model

Prediction based
on the examples


A model
?
?
?
?

Lazy Methods

- Accuracy entails to keep the data and don't use any intermediary model: the best model is the data
- Accuracy requires powerful **local** models with **powerful cross-validation methods**

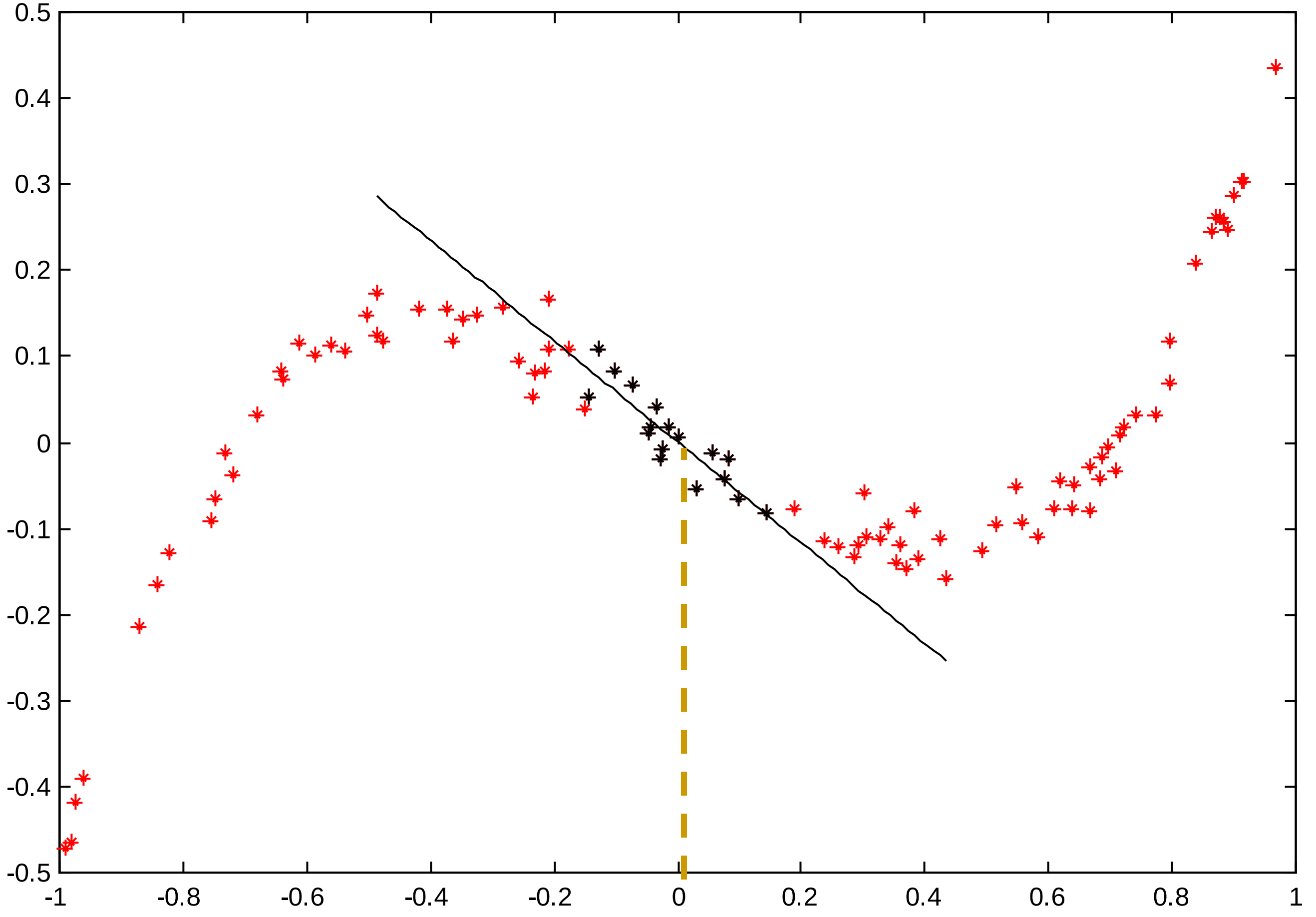
lazy methods is a new trend which
is a revival of an old trend

- Made possible again due to the computer power

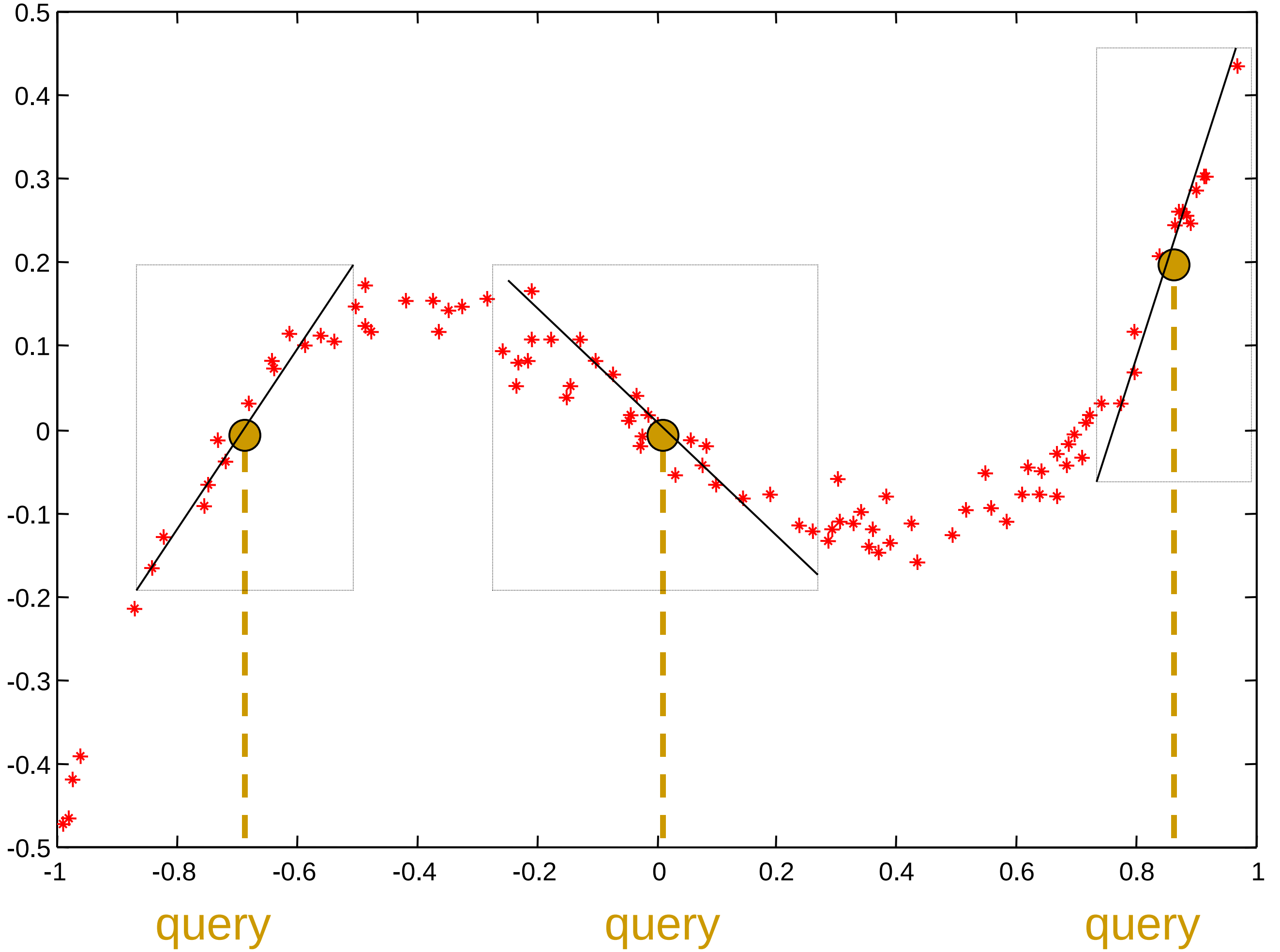
Lazy methods

- A lot of expressions for the same thing:
  - memory-based, instance-based, examples-based,distance-based
  - nearest-neighbour
- lazy for regression, classification and time series prediction
- lazy for quantitative and qualitative features

Local modeling



Prediction with local models



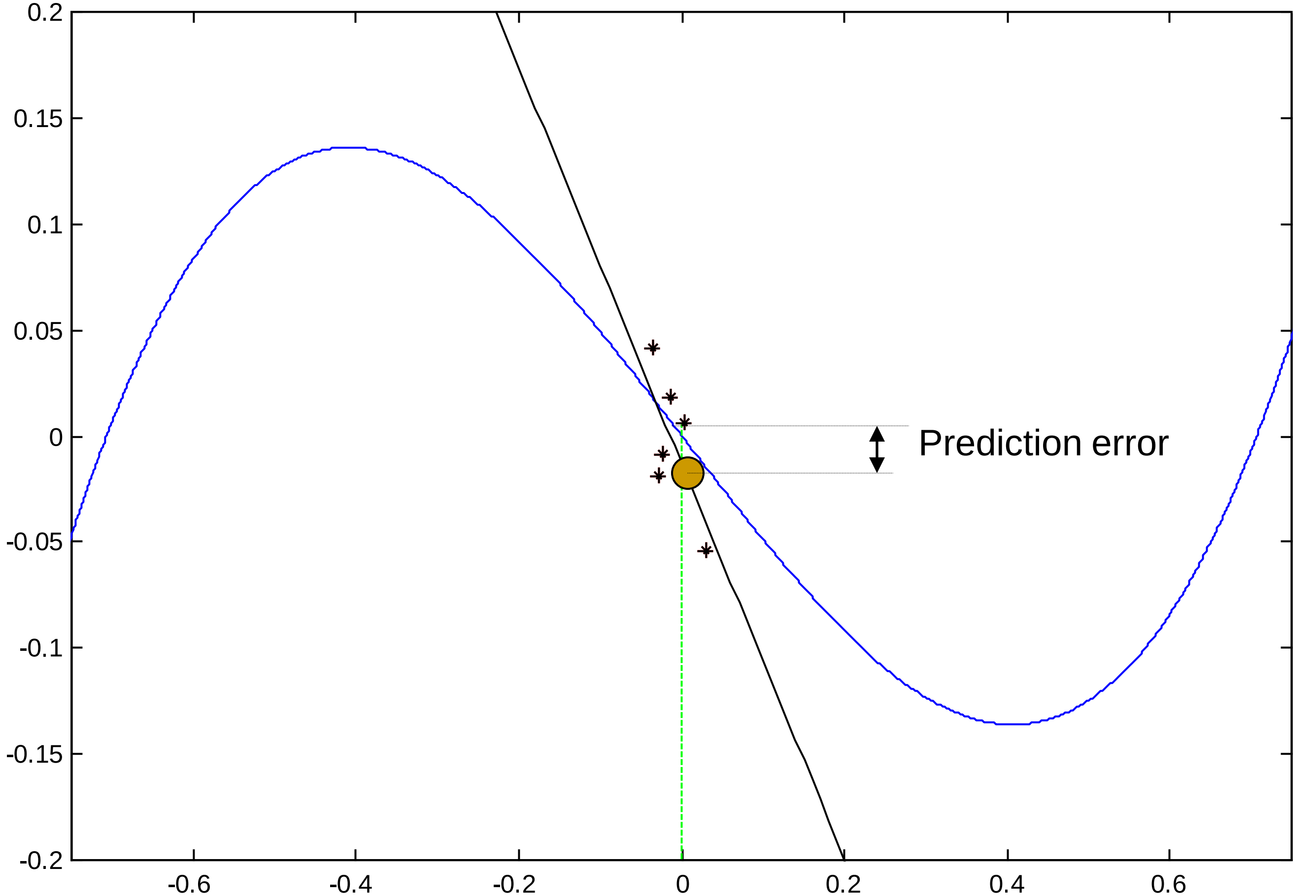
Local modeling procedure

The identification of a local model can be summarized in these steps:

- Compute the distance between the query and the training samples according to a predefined metric.
- Rank the neighbors on the basis of their distance to the query.
- Select a subset of the nearest neighbors according to the bandwidth which measures the size of the neighborhood.
- Fit a local model (e.g. constant, linear,...).

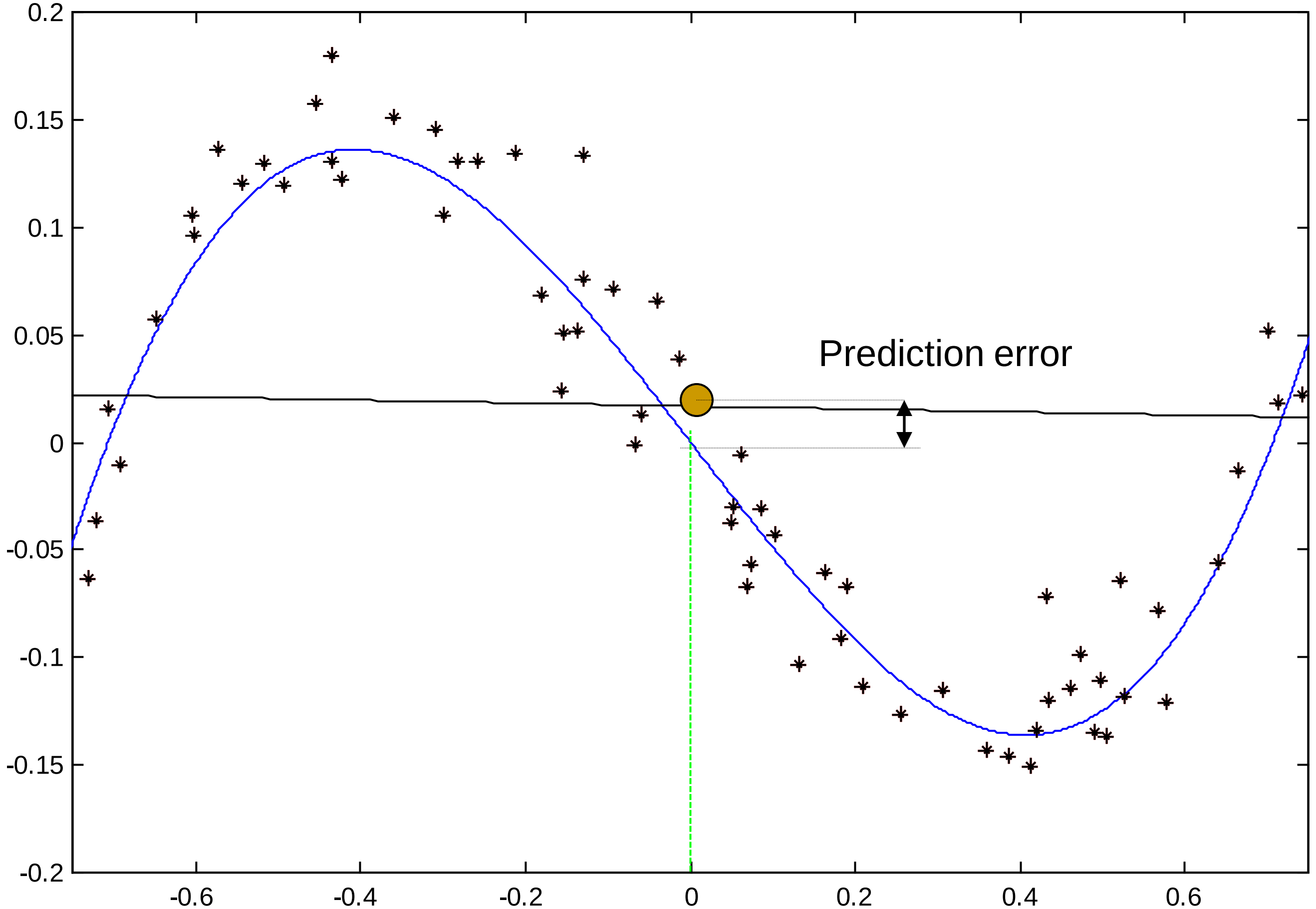
The work focused on the bandwidth selection problem.

Bias/variance trade-off: overfitting



too few neighbors ⇒ overfitting ⇒ large prediction error

Bias/variance trade off: underfitting

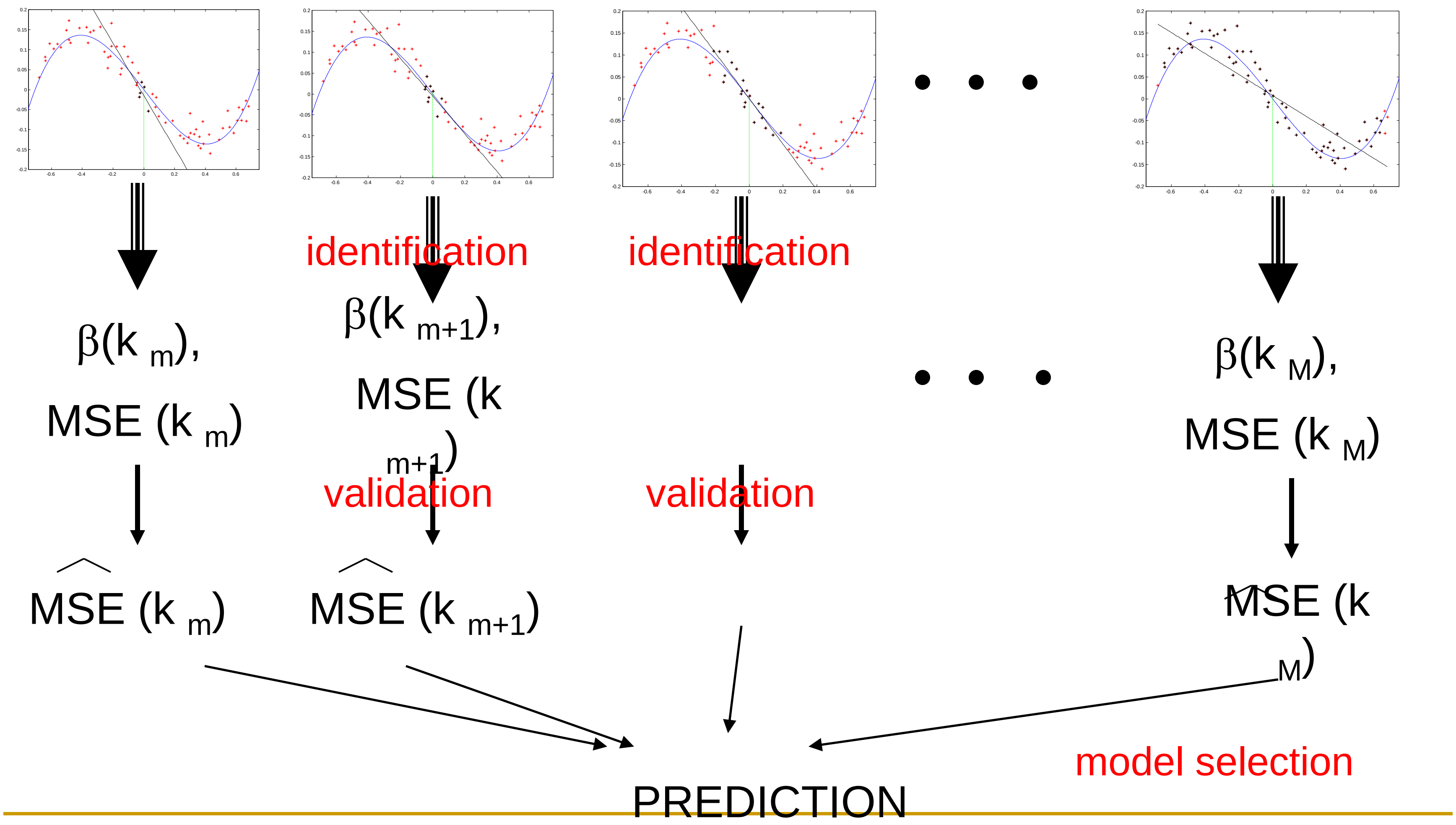


too many neighbors ⇒ underfitting ⇒ large prediction error

Validation croisée: Press

- Fait un leave-one-out sans le faire pour les modèles linéaires
- Un gain computationnel énorme
- Rend possible une des validations croisées les plus puissantes à un prix computationel infime.

Data-driven bandwidth selection



Advantages

- No assumption of uniformity
- Justified in real life
- Adaptive
- Simple

From local learning to Lazy Learning (LL)

- By speeding up the local learning procedure, we can delay the learning procedure to the moment when a prediction in a query point is required (query-by-query learning).

- This method is called lazy since the whole learning procedure is deferred until a prediction is required.

- Example of non lazy methods (eager) are neural networks where learning is performed in advance, the fitted model is stored and data are discarded.

Static benchmarks

- **Datasets**: 15 real and 8 artificial datasets from the ML repository.
- **Methods**: Lazy Learning, Local modeling, Feed Forward Neural Networks, Mixtures of Experts, Neuro Fuzzy, Regression Trees (Cubist).
- **Experimental methodology**: 10-fold cross-validation.
- **Results**: Mean absolute error, relative error, paired t-test.

Observed data

Dataset	No. examples	No. inputs
Housing	330	8
Cpu	506	13
Prices	209	6
Mpg	159	16
Servo	392	7
Ozone	167	8
Bodyfat	252	13
Pool	253	3
Energy	2444	5
Breast	699	9
Abalone	4177	10
Sonar	208	60
Bupa	345	6
Iono	351	34
Pima	768	8

Artificial data

Dataset	No. examples	No. inputs
Kin_8nh	8192	8
Kin_8fm	8192	8
Kin_8nm	8192	8
Kin_32fh	8192	32
Kin_32nh	8192	32
Kin_32fm	8192	32
Kin_32	8192	32

Experimental results: paired comparison (I)

Each method compared with all the others (9*23 =207 comparisons)

Method	No. times significantly worse
LL linear	74
LL constant	96
LL combination	23
Local modeling linear	58
Local modeling constant	81
Cubist	40
Feed Forward NN	53
Mixtures of experts	80
Local Model Network (fuzzy)	132
Local Model Network (k-mean)	145

The lower, the better !!

Experimental results: paired comparison (II)

Each method compared with all the others (9*23 = 207 comparisons)

Method	No. times significantly better
LL linear	80
LL constant	59
LL combination	129
Local modeling linear	89
Local modeling constant	74
Cubist	110
Feed Forward NN	116
Mixtures of experts	72
Local Model Network (fuzzy)	32
Local Model Network (k-mean)	21

The larger, the better !!

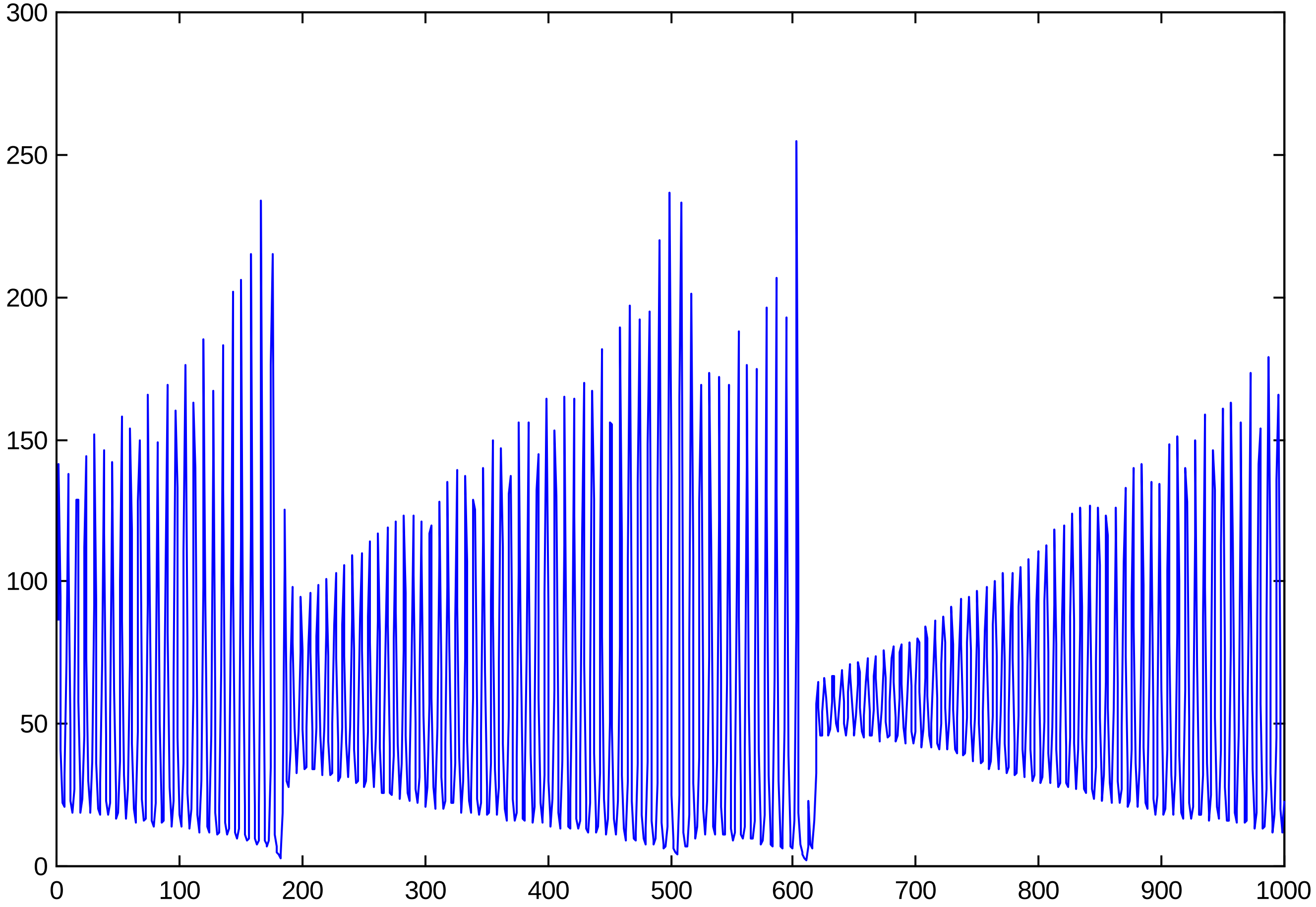
Lazy Learning for dynamic tasks

- long horizon forecasting based on the iteration of a LL one-step-ahead predictor.
- Nonlinear control
  - Lazy Learning inverse/forward control.
  - Lazy Learning self-tuning control.
  - Lazy Learning optimal control.

Dynamic benchmarks

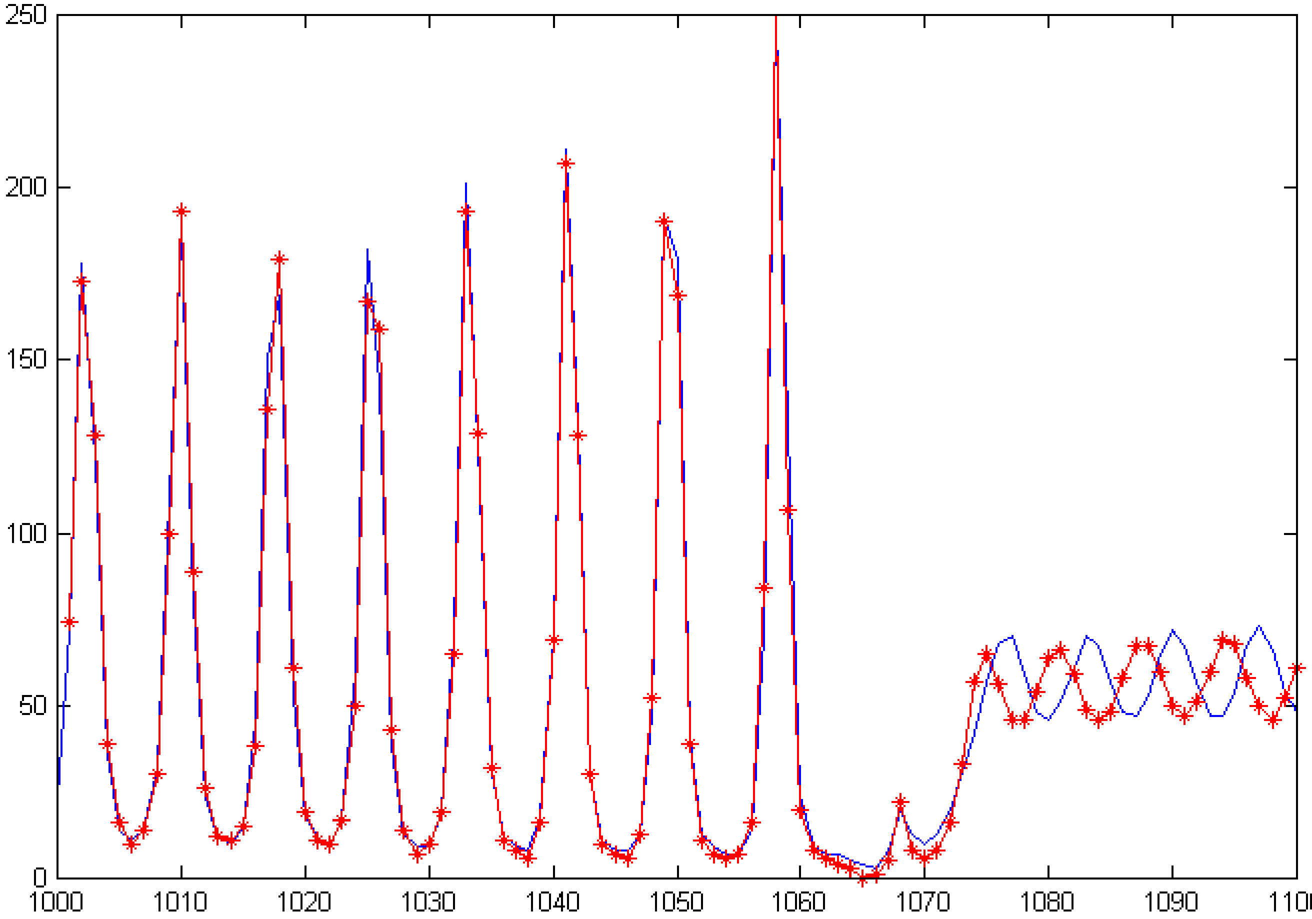
- Multi-step-ahead prediction:
  - Benchmarks: Mackey Glass and 2 Santa Fe time series
  - Referential methods: recurrent neural networks.
- Nonlinear identification and adaptive control:
  - Benchmarks: Narendra nonlinear plants and bioreactor.
  - Referential methods: neuro-fuzzy controller, neural controller, linear controller.

Santa Fe time series



Task: predict the continuation of the series for the next 100 steps.

Lazy Learning prediction



LL is able to predict the abrupt change around t =1060 !

Awards in international competitions

- Data analysis competition: awarded as a runner-up among 21 participants at the 1999 *CoIL International Competition* on *Protecting rivers and streams by monitoring chemical concentrations and algae communities*.
- Time series competition: ranked second among 17 participants to the *International Competition on Time Series* organized by the *International Workshop on Advanced Black-box techniques for nonlinear modeling* in Leuven, Belgium