# What is the place of Machine Learning between Pattern Recognition and Optimization?

Laurent Miclet[1] Antoine Cornuéjols[2]

[1]IRISA-ENSSAT
Lannion (France)

[2]AgroParisTech
Paris (France)

TML Workshop 2008, St Etienne

# Introduction : a paradox in Machine Learning

## Paradox, version 1

- Learning is an matter of highly abstract intelligence
- A learning machine has to be artificially intelligent
- Machines that learn merely run programs designed by intelligent humans (statisticians, optimisers)

## Paradox, version 2

- We teach Machine Learning to Computer Science students
- Computer Science is the science of algorithms (and Computer Science students dislike math)
- Machine Learning is presently a science of equations

# Pattern Recognition, Artificial Intelligence and Theories of Learning : the early times (70's)

The landmark teaching textbooks.

## Pattern Recognition

1. Minsky, M. and Papert, S. *Perceptrons : An introduction to Computational Geometry* (1962).
2. Duda, R., and Hart, R. (1973). *Pattern recognition and scene analysis.*
3. Fu, K. S. (1974). *Syntactic methods in pattern recognition.*
4. Fukunaga, K. (1972). *Introduction to statistical pattern recognition.*

These books were filled with Machine Learning, but the word was not employed.

# The early times

Some classical textbooks.

## Artificial Intelligence

1. Nilsson, N. (1971). *Problem solving methods in artificial intelligence.*
2. Winston, P. H. (1977). *Artificial intelligence.*
3. Raphael, B. (1976). *The thinking computer.*

Machine Learning was not yet quite perceived as a subfield of Artificial Intelligence.

# The early times

These were not really textbooks.

## Theories of Learning

1. Vapnik, V. and Chervonenkis, A. (1974). *Theory of pattern recognition* (in russian).
2. Vapnik, V. (1982). *Estimation of dependences based on empirical data.*
3. L. Valiant, D. Angluin, R. Solomonoff, D. Osherson, etc.

Theories of Statistical Learning had no actual connection with PR nor IA.

# More history : the 80's

Symbolic Machine Learning is coming out as a subfield of AI.
There is a strong reference to Cognitive Science.

## Symbolic Machine Learning

1. Kodratoff, Y. (1986). *Leçons d'apprentissage symbolique automatique.*
2. Kodratoff, Y. (1989). *Introduction to machine learning.*
3. Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (Eds.). (1983). *Machine learning : An artificial intelligence approach.*

## ML, not AI

- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and regression trees.*

# Recent history : the 90's

Machine Learning stands by itself.

## ML,ML,ML !

- Weiss, S., and Kulikowski, C. (1991). *Computer systems that learn.*
- Hutchinson, A. (1994). *Algorithmic learning.*
- Langley, P. (1995). *Elements of machine learning.*
- Mitchell, T. (1997). *Machine learning.*

# How did we teach ML in the 90's ?

## What I wished my students to master (oops ! to have hints on)

- Introduction to natural and artificial learning (Cognitive Science and AI aspects).
- Learning as heuristic search in a ordered discrete space (Version Spaces, Inductive Logic Programming, Grammatical Inference).
- Learning as minimizing some empirical error (Hyperplanes, Neural Networks, Decision Trees).
- Bayesian decision and Clustering.
- Theory and meta-learning (Bias-variance, regularization, Cross-validation, boosting, *pac* learning).

Genetic Learning and Reinforcement Learning could also be included.

# The Present : Machine Learning is basically equations

What do we (and A. Ng) teach nowadays ?

- Introduction (1 class) Basic concepts.
- Supervised learning. (6 classes) Supervised learning setup. LMS. Logistic regression. Perceptron. Exponential family. Generative learning algorithms. Gaussian discriminant analysis. Naive Bayes. Support vector machines. Model selection and feature selection. Ensemble methods : Bagging, boosting, ECOC.
- Learning theory. (3 classes) Bias/variance tradeoff. Union and Chernoff/Hoeffding bounds. VC dimension. Worst case (online) learning. Advice on using learning algorithms.
- Unsupervised learning. (5 classes) Clustering. Kmeans. EM. Mixture of Gaussians. Factor analysis. PCA. MDS. pPCA. Independent components analysis (ICA).
- Reinforcement learning and control. (4 classes) MDPs. Bellman equations. Value iteration. Policy iteration. Linear quadratic regulation (LQR). LQG. Q-learning. Value function approximation. Policy search. Reinforce. POMDPs.

# Present textbooks

## Do I have to advertize for these books ?

- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning.*
- Cornuéjols, A., Miclet, L. (2002). *Apprentissage artificiel : concepts et algorithmes.*
- Alpaydin, E. (2004). *Introduction to machine learning.*
- Bishop, C. (2006). *Pattern recognition and machine learning.*

**C. Bishop's warning** *:*
*"...a good grasp of calculus, linear algebra and probability theory is essential for a clear understanding of modern PR and ML techniques."*

# Present students

## To whom do I teach ?

- No French student in Computer Science can feel comfortable with C. Bishop's book.
- Have you tried yourself to solve the exercises ?

## Good news on my mail, August 2007 :

*I am pleased to announce that the complete set of solutions for all "www" exercises in "Pattern Recognition and Machine Learning" is now available for free download as a PDF file :*

`http ://research.microsoft.com/∼cmbishop/PRML/prml-web-sol-2007-08-03.pdf`

# Present students

## To whom do I teach ?

- No French student in Computer Science can feel comfortable with C. Bishop's book.
- Have you tried yourself to solve the exercises ?

## Good news on my mail, August 2007 :

*I am pleased to announce that the complete set of solutions for all "www" exercises in "Pattern Recognition and Machine Learning" is now available for free download as a PDF file :*

`http ://research.microsoft.com/∼cmbishop/PRML/prml-web-sol-2007-08-03.pdf`

# To sum up

## Is ML still a Computer Science matter ?

- Computer Science (or *Informatique*) is seen by French students as the math-free science
- The math level of ML has raised (SVM, boosting, EM, reinforcement, etc.)

## ML between Pattern Recognition and Optimization

- Structural Pattern Recognition and Symbolic ML (the Computer Science side of ML) are less efficient on large data bases
- Statistical Pattern Recognition has created a lot of new concepts and achieved wonderful results
- Optimization is the spine of modern ML

# To sum up

## Is ML still a Computer Science matter ?

- Computer Science (or *Informatique*) is seen by French students as the math-free science
- The math level of ML has raised (SVM, boosting, EM, reinforcement, etc.)

## ML between Pattern Recognition and Optimization

- Structural Pattern Recognition and Symbolic ML (the Computer Science side of ML) are less efficient on large data bases
- Statistical Pattern Recognition has created a lot of new concepts and achieved wonderful results
- Optimization is the spine of modern ML

# Is Machine Learning a subfield of Statistics and Optimization ?

- Many courses in Statistics include sections on methods of classification, incorporating SVMs, boosting, decision trees and clustering.
- Recent advances in ML methods, such as SVMs or Neural Networks, are optimization methods.

How much room will Machine Learning occupy in the future between statistics and optimization ?

# Scenario 1 : back to algorithms and to computer science

## New trends in computer science

- Upcoming of "life-long" systems. Shift away to "on-line" learning
- Embedding in distributed appliances with limited resources.

## Consequence for ML

- Replace the (fictious) IID data by time-dependent data and data-streams.
- Emerging field of "autonomous computing".

Statistics could cease to play a central role in Machine Learning, while optimization would likely continue to be important, but with respect to the algorithmic resources

# Scenario 2 : Cognitive Science comes back

Nowadays, there is hardly any connection left between Artificial Intelligence and Cognitive Science.
Could we then imagine a future in which the following subjects would be central in the teaching of ML ?

- Perception
- Spatial reasoning
- Sensory-motor coordination
- Biological information-processing
- Theory of the development of complex systems
- Language acquisition
- Recent findings about alien forms of life

# Scenario 3 : Applications first

Devote an increased research activity to challenges and new applications.
Machine Learning would become a blend of engineering methodology, computer science and statistics.

- Datawarehouses
- Sampling for very large databases
- Techniques for dimensionality reduction
- Heuristics for pre-processing and post-processing...

Thanks for your attention !