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

Antoine Cornuéjols antoine.cornuejols@agroparistech.fr
Laurent Miclet miclet@enssat.fr

Keywords: teaching Pattern Recognition, Machine Learning, and Optimization

Abstract

The authors of this paper were co-authors of a book written in French in 2002, which expounded their view of Machine Learning as a specific field. Even though it appears that in recent years there is an increasing overlap between Machine Learning curricula and those of Statistics (Pattern Recognition) and Optimization, we still believe that Machine Learning has retained its own personality and is not doomed to dissolve entirely into the field of mathematics. In this paper, we will broaden the scope of our investigation by looking at the teaching of Machine Learning over three decades. We also invite the reader to contemplate possible futures by suggesting three extreme scenarii.

1. Introduction: who's talking of what?

We have been researchers in Machine Learning (ML) and Pattern Recognition (PR) for mmh! years, and have taught ML extensively, as well as other related areas: PR, Artificial Intelligence (AI), Speech Recognition, Optimization. Our students are generally at Masters degree level or final year (5th year) of "Ecoles d'Ingénieurs".

In the first section, we will give a brief overview as to how what we nowadays call Machine Learning was taught in France over the last decades. This is not exactly the same story as that of research in ML. Teaching is mainly based on textbooks, and produced new textbooks: this is why we will primarily refer to books, not to seminal papers, history of new concepts, nor scientific or practical results.

In the second section, we will examine the background required for a student to acquire a knowledge of Machine Learning today, and ponder on whether this background is still appropriate for a curriculum in Computer Science.

Finally, we will describe three future scenarii for Machine Learning, bearing in mind that "Prediction is very difficult, especially about the future" to quote N. Bohr.

2. The past was hectic

2.1. A brief account of the French story

2.1.1. The 70's: Teaching Machine Learning without realizing it

In the 70's, teaching what we call nowadays "Machine Learning" was mainly a matter of teaching Pattern Recognition courses. The book *Pattern Recognition and Scene Analysis* (Duda & Hart, 1973) was used extensively, as well as the *Introduction to Statistical Pattern Recognition* (Fukunaga, 1972). Some of us began to get interested in new methods, for example, *Syntactic* or *Structural Pattern Recognition* (Fu, 1974). All of these books were filled with information on what we now call Machine Learning, but the word "learning" was scarcely employed in the texts.

In those ancient times, Machine Learning was not perceived as a sub-field of Artificial Intelligence. AI textbooks in the seventies (Nilsson, 1971; Raphael, 1976), on which most teachers relied, hardly mentioned learning. An interesting exception was (Winston, 1977), which had a few sections on the subject (mainly: analogy and learning the "arch" concept).

2.1.2. The 80's: Teaching Symbolic Machine Learning or Pattern Recognition.

We believe that Machine Learning as a sub-field of AI *per se* started to be taught in France in the 80's, whereas PR continued to be taught as before, albeit including new chapters on Neural Networks (NN) and Hidden Markov Models (HMM). As far as we know, the first textbook totally devoted to Machine Learning is French, (Kodratoff, 1986) and deliberately covers only Symbolic Learning. It was soon to be translated into English (Kodratoff, 1989). Its position stated clearly that ML is part of AI and that AI is based on logic.

Simultaneously, AI textbooks began to include some Learning[1], but it was not before the 90's that every AI book would have a section which included symbolic as well as numerical learning techniques. However, whether in the domain of AI or on the theoretical aspects of ML, research was blossoming. The *Machine Learning Journal* was created in 1986.

During the decade, more and more books on Pattern Recognition appeared, including chapters on the increasingly important fields of NN and sometimes on HMM (the latter was at that period conquering Speech Recognition). But they still considered ML as a subfield of AI.

The ML researcher's first bible appeared around this time (Michalski et al., 1983). It was soon to be followed by two other volumes (the first and the second would be translated into French in 1993) and the comprehensive compilation *Readings in Machine Learning* (Schavlik & Dietterich, 1990), with enlightening texts introducing the chapters.

Another important event was the publication of a book on Decision and Regression Trees, familiarly called CART (Breiman et al., 1984). This was to become a landmark in rule learning for binary and nominal data, giving the reader some concepts in statistical learning theory: bias-variance dilemma, validation set, error estimation, cross-validation, etc. We believe that it was a revelation of the statistical side of machine learning for many an AI specialist.

2.1.3. The 90's: Teaching the entire Machine Learning field

Time had come for ML to stand alone. Almost invisible in AI as well as in PR textbooks, ML had nevertheless enough unity and material to become an independent teaching field. And so it did. Depending on the background and the interests of the professor, the focus of the courses would vary, but a typical Masters course would include

1. Introduction to natural and artificial learning (Cognitive Science and AI aspects).

2. Learning as heuristic search in a ordered discrete space (Version Spaces, Inductive Logic Programming, Grammatical Inference).

3. Learning as minimizing some empirical error (Hyperplanes, Neural Networks, Decision Trees).

4. Bayesian decision and Clustering.

5. Theory and meta-learning (Bias-variance, regularization, Cross-validation, boosting, pac learning).

Issues like Genetic Learning and Reinforcement Learning could also be included.

Textbooks in English appeared during this period, presenting Machine Learning as a field of its own: (Weiss & Kulikowski, 1991; Hutchinson, 1994; Langley, 1995; Mitchell, 1997).

We can also note that he first books on Computational Learning Theory appear at the turn of the nineties (Kearns, 1990; Anthony & Biggs, 1992), with a remarkable exception: the first book by Vapnik had been published in 1982.

2.2. One side of the American story

On the two sides of the Atlantic, university departments were organized differently, and subsequently, this affected the curricula. In many universities, the AI departments in particular were associated with, or even incorporated into, the Cognitive Science Department. As a result, at least in the 80's, it was traditional, if not compulsory, that a considerable share of course time for a Masters or Ph.D degree in Artificial Intelligence, was in the domain of Cognitive Science. This would often include courses on experimental methods in cognitive psychology (e.g. protocols to measure the subject's activities), on reasoning and on memory. "Learning" *per se* was often studied in Memory courses only. Overall, there were strong links between Artificial Intelligence and the Cognitive Science fields, where ideas, concepts and empirical studies circulated freely between the different fields.

As far as we know, these links tended to slacken during the 90's, following the trends of the nascent "Machine Learning" field towards more mathematics and statistics and more application-oriented focus.

[1] When, in 1981, Avron Barr, Edward Feigenbaum, and Paul Cohen published their landmark mutivolume *The Handbook of Artificial Intelligence*, (Barr et al., 1981), the main sections were devoted to natural-language processing, cognitive models/logic, planning and problem solving, vision (in robotics), and machine learning; plus core methods of search and knowledge representation.

3. A Tenuous Present

3.1. To whom did we teach in the past?

A typical Machine Learning course around 1990 was designed for students in Computer Science (the AI side of ML) or Electrical Engineering (the PR side of ML) and required classical mathematical prerequisite knowledge from the student: logics (ILP), automata and grammars (structural PR), basic probability and statistics theory (bayesian decision), notions of linear algebra and optimization (Hyperplanes, NN). This was more or less included in the core curriculum for a Masters student in Computer Science, and was merely routine for a student in *Ecole d'Ingénieurs*, because of the *Classes Préparatoires*. HMM learning, based on the EM algorithm, was maybe a bit harder to swallow. But, as a computer scientist, the student would appreciate the challenge of some fancy algorithms: Divide and Conquer (Decision Trees), Dynamic Programming (HMM), Backtracking and semi-greedy heuristics (ILP, Grammatical inference).

Now, let's consider the equivalent ML course in a Masters programme in Computer Science in 2008. The picture has been transformed by two phenomena:

1. Student math background is weaker
2. Machine Learning tools are harder to understand.

This weaker math background may be specific to France, and it is based on easily identified factors: the math level required for the French Baccalauréat is decreasing as Computer Science enlarges its scope, with the result that the student has less time for mathematical problems. Anyway . . .

Secondly, a state-of-the-art Machine Learning course has now to spend most of its time on issues such as boosting, kernel methods, Support Vectors Machines, graphical bayesian models, and indeed has to explain how and why they work, and how they connect.

3.2. What do we teach now?

The development of powerful tools in Machine Learning research has been a gradual process; an interesting experiment is a visit to the ICML conferences website (ICML, 2008) where we may observe this current transformation. Nowadays, a typical ICML or ECML conference is basically devoted to Optimisation (and secondarily to Statistics), and their French competitor CAp (Conférence d'Apprentissage) could be called COp , without shocking an expert in Applied Mathematics.

For current teaching practices, the website of V. Honavar[2] gives an index of ML courses in English around the world, and we believe that the following program of the course by Andrew Ng at Stanford is quite representative:

- 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. K-means. 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.

Concerning the teaching material, some important recent textbooks (Hastie et al., 2001) and (Bishop, 2006) (the latter is now frequently given as the reference of the courses), are explicitly presented as based on statistics and optimization[3]. Bishop warns the student in Computer Science:

> (. . .) a good grasp of calculus, linear algebra and probability theory is essential for a clear understanding of modern PR and ML techniques.

Indeed, from Bishop's point of view, ML is back to its roots: Statistical Pattern Recognition. The addenda of his book, designed to produce self-contained material, are also very interesting to list : Data Sets, Probability Distributions, Properties of Matrices, Calculus of Variations and Lagrange Multipliers.

[2] http://www.cs.iastate.edu/ honavar/ Courses/cs673/machine-learning-courses.html

[3] The book (Alpaydin, 2004), at the contrary, still aims at presenting the field as a whole.

3.3. To whom should we teach ?

It is doubtful whether a French Computer Science student, even at the Masters level, has attained such a background in math, not to mention his (or her[4]) tastes. Computer Science, in French *Informatique*, is often perceived by our students as an enjoyable example of a science with no math.

As a good example, take the exercises of any chapter of Bishop' textbook and wonder what is the probability β that a proportion greater than α can be solved by a Master level Computer Science student in your country. In France, we believe that the answer given in Figure 1 is not far from the truth.

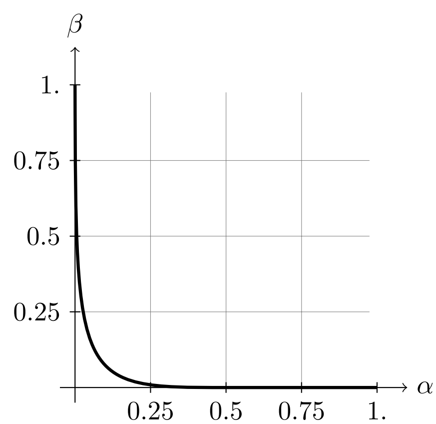

Figure 1. The estimated probability β that a proportion greater than α of the exercises of the book (Bishop, 2006) can be solved by a Master of Computer Science student. This estimation has been obtained from no data at all, thanks to a method which is not described in this book.

We all know that a computer scientist (or: *informaticien*) is fond of algorithms. One of us heard once C. de la Higuera say that ML (and especially Grammatical Inference) was the last interesting field in AI, because it still contained algorithms. But that was in the early nineties. These days have gone: algorithms have disappeared (see section 4.1), we need to teach equations.

And equations have become harder to swallow. It might be that the way out of this dilemma is to put more emphasis on hands on experiments with algorithms and on methodological aspects. The idea is that the students learn less about the algorithms and their mathematical foundations, and more about the proper way to organize the data mining process and to evaluate the results obtained. Indeed, these good practices are general and should easily transfer from one context to another. Or maybe we should chose students outside of the computer science departments.

[4]Very rarely.

4. The future is still unclear

We feel that nowadays it is common to have trouble defining what exactly is the difference between the field of statistics and that of Machine Learning. Indeed, the terminology and the culture may seem strangely alien at times, but, at the core, the commonalties seem prevalent. Many courses in Statistics include sections on methods of classification, incorporating SVMs, boosting, decision trees and so on, and on methods of what we call unsupervised learning (i.e. clustering). At the same time, recent advances in learning methods, such as SVMs or Neural Networks, seem to result from a new understanding of optimization methods. How much room will Machine Learning occupy in the future between statistics (Pattern Recognition) and optimization? And, consequently, on which subjects and fundamental methods should our teaching be based?

There is an interesting area for speculation here. For the sake of argument and in order to stimulate discussions, we have chosen to present the following alternative scenarios.

4.1. Scenario: "a return to algorithms and to computer science"

The gigantic commotion following the publication of the celebrated paper by Valiant in 1984 (Valiant, 1984) was that the notion of algorithm had all but disappeared in the new PAC framework. Up to this point, AI and ML papers presented algorithmic methods to solve a problem and/or new ways to represent knowledge and to reason over it... Suddenly, the representation was restricted to some weak logical expressions (e.g. k-CNF) and soon its role was replaced by the version space's characteristics. At the same time, the learning algorithm was reduced to a fictive entity able to compute the version space at all times, and to pick the best hypothesis if necessary. The computational restrictions of the framework that was part of the original papers were soon forgotten. In a way, the statistical theory of learning that followed was just a descendant or, rather, an extension, of this stance, even if the SVM in particular have come out of it. We are still very much in the situation where learning could be seen as some applied part of Lebesgue' theory of measure and that of his colleagues.

There are, however, reasons to believe that new concerns could alter this course and demand new fundamental studies, therefore, changing the Machine Learning courses of the future. The reasons are twofold. First, the artificial learners of the future will increasingly be part of "life-long" systems. Second, they will be embedded in many diverse and distributed appliances with limited resources both in terms of computational power and of communication bandwidth. The consequences of the first trend are that the paradigm will shift away from "one-shot" learning to "one-line" learning and from the (mostly fictitious) independently and identically distributed data sets to time-dependent data and data-streams. The second trend will force us to reconsider the importance of computational resources and of reasoning about the allocation of limited resources with respect to a task to be performed in isolation or in a team, a field emerging as "autonomous computing".

In this scenario, statistics should cease to play a central role in the development and understanding of Machine Learning, while optimization would likely continue to be important; but with respect to the algorithmic resources and much less with respect to the exploration of a continuous search space as it is now. What then should be the new subject matters of a curriculum in Machine Learning?

- Theory of information
- Time series
- Bounded rationality
- Algorithms "on the flight" for Datastreams
- Markov decision processes
- Random walks and path integrals
- Non standard computing
- ...

The list is obviously open to all suggestions.

4.2. Scenario: "Cognitive science comes back"

Nowadays, there is hardly any connection left between Artificial Intelligence and Cognitive Science. However, Cognitive Science is making progress, harvesting a whole lot of new facts about the functioning of the brain, and examining new ideas about situated cognition, educational science and psychological development in general. Right now, Machine Learning has little to say about these facts and issues[5], but, fundamentally, it cannot ignore these forever. Even if the study of the brain could be completely circumvented by Machine Learning as an unnecessary source of parables and analogies, educational sciences and developmental psychology should have an impact on Machine Learning, if only because future machines will be part of our everyday life.

Could we then imagine a future in which the following subjects would be central in the teaching of machine learning?

- 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
- ...

4.3. Scenario: "Challenges and applications are on the way out"

In view of the difficulty to identify a core of techniques and concepts proper to Machine Learning in contrast to other related fields, one temptation might be to devote increased research activity to challenges and new applications in order to show the practical value of Machine Learning. Machine Learning would then become a blend of engineering methodology, computer science and statistics. Supposing that Machine Learning would continue to exist as an independent field, the following courses could be taught.

- Datawarehouses
- Sampling for very large databases
- Techniques for dimensionality reduction
- Heuristics for pre-processing and post-processing data
- Applicability of quantum computing
- ...

[5] Except maybe in the case of Vision. For instance, sensory coding theory, independent component analysis and other dimension reduction techniques seem to shade light on fundamental laws and, in turn, may be inspired by the understanding of the brain.

5. Conclusion

During the seventies and eighties, times that seem almost prehistoric to our students, cognitive scientists made daily incursions into the field of artificial intelligence and machine learning by ways of concepts such as memory types (procedural vs. declarative, or short-term vs. long-term), semantic networks, and so on. They felt at home giving conferences and seminars about AI and ML, and exchanges were symmetrical. Now, it often feels as if statisticians and specialists of optimization behave like slightly annoying guests who openly give their opinion on our furniture and our collectors items. However, it is quite possible that, in twenty years time, this might seem as outdated as the complicity with cognitive science appears now. These strong but varying partnerships are undoubtedly a testimony of the appeal and central position of the fundamental interrogations behind machine learning rather than a mark of weakness. These questions touch the nature of memory and the nature of information, the condition for transfer from one experience to new situations, the (self-)organization and use of multiple types of knowledge, and many more. We have every reason to expect that many generations of researchers will work on these problems, and that Machine Learning curricula will profoundly evolve over time.

For the near future, it is likely that none of the scenarii outlined above will happen as such. What we have learned in Machine Learning is that (probably approximately) predicting is a very risky business which requires the best specialists. This is why you are encouraged to order now the third edition[6] of a famous book on Machine Learning first published in 2002 (Cornuéjols & Miclet, 2002). The authors do not know yet what will be inside, but guarantee that you will find there the answer to that nagging question: "What is the place of Machine Learning between Pattern Recognition and Optimization?".

[6] To come approximately in 2013, or 2014 or 2015.

References

Alpaydin, E. (2004). *Introduction to machine learning*. MIT Press.

Anthony, M., & Biggs, N. (1992). *Computational learning theory*. Cambridge University Press.

Barr, A., Cohen, P., & Feigenbaum, E. (1981). *The handbook of artificial intelligence*. Addison-Wesley. New edition: 1990.

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.

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

Cornuéjols, A., & Miclet, L. (2002). *Apprentissage artificiel : concepts et algorithmes*. Eyrolles. Seconde édition : 2008 (à paraître).

Duda, R., & Hart, R. (1973). *Pattern recognition and scene analysis*. Wiley.

Fu, K. S. (1974). *Syntactic methods in pattern recognition*. Academic Press.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. Academic Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. Springer.

Hutchinson, A. (1994). *Algorithmic learning*. Clarendon Press.

ICML (2008). Hyperlink "http://www.icml2006.org/icml2006/" http://www.icml2006.org/icml2006/ other/pastconferences.html.

Kearns, M. (1990). *The computational complexity of machine learning*. MIT Press.

Kodratoff, Y. (1986). *Leçons d'apprentissage symbolique automatique*. Cepaduès.

Kodratoff, Y. (1989). *Introduction to machine learning*. Morgan Kaufmann.

Langley, P. (1995). *Elements of machine learning*. Morgan Kaufmann.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (1983). *Machine learning: An artificial intelligence approach*. Tioga Publishing Company.

Mitchell, T. (1997). *Machine learning*. McGraw Hill.

Nilsson, N. (1971). *Problem solving methods in artificial intelligence*. McGraw-Hill.

Raphael, B. (1976). *The thinking computer*. Freeman.

Schavlik, R., & Dietterich, T. (Eds.). (1990). *Readings in machine learning*. Morgan Kaufmann.

Valiant, L. (1984). A theory of the learnable. *Communications of the ACM*, *27(11)*, 1134–1142.

Weiss, S., & Kulikowski, C. (1991). *Computer systems that learn*. Morgan Kaufmann.

Winston, P. H. (1977). *Artificial intelligence*. Reading.