Review of *Machine Learning*, written by A. Abu-Hanna.
Appeared in *Artificial Intelligence in Medicine*, Elsevier, vol. 16, 1999, pp. 201–204.

Tom M. Mitchell
Machine Learning
McGraw-Hill, 1997
ISBN 0-07-115467-1
Paperback, 414 pages
Price £21.99

An old dream of Machine Learning has been to teach computers how to solve problems by merely 'showing' them selected instances of their solutions. This dream reflects one issue in Machine Learning, a field which has become very popular within Artificial Intelligence (AI) research and its applications.

In general, Machine learning is concerned with the question of how to automatically improve the performance of some task (e.g. playing chess or diagnosing patient diseases) based on experience. The aim of the learner is to improve, and initially build up, its problem solving capability in carrying out the task based on the training experience and some performance measure such as the number of games won or of correct diagnoses of (new) patients. Conceptually, the learner is building an improved body of knowledge of how to perform the task. In practice this knowledge can be represented explicitly, e.g. as declarative rules, or implicitly, e.g. as variable weights in a formula.

There is a great variety in learning algorithms in terms of their knowledge representation formalisms, the orientation of inductive versus analytic reasoning, the (un)supervised nature of induction, and the interaction with the environment (control and feedback). There is also a wide variety of disciplines that influence the field as a whole. Besides AI, with topics such as representation and search, some other influential disciplines are computer science (algorithms and their complexity), statistics (evaluating hypotheses), cognition (learning in humans and animals), philosophy of science (scientific discovery), and neurobiology (neural models).

In Medicine, Machine Learning can play an important role in analysing medical databases, often cast as classification problems, to support decision making. For example, models can be built that help explain drug interactions or surgery complications, identify risk groups, select probable diagnoses or appropriate treatments, and provide prognostic information.

In this light, Machine Learning approaches complement traditional decision analytic and statistical approaches and enrich them with a wide range of (explicit) representations.

The idea behind the book 'Machine Learning' is to provide a single self-contained introduction to the main approaches to Machine Learning. It is intended for graduate and advanced undergraduate students, and researchers and developpers in disciplines such as computer science, engineering, statistics, and the social sciences. The book aims at a balanced presentation of theory and practice and is meant to make only few assumptions about the background of the reader by attempting to introduce basic relevant concepts from disciplines such as statistics, artificial intelligence, and information theory as the need arises. These are challenging requirements but in my opinion the author has succeeded in providing an excellent book on the subject that meets its objectives.

The book consists of 13 chapters. Thematically, it consists of an overview of Machine Learning; a framework of concept learning; inductive learning based on decision trees, neural networks, Bayesian learning, genetic algorithms, and sets of declarative rules; computational learning theory; instance-based learning; analytical learning and its combination with inductive learning; and reinforcement learning. Below are impressions about the book's chapters.

The first chapter puts the subject of Machine Learning in perspective and illustrates formulations and design issues of Machine Learning problems. Perhaps the author could offer a more explicit typology of inductive versus analytic approaches and, within induction, supervised and unsupervised approaches. The second chapter provides a framework for concept learning and the organisation of the hypothesis space. Not surprisingly, Mitchell's own work on version spaces is used as a conceptual framework to understand search in a generalization-specialization hypothesis space. This is perhaps the most essential chapter in the book in order to understand, conceptually, the other chapters. Goodies in this chapter include the intuitive explanation of version spaces and the answer to the question why unbiased learners cannot generalize.

Chapter 3 deals with decision trees where issues of bias, preference of concise hypotheses, overfitting, and entropy are crisply explained. The interpretation of entropy as a specification of minimum number of bits could have been elaborated more. Chapter 4 on neural networks, one of the largest chapters, provides a good explanation, albeit somewhat more demanding, of the backpropagation algorithm. The chapter extensively covers issues in neural networks and includes a great illuminating example of face recognition.

Chapter 5 is an introduction to statistical methods used in the evaluation of hypotheses. It is a necessary chapter which includes the most important relevant results from statistics. My guess is that this chapter will not constitute a substitute for a book on statistics to novices on the subject. Chapter 6, Bayesian learning, is very valuable not only for its well structured build-up and its introduction to a probabilistic approach to inference but also because it illuminates the material that has been covered thus far under a Bayesian light. Although Bayesian belief networks received a separate section, there is a feeling that it could have been further elaborated on perhaps at the cost of the somewhat more demanding description of gradient ascent training. The example of classifying text is a truly illustrative one for the practicality of Bayesian learning.

Chapter 7 on computational learning theory is a pleasant surprise (that is, if one expects a difficult chapter). It is a successful attempt to present the essentials of a seemingly difficult subject in an intuitive and understandable manner without sacrificing accuracy. The conceptual tools that the reader acquires in chapter 2 come in quite handy here in understanding the PAC and mistake bound frameworks.

Chapter 8, on instance-based learning is quite accessible and the idea of delaying inference till query time and its tradeoffs comes across clearly. The chapter could however concentrate more on the practical problem of indexing and structuring instances. A treatment and illustration of the kd-tree algorithm should have been incorporated in this chapter.

Chapter 9 on genetic algorithms is also very accessible and provides clear explanations of the motivation behind this approach and many illustrations of the basic concepts. The discussion of Holland's schema theorem could perhaps be better accompanied by an illustration.

The induction of rule sets is the subject of chapter 10 which introduces various inductive algorithms including those of Horn clauses. The chapter is very readable and builds on earlier concepts. A good appreciation of the material, though, would require some background in formal logic and logic programming on the part of the reader.

Chapters 11 and 12 treat analytical learning and its combination with inductive learning. These are excellently written chapters with insightful comments on the interrelationship and tradeoff between background knowledge and search. Chapter 12 is read with a certain amount of excitement (and probably written in this spirit). The regression of literals through a

Horn clause in chapter 11 has perhaps gone a bit too fast.

The last chapter, reinforcement learning, deals with situations where the learner gets delayed, indirect rewards to its actions and completes the spectrum of learning approaches discussed in the book. Q learning is well introduced and illustrated and the relationship with dynamic programming is explicitly indicated.

The book is very well-written. It is clear that it is based on a long teaching experience: there are no conceptual gaps in the text, the cohesion of the material within the chapters is high, the difficulty and style of the chapters are roughly constant, and there are valuable cross references between chapters –unlike the situation in many edited books. Perhaps more importantly it is an enjoyable book to read! The algorithms in the book are clear and well-illustrated by accessible examples and real applications and their potential are discussed in a non-dogmatic but convincing manner. There are also on-line data and software on the Internet for the reader, which I only briefly inspected.

The main strengths of the book not only include its crisp handling of representative topics and algorithms in machine learning but also its provision of powerful conceptual frameworks consistently used throughout the book with which the reader can *understand* the various algorithms and approaches and make comparisons between them. Decision trees, neural networks, genetic algorithms and other approaches are all viewed in terms of their search behavior. Moreover the conceptualization of learning as search in a hypothesis space helps understand computational theory results as far as the relationship between the sizes of the hypothesis space and the training sample is concerned. Similarly, the effect of using background knowledge on the search space is emphasized and well explained which is valuable in understanding the tradeoffs between knowing and searching, also when knowledge is incomplete or incorrect. Moreover, the book casts various algorithms that do not necessarily manipulate probabilities explicitly in Bayesian terms to clarify the conditions and assumptions that underlie these algorithms.

In any book about machine learning there are difficult choices to be made, there are just too many topics and disciplines involved in order to cover them extensively. This book strikes a successful balance between the various considerations. Yet, I believe that an important enhancement to the coverage of the subject is the inclusion of a separate chapter on unsupervised learning. The book touches on this topic in the context of the EM algorithm in the chapter on Bayesian learning but, I feel, it is underrepresented. A separate chapter on this topic could be introduced early with a system like

Cobweb as an example.

In conclusion, this is a truly outstanding book on machine learning which will contribute to the consolidation of this dynamic field by crisply presenting the major algorithms and approaches and, perhaps more importantly, by providing conceptual frameworks for understanding them. As textbook it is most suitable for advanced students of artificial intelligence, computer science and related disciplines. It belongs on the shelves of researchers in the (medical) AI community with interest in data analysis and machine learning and is also very useful for people in the datamining community who would like to understand the background behind typical datamining algorithms such as the induction of association rules or foreign keys in relational databases.

Ameen Abu-Hanna
Department of Medical Informatics, University of Amsterdam
Meibergdreef 15
1105 AZ Amsterdam
The Netherlands
E-mail: a.abu-hanna@amc.uva.nl
Tel: +31 20 5665959
Fax: +31 20 6919840