

Book Reviews

Ethem Alpaydin. *Introduction to Machine Learning* (Adaptive Computation and Machine Learning Series). The MIT Press, 2004. ISBN: 0 262 01211 1 Price £32.95 / \$50.00 (hardcover). xxx+415 pages.
doi:10.1017/S1351324906004438

Synopsis

This book provides students, researchers and application developers with an easily comprehensible and comprehensive introduction to the theory of machine learning. After studying this book, the reader should have a grasp of the most important sub-topics of machine learning, and be capable of implementing any of the topics covered either as stand-alone software tools, or as analytical techniques for research.

This book is structured primarily as a textbook. It should be a valuable teaching aid for grad or undergrad courses in machine learning, and an excellent supplement for courses in statistical natural language processing or other ML-intensive disciplines such as astrostatistics. It is an excellent resource for engineers and researchers, and can be used for self-study by dedicated individuals that have some familiarity with AI or machine learning.

All chapters conform to a consistent structure, which enables the reader to effortlessly orient themselves as they progress through the book. Each chapter (1) starts with a motivating introduction which situates the topic in the context of what has been discussed earlier. (2) Then provides principles or simple examples that equip the reader with the concepts required to understand the topic. (3) This is followed by an explanation of how the techniques in this chapter are applied to automatic classification. Typically, more sophisticated variants of this technique are also presented. (4) Algorithms are then presented (not provided in earlier chapters). (5) An explanation of how this technique can be applied to regression follows. (6) Advanced methods of boosting accuracy or enhancing efficiency that would make these techniques more valuable in production systems are given. (7) This is followed by a set of notes that gives a historical account of the development of the technique and valuable references to the most useful resources where the reader can learn more. (8) A set of exercises that are insightful and well thought-out, and go beyond simply requiring the reader to parrot what they have read are also given. These exercises help the reader to consolidate what they have covered and give them the feeling of satisfaction and confidence that comes through applying the principals covered to cases that are ‘just beyond’ the scope of the contents of this chapter. (9) Each chapter then concludes with a set of references mentioned earlier in the text (mostly in the notes sections).

The book is cohesive and easy to follow. The author has tried to be as gentle as possible with the reader, and has taken pains to highlight common

misunderstandings. Instead of taking the approach of many other books, which is to 'show it as it is' and let the reader fend for themselves, Alpaydin has taken the reader by the hand and helped them avoid the common pitfalls that he must have observed in his years of teaching. In the chapter introduction and end-notes sections, the author nicely ties all of the topics together so a deeper understanding is obtained. The choice of graphs and visual aids is also designed to inform rather than impress. The temptation to jazz up the appearance with polychromatic graphs has been resisted, making space for the minimalist viewgraphs that serve well to illustrate the topic under discussion.

The book also employed the successful LaTeX templates of Manning & Schutze (1999), and as important terms are introduced, these are printed in the margin which makes it easy to scan the text for topics of interest. These terms are also listed in the index. A companion web-site provides a complete set of Powerpoint and PDF slides that can be used by instructors, as well as an updated errata list.

The only gripe that I have is that the references available at the end of each chapter are not collected together and presented at the end of the book. Perhaps the author may choose to maintain a current bibliography on their web site instead?

Contents

The book is divided into 16 chapters:

Chapter 1 motivates interest in machine learning and provides an overview in plain English of association rule learning, regression, classification, unsupervised learning and reinforcement learning. This is a very gentle introduction that highlights many useful applications, and matches key concepts to the jargon of the ML field. The reader is also provided with some intuition on the limitations of these techniques. At the end of this chapter, the user should already have enough knowledge of ML to at least talk authoritatively on this topic with laymen.

Chapter 2 covers supervised learning techniques. The text starts from the simplest binary class cases and proceeds to multi-class problems and concludes with a section on regression. The reader is also introduced to the VC dimension, PAC learning and the effect of noise on errors. The chapter concludes with a discussion on the impact of training data and model complexity, and the issues of overfitting and underfitting.

Chapter 3 discusses probability theory as the framework for making decisions under uncertainty. The introduction to Bayesian decision theory is well written and this framework is extended to encompass risk, misclassification cost and utility theory. A good introduction is also given to Bayesian Networks and the Naïve Bayes (NB) classifier. I felt that NB, which is a really important classifier is not given the amount of space that it deserves, as it only briefly mentioned. However, other readers would probably feel the same about their own favorite techniques. The chapter concludes with a discussion of support and confidence, which are critical to methods dealing with association rule extraction (for market basket analysis).

Chapter 4 focuses on parametric methods to estimate probabilities from a given training set. The chapter starts with an overview of the maximum likelihood estimate, and rapidly moves to discuss log-likelihood for 2-class, multi-class and Gaussian

cases. The chapter concludes with an excellent section on the bias/variance dilemma and procedures for optimal model selection.

Chapter 5 extends the univariate methods introduced in the previous chapter to multivariate methods (which have a large number of input variables or features). Important concepts such as parametric estimation, MV classification and regression are covered. Means of tuning the complexity, gracefully handling missing data and other useful advice are also given.

Chapter 6 starts with a compelling case for both feature selection and feature extraction, both of which can help with dimensionality reduction. Next, both forward feature selection and backward selection with all major variants are covered. There is also a large section devoted to principal component analysis and factor analysis. The examples in this chapter make these intricate concepts much easier to understand and visualize. There is also a section on linear discriminant analysis, which I felt should have been better located in chapter 10.

Chapter 7 covers k-means clustering and the expectation maximization algorithm. The discussion in the chapter is quite helpful, as these techniques are easy to employ, but not easy to employ successfully without strategies to deal with inseparable cases, bad start points etc. The chapter concludes with a short section on hierarchical clustering.

Chapter 8 covers non-parametric techniques for machine learning which are useful when we cannot make assumptions about the underlying distribution, or come up with workable parametric models. In AI, these techniques are referred to as ‘case-based-reasoning’ and they have become quite popular recently as there is finally sufficient computational resources available to make these workable.

Chapter 9 covers decision trees, i.e. learning algorithms that build a tree from a given labeled training sample. Methods of converting these trees to a set of simple rules are also covered. Important algorithms such as CART and RIPPER are introduced as well as means of reducing overfitting.

Chapter 10 covers discriminant based approaches that are quite popular in contemporary pattern recognition problems. The author starts with an introduction to linear discriminants, covers logistic discriminants and then concludes the chapter with an introduction to support vector machines. This chapter is not totally self-contained as this topic was first introduced in chapter 4, which covered Fischer’s linear discriminant.

Chapter 11 focuses on artificial neural networks (ANN), and explains the perceptron, multilayer perceptron (MLP), back-propagation algorithm and the importance of the hidden layer.

Chapter 12 dwells on extensions to MLP. The content includes the online version of k-means, adaptive resonance theory, self-organizing maps, the radial basis function network and mixture of experts framework.

Chapter 13 introduces Hidden Markov Models, which are suited for tasks where observations are dependent on previous states. The chapter starts out by introducing discrete Markov processes, and extends these to cover HMMs. The forward-backward, the Viterbi and Baum-Welch algorithms are also explained. The discussion contains an interesting overview of current applications built using HMMs.

Chapter 14 should be of interest to researchers as it provides tools to assess the expected error rate of classifiers, and compare two or more classifiers. Concepts covered include crossvalidation, bootstrapping, interval estimation and hypothesis testing. Different parametric and non-parametric tests are also covered in this chapter.

Chapter 15 shows how multiple learners can be combined to achieve a higher performance. Key topics include voting, bagging, boosting, mixture of experts, stacked generalization and cascading.

Chapter 16 covers reinforcement learning, which is an online technique for ‘learning with a critic’ that provides feedback after the event, as opposed to supervised learning which provided feedback along with the input (during training). These unsupervised learning techniques show a lot of promise and have been developing rapidly recently.

There is also an appendix at the end that covers the core probability and statistics concept that are relevant to this book. This chapter helps to make the books much more cohesive, and self-contained.

Although much of this content has earlier been covered by Mitchell (1997) and Hastie et al. (2003), and less so by Russell and Norvig (2002), this book still stands out. Mitchell (1997) is the closest substitute for this book, and worth keeping in your library. It used to be my favorite recommendation for someone who wanted to learn about ML. However, there have been many advances in the field since ’97, due to which I now recommend Alpaydin’s book instead. Hastie et al. (2003) provides a statistician’s take on machine learning, and is not only relatively current, but also very densely packed with useful information and serves as an excellent reference for almost every important technique or theory in machine learning. On the down side, it is structured poorly, and written in a difficult-to-understand manner, and exposing a beginner to it would only serve to demoralize them. Hastie et al. (2003) best works as a reference as it is an advanced text. Russell & Norvig (2002) covers important aspects of ML as well as many related concepts such as knowledge representation, and different search heuristics. However, that book is almost three times the size of Alpaydin (2004).

The book presupposes ability in programming, probability, calculus, and linear algebra. It is by no means a ‘machine learning for dummies’, as the field itself is quite technical. Also, many of the techniques have been only briefly described in a few pages, whereas entire books may have been written about them elsewhere. Having said that, I can recommend this book to anyone moving into the machine learning area, and looking for a good introductory text. It packs a lot of learning into a parsimonious 415 pages.

References

- Hastie, T., Tibshirani, R. & Friedman, J. H. (2003). *The Elements of Statistical Learning*. Springer. ISBN: 0387952845.
- Manning, C. D. & Shutze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press. ISBN: 0262133601.

Russell, S. J. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall; 2nd edition. ISBN: 0137903952.

Reviewer:

Shahzad Khan
Computer Laboratory
University of Cambridge
shahzad.khan@cl.cam.ac.uk

Manny Rayner, Beth Ann Hockey, Pierrette Bouillon, *Putting Linguistics into Speech Recognition. The Regulus Grammar Compiler*. With contributions by Nikos Chatzichrisafis, Marianne Santaholma & Marianne Starlander. Foreword by Aravind K. Joshi. Appeared in the series CSLI Studies in Computational Linguistics (series edited by Ann Copestake). Stanford, California, CSLI. ISBN 1-57586-526-2 (paperback), xiv+305 pages. Price: US\$25.00, £17.99. doi:10.1017/S1351324906004426

Automated speech recognition systems achieve high performance by using language models to constrain the search space. The most common approaches are data-driven and based on N-grams models, computed on the basis of large corpora. The drawback of this method is the need for a relatively large training corpus; it also faces difficulties in capturing non-local linguistic constraints. The book by Rayner, Hockey and Bouillon introduces and advocates a radically different approach: rule-based language modelling, employing a linguistically motivated grammar to produce accurate models based on context-free grammars. This has some obvious advantages: it is less dependent on large quantities of training data, and can deal with non-local constraints. The book illustrates how to do this successfully with the Regulus grammar compiler (a software package developed by the authors), using a domain-independent grammar for English that can be specialised for particular application domains. But it does more than that: it also demonstrates how to apply the resulting language models in practical work such as spoken dialogue systems and speech translation tools. Finally, the authors argue that grammar-based language models are serious challengers for corpus-based models.

The book comprises thirteen chapters, arranged in two parts. Part I (Chapter 2–7) is the practical, hands-on part of the book—it shows what grammar-based language modelling involves, and how to use it in practical applications. Part II (Chapter 8–12) presents the theoretical background and motivation—it illustrates how the compiler works, how the English grammar is organised, and evaluates the approach.

Chapter 1 is a gentle introduction to speech recognition and the Regulus system. It explains the difference between the corpus-based approach and the grammar-based approach to language modelling. The major shortcoming of the latter is that, even with linguistic expertise, hand-crafted grammars are tedious to code and hard to maintain, because of the restricted formalism (a subset of context-free grammars) that is used by current speech recognition systems such as Nuance. The Regulus grammar compiler does away with this using a principled way of compiling a