

## BOOK REVIEW

### Pattern Recognition and Machine Learning

Christopher M. Bishop, 73 pp., ISBN 0-387-31073-8, Springer, New York (2006), \$74.95 hardcover.

*Reviewed by* Nasser M. Nasrabadi, U.S. Army Research Laboratory, Adelphi, Maryland

This book provides an introduction to the field of pattern recognition and machine learning. It gives an overview of several basic and advanced topics in machine learning theory. The book is definitely valuable to scientists and engineers who are involved in developing machine learning tools applied to signal and image processing applications. This book is also suitable for courses on machine learning and pattern recognition, designed for advanced undergraduates or PhD students. No previous knowledge of machine learning concepts or algorithms is assumed, but readers need some knowledge of calculus and linear algebra. The book is complemented by a great deal of additional supports for instructors and students. The supports include solutions to the exercises in each chapter, the example data sets used throughout the book and the forthcoming companion book that deals with practical and software implementations of the key algorithms. A strong point of this book is that the mathematical expressions or algorithms are usually accompanied with colorful graphs and figures. This definitely helps to communicate the concepts much better to the students or the interested researchers than pure description of the algorithms. The book also provides an interesting short biography of the key scientists and mathematicians who have contributed historically to the basic mathematical concepts and methods in each chapter.

This book consists of 14 chapters covering the basic concepts of the probability theory, classical linear regression, binary discrimination or classification, neural networks, and advanced topics such as kernel methods, Bayesian graphical models, variational inference, Monte Carlo sampling methods, hidden Markov models, and fusion of classifiers. Chapter 1 introduces the basics of machine learning and classical pattern recognition by introducing two examples—recognition of handwritten digits and polynomial curve fitting. Using

these two examples, basic concepts and terms in machine learning literature are reviewed. Machine learning concepts such as Bayes' theorem, overfitting phenomena, model selection, the curse of dimensionality, decision and information theory are introduced. Chapter 2 is dedicated to the exploration of some particular examples of probability density functions and their properties. In this chapter the parametric probability distributions that are reviewed are binomial and multinomial distributions for discrete random variables and the classical Gaussian distribution for continuous random variables. The nonparametric density estimation based on Parzen window or a kernel function is also covered in this chapter. The next two chapters, Chapters 3 and 4, discuss the classical linear regression and binary classification (logistic regression) based on linear models, respectively. Least-squares and Bayesian-based solutions to the parameters of a linear regression model are given in Chapter 3. The idea of regularized least squares to avoid overfitting is also given in Chapter 3. In addition, the concept of bias-variance decomposition is introduced in this chapter in order to provide flexible models that have the best balance between bias and variance. In the case of classification, the linear discriminant functions and the well-known Fisher's linear discriminant are provided in Chapter 4. The concepts of probabilistic generative and discriminative models are also reviewed in this chapter. Chapter 4 concludes with Bayesian treatment of logistic regression and also introduces the concept of Laplace approximation which fits a Gaussian distribution centered at the mode of a given posterior distribution in order to provide a Bayesian inference for the logistic regression. An alternative method to perform classification is to use neural networks motivated by biological systems. Feed-forward neural networks and their training procedures are discussed in Chapter 5. Use of regularization in neural networks to reduce the number of weights as well as Bayesian neural networks are described in this chapter.

Most of the engineers in signal and image processing are more or less familiar with the topics covered in the first four chapters of this book. In the following chapters more advanced concepts in machine learning and statistics are introduced. Chapter 6 introduces kernel functions, the

idea of kernel trick, kernel-based regression using dual representation, and kernel-based modeling of the probability distribution. The role of kernels in probabilistic discriminative models, leading to the framework of Gaussian processes for regression and classification, is also provided in Chapter 6. Sparse kernel machines, also known as maximum margin classifiers, such as support vector machines (SVM) and relevance vector machines (RVM), are reviewed in Chapter 7. The concept of maximizing the margin in the design of SVM is described. However, more recent extensions of SVM such as the least-squares SVM<sup>1</sup> are not discussed. Probabilistic graphical models represented by directed or undirected graph are covered in Chapter 8. Chapter 8 is quite long and an excellent review of Bayesian networks and Markov networks is provided in detail. Polynomial regression is used as an illustration for Bayesian networks. Image denoising is used as an example of Markov networks. At the end of this chapter, inference in graphical models is used as an efficient method for inference. In Chapter 9 the well-known K-means clustering is first reviewed. In addition, the Bayesian treatment of Gaussian mixture models and the use of the expectation-maximization (EM) algorithm to find their parameters is described. Chapters 10 and 11 deal with stochastic or deterministic approximations of posterior distribution used in probabilistic graphical models that were previously covered in Chapter 8. Although, these approximation techniques are well known in the stochastic literature,<sup>2,3</sup> Bishop has been able to give an excellent review of these topics with examples on linear and logistic regressions. Also, the author's personal research experience on variational inference in graphical models is reflected in this chapter. Chapter 12 covers topics such as principal component analysis (PCA), probabilistic PCA, factor analysis, kernel PCA, independent component analysis, and a short summary of the more recent techniques on modeling nonlinear manifolds. An introduction to Markov models and hidden Markov models (HMM), which are well known in signal and image processing literature, is given in Chapter 13. Linear dynamic systems (LDS), such as Kalman filtering and inference/learning in LDS, are also covered in this chapter. Finally, Chapter 14 deals with combining

---

different models or experts. Classical techniques such as committees, boosting, and mixture of experts are reviewed in Chapter 14.

A comparison of this book with other texts in this area is instructive. Almost all the topics covered in this book can be found in more detail in other references. In fact, in this book each chapter cites several excellent references for the source or for more detailed information. The closest books in terms of topics covered by

Bishop are the books by Hastie *et al.*<sup>4</sup> and Mackay.<sup>5</sup>

### References

1. J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, New Jersey (2002).
2. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, Florida (2003).
3. C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Verlag, New York (2004).
4. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Verlag, New York (2002).
5. D. J. C. Mackay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge, UK (2003).

*Nasser M. Nasrabadi is a senior research scientist at the U.S. Army Research Laboratory in the field of EO/IR image processing. His current research interests are machine learning kernel-based signal and image processing.*