IDEA GROUP PUBLISHING
701 E Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com
ITB7256

Chapter VII

Bayesian Learning

Paula Macrossan
University of New England, Australia

Kerrie Mengersen
University of Newcastle, Australia

Learning from the Bayesian perspective can be described simply as the modification of opinion based on experience. This is in contrast to the Classical or "frequentist" approach that begins with no prior opinion, and inferences are based strictly on information obtained from a random sample selected from the population. An Internet search will quickly provide evidence of the growing popularity of Bayesian methods for data mining in a plethora of subject areas, from agriculture to genetics, engineering, and finance, to name a few. However, despite acknowledged advantages of the Bayesian approach, it is not yet routinely used as a tool for knowledge development. This is, in part, due to a lack of awareness of the language, mechanisms and interpretation inherent in Bayesian modeling, particularly for those trained under a foreign paradigm. The aim of this chapter is to provide a gentle introduction to the topic from the KDD perspective. The concepts involved in Bayes' Theorem are introduced and reinforced through the application of the Bayesian framework to three traditional statistical and/or machine learning examples: a simple probability experiment involving coin tossing, Bayesian linear regression and Bayesian neural network learning. Some of the problems associated with the practical aspects of the implementation of Bayesian learning are then detailed, and various software freely available on the Internet is introduced. The advantages of the Bayesian approach to learning and inference, its impact on diverse scientific fields and its present applications are identified.

INTRODUCTION

Learning from the Bayesian perspective can be described very simply as the modification of opinion based on experience. The Bayesian approach to learning combines a prior subjective opinion with new information, in the form of data, to develop a revised opinion. This process can be repeated any number of times and occurs in a cyclical fashion, with the revised opinion becoming the prior opinion with the arrival of new data. This is in contrast to the classical, traditional or "frequentist" approach that begins with no prior opinion, and inferences are based strictly on information obtained from the data.

This chapter is intended to provide a gentle introduction of Bayesian methods by beginning with simple examples and explanations and later reviewing more complex applications in data mining. The concepts involved in Bayes' Theorem are established with the example case of a simple coin-tossing experiment, after which the typical hierarchical methods used in Bayesian modeling are described. Graphical models, in which the essential structures implicit in a model are defined, are introduced as a convenient method of breaking down a complex model into simple components that provide the basis for Bayesian computation. Bayesian linear regression is examined to illustrate the Bayesian approach to a well-known statistical technique. The application of Bayesian inference to hybrid methods of machine learning is illustrated with the discussion of Bayesian neural network learning. Some of the problems and pitfalls associated with the practical implementation of the Bayesian framework are discussed, after which the reader is then introduced to powerful software for Bayesian inference and diagnostics freely available on the Internet. In conclusion, the present and predicted future impact of the Bayesian approach to learning and inference on diverse scientific fields is discussed. Some texts are recommended for further reading.

BACKGROUND

The essential characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty of parameter estimates of interest in scientific analysis. Bayesian statistics are used to analyse the plausibility of alternative hypotheses that are represented by probabilities, and inference is performed by evaluating these probabilities and making consequent decisions or estimates.

There has been a long-standing debate between Bayesian and classical statisticians regarding the approach to data analysis and knowledge acquisition. The interested reader can pursue this largely philosophical discussion by referring to Press and Tanur (2001). Recently there has been a revival of interest, fuelled by pragmatic rather than philosophical concerns, in the Bayesian approach in all areas of scientific research. It is no coincidence that this revival has occurred in parallel with the burgeoning of computer technology, allowing computationally intensive algorithms to be implemented easily and inexpensively. The most far-reaching of these algorithms is the Markov Chain Monte Carlo (MCMC) approach for simulat-