Information Theory, Inference, and Learning Algorithms

David J.C. MacKay

CAMBRIDGE
UNIVERSITY PRESS

Information Theory, Inference, and Learning Algorithms

David J.C. MacKay
mackay@mrao.cam.ac.uk

Version 6.8 (second printing) March 14, 2004

Please send feedback on this book via
http://www.inference.phy.cam.ac.uk/mackay/itila/

This book (version 6.0) was published by C.U.P. in September 2003. It will remain viewable on-screen on the above website, in postscript, djvu, and pdf formats.

In the second printing (version 6.6) minor typos were corrected, and the book design was slightly altered to modify the placement of section numbers.

(C.U.P. replace this page with their own page ii.)

Contents

Preface

This book is aimed at senior undergraduates and graduate students in Engineering, Science, Mathematics, and Computing. It expects familiarity with calculus, probability theory, and linear algebra as taught in a first- or second-year undergraduate course on mathematics for scientists and engineers.

Conventional courses on information theory cover not only the beautiful *theoretical* ideas of Shannon, but also *practical* solutions to communication problems. This book goes further, bringing in Bayesian data modelling, Monte Carlo methods, variational methods, clustering algorithms, and neural networks.

Why unify information theory and machine learning? Because they are two sides of the same coin. In the 1960s, a single field, cybernetics, was populated by information theorists, computer scientists, and neuroscientists, all studying common problems. Information theory and machine learning still belong together. Brains are the ultimate compression and communication systems. And the state-of-the-art algorithms for both data compression and error-correcting codes use the same tools as machine learning.

How to use this book

The essential dependencies between chapters are indicated in the figure on the next page. An arrow from one chapter to another indicates that the second chapter requires some of the first.

Within Parts I, II, IV, and V of this book, chapters on advanced or optional topics are towards the end. All chapters of Part III are optional on a first reading, except perhaps for Chapter 16 (Message Passing).

The same system sometimes applies within a chapter: the final sections often deal with advanced topics that can be skipped on a first reading. For example in two key chapters – Chapter 4 (The Source Coding Theorem) and Chapter 10 (The Noisy-Channel Coding Theorem) – the first-time reader should detour at section 4.5 and section 10.4 respectively.

Pages vii–x show a few ways to use this book. First, I give the roadmap for a course that I teach in Cambridge: 'Information theory, pattern recognition, and neural networks'. The book is also intended as a textbook for traditional courses in information theory. The second roadmap shows the chapters for an introductory information theory course and the third for a course aimed at an understanding of state-of-the-art error-correcting codes. The fourth roadmap shows how to use the text in a conventional course on machine learning.