An Introduction to Computational Learning Theory

An Introduction to Computational Learning Theory

Michael J. Kearns
Umesh V. Vazirani

The MIT Press
Cambridge, Massachusetts
London, England

This book was typeset by the authors and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Kearns, Michael J.
An introduction to computational learning theory / Michael J. Kearns, Umesh V. Vazirani.
p. cm.
Includes bibliographical references and index.
ISBN 0-262-11193-4
1. Machine learning. 2. Artificial intelligence. 3. Algorithms. 4. Neural networks. I. Vazirani, Umesh Virkumar. II. Title.
Q325.5.K44 1994
006.3—dc20 94-16588
CIP

10 9 8 7 6

Contents

Preface xi

1 The Probably Approximately Correct Learning Model 1

1.1 A Rectangle Learning Game 1

1.2 A General Model 6

1.2.1 Definition of the PAC Model 7

1.2.2 Representation Size and Instance Dimension 12

1.3 Learning Boolean Conjunctions 16

1.4 Intractability of Learning 3-Term DNF Formulae 18

1.5 Using 3-CNF Formulae to Avoid Intractability 22

1.6 Exercises 26

1.7 Bibliographic Notes 28

2 Occam's Razor 31

2.1 Occam Learning and Succinctness 33

2.2 Improving the Sample Size for Learning Conjunctions 37
2.3 Learning Conjunctions with Few Relevant Variables 38
2.4 Learning Decision Lists 42
2.5 Exercises 44
2.6 Bibliographic Notes 46

3 The Vapnik-Chervonenkis Dimension 49

3.1 When Can Infinite Classes Be Learned with a Finite Sample? 49
3.2 The Vapnik-Chervonenkis Dimension 50
3.3 Examples of the VC Dimension 51
3.4 A Polynomial Bound on $|\Pi_{\mathcal{C}}(S)|$ 54
3.5 A Polynomial Bound on the Sample Size for PAC Learning 57
3.5.1 The Importance of ϵ-Nets 57
3.5.2 A Small ϵ-Net from Random Sampling 59
3.6 Sample Size Lower Bounds 62
3.7 An Application to Neural Networks 64
3.8 Exercises 67
3.9 Bibliographic Notes 70

4 Weak and Strong Learning 73

4.1 A Relaxed Definition of Learning? 73
4.2 Boosting the Confidence 76

4.3 Boosting the Accuracy 78

4.3.1 A Modest Accuracy Boosting Procedure 79

4.3.2 Error Analysis for the Modest Procedure 81

4.3.3 A Recursive Accuracy Boosting Algorithm 85

4.3.4 Bounding the Depth of the Recursion 88

4.3.5 Analysis of Filtering Efficiency 89

4.3.6 Finishing Up 96

4.4 Exercises 101

4.5 Bibliographic Notes 102

5 Learning in the Presence of Noise 103

5.1 The Classification Noise Model 104

5.2 An Algorithm for Learning Conjunctions from Statistics 106

5.3 The Statistical Query Learning Model 108

5.4 Simulating Statistical Queries in the Presence of Noise 111

5.4.1 A Nice Decomposition of P_χ 112

5.4.2 Solving for an Estimate of P_χ 114

5.4.3 Guessing and Verifying the Noise Rate 115

5.4.4 Description of the Simulation Algorithm 117

5.5 Exercises 119

5.6 Bibliographic Notes 121

6 Inherent Unpredictability **123**

6.1 Representation Dependent and Independent Hardness 123

6.2 The Discrete Cube Root Problem 124

6.2.1 The Difficulty of Discrete Cube Roots 126

6.2.2 Discrete Cube Roots as a Learning Problem 128

6.3 Small Boolean Circuits Are Inherently Unpredictable 131

6.4 Reducing the Depth of Inherently Unpredictable Circuits 133

6.4.1 Expanding the Input 135

6.5 A General Method and Its Application to Neural Networks 139

6.6 Exercises 140

6.7 Bibliographic Notes 141

7 Reducibility in PAC Learning **143**

7.1 Reducing DNF to Monotone DNF 144

7.2 A General Method for Reducibility 147

7.3 Reducing Boolean Formulae to Finite Automata 149

7.4 Exercises 153

7.5 Bibliographic Notes 154

8 Learning Finite Automata by Experimentation **155**

8.1 Active and Passive Learning 155

8.2 Exact Learning Using Queries 158

8.3 Exact Learning of Finite Automata 160

8.3.1 Access Strings and Distinguishing Strings 160

8.3.2 An Efficiently Computable State Partition 162

8.3.3 The Tentative Hypothesis $\hat{M}$ 164

8.3.4 Using a Counterexample 166

8.3.5 The Algorithm for Learning Finite Automata 169

8.3.6 Running Time Analysis 171

8.4 Learning without a Reset 174

8.4.1 Using a Homing Sequence to Learn 176

8.4.2 Building a Homing Sequence Using Oversized Generalized Classification Trees 178

8.4.3 The No-Reset Algorithm 181

8.4.4 Making Sure L_σ Builds Generalized Classification Trees 182

8.5 Exercises 185

8.6 Bibliographic Notes 186

9 Appendix: Some Tools for Probabilistic Analysis 189

9.1 The Union Bound 189

9.2 Markov's Inequality 189

9.3 Chernoff Bounds 190

Bibliography **193**

Index **205**

Preface

In the Fall term of 1990, we jointly taught a graduate seminar in computational learning theory in the computer science department of the University of California at Berkeley. The material that is presented here has its origins in that course, both in content and exposition. Rather than attempt to give an exhaustive overview of this rapidly expanding and changing area of research, we have tried to carefully select fundamental topics that demonstrate important principles that may be applicable in a wider setting than the one examined here. In the technical sections, we have tried to emphasize intuition whenever possible, while still providing precise arguments.

The book is intended for researchers and students in artificial intelligence, neural networks, theoretical computer science and statistics, and anyone else interested in mathematical models of learning. It is appropriate for use as the central text in a specialized seminar course, or as a supplemental text in a broader course that perhaps also studies the viewpoints taken by artificial intelligence and neural networks. While Chapter 1 lays a common foundation for all the subsequent material, the later chapters are essentially self-contained and may be read selectively and in any order. Exercises are provided at the end of each chapter.

Some brief comments on the expected background of the reader are appropriate here. Familiarity with some basic tools of the formal analysis of algorithms is necessary, as is familiarity with only the most elementary notions of complexity theory, such as *NP*-completeness. For the reader unfamiliar with these topics, the books of Cormen, Leiserson and Rivest [27], Garey and Johnson [38] and Aho, Hopcroft and Ullman [2] provide classic background reading. Some background in probability theory and statistics is desirable but not necessary. In an Appendix in Chapter 9 we have gathered in one place the simple tools of probability theory that we will invoke repeatedly throughout our study.

We are deeply indebted to many colleagues for the advice, feedback and support they gave to us during the writing of this book. We are especially grateful to Ron Rivest of M.I.T. for using preliminary versions of the book for two years as a text in his machine learning course. The comments that resulted from this course were invaluable, and we thank Jay Alsam of M.I.T. for improving several derivations.

We give warm thanks to Dana Angluin of Yale for a detailed critique of a preliminary version. We incorporated practically all of her suggestions, and they greatly improved the presentation. We are very grateful to Les Valiant of Harvard for his many comments and continuing support of the project.

For many suggested improvements and discussions of the material, we thank Scott Decatur of Harvard, John Denker of Bell Labs, Sally Goldman of Washington University, David Haussler of U.C. Santa Cruz, Esther Levin of Bell Labs, Marina Meila of M.I.T., Fernando Pereira of Bell Labs, Stuart Russell of U.C. Berkeley, Rob Schapire of Bell Labs, Donna Slonim of M.I.T., and Manfred Warmuth of U.C. Santa Cruz. Thanks to Danuta Sowinska-Khan and the Art Department of Bell Labs for their preparation of the figures.

We give warm thanks to Terry Ehling of The MIT Press for bringing this project to fruition, and for her enthusiastic support from beginning to end.

An Introduction to Computational Learning Theory