

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

Neural Network Learning: Theoretical Foundations

This book describes recent theoretical advances in the study of artificial neural networks. It explores probabilistic models of supervised learning problems, and addresses the key statistical and computational questions. Research on pattern classification with binary-output networks is surveyed, including a discussion of the relevance of the Vapnik-Chervonenkis dimension. Estimates of this dimension are calculated for several neural network models. A model of classification by real-output networks is developed, and the usefulness of classification with a large margin is demonstrated. The authors explain the role of scale-sensitive versions of the Vapnik-Chervonenkis dimension in large margin classification, and in real estimation. They also discuss the computational complexity of neural network learning, describing a variety of hardness results, and outlining two efficient constructive learning algorithms. The book is self-contained and is intended to be accessible to researchers and graduate students in computer science, engineering, and mathematics.

Martin Anthony is Reader in Mathematics and Executive Director of the Centre for Discrete and Applicable Mathematics at the London School of Economics and Political Science.

Peter Bartlett is a Senior Fellow at the Research School of Information Sciences and Engineering at the Australian National University.

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett



Cambridge University Press
978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations
Martin Anthony and Peter L. Bartlett
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521118620

© Cambridge University Press 1999

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 1999
Reprinted 2001, 2002
This digitally printed version 2009

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Anthony, Martin.

Learning in neural networks : theoretical foundations /
Martin Anthony and Peter L. Bartlett.
p. cm.

Includes bibliographical references.
ISBN 0 521 57353 X (hardcover)

1. Neural networks (Computer science). I. Bartlett, Peter L.,
1966- . II. Title.
QA76.87.A58 1999
006.3'2-dc21 98-53260 CIP

ISBN 978-0-521-57353-5 hardback
ISBN 978-0-521-11862-0 paperback

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

To Colleen, Selena and James.

Contents

<i>Preface</i>	<i>page</i>	<i>xiii</i>
1 Introduction		1
1.1 Supervised learning		1
1.2 Artificial neural networks		2
1.3 Outline of the book		7
1.4 Bibliographical notes		9
Part one: Pattern Classification with Binary-Output		
Neural Networks		
2 The Pattern Classification Problem		11
2.1 The learning problem		13
2.2 Learning finite function classes		19
2.3 Applications to perceptrons		22
2.4 Restricted model		23
2.5 Remarks		25
2.6 Bibliographical notes		27
3 The Growth Function and VC-Dimension		29
3.1 Introduction		29
3.2 The growth function		29
3.3 The Vapnik-Chervonenkis dimension		35
3.4 Bibliographical notes		41
4 General Upper Bounds on Sample Complexity		42
4.1 Learning by minimizing sample error		42
4.2 Uniform convergence and learnability		43
4.3 Proof of uniform convergence result		45
4.4 Application to the perceptron		50
4.5 The restricted model		52
4.6 Remarks		53
4.7 Bibliographical notes		58

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

viii	<i>Contents</i>	
5	General Lower Bounds on Sample Complexity	59
5.1	Introduction	59
5.2	A lower bound for learning	59
5.3	The restricted model	65
5.4	VC-dimension quantifies sample complexity	69
5.5	Remarks	71
5.6	Bibliographical notes	72
6	The VC-Dimension of Linear Threshold Networks	74
6.1	Feed-forward neural networks	74
6.2	Upper bound	77
6.3	Lower bounds	80
6.4	Sigmoid networks	83
6.5	Bibliographical notes	85
7	Bounding the VC-Dimension using Geometric Techniques	86
7.1	Introduction	86
7.2	The need for conditions on the activation functions	86
7.3	A bound on the growth function	89
7.4	Proof of the growth function bound	92
7.5	More on solution set components bounds	102
7.6	Bibliographical notes	106
8	Vapnik-Chervonenkis Dimension Bounds for Neural Networks	108
8.1	Introduction	108
8.2	Function classes that are polynomial in their parameters	108
8.3	Piecewise-polynomial networks	112
8.4	Standard sigmoid networks	122
8.5	Remarks	128
8.6	Bibliographical notes	129
	Part two: Pattern Classification with Real-Output Networks	131
9	Classification with Real-Valued Functions	133
9.1	Introduction	133
9.2	Large margin classifiers	135
9.3	Remarks	138
9.4	Bibliographical notes	138
10	Covering Numbers and Uniform Convergence	140
10.1	Introduction	140
10.2	Covering numbers	140
10.3	A uniform convergence result	143
10.4	Covering numbers in general	147
10.5	Remarks	149

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

	<i>Contents</i>	ix
10.6	Bibliographical notes	150
11	The Pseudo-Dimension and Fat-Shattering Dimension	151
11.1	Introduction	151
11.2	The pseudo-dimension	151
11.3	The fat-shattering dimension	159
11.4	Bibliographical notes	163
12	Bounding Covering Numbers with Dimensions	165
12.1	Introduction	165
12.2	Packing numbers	165
12.3	Bounding with the pseudo-dimension	167
12.4	Bounding with the fat-shattering dimension	174
12.5	Comparing the two approaches	181
12.6	Remarks	182
12.7	Bibliographical notes	183
13	The Sample Complexity of Classification Learning	184
13.1	Large margin SEM algorithms	184
13.2	Large margin SEM algorithms as learning algorithms	185
13.3	Lower bounds for certain function classes	188
13.4	Using the pseudo-dimension	191
13.5	Remarks	191
13.6	Bibliographical notes	192
14	The Dimensions of Neural Networks	193
14.1	Introduction	193
14.2	Pseudo-dimension of neural networks	194
14.3	Fat-shattering dimension bounds: number of parameters	196
14.4	Fat-shattering dimension bounds: size of parameters	203
14.5	Remarks	213
14.6	Bibliographical notes	216
15	Model Selection	218
15.1	Introduction	218
15.2	Model selection results	220
15.3	Proofs of the results	223
15.4	Remarks	225
15.5	Bibliographical notes	227

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

x

Contents

	Part three: Learning Real-Valued Functions	229
16	Learning Classes of Real Functions	231
16.1	Introduction	231
16.2	The learning framework for real estimation	232
16.3	Learning finite classes of real functions	234
16.4	A substitute for finiteness	236
16.5	Remarks	239
16.6	Bibliographical notes	240
17	Uniform Convergence Results for Real Function Classes	241
17.1	Uniform convergence for real functions	241
17.2	Remarks	245
17.3	Bibliographical notes	246
18	Bounding Covering Numbers	247
18.1	Introduction	247
18.2	Bounding with the fat-shattering dimension	247
18.3	Bounding with the pseudo-dimension	250
18.4	Comparing the different approaches	254
18.5	Remarks	255
18.6	Bibliographical notes	256
19	Sample Complexity of Learning Real Function Classes	258
19.1	Introduction	258
19.2	Classes with finite fat-shattering dimension	258
19.3	Classes with finite pseudo-dimension	260
19.4	Results for neural networks	261
19.5	Lower bounds	262
19.6	Remarks	265
19.7	Bibliographical notes	267
20	Convex Classes	269
20.1	Introduction	269
20.2	Lower bounds for non-convex classes	270
20.3	Upper bounds for convex classes	277
20.4	Remarks	280
20.5	Bibliographical notes	282
21	Other Learning Problems	284
21.1	Loss functions in general	284
21.2	Convergence for general loss functions	285
21.3	Learning in multiple-output networks	286
21.4	Interpolation models	289
21.5	Remarks	295
21.6	Bibliographical notes	296

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

<i>Contents</i>		xi
	Part four: Algorithmics	297
22	Efficient Learning	299
22.1	Introduction	299
22.2	Graded function classes	299
22.3	Efficient learning	301
22.4	General classes of efficient learning algorithms	302
22.5	Efficient learning in the restricted model	305
22.6	Bibliographical notes	306
23	Learning as Optimization	307
23.1	Introduction	307
23.2	Randomized algorithms	307
23.3	Learning as randomized optimization	311
23.4	A characterization of efficient learning	312
23.5	The hardness of learning	312
23.6	Remarks	314
23.7	Bibliographical notes	315
24	The Boolean Perceptron	316
24.1	Introduction	316
24.2	Learning is hard for the simple perceptron	316
24.3	Learning is easy for fixed fan-in perceptrons	319
24.4	Perceptron learning in the restricted model	322
24.5	Remarks	328
24.6	Bibliographical notes	329
25	Hardness Results for Feed-Forward Networks	331
25.1	Introduction	331
25.2	Linear threshold networks with binary inputs	331
25.3	Linear threshold networks with real inputs	335
25.4	Sigmoid networks	337
25.5	Remarks	338
25.6	Bibliographical notes	339
26	Constructive Learning Algorithms for Two-Layer Networks	342
26.1	Introduction	342
26.2	Real estimation with convex combinations	342
26.3	Classification learning using boosting	351
26.4	Bibliographical notes	355
	<i>Appendix 1</i> Useful Results	357
	<i>Bibliography</i>	365
	<i>Author index</i>	379
	<i>Subject index</i>	382

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

Preface

Results from computational learning theory are important in many aspects of machine learning practice. Understanding the behaviour of systems that learn to solve information processing problems (like pattern recognition and prediction) is crucial for the design of effective systems. In recent years, ideas and techniques in computational learning theory have matured to the point where theoretical advances are now contributing to machine learning applications, both through increased understanding and through the development of new practical algorithms.

In this book, we concentrate on statistical and computational questions associated with the use of rich function classes, such as artificial neural networks, for pattern recognition and prediction problems. These issues are of fundamental importance in machine learning, and we have seen several significant advances in this area in the last decade. The book focuses on three specific models of learning, although the techniques, results, and intuitions we obtain from studying these formal models carry over to many other situations.

The book is aimed at researchers and graduate students in computer science, engineering, and mathematics. The reader is assumed to have some familiarity with analysis, probability, calculus, and linear algebra, to the level of an early undergraduate course. We remind the reader of most definitions, so it should suffice just to have met the concepts before.

Most chapters have a 'Remarks' section near the end, containing material that is somewhat tangential to the main flow of the text. All chapters finish with a 'Bibliographical Notes' section giving pointers to the literature, both for the material in the chapter and related results. However these sections are not exhaustive.

It is a pleasure to thank many colleagues and friends for their contri-

Cambridge University Press

978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations

Martin Anthony and Peter L. Bartlett

Frontmatter

[More information](#)

xiv

Preface

Contributions to this book. Thanks, in particular, to Phil Long for carefully and thoroughly reading the book, and making many helpful suggestions, and to Ron Meir for making many thoughtful comments on large sections of the book. Jon Baxter considerably improved the results in Chapter 5, and made several useful suggestions that improved the presentation of topics in Chapter 7. Gábor Lugosi suggested significant improvements to the results in Chapter 4. Thanks also to James Ashton, Shai Ben-David, Graham Brightwell, Mostefa Golea, Ying Guo, Ralf Herbrich, Wee Sun Lee, Frederic Maire, Shie Mannor, Llew Mason, Michael Schmitt and Ben Veal for comments, corrections, and suggestions. It is also a pleasure to thank the many collaborators and colleagues who have influenced the way we think about the topics covered in this book: Andrew Barron, Jon Baxter, Shai Ben-David, Norman Biggs, Soura Dasgupta, Tom Downs, Paul Fischer, Marcus Frean, Yoav Freund, Mostefa Golea, Dave Helmbold, Klaus Höffgen, Adam Kowalczyk, Sanjeev Kulkarni, Wee Sun Lee, Tamás Linder, Phil Long, David Lovell, Gábor Lugosi, Wolfgang Maass, Llew Mason, Ron Meir, Eli Posner, Rob Schapire, Bernhard Schölkopf, John Shawe-Taylor, Alex Smola, and Bob Williamson. We also thank Roger Astley of Cambridge University Press for his support and for his efficient handling of this project.

Parts of the book were written while Martin Anthony was visiting the Australian National University, supported by the Royal Society and the Australian Telecommunications and Electronics Research Board, and while Peter Bartlett was visiting the London School of Economics. Martin Anthony's research has also been supported by the European Union (through the 'Neurocolt' and 'Neurocolt 2' ESPRIT projects) and the Engineering and Physical Sciences Research Council. Peter Bartlett's research has been supported by the Australian Research Council and the Department of Industry, Science and Tourism. We are grateful to these funding bodies and to our respective institutions for providing the opportunities for us to work on this book.

We thank our families, particularly Colleen and Selena, for their help, encouragement and tolerance over the years.

Martin Anthony and Peter Bartlett
London and Canberra
March 1999.