Contents

	Preface	*page* xiii
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	11
2	The Pattern Classification Problem	13
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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations
Martin Anthony and Peter L. Bartlett
Table of Contents
More information

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations
Martin Anthony and Peter L. Bartlett
Table of Contents
More information

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

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations
Martin Anthony and Peter L. Bartlett
Table of Contents
More information

	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

© in this web service Cambridge University Press www.cambridge.org

Cambridge University Press
978-0-521-11862-0 - Neural Network Learning: Theoretical Foundations
Martin Anthony and Peter L. Bartlett
Table of Contents
More information

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
Appendix 1 Useful Results 357
Bibliography 365
Author index 379
Subject index 382

© in this web service Cambridge University Press www.cambridge.org