

Robert E. Schapire
Yoav Freund

Boosting

Foundations and Algorithms

Contents

Series Foreword	xi
Preface	xiii
1 Introduction and Overview	1
1.1 Classification Problems and Machine Learning	2
1.2 Boosting	4
1.3 Resistance to Overfitting and the Margins Theory	14
1.4 Foundations and Algorithms	17
<i>Summary</i>	19
<i>Bibliographic Notes</i>	19
<i>Exercises</i>	20
I CORE ANALYSIS	21
2 Foundations of Machine Learning	23
2.1 A Direct Approach to Machine Learning	24
2.2 General Methods of Analysis	30
2.3 A Foundation for the Study of Boosting Algorithms	43
<i>Summary</i>	49
<i>Bibliographic Notes</i>	49
<i>Exercises</i>	50
3 Using AdaBoost to Minimize Training Error	53
3.1 A Bound on AdaBoost's Training Error	54
3.2 A Sufficient Condition for Weak Learnability	56
3.3 Relation to Chernoff Bounds	60
3.4 Using and Designing Base Learning Algorithms	62
<i>Summary</i>	70
<i>Bibliographic Notes</i>	71
<i>Exercises</i>	71

4	Direct Bounds on the Generalization Error	75
4.1	Using VC Theory to Bound the Generalization Error	75
4.2	Compression-Based Bounds	83
4.3	The Equivalence of Strong and Weak Learnability	86
	<i>Summary</i>	88
	<i>Bibliographic Notes</i>	89
	<i>Exercises</i>	89
5	The Margins Explanation for Boosting's Effectiveness	93
5.1	Margin as a Measure of Confidence	94
5.2	A Margins-Based Analysis of the Generalization Error	97
5.3	Analysis Based on Rademacher Complexity	106
5.4	The Effect of Boosting on Margin Distributions	111
5.5	Bias, Variance, and Stability	117
5.6	Relation to Support-Vector Machines	122
5.7	Practical Applications of Margins	128
	<i>Summary</i>	132
	<i>Bibliographic Notes</i>	132
	<i>Exercises</i>	134
II	FUNDAMENTAL PERSPECTIVES	139
6	Game Theory, Online Learning, and Boosting	141
6.1	Game Theory	142
6.2	Learning in Repeated Game Playing	145
6.3	Online Prediction	153
6.4	Boosting	157
6.5	Application to a "Mind-Reading" Game	163
	<i>Summary</i>	169
	<i>Bibliographic Notes</i>	169
	<i>Exercises</i>	170
7	Loss Minimization and Generalizations of Boosting	175
7.1	AdaBoost's Loss Function	177
7.2	Coordinate Descent	179
7.3	Loss Minimization Cannot Explain Generalization	184
7.4	Functional Gradient Descent	188
7.5	Logistic Regression and Conditional Probabilities	194
7.6	Regularization	202
7.7	Applications to Data-Limited Learning	211
	<i>Summary</i>	219
	<i>Bibliographic Notes</i>	219
	<i>Exercises</i>	220

8	Boosting, Convex Optimization, and Information Geometry	227
8.1	Iterative Projection Algorithms	228
8.2	Proving the Convergence of AdaBoost	243
8.3	Unification with Logistic Regression	252
8.4	Application to Species Distribution Modeling	255
	<i>Summary</i>	260
	<i>Bibliographic Notes</i>	262
	<i>Exercises</i>	263
III	ALGORITHMIC EXTENSIONS	269
9	Using Confidence-Rated Weak Predictions	271
9.1	The Framework	273
9.2	General Methods for Algorithm Design	275
9.3	Learning Rule-Sets	287
9.4	Alternating Decision Trees	290
	<i>Summary</i>	296
	<i>Bibliographic Notes</i>	297
	<i>Exercises</i>	297
10	Multiclass Classification Problems	303
10.1	A Direct Extension to the Multiclass Case	305
10.2	The One-against-All Reduction and Multi-label Classification	310
10.3	Application to Semantic Classification	316
10.4	General Reductions Using Output Codes	320
	<i>Summary</i>	333
	<i>Bibliographic Notes</i>	333
	<i>Exercises</i>	334
11	Learning to Rank	341
11.1	A Formal Framework for Ranking Problems	342
11.2	A Boosting Algorithm for the Ranking Task	345
11.3	Methods for Improving Efficiency	351
11.4	Multiclass, Multi-label Classification	361
11.5	Applications	364
	<i>Summary</i>	367
	<i>Bibliographic Notes</i>	369
	<i>Exercises</i>	369

IV	ADVANCED THEORY	375
12	Attaining the Best Possible Accuracy	377
12.1	Optimality in Classification and Risk Minimization	378
12.2	Approaching the Optimal Risk	382
12.3	How Minimizing Risk Can Lead to Poor Accuracy	398
	<i>Summary</i>	406
	<i>Bibliographic Notes</i>	406
	<i>Exercises</i>	407
13	Optimally Efficient Boosting	415
13.1	The Boost-by-Majority Algorithm	416
13.2	Optimal Generalization Error	432
13.3	Relation to AdaBoost	448
	<i>Summary</i>	453
	<i>Bibliographic Notes</i>	453
	<i>Exercises</i>	453
14	Boosting in Continuous Time	459
14.1	Adaptiveness in the Limit of Continuous Time	460
14.2	BrownBoost	468
14.3	AdaBoost as a Special Case of BrownBoost	476
14.4	Experiments with Noisy Data	483
	<i>Summary</i>	485
	<i>Bibliographic Notes</i>	486
	<i>Exercises</i>	486
	Appendix: Some Notation, Definitions, and Mathematical Background	491
A.1	General Notation	491
A.2	Norms	492
A.3	Maxima, Minima, Suprema, and Infima	493
A.4	Limits	493
A.5	Continuity, Closed Sets, and Compactness	494
A.6	Derivatives, Gradients, and Taylor's Theorem	495
A.7	Convexity	496
A.8	The Method of Lagrange Multipliers	497
A.9	Some Distributions and the Central Limit Theorem	498
	Bibliography	501
	Index of Algorithms, Figures, and Tables	511
	Subject and Author Index	513