# 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  
  Summary  19  
  Bibliographic Notes  19  
  Exercises  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  
  Summary  49  
  Bibliographic Notes  49  
  Exercises  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  
  Summary  70  
  Bibliographic Notes  71  
  Exercises  71
# 4 Direct Bounds on the Generalization Error

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  
**Summary** 88  
**Bibliographic Notes** 89  
**Exercises** 89

# 5 The Margins Explanation for Boosting’s Effectiveness

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  
**Summary** 132  
**Bibliographic Notes** 132  
**Exercises** 134

## II FUNDAMENTAL PERSPECTIVES

# 6 Game Theory, Online Learning, and Boosting

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  
**Summary** 169  
**Bibliographic Notes** 169  
**Exercises** 170

# 7 Loss Minimization and Generalizations of Boosting

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  
**Summary** 219  
**Bibliographic Notes** 219  
**Exercises** 220
### Contents

#### 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

**Summary** 260

**Bibliographic Notes** 262

**Exercises** 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

**Summary** 296

**Bibliographic Notes** 297

**Exercises** 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

**Summary** 333

**Bibliographic Notes** 333

**Exercises** 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

**Summary** 367

**Bibliographic Notes** 369

**Exercises** 369