

CONTENTS

| | |
|--|------------|
| Preface | xii |
| Acknowledgements | xiv |
| Notation | xvi |
| | |
| 1 Introduction and Overview | 1 |
| 1.1 Overview, 2 | |
| 1.2 Historical Perspective, 5 | |
| 1.3 Growth of Empirical Knowledge, 10 | |
| 1.4 Mathematical Representation of Empirical Knowledge, 15 | |
| 1.5 Summary and Discussion, 22 | |
| 1.6 Bibliographic Notes, 25 | |
| References, 26 | |
| Problems, 27 | |
| | |
| 2 Basic Learning Approaches and Complexity Control | 32 |
| 2.1 Data Encoding and Preprocessing, 34 | |
| 2.2 Terminology and Common Learning Tasks, 36 | |
| 2.3 Basic Learning Approaches, 47 | |
| 2.4 Generalization and Complexity Control, 55 | |
| 2.5 Application Example, 62 | |
| 2.6 Summary and Bibliographic Notes, 65 | |
| References, 67 | |
| Problems, 68 | |

| | |
|--|------------|
| 3 Philosophical Perspectives | 73 |
| 3.1 Brief Overview of the Philosophy of Science, 75 | |
| 3.2 Epistemology and Treatment of Uncertainty, 77 | |
| 3.3 Acquisition of Knowledge and Inference, 80 | |
| 3.4 Empirical Risk Minimization and Maximum Likelihood, 86 | |
| 3.5 Occam's Razor, 88 | |
| 3.6 Popper's Falsifiability, 91 | |
| 3.7 Bayesian Inference, 96 | |
| 3.8 Principle of Multiple Explanations and Bayesian Averaging, 102 | |
| 3.9 Social Systems and Philosophy, 106 | |
| 3.10 Summary and Bibliographic Notes, 112 | |
| References, 114 | |
| Problems, 115 | |
| | |
| 4 Statistical Learning Theory | 118 |
| 4.1 Motivation and Modeling Assumptions, 119 | |
| 4.2 Analysis of Empirical Risk Minimization, 128 | |
| 4.3 VC-dimension, 133 | |
| 4.4 Generalization Bounds, 137 | |
| 4.5 Structural Risk Minimization, 145 | |
| 4.6 Summary and Discussion, 152 | |
| 4.7 Bibliographic Notes, 155 | |
| References, 156 | |
| Problems, 156 | |
| | |
| 5 Statistical Methods | 161 |
| 5.1 Statistical Methodology and Basic Methods, 162 | |
| 5.2 Taxonomy of Nonlinear Methods, 175 | |
| 5.3 Decision Trees, 180 | |
| 5.4 Additive Modeling and Projection Pursuit, 191 | |
| 5.5 Greedy Feature Selection, 196 | |
| 5.6 Signal Denoising and Orthogonal Estimators, 202 | |
| 5.7 Summary and Discussion, 208 | |
| 5.8 Bibliographic Notes, 210 | |
| References, 211 | |
| Problems, 212 | |

| | |
|---|------------|
| 6 Neural Network Methods | 215 |
| 6.1 Neural Networks: Motivation, Terminology and History, 216 | |
| 6.2 Gradient Descent and Backpropagation, 221 | |
| 6.3 Multilayer Perceptron (MLP) Networks, 226 | |
| 6.4 Radial Basis Function (RBF) Networks, 234 | |
| 6.5 Methods for Unsupervised Learning, 240 | |
| 6.6 Self-Organizing Maps, 247 | |
| 6.7 SOM Application: Similarity between European Languages, 256 | |
| 6.8 Summary and Bibliographic Notes, 260 | |
| References, 265 | |
| Problems, 266 | |
| 7 Support Vector Machines | 270 |
| 7.1 Margin-Based Loss Functions, 274 | |
| 7.2 Linear SVM Classifiers, 284 | |
| 7.3 Nonlinear SVM Classifiers, 292 | |
| 7.4 Practical Issues for SVM Classification, 301 | |
| 7.5 SVM Regression, 308 | |
| 7.6 Summary and Discussion, 317 | |
| 7.7 Bibliographic Notes, 319 | |
| References, 320 | |
| Problems, 322 | |
| 8 Combining Methods and Ensemble Learning | 327 |
| 8.1. Committee of Networks and Stacking, 330 | |
| 8.2. Bagging, 339 | |
| 8.3. Boosting, 345 | |
| 8.4. Summary and Discussion, 356 | |
| 8.5. Bibliographic Notes, 358 | |
| References, 360 | |
| Problems, 361 | |
| 9 Nonstandard Learning Approaches | 364 |
| 9.1 Transduction and Semi-Supervised Learning, 367 | |
| 9.2 Universum Learning, 381 | |
| 9.3 Learning Using Privileged Information, 389 | |
| 9.4 Multi-Task Learning, 396 | |

| |
|---------------------------------|
| 9.5 Summary and Discussion, 403 |
| 9.6 Bibliographic Notes, 405 |
| References, 408 |
| Problems, 409 |

| | |
|---|------------|
| 10 Methodological Issues and Case Studies | 413 |
| 10.1 Human Modeling Biases and Problem Formalization, 415 | |
| 10.2 Feature Selection, 419 | |
| 10.3 Case Study: Real-Time Face Detection, 425 | |
| 10.4 Case Study: Market Timing of International Mutual Funds, 432 | |
| 10.5 Case Study: Prediction of Transplant-Related Mortality, 437 | |
| 10.6 Case Study: Modeling of Power Consumption in Commercial Buildings, 442 | |
| 10.7 Summary: Philosophical Aspects of Predictive Data Modeling, 449 | |
| 10.8 Bibliographic Notes, 451 | |
| References, 452 | |
| Problems, 453 | |
| Index | 455 |