

Editors

Michael Berthold • David J. Hand

Intelligent Data Analysis

An Introduction

2nd revised and extended Edition



Springer

Table of Contents

1. Introduction	1
1.1 Why "Intelligent Data Analysis"?	1
1.2 How the Computer Is Changing Things	4
1.3 The Nature of Data	8
1.4 Modern Data Analytic Tools	12
1.5 Conclusion	14
2. Statistical Concepts	17
2.1 Introduction	17
2.2 Probability	18
2.3 Sampling and Sampling Distributions	29
2.4 Statistical Inference	33
2.5 Prediction and Prediction Error	46
2.6 R sampling	57
2.7 Conclusion	68
3. Statistical Methods	69
3.1 Introduction	69
3.2 Generalized Linear Models	70
3.3 Special Topics in Regression Modelling	93
3.4 Classical Multivariate Analysis	100
3.5 Conclusion	129
4. Bayesian Methods	131
4.1 Introduction	131
4.2 The Bayesian Paradigm	132
4.3 Bayesian Inference	135
4.4 Bayesian Modelling	143
4.5 Bayesian Networks	153
4.6 Conclusion	167
5. Support Vector and Kernel Methods	169
5.1 Example: Kernel Perceptron	170
5.2 Overfitting and Generalization Bounds	176
5.3 Support Vector Machines	181

5.4 Kernel PCA and CCA	194
5.5 Conclusion	196
6. Analysis of Time Series	199
6.1 Introduction	199
6.2 Linear Systems Analysis	202
6.3 Nonlinear Dynamics Basics	207
6.4 Delay-Coordinate Embedding	213
6.5 Examples	218
6.6 Conclusion	226
7. Rule Induction	229
7.1 Introduction	229
7.2 Propositional rule learning	232
7.3 Rule learning as search	236
7.4 Evaluating the quality of rules	242
7.5 Propositional rule induction at work	246
7.6 Learning first-order rules	250
7.7 Some ILP Systems at work	262
7.8 Conclusion	267
8. Neural Networks	269
8.1 Introduction	269
8.2 Fundamentals	270
8.3 Multilayer Feedforward Neural Networks	278
8.4 Learning and Generalization	283
8.5 Radial Basis Function Networks	292
8.6 Competitive Learning	300
8.7 Principal Components Analysis and Neural Networks	307
8.8 Time Series Analysis	312
8.9 Conclusion	319
9. Fuzzy Logic	321
9.1 Introduction	321
9.2 Basics of Fuzzy Sets and Fuzzy Logic	322
9.3 Extracting Fuzzy Models from Data	336
9.4 Fuzzy Decision Trees	346
9.5 Conclusion	350
10. Stochastic Search Methods	351
10.1 Introduction	351
10.2 Stochastic Search by Simulated Annealing	354
10.3 Stochastic, Adaptive Search by Evolution	360
10.4 Evolution Strategies	362
10.5 Genetic Algorithms	374

10.6 Genetic Programming	390
10.7 Conclusion	400
11. Visualization	403
11.1 Introduction	403
11.2 Classification of Visual Data Analysis Techniques	405
11.3 Data Type to be Visualized	406
11.4 Visualization Techniques	411
11.5 Interaction Techniques	414
11.C Specific Visual Data Analysis Techniques	418
11.7 Conclusion	426
12. Systems and Applications	429
12.1 Introduction	429
12.2 Diversity of IDA Applications	430
12.3 Several Development Issues	436
12.4 Conclusion	442
Appendix A: Tools	445
A.1 Tools for statistical analysis	445
A.2 Tools for exploration/modcling	447
A.3 Tools for Text and Web Mining	454
A.4 Data Analysis Suites	456
A.5 Conclusion	464
Appendix B: Information-Theoretic Tree and Rule Induction	465
B.1 Information and Uncertainty	465
B.2 Decision Tree Induction	468
B.3 Rule Induction	470
References	475
Index	501
Author Addresses	513