Table of Contents

Foreword xi

Acknowledgement xv

Part 1 Data Mining and Knowledge Discovery Process **1**

Chapter 1. Introduction 3
1. What is Data Mining? 3
2. How does Data Mining Differ from Other Approaches? 5
3. Summary and Bibliographical Notes 6
4. Exercises 7

Chapter 2. The Knowledge Discovery Process 9
1. Introduction 9
2. What is the Knowledge Discovery Process? 10
3. Knowledge Discovery Process Models 11
4. Research Issues 19
5. Summary and Bibliographical Notes 20
6. Exercises 24

Part 2 Data Understanding **25**

Chapter 3. Data 27
1. Introduction 27
2. Attributes, Data Sets, and Data Storage 27
3. Issues Concerning the Amount and Quality of Data 37
4. Summary and Bibliographical Notes 44
5. Exercises 46

Chapter 4. Concepts of Learning, Classification, and Regression 49
1. Introductory Comments 49
2. Classification 55
3. Summary and Bibliographical Notes 65
4. Exercises 66

Chapter 5. Knowledge Representation 69
1. Data Representation and their Categories: General Insights 69
2. Categories of Knowledge Representation 71
3. Granularity of Data and Knowledge Representation Schemes 76
4. Sets and Interval Analysis 77
5. Fuzzy Sets as Human-Centric Information Granules 78

6. Shadowed Sets 82
7. Rough Sets 84
8. Characterization of Knowledge Representation Schemes 86
9. Levels of Granularity and Perception Perspectives 87
10. The Concept of Granularity in Rules 88
11. Summary and Bibliographical Notes 89
12. Exercises 90

Part 3 Data Preprocessing **93**

Chapter 6. Databases, Data Warehouses, and OLAP 95
1. Introduction 95
2. Database Management Systems and SQL 95
3. Data Warehouses 106
4. On-Line Analytical Processing (OLAP) 116
5. Data Warehouses and OLAP for Data Mining 127
6. Summary and Bibliographical Notes 128
7. Exercises 130

Chapter 7. Feature Extraction and Selection Methods 133
1. Introduction 133
2. Feature Extraction 133
3. Feature Selection 207
4. Summary and Bibliographical Notes 228
5. Exercises 230

Chapter 8. Discretization Methods 235
1. Why Discretize Data Attributes? 235
2. Unsupervised Discretization Algorithms 237
3. Supervised Discretization Algorithms 237
4. Summary and Bibliographical Notes 253
5. Exercises 254

Part 4 Data Mining: Methods for Constructing Data Models **255**

Chapter 9. Unsupervised Learning: Clustering 257
1. From Data to Information Granules or Clusters 257
2. Categories of Clustering Algorithms 258
3. Similarity Measures 258
4. Hierarchical Clustering 260
5. Objective Function-Based Clustering 263
6. Grid - Based Clustering 272
7. Self-Organizing Feature Maps 274
8. Clustering and Vector Quantization 279
9. Cluster Validity 280
10. Random Sampling and Clustering as a Mechanism of Dealing with Large Datasets 284
11. Summary and Biographical Notes 286
12. Exercises 287

Chapter 10. Unsupervised Learning: Association Rules 289
1. Introduction 289
2. Association Rules and Transactional Data 290
3. Mining Single Dimensional, Single-Level Boolean Association Rules 295

4. Mining Other Types of Association Rules 301
5. Summary and Bibliographical Notes 304
6. Exercises 305

Chapter 11. Supervised Learning: Statistical Methods 307
1. Bayesian Methods 307
2. Regression 346
3. Summary and Bibliographical Notes 375
4. Exercises 376

Chapter 12. Supervised Learning: Decision Trees, Rule Algorithms, and Their Hybrids 381
1. What is Inductive Machine Learning? 381
2. Decision Trees 388
3. Rule Algorithms 393
4. Hybrid Algorithms 399
5. Summary and Bibliographical Notes 416
6. Exercises 416

Chapter 13. Supervised Learning: Neural Networks 419
1. Introduction 419
2. Biological Neurons and their Models 420
3. Learning Rules 428
4. Neural Network Topologies 431
5. Radial Basis Function Neural Networks 431
6. Summary and Bibliographical Notes 449
7. Exercises 450

Chapter 14. Text Mining 453
1. Introduction 453
2. Information Retrieval Systems 454
3. Improving Information Retrieval Systems 462
4. Summary and Bibliographical Notes 464
5. Exercises 465

Part 5 Data Models Assessment **467**

Chapter 15. Assessment of Data Models 469
1. Introduction 469
2. Models, their Selection, and their Assessment 470
3. Simple Split and Cross-Validation 473
4. Bootstrap 474
5. Occam's Razor Heuristic 474
6. Minimum Description Length Principle 475
7. Akaike's Information Criterion and Bayesian Information Criterion 476
8. Sensitivity, Specificity, and ROC Analyses 477
9. Interestingness Criteria 484
10. Summary and Bibliographical Notes 485
11. Exercises 486

Part 6 Data Security and Privacy Issues **487**

Chapter 16. Data Security, Privacy and Data Mining 489
1. Privacy in Data Mining 489
2. Privacy Versus Levels of Information Granularity 490

3. Distributed Data Mining 491
4. Collaborative Clustering 492
5. The Development of the Horizontal Model of Collaboration 494
6. Dealing with Different Levels of Granularity in the Collaboration Process 498
7. Summary and Biographical Notes 499
8. Exercises 501

Part 7 Overview of Key Mathematical Concepts 503

Appendix A. Linear Algebra 505
1. Vectors 505
2. Matrices 519
3. Linear Transformation 540

Appendix B. Probability 547
1. Basic Concepts 547
2. Probability Laws 548
3. Probability Axioms 549
4. Defining Events With Set–Theoretic Operations 549
5. Conditional Probability 551
6. Multiplicative Rule of Probability 552
7. Random Variables 553
8. Probability Distribution 555

Appendix C. Lines and Planes in Space 567
1. Lines on Plane 567
2. Lines and Planes in a Space 569
3. Planes 572
4. Hyperplanes 575

Appendix D. Sets 579
1. Set Definition and Notations 579
2. Types of Sets 581
3. Set Relations 585
4. Set Operations 587
5. Set Algebra 590
6. Cartesian Product of Sets 592
7. Partition of a Nonempty Set 596

Index 597

http://www.springer.com/978-0-387-33333-5

Data Mining
A Knowledge Discovery Approach
Cios, K.J.; Pedrycz, W.; Swiniarski, R.W.; Kurgan, L.
2007, XV, 606 p., Hardcover
ISBN: 978-0-387-33333-5