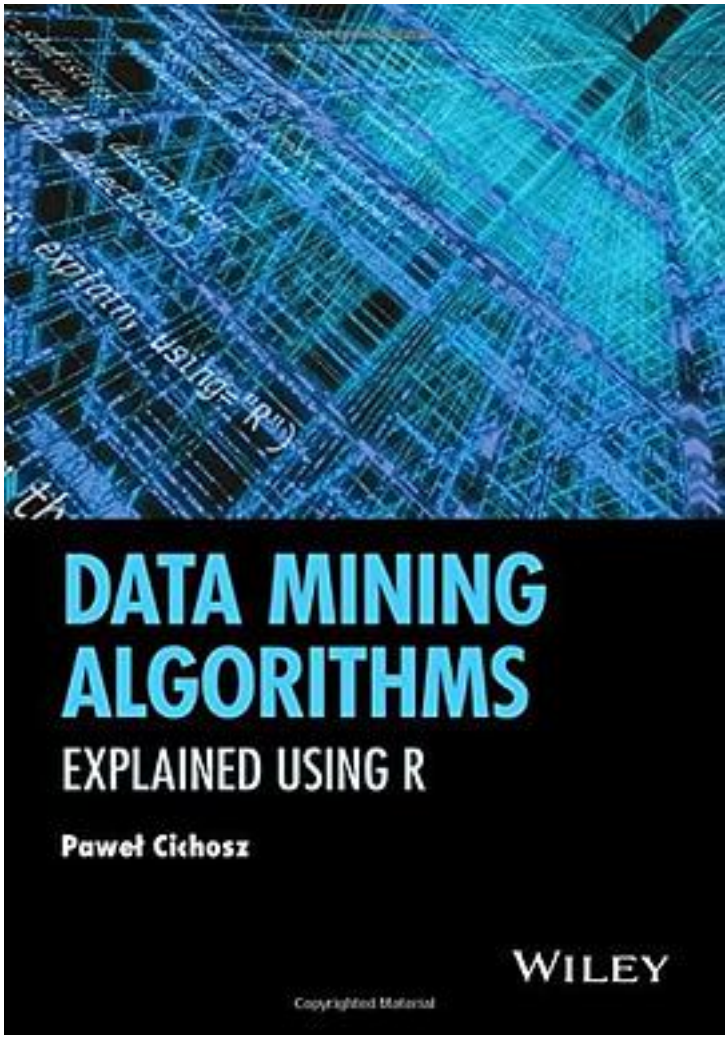
Data Mining Algorithms



Data Mining Algorithms_下载链接1_

著者:Pawel Cichosz

出版者:Wiley

出版时间:2015-1-27

装帧:Hardcover

isbn:9781118332580

Data Mining Algorithms is a practical, technically-oriented guide to data mining

algorithms that covers the most important algorithms for building classification, regression, and clustering models, as well as techniques used for attribute selection and transformation, model quality evaluation, and creating model ensembles. The author presents many of the important topics and methodologies widely used in data mining, whilst demonstrating the internal operation and usage of data mining algorithms using examples in R.

作者介绍:

目录:
Acknowledgements xix
Preface xxi
References xxxi
Part I Preliminaries 1
1 Tasks 3
1.1 Introduction 3
1.2 Inductive learning tasks 5
1.3 Classification 9
1.4 Regression 14
1.5 Clustering 16
1.6 Practical issues 19
1.7 Conclusion 20
1.8 Further readings 21
References 22
2 Basic statistics 23
2.1 Introduction 23
2.2 Notational conventions 24
2.3 Basic statistics as modeling 24
2.4 Distribution description 25
2.5 Relationship detection 47
2.6 Visualization 62
2.7 Conclusion 65
2.8 Further readings 66
References 67
Part II Classification 69
3 Decision trees 71
3.1 Introduction 71
3.2 Decision tree model 72
3.3 Growing 76
3.4 Pruning 90
3.5 Prediction 103
3.6 Weighted instances 105
3.7 Missing value handling 106
3.8 Conclusion 114
3.9 Further readings 114
References 116
4 Naïve Bayes classifier 118
4.1 Introduction 118
4.2 Bayes rule 118
4.3 Classification by Bayesian inference 120
4.4 Practical issues 125
4.5 Conclusion 131
4.6 Further readings 131

References 132
5 Linear classification 134
5.1 Introduction 134
5.2 Linear representation 136
5.3 Parameter estimation 145
5.4 Discrete attributes 154
5.5 Conclusion 155
5.6 Further readings 156
References 157
6 Misclassification costs 159
6.1 Introduction 159
6.2 Cost representation 161
6.3 Incorporating misclassification costs 164
6.4 Effects of cost incorporation 176
6.5 Experimental procedure 180
6.6 Conclusion 184
6.7 Further readings 185
References 187
7 Classification model evaluation 189
7.1 Introduction 189
7.2 Performance measures 190
7.3 Evaluation procedures 213
7.4 Conclusion 231
7.5 Further readings 232
References 233
Part III Regression 235
8 Linear regression 237
8.1 Introduction 237
8.2 Linear representation 238
8.3 Parameter estimation 242
8.4 Discrete attributes 250
8.5 Advantages of linear models 251
8.6 Beyond linearity 252
8.7 Conclusion 258
8.8 Further readings 258
References 259
9 Regression trees 261
9.1 Introduction 261
9.2 Regression tree model 262
9.3 Growing 263
9.4 Pruning 274
9.5 Prediction 277
9.6 Weighted instances 278
9.7 Missing value handling 279
9.8 Piecewise linear regression 284
9.9 Conclusion 292
9.10 Further readings 292
References 293
10 Regression model evaluation 295
10.1 Introduction 295
10.2 Performance measures 296
10.3 Evaluation procedures 303
10.4 Conclusion 309
10.5 Further readings 309

References 310
Part IV Clustering 311
11 (Dis)similarity measures 313
11.1 Introduction 313
11.2 Measuring dissimilarity and similarity 313
11.3 Difference-based dissimilarity 314
11.4 Correlation-based similarity 321
11.5 Missing attribute values 324
11.6 Conclusion 325
11.7 Further readings 325
References 326
12 k-Centers clustering 328
12.1 Introduction 328
12.2 Algorithm scheme 330
12.3 k-Means 334
12.4 Beyond means 338
12.5 Beyond (fixed) k 342
12.6 Explicit cluster modeling 343
12.7 Conclusion 345
12.8 Further readings 345
References 347
13 Hierarchical clustering 349
13.1 Introduction 349
13.2 Cluster hierarchies 351
13.3 Agglomerative clustering 353
13.4 Divisive clustering 361
13.5 Hierarchical clustering visualization 364
13.6 Hierarchical clustering prediction 366
13.7 Conclusion 369
13.8 Further readings 370
References 371
14 Clustering model evaluation 373
14.1 Introduction 373
14.2 Per-cluster quality measures 376
14.3 Overall quality measures 385
14.4 External quality measures 393
14.5 Using quality measures 397
14.6 Conclusion 398
14.7 Further readings 398
References 399
Part V Getting Better Models 401
15 Model ensembles 403
15.1 Introduction 403
15.2 Model committees 404
15.3 Base models 406
15.4 Model aggregation 420
15.5 Specific ensemble modeling algorithms 431
15.6 Quality of ensemble predictions 448
15.7 Conclusion 449
15.8 Further readings 450
References 451
16 Kernel methods 454
16.1 Introduction 454
16.2 Support vector machines 457

16.3 Support vector regression 473
16.4 Kernel trick 482
16.5 Kernel functions 484
16.6 Kernel prediction 487
16.7 Kernel-based algorithms 489
16.8 Conclusion 494
16.9 Further readings 495
References 496
17 Attribute transformation 498
17.1 Introduction 498
17.2 Attribute transformation task 499
17.3 Simple transformations 504
17.4 Multiclass encoding 510
17.5 Conclusion 521
17.6 Further readings 521
References 522
18 Discretization 524
18.1 Introduction 524
18.2 Discretization task 525
18.3 Unsupervised discretization 530
18.4 Supervised discretization 533
18.5 Effects of discretization 551
18.6 Conclusion 553
18.7 Further readings 553
References 556
19 Attribute selection 558
19.1 Introduction 558
19.2 Attribute selection task 559
19.3 Attribute subset search 562
19.4 Attribute selection filters 568
19.5 Attribute selection wrappers 588
19.6 Effects of attribute selection 593
19.7 Conclusion 598
19.8 Further readings 599
References 600
20 Case studies 602
20.1 Introduction 602
20.2 Census income 605
20.3 Communities and crime 631
20.4 Cover type 640
20.5 Conclusion 654
20.6 Further readings 655
References 655
Closing 657
A Notation 659
A.1 Attribute values 659
A.2 Data subsets 659
A.3 Probabilities 660
B R packages 661
B.1 CRAN packages 661
B.2 DMR packages 662
B.3 Installing packages 663
References 664
C Datasets 666

Index 667
・・・・・・・(收起)

Data Mining Algorithms_下载链接1_

标签

数据挖掘

R

DM

机器学习

Statistics

评论

基础教科书，难度不大，讲的挺细，700多页。

Data Mining Algorithms_下载链接1_

书评

Data Mining Algorithms_下载链接1_