

Contents

Chapter 1

Advanced Modelling Paradigms in Data Mining	1
Dawn E. Holmes, Jeffrey Tweedale, Lakhmi C. Jain	
1 Introduction	1
2 Foundations	1
2.1 Statistical Modelling	2
2.2 Predictions Analysis	2
2.3 Data Analysis	3
2.4 Chains of Relationships	3
3 Intelligent Paradigms	4
3.1 Bayesian Analysis	4
3.2 Support Vector Machines	4
3.3 Learning	5
4 Chapters Included in the Book	5
5 Conclusion	6
References	7

Chapter 2

Data Mining with Multilayer Perceptrons and Support Vector Machines	9
Paulo Cortez	
1 Introduction	9
2 Supervised Learning	10
2.1 Classical Regression	11
2.2 Multilayer Perceptron	11
2.3 Support Vector Machines	13
3 Data Mining	14
3.1 Business Understanding	14
3.2 Data Understanding	14
3.3 Data Preparation	15
3.4 Modeling	15
3.5 Evaluation	18
3.6 Deployment	18

4	Experiments	19
4.1	Classification Example	19
4.2	Regression Example	21
5	Conclusions and Further Reading	23
	References	23

Chapter 3

Regulatory Networks under Ellipsoidal Uncertainty – Data Analysis and Prediction by Optimization Theory and Dynamical Systems		27
Erik Kropat, Gerhard-Wilhelm Weber, Chandra Sekhar Pedamallu		
1	Introduction	27
2	Ellipsoidal Calculus	30
2.1	Ellipsoidal Descriptions	30
2.2	Affine Transformations	31
2.3	Sums of Two Ellipsoids	31
2.4	Sums of K Ellipsoids	31
2.5	Intersection of Ellipsoids	32
3	Target-Environment Regulatory Systems under Ellipsoidal Uncertainty	33
3.1	The Time-Discrete Model	33
3.2	Algorithm	37
4	The Regression Problem	40
4.1	The Trace Criterion	43
4.2	The Trace of the Square Criterion	43
4.3	The Determinant Criterion	44
4.4	The Diameter Criterion	44
4.5	Optimization Methods	45
5	Mixed Integer Regression Problem	47
6	Conclusion	49
	References	50

Chapter 4

A Visual Environment for Designing and Running Data Mining Workflows in the Knowledge Grid		57
Eugenio Cesario, Marco Lackovic, Domenico Talia, Paolo Trunfio		
1	Introduction	57
2	The Knowledge Grid	58
3	Workflow Components	60
4	The DIS3GNO System	63
5	Execution Management	65
6	Use Cases and Performance	67
6.1	Parameter Sweeping Workflow	67
6.2	Ensemble Learning Workflow	70

7	Related Work	72
8	Conclusions	74
	References	74

Chapter 5

Formal Framework for the Study of Algorithmic Properties of Objective Interestingness Measures		77
Le Bras Yannick, Lenca Philippe, Stéphane Lallich		
1	Introduction	77
2	Scientific Landscape	79
	2.1 Database	79
	2.2 Association Rules	81
	2.3 Interestingness Measures	82
3	A Framework for the Study of Measures	83
	3.1 Adapted Functions of Measure	84
	3.2 Expression of a Set of Measures	87
4	Application to Pruning Strategies	88
	4.1 All-Monotony	89
	4.2 Universal Existential Upward Closure	90
	4.3 Optimal Rule Discovery	92
	4.4 Properties Verified by the Measures	94
	Conclusion	94
	References	95

Chapter 6

Nonnegative Matrix Factorization: Models, Algorithms and Applications		99
Zhong-Yuan Zhang		
1	Introduction	99
2	Standard NMF and Variations	101
	2.1 Standard NMF	101
	2.2 Semi-NMF ([22])	103
	2.3 Convex-NMF ([22])	103
	2.4 Tri-NMF ([23])	103
	2.5 Kernel NMF ([24])	104
	2.6 Local Nonnegative Matrix Factorization, LNMF ([25,26])	104
	2.7 Nonnegative Sparse Coding, NNSC ([28])	104
	2.8 Spares Nonnegative Matrix Factorization, SNMF ([29,30,31])	104
	2.9 Nonnegative Matrix Factorization with Sparseness Constraints, NMFSC ([32])	105
	2.10 Nonsmooth Nonnegative Matrix Factorization, nsNMF ([15])	105
	2.11 Sparse NMFs: SNMF/R, SNMF/L ([33])	106

2.12	CUR Decomposition ([34])	106
2.13	Binary Matrix Factorization, BMF ([20,21])	106
3	Divergence Functions and Algorithms for NMF	106
3.1	Divergence Functions	108
3.2	Algorithms for NMF	109
4	Applications of NMF	115
4.1	Image Processing	115
4.2	Clustering	116
4.3	Semi-supervised Clustering	116
4.4	Bi-clustering (co-clustering)	117
4.5	Financial Data Mining	118
5	Relations with Other Relevant Models	118
5.1	Relations between NMF and K-means	119
5.2	Relations between NMF and PLSI	120
6	Conclusions and Future Works	126
	Appendix	127
	References	131

Chapter 7

Visual Data Mining and Discovery with Binarized Vectors		135
Boris Kovalerchuk, Florian Delizy, Logan Riggs, Evgenii Vityaev		
1	Introduction	136
2	Method for Visualizing Data	138
3	Visualization for Breast Cancer Diagnostics	145
4	General Concept of Using MDF in Data Mining	147
5	Scaling Algorithms	148
5.1	Algorithm with Data-Based Chains	148
5.2	Algorithm with Pixel Chains	149
6	Binarization and Monotonization	152
7	Monotonization	154
8	Conclusion	155
	References	155

Chapter 8

A New Approach and Its Applications for Time Series Analysis and Prediction Based on Moving Average of n^{th}-Order Difference		157
Yang Lan, Daniel Neagu		
1	Introduction	157
2	Definitions Relevant to Time Series Prediction	159
3	The Algorithm of Moving Average of n^{th} -order Difference for Bounded Time Series Prediction	161
4	Finding Suitable Index m and Order Level n for Increasing the Prediction Precision	168
5	Prediction Results for Sunspot Number Time Series	170

6	Prediction Results for Earthquake Time Series	173
7	Prediction Results for Pseudo-Periodical Synthetic Time Series	175
8	Prediction Results Comparison	177
9	Conclusions	179
10	Appendix	180
	References	182

Chapter 9

Exceptional Model Mining		183
Arno Knobbe, Ad Feelders, Dennis Leman		
1	Introduction	183
2	Exceptional Model Mining	185
3	Model Classes	187
3.1	Correlation Models	187
3.2	Regression Model	188
3.3	Classification Models	189
4	Experiments	192
4.1	Analysis of Housing Data	192
4.2	Analysis of Gene Expression Data	194
5	Conclusions and Future Research	197
References		198

Chapter 10

Online ChiMerge Algorithm		199
Petri Lehtinen, Matti Saarela, Tapio Elomaa		
1	Introduction	199
2	Numeric Attributes, Decision Trees, and Data Streams	201
2.1	VFDT and Numeric Attributes	201
2.2	Further Approaches	202
3	ChiMerge Algorithm	204
4	Online Version of ChiMerge	205
4.1	Time Complexity of Online ChiMerge	208
4.2	Alternative Approaches	209
5	A Comparative Evaluation	210
6	Conclusion	213
References		214

Chapter 11

Mining Chains of Relations		217
Foto Afrati, Gautam Das, Aristides Gionis, Heikki Mannila, Taneli Mielikäinen, Panayiotis Tsaparas		
1	Introduction	217
2	Related Work	219
3	The General Framework	220

3.1	Motivation	222
3.2	Problem Definition	223
3.3	Examples of Properties.....	225
3.4	Extensions of the Model.....	227
4	Algorithmic Tools	229
4.1	A Characterization of Monotonicity.....	230
4.2	Integer Programming Formulations	231
4.3	Case Studies	233
5	Experiments	238
5.1	Datasets	238
5.2	Problems	239
6	Conclusions.....	241
	References	243
	Author Index	247