

Contents

Preface — V

Acknowledgment — VII

1 Introduction — 1

- 1.1 Objectives — 3
- 1.2 Intended audience — 3
- 1.3 Features of the chapters — 4
- 1.4 Content of the chapters — 5
- 1.5 Text supplements — 7
- 1.5.1 Datasets — 7

2 Dataset — 9

- 2.1 Big data and data science — 12
- 2.2 Data objects, attributes and types of attributes — 15
 - 2.2.1 Nominal attributes and numeric attributes — 18
- 2.3 Basic definitions of real and synthetic data — 19
 - 2.3.1 Real dataset — 20
 - 2.3.2 Synthetic dataset — 20
- 2.4 Real and synthetic data from the fields of economy and medicine — 26
 - 2.4.1 Economy data: Economy (U.N.I.S.) dataset — 26
 - 2.4.2 MS data and content (neurology and radiology data): MS dataset — 26
 - 2.4.3 Clinical psychology data: WAIS-R dataset — 39
- 2.5 Basic statistical descriptions of data — 54
 - 2.5.1 Central tendency: Mean, median and mode — 55
 - 2.5.2 Spread of data — 57
 - 2.5.3 Measures of data dispersion — 60
 - 2.5.4 Graphic displays — 61
- 2.6 Data matrix versus dissimilarity matrix — 65
- References — 67

3 Data preprocessing and model evaluation — 71

- 3.1 Data quality — 71
- 3.2 Data preprocessing: Major steps involved — 72
 - 3.2.1 Data cleaning and methods — 73
 - 3.2.2 Data cleaning as a process — 74
 - 3.2.3 Data integration and methods — 75
- 3.3 Data value conflict detection and resolution — 81

3.4	Data smoothing and methods —	81
3.5	Data reduction —	83
3.6	Data transformation —	84
3.7	Attribute subset selection —	84
3.8	Classification of data —	88
3.8.1	Definition of classification —	89
3.9	Model evaluation and selection —	91
3.9.1	Metrics for evaluating classifier performance —	92
	References —	101
4	Algorithms —	103
4.1	What is an algorithm? —	104
4.1.1	What is the flowchart of an algorithm? —	104
4.1.2	Fundamental concepts of programming —	105
4.1.3	Expressions and repetition statements —	108
4.2	Image on coordinate systems —	112
4.2.1	Pixel coordinates —	112
4.2.2	Color models —	113
4.3	Statistical application with algorithms: Image on coordinate systems —	114
4.3.1	Statistical application with algorithms: BW image on coordinate systems —	116
4.3.2	Statistical application with algorithms: RGB image on coordinate systems —	130
	References —	144
5	Linear model and multilinear model —	147
5.1	Linear model analysis for various data —	156
5.1.1	Application of economy dataset based on linear model —	156
5.1.2	Linear model for the analysis of MS —	159
5.1.3	Linear model for the analysis of mental functions —	162
5.2	Multilinear model algorithms for the analysis of various data —	165
5.2.1	Multilinear model for the analysis of economy (U.N.I.S.) dataset —	165
5.2.2	Multilinear model for the analysis of MS —	166
5.2.3	Multilinear model for the analysis of mental functions —	169
	References —	171
6	Decision Tree —	173
6.1	Decision tree induction —	174
6.2	Attribute selection measures —	178

6.3	Iterative dichotomiser 3 (ID3) algorithm —	178
6.3.1	ID3 algorithm for the analysis of various data —	185
6.4	C4.5 algorithm —	195
6.4.1	C4.5 Algorithm for the analysis of various data —	199
6.5	CART algorithm —	208
6.5.1	CART algorithm for the analysis of various data —	217
	References —	226
7	Naive Bayesian classifier —	229
7.1	Naive Bayesian classifier algorithm (and its types) for the analysis of various data —	237
7.1.1	Naive Bayesian classifier algorithm for the analysis of economy (U.N.I.S.) —	237
7.1.2	Algorithms for the analysis of multiple sclerosis —	241
7.1.3	Naive Bayesian classifier algorithm for the analysis of mental functions —	246
	References —	250
8	Support vector machines algorithms —	251
8.1	The case with data being linearly separable —	251
8.2	The case when the data are linearly inseparable —	257
8.3	SVM algorithm for the analysis of various data —	260
8.3.1	SVM algorithm for the analysis of economy (U.N.I.S.) Data —	260
8.3.2	SVM algorithm for the analysis of multiple sclerosis —	263
8.3.3	SVM algorithm for the analysis of mental functions —	267
	References —	270
9	<i>k</i>-Nearest neighbor algorithm —	273
9.1	<i>k</i> -Nearest algorithm for the analysis of various data —	277
9.1.1	<i>k</i> -Nearest algorithm for the analysis of Economy (U.N.I.S.) —	277
9.1.2	<i>k</i> -Nearest algorithm for the analysis of multiple sclerosis —	280
9.1.3	<i>k</i> -Nearest algorithm for the analysis of mental functions —	284
	References —	287
10	Artificial neural networks algorithm —	289
10.1	Classification by backpropagation —	289
10.1.1	A multilayer feed-forward neural network —	290
10.2	Feed-forward backpropagation (FFBP) algorithm —	291
10.2.1	FFBP algorithm for the analysis of various data —	297
10.3	LVQ algorithm —	311
10.3.1	LVQ algorithm for the analysis of various data —	313
	References —	321

11	Fractal and multifractal methods with ANN — 323
11.1	Basic descriptions of fractal — 323
11.2	Fractal dimension — 334
11.3	Multifractal methods — 335
11.3.1	Two-dimensional fractional Brownian motion — 335
11.3.2	Hölder regularity — 337
11.3.3	Fractional Brownian motion — 338
11.4	Multifractal analysis with LVQ algorithm — 352
11.4.1	Polynomial Hölder function with LVQ algorithm for the analysis of various data — 355
11.4.2	Exponential Hölder function with LVQ algorithm for the analysis of various data — 363
	References — 370
Index	— 375