
Contents

1	Introduction	1
1.1	It's All About Data	1
1.2	Data Analytics, Data Mining, and Knowledge Discovery	2
	References	3
2	Data and Relations	5
2.1	The Iris Data Set	5
2.2	Data Scales	8
2.3	Set and Matrix Representations	10
2.4	Relations	11
2.5	Dissimilarity Measures	11
2.6	Similarity Measures	14
2.7	Sequence Relations	16
2.8	Sampling and Quantization	18
	Exercises	20
	References	21
3	Data Preprocessing	23
3.1	Error Types	23
3.2	Error Handling	26
3.3	Filtering	27
3.4	Data Transformation	33
3.5	Data Integration	35
	Exercises	36
	References	37
4	Data Visualization	39
4.1	Diagrams	39
4.2	Principal Component Analysis	41
4.3	Multidimensional Scaling	45
4.4	Sammon Mapping	50
4.5	Auto-Encoder	53

4.6	Histograms.....	54
4.7	Spectral Analysis	57
	Exercises	60
	References.....	61
5	Correlation	63
5.1	Linear Correlation	63
5.2	Correlation and Causality	65
5.3	Chi-Square Test for Independence	67
	Exercises	69
	References.....	70
6	Regression	71
6.1	Linear Regression	71
6.2	Linear Regression with Nonlinear Substitution.....	76
6.3	Robust Regression.....	76
6.4	Neural Networks	77
6.5	Radial Basis Function Networks	82
6.6	Cross-Validation.....	85
6.7	Feature Selection	87
	Exercises	87
	References.....	89
7	Forecasting	91
7.1	Finite State Machines	91
7.2	Recurrent Models	93
7.3	Autoregressive Models.....	94
	Exercises	95
	References.....	96
8	Classification	99
8.1	Classification Criteria	99
8.2	Naive Bayes Classifier	103
8.3	Linear Discriminant Analysis.....	106
8.4	Support Vector Machine	108
8.5	Nearest Neighbor Classifier.....	110
8.6	Learning Vector Quantization	111
8.7	Decision Trees.....	112
	Exercises	117
	References.....	119
9	Clustering	121
9.1	Cluster Partitions	121
9.2	Sequential Clustering	123
9.3	Prototype-Based Clustering.....	125

9.4	Fuzzy Clustering	128
9.5	Relational Clustering.....	133
9.6	Cluster Tendency Assessment	137
9.7	Cluster Validity	138
9.8	Self-organizing Map	139
	Exercises	140
	References.....	141
10	Solutions	145
	Exercises from Chap. 2	145
	Exercises from Chap. 3	147
	Exercises from Chap. 4	150
	Exercises from Chap. 5	154
	Exercises from Chap. 6	156
	Exercises from Chap. 7	161
	Exercises from Chap. 8	164
	Exercises from Chap. 9	172
11	Brief Review of Some Optimization Methods	177
	11.1 Optimization with Derivatives	177
	11.2 Gradient Descent	178
	11.3 Lagrange Optimization.....	179
	References.....	181
	Index.....	183