

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
Problems	21
References	22
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
Problems	36
References	36
4 Data Visualization	37
4.1 Diagrams	37
4.2 Principal Component Analysis	39
4.3 Multidimensional Scaling	43
4.4 Sammon Mapping	47

4.5	Auto-encoder	51
4.6	Histograms	51
4.7	Spectral Analysis	54
	Problems	58
	References	59
5	Correlation	61
5.1	Linear Correlation	61
5.2	Correlation and Causality	63
5.3	Chi-Square Test for Independence	64
	Problems	67
	References	68
6	Regression	69
6.1	Linear Regression	69
6.2	Linear Regression with Nonlinear Substitution	74
6.3	Robust Regression	75
6.4	Neural Networks	75
6.5	Radial Basis Function Networks	81
6.6	Cross-Validation	82
6.7	Feature Selection	85
	Problems	86
	References	87
7	Forecasting	89
7.1	Finite State Machines	89
7.2	Recurrent Models	91
7.3	Autoregressive Models	92
	Problems	93
	References	94
8	Classification	95
8.1	Classification Criteria	95
8.2	Naive Bayes Classifier	99
8.3	Linear Discriminant Analysis	102
8.4	Support Vector Machine	104
8.5	Nearest Neighbor Classifier	106
8.6	Learning Vector Quantization	107
8.7	Decision Trees	108
	Problems	113
	References	114

9 Clustering	117
9.1 Cluster Partitions	117
9.2 Sequential Clustering.....	119
9.3 Prototype-Based Clustering	121
9.4 Fuzzy Clustering.....	123
9.5 Relational Clustering	129
9.6 Cluster Tendency Assessment	133
9.7 Cluster Validity	134
9.8 Self-organizing Map.....	135
Problems	137
References	137
A Brief Review of Some Optimization Methods	141
A.1 Optimization with Derivatives	141
A.2 Gradient Descent	142
A.3 Lagrange Optimization	143
References	145
Solutions	147
Index	155