

Contents

CHAPTER

I. INTRODUCTION	1
I.1. Clustering	1
I.1. Examples of Clustering	1
I.3. Functions of Clustering	6
I.4. Statistics and Data Analysis	7
I.5. Types of Data	9
I.6. Clustering Structures and Models	11
I.7. Algorithms	11
I.8. Interpretation and Evaluation of Clusters	12
I.9. Using This Book	13
References	14
1. PROFILES	28
1.1. Introduction	28
1.2. Profiles Algorithm	29
1.3. Profiles of City Crime	29
1.4. Rank Profiles Algorithm	32
1.5. Linearly Optimal Profiles Algorithm	34
1.6. Linearly Optimal Profiles of Crime Data	36
1.7. Things To Do	38
References	40
Programs	40
2. DISTANCES	58
2.1. Introduction	58
2.2. Euclidean Distances	58
2.3. Relations Between Variables	62
2.4. Disguised Euclidean Distances	64
2.5. Other Distances	64
2.6. Plotting Distances To Detect Clusters	65

2.7.	Things To Do	66
	References	67
	Programs	69
3.	QUICK PARTITION ALGORITHMS	74
3.1.	Introduction	74
3.2.	Leader Algorithm	75
3.3.	Leader Algorithm Applied To Jigsaw Puzzle	76
3.4.	Properties of Leader Algorithm	77
3.5.	Sorting Algorithm	78
3.6.	Sorting Algorithm Applied To Jigsaw Puzzle	79
3.7.	Properties of Sorting Algorithm	80
3.8.	Things To Do	80
	Programs	83
4.	THE <i>K</i> -MEANS ALGORITHM	84
4.1.	Introduction	84
4.2.	<i>K</i> -means Algorithm	85
4.3.	<i>K</i> -means Applied To Food Nutrient Data	87
4.4.	Analysis of Variance	89
4.5.	Weights	91
4.6.	Other Distances	92
4.7.	The Shape of <i>K</i> -means Clusters	93
4.8.	Significance Tests	97
4.9.	Things To Do	100
	References	105
	Programs	107
5.	MIXTURES	113
5.1.	Introduction	113
5.2.	Normal Mixture Algorithm	116
5.3.	Normal Mixture Algorithm Applied To New Haven School Scores	118
5.4.	Things To Do	120
	Programs	125
6.	PARTITION BY EXACT OPTIMIZATION	130
6.1.	Introduction	130
6.2.	Fisher Algorithm	130
6.3.	Fisher Algorithm Applied To Olympic Times	132
6.4.	Significance Testing and Stopping Rules	135
6.5.	Time and Space	137
6.6.	Things To Do	138
	References	140
	Programs	141

7.	THE DITTO ALGORITHM	143
	7.1. Introduction	143
	7.2. Ditto Algorithm	143
	7.3. Application of Ditto Algorithm To Wines	145
	7.4. Things To Do	147
	Programs	148
8.	DRAWING TREES	155
	8.1. Definition of a Tree	155
	8.2. Reordering To Contiguous Clusters	156
	8.3. Application of Reordering To Animal Clusters	156
	8.4. Naming Clusters	156
	8.5. <i>I</i> Representation of Clusters With Diameters	157
	8.6. <i>I</i> Representation of Animal Clusters	157
	8.7. Trees and Directed Graphs	158
	8.8. Linear Representations of Trees	158
	8.9. Trees and Distances	160
	8.10. Block Representations of Trees	162
	8.11. Things To Do	163
	References	164
	Programs	165
9.	QUICK TREE CALCULATION	169
	9.1. Introduction	169
	9.2. Leader Algorithm For Trees	169
	9.3. Tree-leader Algorithm Applied To Mammals' Teeth	171
	9.4. Things To Do	172
	Programs	176
10.	TRIADS	177
	10.1. Introduction	177
	10.2. Triads Algorithm	177
	10.3. Triads Algorithm Applied to Hardware	179
	10.4. Properties of Triads Algorithm	181
	10.5. Triads-leader Algorithm	181
	10.6. Application of Triads-leader Algorithm To Expectation of Life	182
	10.7. Remarks On Triads-leader Algorithm	184
	10.8. Things To Do	185
	References	187
	Programs	187
11.	SINGLE-LINKAGE TREES	191
	11.1. Introduction	191
	11.2. Single-linkage Algorithm	191
	11.3. Application of Single-linkage Algorithm To Airline Distances	193
	11.4. Computational Properties of Single Linkage	195
	11.5. Spiral Search Algorithm	196

11.6.	Application of Spiral Search Algorithm To Births and Deaths	196
11.7.	Single-linkage Clusters From Partitions	199
11.8.	Joining and Splitting.	199
11.9.	Ultrametrics	200
11.10.	Strung-out Clusters	200
11.11.	Minimum Spanning Trees	201
11.12.	Reality of Clusters	202
11.13.	Density-contour Tree	205
11.14.	Densities and Connectedness, Distances Given	208
11.15.	Things To Do	209
	References	212
	Programs	214
12.	DISTANCE AND AMALGAMATION ALGORITHMS	216
12.1.	Introduction	216
12.2.	Joining Algorithm	216
12.3.	Joining Algorithm Applied To Ivy League Football	218
12.4.	Remarks On Joining Algorithm	222
12.5.	Adding Algorithm	222
12.6.	Adding Algorithm Applied To Questionnaire	223
12.7.	Things To Do	227
	References	230
	Programs	230
13.	MINIMUM MUTATION METHODS	233
13.1.	Introduction	233
13.2.	Minimum Mutation Fits	233
13.3.	Application of Minimum Mutation Algorithm To Cerci In Insects	235
13.4.	Some Probability Theory For the Number of Mutations	236
13.5.	Reduced Mutation Tree	237
13.6.	Application of Reduced Mutation Algorithm To Amino Acid Sequences	239
13.7.	Things To Do	241
	References	245
	Programs	246
14.	DIRECT SPLITTING	251
14.1.	Introduction	251
14.2.	Binary Splitting Algorithm	252
14.3.	Application of Binary Splitting Algorithm To Voting Data With Missing Values.	255
14.4.	One-way Splitting Algorithm	257
14.5.	One-way Splitting Algorithm Applied To Republican Percentages	258
14.6.	Two-way Splitting Algorithm	260
14.7.	Two-way Splitting Algorithm Applied To Republican Vote for President	262

14.8.	Things To Do	264
	References	268
	Programs	269
15.	DIRECT JOINING	278
15.1.	Introduction	278
15.2.	Two-way Joining Algorithm	280
15.3.	Application of Two-way Joining Algorithm To Candida	282
15.4.	Generalizations of Two-way Joining Algorithm	284
15.5.	Significance Tests for Outcomes of Two-way joining Algorithm	285
15.6.	Direct Joining Algorithm for Variables on Different Scales	286
15.7.	Things To Do	288
	Programs	293
16.	SIMULTANEOUS CLUSTERING AND SCALING	299
16.1.	Introduction	299
16.2.	Scaling Ordered Variables	300
16.3.	Scaling Ordered Variables Applied To U.N. Questions	302
16.4.	Joiner-scaler	303
16.5.	Application of Joiner-scaler Algorithm To U.N. Votes	305
16.6.	Things To Do	308
	References	312
17.	FACTOR ANALYSIS	313
17.1.	Introduction	313
17.2.	Sparse Root Algorithm	314
17.3.	Sparse Root Algorithm Applied To Face Measurements	316
17.4.	Remarks on the Sparse Root Algorithm	316
17.5.	Rotation to Simple Structure	318
17.6.	Joining Algorithm for Factor Analysis	319
17.7.	Application of Joining Algorithm To Physical Measurements Data	320
17.8.	Things To Do	323
	References	325
	Programs	326
18.	PREDICTION	330
18.1.	Introduction	330
18.2.	Variance Components Algorithm	332
18.3.	Variance Components Algorithm Applied To Prediction of Leukemia Mortality Rates	333
18.4.	Alternatives To Variance Components Algorithm	337
18.5.	Automatic Interaction Detection	337
18.6.	Application of AID Algorithm To Leukemia Mortality	338
18.7.	Remarks On The AID Algorithm	340
18.8.	Things To Do	341
	Programs	343
	INDEX	347