

Contents

- 1. Introduction** **1**
 - 1. Motivation, 1
 - 2. Types of Data and How to Handle Them, 3
 - 2.1 Interval-Scaled Variables, 4
 - 2.2 Dissimilarities, 16
 - 2.3 Similarities, 20
 - 2.4 Binary Variables, 22
 - 2.5 Nominal, Ordinal, and Ratio Variables, 28
 - 2.6 Mixed Variables, 32
 - 3. Which Clustering Algorithm to Choose, 37
 - 3.1 Partitioning Methods, 38
 - 3.2 Hierarchical Methods, 44
 - 4. A Schematic Overview of Our Programs, 50
 - 5. Computing Dissimilarities with the Program DAISY, 52
 - Exercises and Problems, 63

- 2. Partitioning Around Medoids (Program PAM)** **68**
 - 1. Short Description of the Method, 68
 - 2. How to Use the Program PAM, 72
 - 2.1 Interactive Use and Input, 72
 - 2.2 Output, 80
 - 2.3 Missing Values, 88
 - 3. Examples, 92
 - *4. More on the Algorithm and the Program, 102
 - 4.1 Description of the Algorithm, 102
 - 4.2 Structure of the Program, 104

- *5. Related Methods and References, 108
 - 5.1 The k -Medoid Method and Optimal Plant Location, 108
 - 5.2 Other Methods Based on the Selection of Representative Objects, 110
 - 5.3 Methods Based on the Construction of Central Points, 111
 - 5.4 Some Other Nonhierarchical Methods, 116
 - 5.5 Why Did We Choose the k -Medoid Method?, 117
 - 5.6 Graphical Displays, 119

Exercises and Problems, 123

3. Clustering Large Applications (Program CLARA) 126

- 1. Short Description of the Method, 126
- 2. How to Use the Program CLARA, 127
 - 2.1 Interactive Use and Input, 127
 - 2.2 Output, 130
 - 2.3 Missing Values, 134
- 3. An Example, 139
- *4. More on the Algorithm and the Program, 144
 - 4.1 Description of the Algorithm, 144
 - 4.2 Structure of the Program, 146
 - 4.3 Limitations and Special Messages, 151
 - 4.4 Modifications and Extensions of CLARA, 153
- *5. Related Methods and References, 155
 - 5.1 Partitioning Methods for Large Data Sets, 155
 - 5.2 Hierarchical Methods for Large Data Sets, 157
 - 5.3 Implementing CLARA on a Parallel Computer, 160

Exercises and Problems, 162

4. Fuzzy Analysis (Program FANNY) 164

- 1. The Purpose of Fuzzy Clustering, 164
- 2. How to Use the Program FANNY, 166
 - 2.1 Interactive Use and Input, 167
 - 2.2 Output, 170
- 3. Examples, 175
- *4. More on the Algorithm and the Program, 182
 - 4.1 Description of the Algorithm, 182
 - 4.2 Structure of the Program, 188

- *5. Related Methods and References, 189
 - 5.1 Fuzzy k -Means and the MND2 Method, 189
 - 5.2 Why Did We Choose FANNY?, 191
 - 5.3 Measuring the Amount of Fuzziness, 191
 - 5.4 A Graphical Display of Fuzzy Memberships, 195
- Exercises and Problems, 197

5. Agglomerative Nesting (Program AGNES) 199

- 1. Short Description of the Method, 199
- 2. How to Use the Program AGNES, 208
 - 2.1 Interactive Use and Input, 208
 - 2.2 Output, 209
- 3. Examples, 214
- *4. More on the Algorithm and the Program, 221
 - 4.1 Description of the Algorithm, 221
 - 4.2 Structure of the Program, 223
- *5. Related Methods and References, 224
 - 5.1 Other Agglomerative Clustering Methods, 224
 - 5.2 Comparing Their Properties, 238
 - 5.3 Graphical Displays, 243
- Exercises and Problems, 250

6. Divisive Analysis (Program DIANA) 253

- 1. Short Description of the Method, 253
- 2. How to Use the Program DIANA, 259
- 3. Examples, 263
- *4. More on the Algorithm and the Program, 271
 - 4.1 Description of the Algorithm, 271
 - 4.2 Structure of the Program, 272
- *5. Related Methods and References, 273
 - 5.1 Variants of the Selected Method, 273
 - 5.2 Other Divisive Techniques, 275
- Exercises and Problems, 277

7. Monothetic Analysis (Program MONA) 280

- 1. Short Description of the Method, 280
- 2. How to Use the Program MONA, 283

| | | |
|-----|--|--|
| 2.1 | Interactive Use and Input, 284 | |
| 2.2 | Output, 287 | |
| 3. | Examples, 290 | |
| *4. | More on the Algorithm and the Program, 298 | |
| 4.1 | Description of the Algorithm, 298 | |
| 4.2 | Structure of the Program, 301 | |
| *5. | Related Methods and References, 304 | |
| 5.1 | Association Analysis, 304 | |
| 5.2 | Other Monothetic Divisive Algorithms for Binary Data, 307 | |
| 5.3 | Some Other Divisive Clustering Methods, 308 | |
| | Exercises and Problems, 310 | |

| | | |
|----------------------|---|------------|
| APPENDIX | | 312 |
| 1. | Implementation and Structure of the Programs, 312 | |
| 2. | Running the Programs, 313 | |
| 3. | Adapting the Programs to Your Needs, 316 | |
| 4. | The Program CLUSPLOT, 318 | |
| References | | 320 |
| Author Index | | 332 |
| Subject Index | | 335 |