

Contents

Preface	v
1 An Introduction to Classification and Clustering	1
1.1 Introduction	1
1.2 Reasons for classifying	2
1.3 Numerical methods of classification	4
1.4 What is a cluster?	6
1.5 Examples of the use of clustering	8
1.6 Summary	10
2 The Initial Examination of Multivariate Data	11
2.1 Introduction	11
2.2 Univariate and bivariate views	11
2.3 Accommodating extra variable values on a scattergram	12
2.4 Icon representations of multivariate data	12
2.5 Andrews' plots	21
2.6 The use of derived variables	26
2.7 Summary	33
3 Measurement of Similarity, Dissimilarity and Distance	37
3.1 Introduction	37
3.2 The choice of variables	37
3.3 Similarity measures	40
3.4 Dissimilarity and distance measures	46
3.5 Between group similarity and distance measures	47
3.6 The graphical display of distance matrices	50
3.7 Summary	51
4 Hierarchical Clustering Techniques	55
4.1 Introduction	55
4.2 Agglomerative methods	56
4.3 Lance and Williams' recurrence formula	66
4.4 Properties and problems of agglomerative hierarchical clustering techniques	68
4.5 Global fit of a hierarchical clustering solution	72
4.6 Partitions from a hierarchy—the number of groups problem	73
4.7 Applications of agglomerative techniques	74
4.8 Divisive methods	82
4.9 Summary	88

5	Optimization Methods for Cluster Analysis	91
5.1	Introduction	91
5.2	Clustering criteria	91
5.3	Optimizing the clustering criterion	93
5.4	Properties of, and problems with, optimization clustering techniques	96
5.5	Selecting the number of groups	100
5.6	Applications of optimization methods	102
5.7	Summary	107
6	Mixture Models for Cluster Analysis	109
6.1	Introduction	109
6.2	Finite mixture distributions	109
6.3	Normal mixtures	110
6.4	Estimating g , the number of components in a mixture	114
6.5	Applications of normal mixtures	115
6.6	Mixtures for categorical data—latent class analysis	117
6.7	Mixture models for mixed mode data	122
6.8	Summary	123
7	Other Clustering Techniques	125
7.1	Introduction	125
7.2	Density search clustering techniques	125
7.3	Clumping techniques	130
7.4	Constrained classification	133
7.5	Simultaneous clustering of individuals and variables	134
7.6	Summary	137
8	Some Final Comments and Guidelines	141
8.1	Introduction	141
8.2	Using clustering techniques in practice	141
8.3	Assessing clustering solutions	142
8.4	Comparing classifications	146
8.5	Summary	148
	Appendix A Software	151
A.1	Introduction	151
A.2	BMDP	151
A.3	SPSS	152
A.4	SAS	152
A.5	CLUSTAN	152
A.6	SYSTAT	153
A.7	NTSYS	153
A.8	CLUSTAR	153
	References	155
	Index	165