

Contents

PART I THE MEASUREMENT OF DISSIMILARITY

1. Kinds of Similarity and Dissimilarity

1.1 Terminology	3
1.2 Kinds of similarity and dissimilarity	4
1.3 Similarity and dissimilarity between individuals	6
1.4 Dissimilarity between classes of individuals	8

2. Information Theory

2.1 Background	10
2.2 Information gain	11
2.3 Information radius	12
2.4 Product information radius	14
2.5 Normal information radius	14

3. K -dissimilarity

3.1 Information radius and D -dissimilarity for one attribute	16
3.2 The construction of K -dissimilarity	17
3.3 Conditional definition	17

4. Correlation and Weighting of Attributes

4.1 Introduction	21
4.2 Weighting of attributes	21
4.3 Correlation and redundancy	25

5. Other Measures of Similarity and Dissimilarity

5.1	Introduction	29
5.2	Association measures	30
5.3	Correlation coefficients	31
5.4	Taxonomic distances	31
5.5	Measures of I -distinguishability	33

References to Part I	34
---------------------------------------	-----------

PART II CLUSTER ANALYSIS**6. The Form and Analysis of the Data**

6.1	General considerations	39
6.2	Models, methods, and algorithms	42
6.3	Discussion of cluster methods	44

7. Hierarchic Cluster Methods

7.1	DC's and dendrograms	45
7.2	The ultrametric inequality	49
7.3	Type A cluster methods	51
7.4	Properties of A_S, A_W, A_U, A_C	54
7.5	Other methods	57

8. Non-Hierarchic (Overlapping) Stratified Cluster Methods

8.1	The basis of the generalization	59
8.2	Numerically stratified clusterings	61
8.3	Subdominant methods	63
8.4	Constraints on NSC's	64
8.5	The methods B_k	65
8.6	The methods B_k^c	69
8.7	The methods C_u	71
8.8	Other type C methods	74
8.9	Other non-hierarchic methods	74
8.10	Summary	75

9. An Axiomatic Approach to Cluster Analysis

9.1	Definitions and constructions	77
9.2	The nature of cluster methods	81
9.3	Conditions on cluster methods	83
9.4	Sequences of methods: fitting conditions	89

10. Flat Cluster Methods

10.1 Monotone equivariance	92
10.2 Flat cluster methods: the flatness and reduction theorems	93
10.3 Measures of association	99

11. Evaluation of Clusterings

11.1 The nature of the problems	102
11.2 Comparison of input and output DC's	103
11.3 Scale-free comparison of DC's	106
11.4 Comparison of relations	107
11.5 Random effects: classifiability and stability	110
11.6 Validity of clusters	111

12. Other Approaches to Cluster Analysis

12.1 Introduction	113
12.2 Cluster analysis without a dissimilarity measure	113
12.3 Simple cluster analysis based on a dissimilarity measure	115
12.4 ML-sets and associated clusterings	118
12.5 Conclusions	119

References to Part II	120
--	------------

PART III MATHEMATICAL AND BIOLOGICAL TAXONOMY**13. The Components and Aims of Biological Taxonomy**

13.1 The components of biological taxonomy	127
13.2 The demarcation of taxa of specific rank	130
13.3 Supraspecific classification	134
13.4 Intraspecific classification	144

14. Automatic Classification as a Research Tool

14.1 Introduction	149
14.2 Automatic classification and the creation of taxa	150
14.3 The stability of classifications	154
14.4 Comparison of classifications based on different criteria	157
14.5 Classification and phylogeny	160
14.6 Patterns of intraspecific variation	163
14.7 Data simplification and hypothesis testing	165

15. The Application of Methods of Automatic Classification

15.1 Introduction	167
15.2 Selection of OTU's	167
15.3 Selection of attributes	169
15.4 Estimation of character states	172
15.5 Strategies for the analysis of taxonomic data	174

References to Part III	178
---	------------

Appendices

Appendix 1 Estimation of information radius	189
Appendix 2 Information radius for normal distributions	202
Appendix 3 Computation of B_k	227
Appendix 4 Computation of C_u	235
Appendix 5 A cluster listing method	237
Appendix 6 Fast hierarchic clustering	240
Appendix 7 Illustrative applications	243
Appendix 8 Bibliographic sources	259

Glossary	265
---------------------------	------------

Index of References	277
--------------------------------------	------------

Subject Index	281
--------------------------------	------------