

Contents

Preface	ix
---------------	----

Part I. Tangles

A new paradigm for clusters and types

1. The idea behind tangles	3
1.1 Tangles in the natural sciences	3
1.2 Tangles in the social sciences	7
1.3 Tangles in data science	8
1.4 On the use of mathematical language in this book	12
2. The notion of a tangle	15
2.1 Features that often occur together	16
2.2 Consistency of features	17
2.3 From consistency to tangles	18
2.4 Principal and black hole tangles: two simple examples	21
2.5 Guiding sets and functions	23
3. The two main tangle theorems: an informal preview	27
3.1 The first tangle theorem: how tangles structure our data	27
3.2 The second tangle theorem: structure when there is no tangle ...	28
3.3 The predictive power of tangles	29

Part II. Tangles in Different Contexts

A collection of informal examples

4. Examples from the natural sciences	33
4.1 Expert systems: what is an illness?	34
4.2 DNA tangles: identifying organisms from imperfect data	36
4.3 Drug development: chance discovery or focussed effort?	39
5. Examples from the social sciences	41
5.1 Sociology: discovering mindsets and social groupings	42
5.2 Psychology: understanding the unfamiliar	44
5.3 Analytic philosophy: quantifying family resemblances	46
5.4 Politics and society: appointing representative bodies	49
5.5 Education: combining teaching techniques into methods	50
6. Examples from data science	53
6.1 Indirect clustering by separation: hard, soft or fuzzy	55
6.2 Tangles of pixels: clusters in images	61
6.3 E-commerce: identifying customer and product types	64
6.4 Tangles of words: identifying topics, genres, authors	67
6.5 Semantics: teaching computers meaning	69

Part III. The Mathematics of Tangles

Concepts, theorems, algorithms

7. The formal setup for tangles	75
7.1 Tangles in graphs: how it all began	76
7.2 Tangles of feature systems	78
7.3 Why triples?	81
7.4 The evolution of tangles: hierarchies and order	84
7.5 Cutting corners: submodularity of features and functions	88
7.6 Advanced feature systems	91
7.7 Duality of feature systems	94
8. Tangle theorems	99
8.1 Tree sets of partitions	100
8.2 Tangle-distinguishing feature sets	102
8.3 Tangle-precluding feature sets	107

9. Order functions	111
9.1 Order functions based on point similarity	115
9.2 Advanced order functions based on point similarity	119
9.3 Order functions based on feature similarity	124
9.4 Algebraic order functions	128
9.5 Entropy-based order functions	137
9.6 Entropy-based order functions for the dual system	146
9.7 Order functions based on tangles of the dual system	148
9.8 Tweaking order functions to increase balance of partitions	150
9.9 Using order functions for deliberate bias	153
10. Choosing the feature system	157
10.1 Random partitions, locally optimized	158
10.2 Spectral feature systems	159
10.3 PCA-related feature systems	163
11. Algorithms	169
11.1 Generating the feature systems	169
11.2 Computing the tangles	174
11.3 Computing a tree of tangles	177
11.4 Computing a tree set certifying that no tangles exist	181

Part IV. Applying Tangles

Back to the examples

12. Applying tangles in the natural sciences	187
12.1 Predictions	187
12.2 Expert systems: fuzziness versus structure in medical diagnostics	190
12.3 DNA and protein tangles: aligned and alignment-free	195
12.4 Drug development: targeting tangles of pathogens	195
13. Applying tangles in the social sciences	203
13.1 Sociology: from mindsets to matchmaking	203
13.2 Psychology: diagnostics, new syndromes, and the use of duality .	208
13.3 Political science: appointing representative bodies	210
13.4 Education: devising methods and assigning students	212

14. Applying tangles in data science	215
14.1 Clustering by tangles: hard clusters or none	216
14.2 Clustering by tangles: fuzzy clusters	224
14.3 Tangles are not clusters in the feature space	227
14.4 Tangles versus decision trees and hierarchical search	228
14.5 Clustering by advanced feature tangles, versus spectral clustering	232
14.6 Image segmentation, identification, compression	234
14.7 E-Commerce: customer-product duality	242
14.8 Text tangles: direct and indirect	245
14.9 Semantics and order functions	247
14.10 Acquiring meaning: an interactive thesaurus	248
Notes	251
References	273
Index	277
Symbol index	283