

目 次

まえがき	iii
訳者序文	v
第I章 序 章	1
I.1 クラスタ分析	1
I.2 クラスタ分析の例	3
I.2.1 動物と植物	3
I.2.2 医 学	4
I.2.3 精神医学	6
I.2.4 考古学と人類学	7
I.2.5 植物社会学	7
I.2.6 そ の 他	8
I.3 クラスタ分析の機能	8
I.4 統計学とデータ解析	10
I.5 データの型	12
I.5.1 類 似 性	12
I.5.2 ケース×変数	13
I.5.3 他のデータ構造	14
I.6 クラスタ分析の構造とモデル	15
I.7 アルゴリズム	16
I.7.1 分 類	16
I.7.2 入れ換え	16
I.7.3 結 合	17
I.7.4 分 割	17
I.7.5 加 法	17
I.7.6 探 査	18
I.8 クラスタの解釈と評価	18
I.9 本書の使い方	20

参考文献	21
第1章 プロフィール	40
1.1 はじめに	40
1.2 プロフィール・アルゴリズム	41
1.3 都市犯罪のプロフィール	42
1.4 順位プロフィール・アルゴリズム	46
1.5 線形最適プロフィール・アルゴリズム	47
1.6 犯罪データの線形最適プロフィール	51
1.7 実行事項	53
1.7.1 固有値	53
1.7.2 直線	55
1.7.3 双線	55
1.7.4 顔	55
1.7.5 箱	55
1.7.6 ブロック・ヒストグラム	55
参考文献	56
プログラム	56
第2章 距離	75
2.1 はじめに	75
2.2 ユークリッド距離	75
2.3 変数間の関係	81
2.4 見かけのユークリッド距離	83
2.5 他の距離	84
2.6 クラスターを見つけるための距離のプロット	85
2.7 実行事項	85
2.7.1 変数間の距離	85
2.7.2 距離の分布	87
参考文献	87
プログラム	90
第3章 迅速分割アルゴリズム	96
3.1 はじめに	96
3.2 リーダー・アルゴリズム	97
3.3 ジグソーパズルに適用されたリーダー・アルゴリズム	99

3.4	リーダー・アルゴリズムの性質	102
3.5	ソーティング・アルゴリズム	102
3.6	ジグソーパズルに適用したソーティング・アルゴリズム	102
3.7	ソーティング・アルゴリズムの性質	104
3.8	実行事項	106
3.8.1	リーダー・アルゴリズムの実行	106
3.8.2	改善されたリーダー・アルゴリズム	106
3.8.3	クラスターの数	106
3.8.4	クラスターのサイズ	106
	プログラム	107
第4章 K-平均アルゴリズム		109
4.1	はじめに	109
4.2	K -平均アルゴリズム	112
4.3	食品栄養データに適用した K -平均アルゴリズム	113
4.4	分散分析	116
4.5	ウェイト	119
4.6	他の距離	119
4.7	K -平均クラスターの形	122
4.8	有意性検定	126
4.9	実行事項	131
4.9.1	K -平均の実行	131
4.9.2	K -平均アルゴリズムの諸法	131
4.9.3	全体的最適における限界	133
4.9.4	他の基準	134
4.9.5	漸近理論	135
4.9.6	二段サンプリング	137
4.9.7	対称的パラドックス	137
4.9.8	大きなデータ集合	137
	参考文献	138
	プログラム	140
第5章 混 合		146
5.1	はじめに	146
5.2	正規混合アルゴリズム	149

5.3	ニューヘブンの学校の学業成績に適用した正規混合アルゴリズム	152
5.4	実行事項	154
5.4.1	混合アルゴリズムの実行	154
5.4.2	特異点	156
5.4.3	より確率的なアプローチ	158
5.4.4	木	160
5.4.5	初期値化	160
5.4.6	K -平均との一致	161
	プログラム	161
第6章	厳密な最適化による分割	166
6.1	はじめに	166
6.2	フィッシャー・アルゴリズム	168
6.3	オリンピック記録に適用したフィッシャー・アルゴリズム	169
6.4	有意性検定と停止規則	172
6.5	時間と空間	176
6.6	実行事項	178
6.6.1	フィッシャー・アルゴリズムの実行	178
6.6.2	実数値データ	179
6.6.3	密度の推定	179
6.6.4	逐次分割	179
6.6.5	絶対偏差の和の更新	179
	参考文献	179
	プログラム	180
第7章	ディトー・アルゴリズム	183
7.1	はじめに	183
7.2	ディトー・アルゴリズム	183
7.3	ディトー・アルゴリズムのワインへの適用	186
7.4	実行事項	190
7.4.1	ディトー・アルゴリズムの実行	190
	プログラム	190
第8章	木の描画	197
8.1	木の定義	197

8.2	連続するクラスターへの並べかえ	197
8.3	動物クラスターへの並べかえの適用	197
8.4	クラスターの命名	199
8.5	直径をもったクラスターの I-表示	199
8.6	動物クラスターの I-表示	200
8.7	木と有向グラフ	201
8.8	木の直線表現	202
8.9	木と距離	203
8.10	木のブロック表現	206
8.11	実行事項	207
8.11.1	グラフ	207
8.11.2	木を通過すること	208
8.11.3	自然の木	208
	参考文献	209
	プログラム	210

第9章 迅速な木の計算 214

9.1	はじめに	214
9.2	木に関するリーダー・アルゴリズム	214
9.3	哺乳動物の歯に適用した木リーダー・アルゴリズム	217
9.4	実行事項	220
9.4.1	木リーダー・アルゴリズムの実行	220
9.4.2	分類	220
9.4.3	微分的分類	222
9.4.4	フィルター化のアルゴリズム	222
	プログラム	223

第10章 三つ組 225

10.1	はじめに	225
10.2	三つ組アルゴリズム	226
10.3	クギ類に適用した三つ組アルゴリズム	228
10.4	三つ組アルゴリズムの特徴	229
10.5	三つ組リーダー・アルゴリズム	231
10.6	三つ組リーダー・アルゴリズムの平均余命への適用	232
10.7	三つ組リーダー・アルゴリズムに関する注釈	233

10.8 実行事項	236
10.8.1 三つ組アルゴリズムの実行	236
10.8.2 完全な探索	236
10.8.3 公理	238
10.8.4 順位	238
10.8.5 確率モデル	238
参考文献	238
プログラム	239
第11章 SL 木	243
11.1 はじめに	243
11.2 SL アルゴリズム	245
11.3 SL アルゴリズムの航空距離への適用	245
11.4 SL の計算上の特性	249
11.5 螺旋探索アルゴリズム	249
11.6 螺旋探索アルゴリズムの出生と死亡への適用	250
11.7 分割から見た SL クラスタ	253
11.8 結合と分離	254
11.9 超距離	255
11.10 クラスタの配列	256
11.11 最小生成木	256
11.12 クラスタの現実性	258
11.13 等密度木	261
11.14 距離が与えられた場合の密度と連結性	265
11.15 実行事項	268
11.15.1 SL アルゴリズムの実行	268
11.15.2 等密度と SL	268
11.15.3 モード (mode)	268
11.15.4 密度推定値	269
11.15.5 クラスタ数の減少	271
11.15.6 凸性	271
11.15.7 距離計算を避けること	271
参考文献	271
プログラム	274

第 12 章 距離と融合のアルゴリズム276

12.1 はじめに	276
12.2 結合アルゴリズム	276
12.3 結合アルゴリズムの Ivy フットボールへの適用	278
12.4 結合アルゴリズムに関する所見	284
12.5 加法アルゴリズム	285
12.6 加法アルゴリズムの質問票 (表 12.4) への適用	286
12.7 実行事項	291
12.7.1 距離と融合のアルゴリズムの実行	291
12.7.2 加重平均と加重しない平均	291
12.7.3 分割表	294
12.7.4 遞減する直径	294
12.7.5 確率モデル	294
参考文献	294
プログラム	295

第 13 章 最小変異法298

13.1 はじめに	298
13.2 最小変異適合	299
13.3 最小変異アルゴリズムの昆虫の尾角への適用	301
13.4 変異の数に関する確率理論	302
13.5 縮小変異木	305
13.6 縮小変異アルゴリズムのアミノ酸の列への適用	307
13.7 実行事項	309
13.7.1 最小変異技法の使用	309
13.7.2 一意性	309
13.7.3 確率モデル	310
13.7.4 ディトー分析	310
13.7.5 根を持たない木	313
13.7.6 実変数	313
13.7.7 反復	314
参考文献	315
プログラム	316

第 14 章 直接分割法	321
14.1 はじめに	321
14.2 2成分分割アルゴリズム	324
14.3 欠損値をもった投票データへの2成分分割アルゴリズムの適用	326
14.4 一重分割アルゴリズム	328
14.5 一重分割アルゴリズムの共和党投票率への適用	330
14.6 二重分割アルゴリズム	332
14.7 二重分割アルゴリズムの大統領選における共和党投票への適用	335
14.8 実行事項	339
14.8.1 分割アルゴリズムの実行	339
14.8.2 欠損値	341
14.8.3 誤差分析	341
14.8.4 レンジ	341
14.8.5 2成分分割とレンジ	341
14.8.6 実験	343
14.8.7 一重モデル	344
参考文献	344
プログラム	345
第 15 章 直接結合法	354
15.1 はじめに	354
15.2 二重結合アルゴリズム	358
15.3 二重結合アルゴリズムの酵母菌データへの適用	359
15.4 二重結合アルゴリズムの一般化	362
15.5 二重結合アルゴリズムの結果に対する有意性検定	363
15.6 異なる尺度で測定された変数のための直接結合アルゴリズム	363
15.7 実行事項	366
15.7.1 二重結合アルゴリズムの実行	366
15.7.2 二重クラスター分析モデル	367
15.7.3 距離, 融合とブロック	371
15.7.4 レンジ	372
15.7.5 計算費用	372
15.7.6 同一の行と列	372
15.7.7 最小の数のブロック	373
15.7.8 順序不変性	373

プログラム	374
第 16 章 クラスタ分析と尺度構成の同時化	380
16.1 はじめに	380
16.2 順序変数の尺度構成	382
16.3 国連の議題に適用された順序変数の尺度構成	383
16.4 結合尺度法	386
16.5 結合尺度法アルゴリズムの国連投票データへの適用	389
16.6 実行事項	393
16.6.1 結合尺度法の実行	393
16.6.2 単調な部分列	393
16.6.3 カテゴリーデータ	395
16.6.4 連続データ	395
16.6.5 より大きな一般性	396
16.6.6 メディアン回帰	397
参考文献	397
第 17 章 因子分析	398
17.1 はじめに	398
17.2 SR 法アルゴリズム	400
17.3 SR 法アルゴリズムの顔の測定データへの適用	402
17.4 SR 法アルゴリズムに関する注意	403
17.5 単純構造への回帰	403
17.6 因子分析のための結合アルゴリズム	406
17.7 結合アルゴリズムの身体測定データへの適用	409
17.8 実行事項	413
17.8.1 因子分析アルゴリズムの実行	413
17.8.2 データ行列の直接の因子化	414
参考文献	415
プログラム	417
第 18 章 予 測	422
18.1 はじめに	422
18.2 分散成分アルゴリズム	425
18.3 分散成分アルゴリズムの白血病死亡率への適用	427

18.4	分散成分アルゴリズムに代るもの	431
18.5	AID 法	432
18.6	AID アルゴリズムの白血病死亡率への適用	433
18.7	AID アルゴリズムについての注意	436
18.8	実行事項	438
18.8.1	予測アルゴリズムの実行	438
18.8.2	行または列による予測	438
18.8.3	ブロックモデル	439
18.8.4	メディアン適合	440
	プログラム	440
	索引	445

