

Contents

	Page
<i>Preface</i>	xv
1 Introduction	1
1.1 Cluster Analysis	1
1.2 Classification	4
1.3 Examples	7
1.4 Software	12
1.5 Organization of the Book	13
1.6 Bibliographic Notes	14
2 Model-based Clustering: Basic Ideas	15
2.1 Finite Mixture Models	15
2.2 Geometrically Constrained Multivariate Normal Mixture Models	18
2.3 Estimation by Maximum Likelihood	23
2.4 Initializing the EM Algorithm	31
2.5 Examples with Known Number of Clusters	39
2.6 Choosing the Number of Clusters and the Clustering Model	46
2.7 Illustrative Analyses	60
2.8 Who Invented Model-based Clustering?	71
2.9 Bibliographic Notes	75
3 Dealing with Difficulties	79
3.1 Outliers	79
3.2 Dealing with Degeneracies: Bayesian Regularization	92
3.3 Non-Gaussian Mixture Components and Merging	97
3.4 Bibliographic Notes	105
4 Model-based Classification	109
4.1 Classification in the Probabilistic Framework	109
4.2 Parameter Estimation	113
4.3 Parsimonious Classification Models	114
4.4 Multinomial Classification	119
4.5 Variable Selection	124
4.6 Mixture Discriminant Analysis	126
4.7 Model Assessment and Selection	127

5	Semi-supervised Clustering and Classification	134
5.1	Semi-supervised Classification	134
5.2	Semi-supervised Clustering	141
5.3	Supervised Classification with Uncertain Labels	144
5.4	Novelty Detection: Supervised Classification with Unobserved Classes	154
5.5	Bibliographic Notes	160
6	Discrete Data Clustering	163
6.1	Example Data	163
6.2	The Latent Class Model for Categorical Data	165
6.3	Model-based Clustering for Ordinal and Mixed Type Data	185
6.4	Model-based Clustering of Count Data	190
6.5	Bibliographic Notes	197
7	Variable Selection	199
7.1	Continuous Variable Selection for Model-based Clustering	199
7.2	Continuous Variable Regularization for Model-based Clustering	208
7.3	Continuous Variable Selection for Model-based Classification	210
7.4	Categorical Variable Selection Methods for Model-based Clustering	211
7.5	Bibliographic Notes	215
8	High-dimensional Data	217
8.1	From Multivariate to High-dimensional Data	217
8.2	The Curse of Dimensionality	221
8.3	Earlier Approaches for Dealing with High-dimensional Data	227
8.4	Subspace Methods for Clustering and Classification	238
8.5	Bibliographic Notes	257
9	Non-Gaussian Model-based Clustering	259
9.1	Multivariate t -Distribution	259
9.2	Skew-normal Distribution	267
9.3	Skew- t Distribution	270
9.4	Box-Cox Transformed Mixtures	278
9.5	Generalized Hyperbolic Distribution	282
9.6	Example: Old Faithful Data	285
9.7	Example: Flow Cytometry	287
9.8	Bibliographic Notes	288
10	Network Data	292
10.1	Introduction	292
10.2	Example Data	294
10.3	Stochastic Block Model	298
10.4	Mixed Membership Stochastic Block Model	304
10.5	Latent Space Models	312
10.6	Stochastic Topic Block Model	320
10.7	Bibliographic Notes	329

11	Model-based Clustering with Covariates	331
11.1	Examples	331
11.2	Mixture of Experts Model	333
11.3	Model Assessment	339
11.4	Software	339
11.5	Results	340
11.6	Discussion	348
11.7	Bibliographic Notes	349
12	Other Topics	351
12.1	Model-based Clustering of Functional Data	351
12.2	Model-based Clustering of Texts	363
12.3	Model-based Clustering for Image Analysis	368
12.4	Model-based Co-clustering	373
12.5	Bibliographic Notes	382
	<i>List of R Packages</i>	384
	<i>Bibliography</i>	386
	<i>Author Index</i>	415
	<i>Subject Index</i>	423