

# Contents

<i>Preface</i>	vii
<i>Notation and basic terminology</i>	xi
<i>Abbreviations</i>	xix
1. Introduction	1
1.1 Recognizing the pattern . . . . .	1
1.2 Dissimilarities for representation . . . . .	2
1.3 Learning from examples . . . . .	4
1.4 Motivation of the use of dissimilarity representations . . . . .	8
1.5 Relation to kernels . . . . .	13
1.6 Outline of the book . . . . .	14
1.7 In summary . . . . .	16
2. Spaces	23
2.1 Preliminaries . . . . .	25
2.2 A brief look at spaces . . . . .	28
2.3 Generalized topological spaces . . . . .	32
2.4 Generalized metric spaces . . . . .	46
2.5 Vector spaces . . . . .	56
2.6 Normed and inner product spaces . . . . .	62
2.6.1 Reproducing kernel Hilbert spaces . . . . .	69
2.7 Indefinite inner product spaces . . . . .	71
2.7.1 Reproducing kernel Kreĭn spaces . . . . .	85
2.8 Discussion . . . . .	87
3. Characterization of dissimilarities	89
3.1 Embeddings, tree models and transformations . . . . .	90

3.1.1	Embeddings . . . . .	90
3.1.2	Distorted metric embeddings . . . . .	95
3.2	Tree models for dissimilarities . . . . .	95
3.3	Useful transformations . . . . .	99
3.3.1	Transformations in semimetric spaces . . . . .	99
3.3.2	Direct product spaces . . . . .	102
3.3.3	Invariance and robustness . . . . .	103
3.4	Properties of dissimilarity matrices . . . . .	105
3.4.1	Dissimilarity matrices . . . . .	105
3.4.2	Square distances and inner products . . . . .	116
3.5	Linear embeddings of dissimilarities . . . . .	118
3.5.1	Euclidean embedding . . . . .	118
3.5.2	Correction of non-Euclidean dissimilarities . . . . .	120
3.5.3	Pseudo-Euclidean embedding . . . . .	122
3.5.4	Generalized average variance . . . . .	124
3.5.5	Projecting new vectors to an embedded space . . . . .	125
3.5.6	Reduction of dimension . . . . .	127
3.5.7	Reduction of complexity . . . . .	128
3.5.8	A general embedding . . . . .	129
3.5.9	Spherical embeddings . . . . .	130
3.6	Spatial representation of dissimilarities . . . . .	132
3.6.1	FastMap . . . . .	133
3.6.2	Multidimensional scaling . . . . .	135
3.6.3	Reduction of complexity . . . . .	143
3.7	Summary . . . . .	144
4.	Learning approaches . . . . .	147
4.1	Traditional learning . . . . .	148
4.1.1	Data bias and model bias . . . . .	148
4.1.2	Statistical learning . . . . .	151
4.1.3	Inductive principles . . . . .	154
4.1.3.1	Empirical risk minimization (ERM) . . . . .	156
4.1.3.2	Principles based on Occam's razor . . . . .	160
4.1.4	Why is the statistical approach not good enough for learning from objects? . . . . .	163
4.2	The role of dissimilarity representations . . . . .	166
4.2.1	Learned proximity representations . . . . .	171
4.2.2	Dissimilarity representations: learning . . . . .	172
4.3	Classification in generalized topological spaces . . . . .	175

4.4	Classification in dissimilarity spaces . . . . .	180
4.4.1	Characterization of dissimilarity spaces . . . . .	180
4.4.2	Classifiers . . . . .	185
4.5	Classification in pseudo-Euclidean spaces . . . . .	196
4.6	On generalized kernels and dissimilarity spaces . . . . .	205
4.6.1	Connection between dissimilarity spaces and pseudo-Euclidean spaces . . . . .	209
4.7	Discussion . . . . .	211
5.	Dissimilarity measures . . . . .	215
5.1	Measures depending on feature types . . . . .	216
5.2	Measures between populations . . . . .	228
5.2.1	Normal distributions . . . . .	228
5.2.2	Divergence measures . . . . .	229
5.2.3	Discrete probability distributions . . . . .	233
5.3	Dissimilarity measures between sequences . . . . .	234
5.4	Information-theoretic measures . . . . .	237
5.5	Dissimilarity measures between sets . . . . .	238
5.6	Dissimilarity measures in applications . . . . .	242
5.6.1	Invariance and robustness . . . . .	242
5.6.2	Example measures . . . . .	242
5.7	Discussion and conclusions . . . . .	250
6.	Visualization . . . . .	255
6.1	Multidimensional scaling . . . . .	257
6.1.1	First examples . . . . .	259
6.1.2	Linear and nonlinear methods: examples . . . . .	261
6.1.3	Implementation . . . . .	267
6.2	Other mappings . . . . .	268
6.3	Examples: getting insight into the data . . . . .	274
6.4	Tree models . . . . .	281
6.5	Summary . . . . .	287
7.	Further data exploration . . . . .	289
7.1	Clustering . . . . .	290
7.1.1	Standard approaches . . . . .	290
7.1.2	Clustering on dissimilarity representations . . . . .	295
7.1.3	Clustering examples for dissimilarity representations . . . . .	303

7.2	Intrinsic dimension . . . . .	309
7.3	Sampling density . . . . .	319
7.3.1	Proposed criteria . . . . .	320
7.3.2	Experiments with the NIST digits . . . . .	325
7.4	Summary . . . . .	331
8.	One-class classifiers . . . . .	333
8.1	General issues . . . . .	336
8.1.1	Construction of one-class classifiers . . . . .	337
8.1.2	One-class classifiers in feature spaces . . . . .	341
8.2	Domain descriptors for dissimilarity representations . . . . .	346
8.2.1	Neighborhood-based OCCs . . . . .	348
8.2.2	Generalized mean class descriptor . . . . .	350
8.2.3	Linear programming dissimilarity data description . . . . .	353
8.2.4	More issues on class descriptors . . . . .	359
8.3	Experiments . . . . .	366
8.3.1	Experiment I: Condition monitoring . . . . .	366
8.3.2	Experiment II: Diseased mucosa in the oral cavity . . . . .	374
8.3.3	Experiment III: Heart disease data . . . . .	377
8.4	Conclusions . . . . .	379
9.	Classification . . . . .	383
9.1	Proof of principle . . . . .	384
9.1.1	NN rule vs alternative dissimilarity-based classifiers . . . . .	384
9.1.2	Experiment I: square dissimilarity representations . . . . .	388
9.1.3	Experiment II: the dissimilarity space approach . . . . .	389
9.1.4	Discussion . . . . .	395
9.2	Selection of the representation set: the dissimilarity space approach . . . . .	396
9.2.1	Prototype selection methods . . . . .	398
9.2.2	Experimental setup . . . . .	401
9.2.3	Results and discussion . . . . .	404
9.2.4	Conclusions . . . . .	416
9.3	Selection of the representation set: the embedding approach . . . . .	417
9.3.1	Prototype selection methods . . . . .	418
9.3.2	Experiments and results . . . . .	421
9.3.3	Conclusions . . . . .	428

9.4	On corrections of dissimilarity measures . . . . .	428
9.4.1	Going more Euclidean . . . . .	429
9.4.2	Experimental setup . . . . .	430
9.4.3	Results and conclusions . . . . .	432
9.5	A few remarks on a simulated missing value problem . . .	439
9.6	Existence of zero-error dissimilarity-based classifiers . . .	443
9.6.1	Asymptotic separability of classes . . . . .	444
9.7	Final discussion . . . . .	451
10.	Combining . . . . .	453
10.1	Combining for one-class classification . . . . .	455
10.1.1	Combining strategies . . . . .	456
10.1.2	Data and experimental setup . . . . .	459
10.1.3	Results and discussion . . . . .	462
10.1.4	Summary and conclusions . . . . .	465
10.2	Combining for standard two-class classification . . . . .	466
10.2.1	Combining strategies . . . . .	466
10.2.2	Experiments on the handwritten digit set . . . . .	468
10.2.3	Results . . . . .	470
10.2.4	Conclusions . . . . .	473
10.3	Classifier projection space . . . . .	474
10.3.1	Construction and the use of CPS . . . . .	475
10.4	Summary . . . . .	483
11.	Representation review and recommendations . . . . .	485
11.1	Representation review . . . . .	485
11.1.1	Three generalization ways . . . . .	486
11.1.2	Representation formation . . . . .	489
11.1.3	Generalization capabilities . . . . .	492
11.2	Practical considerations . . . . .	493
11.2.1	Clustering . . . . .	495
11.2.2	One-class classification . . . . .	496
11.2.3	Classification . . . . .	497
12.	Conclusions and open problems . . . . .	503
12.1	Summary and contributions . . . . .	505
12.2	Extensions of dissimilarity representations . . . . .	508
12.3	Open questions . . . . .	510

Appendix A	On convex and concave functions	515
Appendix B	Linear algebra in vector spaces	519
	B.1 Some facts on matrices in a Euclidean space . . . . .	519
	B.2 Some facts on matrices in a pseudo-Euclidean space . . . . .	523
Appendix C	Measure and probability	527
Appendix D	Statistical sidelines	533
	D.1 Likelihood and parameter estimation . . . . .	533
	D.2 Expectation-maximization (EM) algorithm . . . . .	535
	D.3 Model selection . . . . .	536
	D.4 PCA and probabilistic models . . . . .	538
	D.4.1 Gaussian model . . . . .	538
	D.4.2 A Gaussian mixture model . . . . .	539
	D.4.3 PCA . . . . .	541
	D.4.4 Probabilistic PCA . . . . .	542
	D.4.5 A mixture of probabilistic PCA . . . . .	543
Appendix E	Data sets	545
	E.1 Artificial data sets . . . . .	545
	E.2 Real-world data sets . . . . .	549
	<i>Bibliography</i>	561
	<i>Index</i>	599