
Contents

Notations	XV
Acronyms	XVII
1 High-Dimensional Data	1
1.1 Practical motivations	1
1.1.1 Fields of application	2
1.1.2 The goals to be reached	3
1.2 Theoretical motivations	3
1.2.1 How can we visualize high-dimensional spaces?	4
1.2.2 Curse of dimensionality and empty space phenomenon .	6
1.3 Some directions to be explored	9
1.3.1 Relevance of the variables	10
1.3.2 Dependencies between the variables	10
1.4 About topology, spaces, and manifolds	11
1.5 Two benchmark manifolds	14
1.6 Overview of the next chapters	16
2 Characteristics of an Analysis Method	17
2.1 Purpose	17
2.2 Expected functionalities	18
2.2.1 Estimation of the number of latent variables	18
2.2.2 Embedding for dimensionality reduction	19
2.2.3 Embedding for latent variable separation	20
2.3 Internal characteristics	22
2.3.1 Underlying model	22
2.3.2 Algorithm	23
2.3.3 Criterion	23
2.4 Example: Principal component analysis	24
2.4.1 Data model of PCA	24
2.4.2 Criteria leading to PCA	26

2.4.3	Functionalities of PCA	29
2.4.4	Algorithms	31
2.4.5	Examples and limitations of PCA	33
2.5	Toward a categorization of DR methods	37
2.5.1	Hard vs. soft dimensionality reduction	38
2.5.2	Traditional vs. generative model	39
2.5.3	Linear vs. nonlinear model	40
2.5.4	Continuous vs. discrete model	40
2.5.5	Implicit vs. explicit mapping	41
2.5.6	Integrated vs. external estimation of the dimensionality	41
2.5.7	Layered vs. standalone embeddings	42
2.5.8	Single vs. multiple coordinate systems	42
2.5.9	Optional vs. mandatory vector quantization	43
2.5.10	Batch vs. online algorithm	43
2.5.11	Exact vs. approximate optimization	44
2.5.12	The type of criterion to be optimized	44
3	Estimation of the Intrinsic Dimension	47
3.1	Definition of the intrinsic dimension	47
3.2	Fractal dimensions	48
3.2.1	The q -dimension	49
3.2.2	Capacity dimension	51
3.2.3	Information dimension	52
3.2.4	Correlation dimension	53
3.2.5	Some inequalities	54
3.2.6	Practical estimation	55
3.3	Other dimension estimators	59
3.3.1	Local methods	59
3.3.2	Trial and error	60
3.4	Comparisons	62
3.4.1	Data Sets	63
3.4.2	PCA estimator	63
3.4.3	Correlation dimension	63
3.4.4	Local PCA estimator	65
3.4.5	Trial and error	66
3.4.6	Concluding remarks	67
4	Distance Preservation	69
4.1	State-of-the-art	69
4.2	Spatial distances	70
4.2.1	Metric space, distances, norms and scalar product	70
4.2.2	Multidimensional scaling	73
4.2.3	Sammon's nonlinear mapping	82
4.2.4	Curvilinear component analysis	88
4.3	Graph distances	97
4.3.1	Geodesic distance and graph distance	97

4.3.2	Isomap	102
4.3.3	Geodesic NLM	111
4.3.4	Curvilinear distance analysis	114
4.4	Other distances	119
4.4.1	Kernel PCA	120
4.4.2	Semidefinite embedding	125
5	Topology Preservation	133
5.1	State of the art	133
5.2	Predefined lattice	135
5.2.1	Self-Organizing Maps	135
5.2.2	Generative Topographic Mapping	143
5.3	Data-driven lattice	152
5.3.1	Locally linear embedding	152
5.3.2	Laplacian eigenmaps	159
5.3.3	Isotop	165
6	Method comparisons	173
6.1	Toy examples	173
6.1.1	The Swiss roll	173
6.1.2	Manifolds having essential loops or spheres	193
6.2	Cortex unfolding	199
6.3	Image processing	203
6.3.1	Artificial faces	206
6.3.2	Real faces	214
7	Conclusions	225
7.1	Summary of the book	225
7.1.1	The problem	225
7.1.2	A basic solution	226
7.1.3	Dimensionality reduction	226
7.1.4	Latent variable separation	228
7.1.5	Intrinsic dimensionality estimation	229
7.2	Data flow	230
7.2.1	Variable Selection	230
7.2.2	Calibration	230
7.2.3	Linear dimensionality reduction	231
7.2.4	Nonlinear dimensionality reduction	231
7.2.5	Latent variable separation	232
7.2.6	Further processing	232
7.3	Model complexity	232
7.4	Taxonomy	233
7.4.1	Distance preservation	236
7.4.2	Topology preservation	237
7.5	Spectral methods	238
7.6	Nonspectral methods	241

7.7	Tentative methodology	242
7.8	Perspectives	245
A	Matrix Calculus	247
A.1	Singular value decomposition	247
A.2	Eigenvalue decomposition	248
A.3	Square root of a square matrix	248
B	Gaussian Variables	251
B.1	One-dimensional Gaussian distribution	251
B.2	Multidimensional Gaussian distribution	253
B.2.1	Uncorrelated Gaussian variables	254
B.2.2	Isotropic multivariate Gaussian distribution	254
B.2.3	Linearly mixed Gaussian variables	256
C	Optimization	259
C.1	Newton's method	259
C.1.1	Finding extrema	260
C.1.2	Multivariate version	260
C.2	Gradient ascent/descent	261
C.2.1	Stochastic gradient descent	261
D	Vector quantization	263
D.1	Classical techniques	265
D.2	Competitive learning	266
D.3	Taxonomy	266
D.4	Initialization and "dead units"	267
E	Graph Building	269
E.1	Without vector quantization	270
E.1.1	K -rule	270
E.1.2	ϵ -rule	271
E.1.3	τ -rule	271
E.2	With vector quantization	272
E.2.1	Data rule	272
E.2.2	Histogram rule	274
F	Implementation Issues	277
F.1	Dimension estimation	277
F.1.1	Capacity dimension	277
F.1.2	Correlation dimension	277
F.2	Computation of the closest point(s)	279
F.3	Graph distances	280
	References	283

Index 297