René Vidal • Yi Ma • S. Shankar Sastry

Generalized Principal Component Analysis

Springer

Contents

1 Introduction 1
1.1 Modeling Data with a Parametric Model 2
1.1.1 The Choice of a Model Class 3
1.1.2 Statistical Models versus Geometric Models 4
1.2 Modeling Mixed Data with a Mixture Model 6
1.2.1 Examples of Mixed Data Modeling 7
1.2.2 Mathematical Representations of Mixture Models 12
1.3 Clustering via Discriminative or Nonparametric Methods 16
1.4 Noise, Errors, Outliers, and Model Selection 18

Part I Modeling Data with a Single Subspace

2 Principal Component Analysis 25
2.1 Classical Principal Component Analysis (PCA) 25
2.1.1 A Statistical View of PCA 26
2.1.2 A Geometric View of PCA 30
2.1.3 A Rank Minimization View of PCA 34
2.2 Probabilistic Principal Component Analysis (PPCA) 38
2.2.1 PPCA from Population Mean and Covariance 39
2.2.2 PPCA by Maximum Likelihood 40
2.3 Model Selection for Principal Component Analysis 45
2.3.1 Model Selection by Information-Theoretic Criteria 46
2.3.2 Model Selection by Rank Minimization 49
2.3.3 Model Selection by Asymptotic Mean Square Error 51
2.4 Bibliographic Notes 53
2.5 Exercises 54

3 Robust Principal Component Analysis 63
3.1 PCA with Robustness to Missing Entries 64
3.1.1 Incomplete PCA by Mean and Covariance Completion 68
3.1.2 Incomplete PPCA by Expectation Maximization 69

3.1.3 Matrix Completion by Convex Optimization 73
3.1.4 Incomplete PCA by Alternating Minimization..... 78
3.2 PCA with Robustness to Corrupted Entries..... 87
3.2.1 Robust PCA by Iteratively Reweighted Least Squares ... 89
3.2.2 Robust PCA by Convex Optimization 92
3.3 PCA with Robustness to Outliers..... 99
3.3.1 Outlier Detection by Robust Statistics..... 101
3.3.2 Outlier Detection by Convex Optimization 107
3.4 Bibliographic Notes 113
3.5 Exercises..... 115

4 Nonlinear and Nonparametric Extensions..... 123
4.1 Nonlinear and Kernel PCA 126
4.1.1 Nonlinear Principal Component Analysis (NLPCA) 126
4.1.2 NLPCA in a High-dimensional Feature Space..... 128
4.1.3 Kernel PCA (KPCA) 129
4.2 Nonparametric Manifold Learning 133
4.2.1 Multidimensional Scaling (MDS) 134
4.2.2 Locally Linear Embedding (LLE) 135
4.2.3 Laplacian Eigenmaps (LE)..... 138
4.3 K-Means and Spectral Clustering 143
4.3.1 K-Means Clustering 145
4.3.2 Spectral Clustering 148
4.4 Bibliographic Notes 160
4.5 Exercises..... 161
4.A Laplacian Eigenmaps: Continuous Formulation..... 166

Part II Modeling Data with Multiple Subspaces

5 Algebraic-Geometric Methods 171
5.1 Problem Formulation of Subspace Clustering 172
5.1.1 Projectivization of Affine Subspaces 172
5.1.2 Subspace Projection and Minimum Representation 174
5.2 Introductory Cases of Subspace Clustering 176
5.2.1 Clustering Points on a Line 176
5.2.2 Clustering Lines in a Plane..... 179
5.2.3 Clustering Hyperplanes 181
5.3 Subspace Clustering Knowing the Number of Subspaces..... 184
5.3.1 An Introductory Example 184
5.3.2 Fitting Polynomials to Subspaces..... 186
5.3.3 Subspaces from Polynomial Differentiation 188
5.3.4 Point Selection via Polynomial Division 190
5.3.5 The Basic Algebraic Subspace Clustering Algorithm 193
5.4 Subspace Clustering not Knowing the Number of Subspaces 196
5.4.1 Introductory Examples 196

5.4.2 Clustering Subspaces of Equal Dimension 198
5.4.3 Clustering Subspaces of Different Dimensions 200
5.5 Model Selection for Multiple Subspaces 201
5.5.1 Effective Dimension of Samples of Multiple Subspaces . 202
5.5.2 Minimum Effective Dimension of Noisy Samples 204
5.5.3 Recursive Algebraic Subspace Clustering 205
5.6 Bibliographic Notes 207
5.7 Exercises 210

6 Statistical Methods 217
6.1 K-Subspaces 219
6.1.1 K-Subspaces Model 219
6.1.2 K-Subspaces Algorithm 220
6.1.3 Convergence of the K-Subspaces Algorithm 221
6.1.4 Advantages and Disadvantages of K-Subspaces 222
6.2 Mixture of Probabilistic PCA (MPPCA) 222
6.2.1 MPPCA Model 223
6.2.2 Maximum Likelihood Estimation for MPPCA 223
6.2.3 Maximum a Posteriori (MAP) Estimation for MPPCA .. 226
6.2.4 Relationship between K-Subspaces and MPPCA 228
6.3 Compression-Based Subspace Clustering 231
6.3.1 Model Estimation and Data Compression 231
6.3.2 Minimium Coding Length via Agglomerative Clustering 233
6.3.3 Lossy Coding of Multivariate Data 238
6.3.4 Coding Length of Mixed Gaussian Data 242
6.4 Simulations and Applications 247
6.4.1 Statistical Methods on Synthetic Data 247
6.4.2 Statistical Methods on Gene Expression Clustering, Image Segmentation, and Face Clustering ... 254
6.5 Bibliographic Notes 258
6.6 Exercises 261
6.A Lossy Coding Length for Subspace-like Data 263

7 Spectral Methods 267
7.1 Spectral Subspace Clustering 268
7.2 Local Subspace Affinity (LSA) and Spectral Local Best-Fit Flats (SLBF) 270
7.3 Locally Linear Manifold Clustering (LLMC) 274
7.4 Spectral Curvature Clustering (SCC) 276
7.5 Spectral Algebraic Subspace Clustering (SASC) 279
7.6 Simulations and Applications 281
7.6.1 Spectral Methods on Synthetic Data 281
7.6.2 Spectral Methods on Face Clustering 285
7.7 Exercises 289

8 Sparse and Low-Rank Methods 291
8.1 Self-Expressiveness and Subspace-Preserving Representations ... 294
8.1.1 Self-Expressiveness Property 294
8.1.2 Subspace-Preserving Representation 296
8.2 Low-Rank Subspace Clustering (LRSC) 297
8.2.1 LRSC with Uncorrupted Data 297
8.2.2 LRSC with Robustness to Noise 302
8.2.3 LRSC with Robustness to Corruptions 308
8.3 Sparse Subspace Clustering (SSC) 310
8.3.1 SSC with Uncorrupted Data 310
8.3.2 SSC with Robustness to Outliers 324
8.3.3 SSC with Robustness to Noise 326
8.3.4 SSC with Robustness to Corrupted Entries 330
8.3.5 SSC for Affine Subspaces 332
8.4 Simulations and Applications 333
8.4.1 Low-Rank and Sparse Methods on Synthetic Data 333
8.4.2 Low-Rank and Sparse Methods on Face Clustering 336
8.5 Bibliographic Notes 344
8.6 Exercises 345

Part III Applications

9 Image Representation 349
9.1 Seeking Compact and Sparse Image Representations 349
9.1.1 Prefixed Linear Transformations 350
9.1.2 Adaptive, Overcomplete, and Hybrid Representations ... 351
9.1.3 Hierarchical Models for Multiscale Structures. 353
9.2 Image Representation with Multiscale Hybrid Linear Models 354
9.2.1 Linear versus Hybrid Linear Models 354
9.2.2 Multiscale Hybrid Linear Models 361
9.2.3 Experiments and Comparisons 365
9.3 Multiscale Hybrid Linear Models in Wavelet Domain 369
9.3.1 Imagery Data Vectors in the Wavelet Domain 369
9.3.2 Hybrid Linear Models in the Wavelet Domain 371
9.3.3 Comparison with Other Lossy Representations 372
9.4 Bibliographic Notes 376

10 Image Segmentation 377
10.1 Basic Models and Principles 378
10.1.1 Problem Formulation 378
10.1.2 Image Segmentation as Subspace Clustering 380
10.1.3 Minimum Coding Length Principle 381
10.2 Encoding Image Textures and Boundaries 382
10.2.1 Construction of Texture Features 382
10.2.2 Texture Encoding 383
10.2.3 Boundary Encoding 384

10.3 Compression-Based Image Segmentation 386
10.3.1 Minimizing Total Coding Length 386
10.3.2 Hierarchical Implementation 387
10.3.3 Choosing the Proper Distortion Level 389
10.4 Experimental Evaluation 392
10.4.1 Color Spaces and Compressibility 392
10.4.2 Experimental Setup 394
10.4.3 Results and Discussions 395
10.5 Bibliographic Notes 399

11 Motion Segmentation 401
11.1 The 3D Motion Segmentation Problem 402
11.2 Motion Segmentation from Multiple Affine Views 405
11.2.1 Affine Projection of a Rigid-Body Motion 405
11.2.2 Motion Subspace of a Rigid-Body Motion 406
11.2.3 Segmentation of Multiple Rigid-Body Motions 406
11.2.4 Experiments on Multiview Motion Segmentation 407
11.3 Motion Segmentation from Two Perspective Views 413
11.3.1 Perspective Projection of a Rigid-Body Motion 414
11.3.2 Segmentation of 3D Translational Motions 415
11.3.3 Segmentation of Rigid-Body Motions 416
11.3.4 Segmentation of Rotational Motions or Planar Scenes ... 417
11.3.5 Experiments on Two-View Motion Segmentation 418
11.4 Temporal Motion Segmentation 421
11.4.1 Dynamical Models of Time-Series Data 422
11.4.2 Experiments on Temporal Video Segmentation 423
11.4.3 Experiments on Segmentation of Human Motion Data 425
11.5 Bibliographical Notes 428

12 Hybrid System Identification 431
12.1 Problem Statement 433
12.2 Identification of a Single ARX System 434
12.3 Identification of Hybrid ARX Systems 438
12.3.1 The Hybrid Decoupling Polynomial 439
12.3.2 Identifying the Hybrid Decoupling Polynomial 440
12.3.3 Identifying System Parameters and Discrete States 443
12.3.4 The Basic Algorithm and Its Extensions 445
12.4 Simulations and Experiments 446
12.4.1 Error in the Estimation of the Model Parameters 447
12.4.2 Error as a Function of the Model Orders 447
12.4.3 Error as a Function of Noise 448
12.4.4 Experimental Results on Test Data Sets 449
12.5 Bibliographic Notes 450

13 Final Words 453
13.1 Unbalanced and Multimodal Data 454
13.2 Unsupervised and Semisupervised Learning 454
13.3 Data Acquisition and Online Data Analysis 455
13.4 Other Low-Dimensional Models 456
13.5 Computability and Scalability 457
13.6 Theory, Algorithms, Systems, and Applications 459

A Basic Facts from Optimization 461
A.1 Unconstrained Optimization 461
A.1.1 Optimality Conditions 462
A.1.2 Convex Set and Convex Function 462
A.1.3 Subgradient 464
A.1.4 Gradient Descent Algorithm 465
A.1.5 Alternating Direction Minimization 466
A.2 Constrained Optimization 468
A.2.1 Optimality Conditions and Lagrangian Multipliers 468
A.2.2 Augmented Lagrange Multipler Methods 470
A.2.3 Alternating Direction Method of Multipliers 471
A.3 Exercises 474

B Basic Facts from Mathematical Statistics 475
B.1 Estimation of Parametric Models 475
B.1.1 Sufficient Statistics 476
B.1.2 Mean Square Error, Efficiency, and Fisher Information 477
B.1.3 The Rao–Blackwell Theorem and Uniformly Minimum-Variance Unbiased Estimator 479
B.1.4 Maximum Likelihood (ML) Estimator 480
B.1.5 Consistency and Asymptotic Efficiency of the ML Estimator 481
B.2 ML Estimation for Models with Latent Variables 485
B.2.1 Expectation Maximization (EM) 486
B.2.2 Maximum a Posteriori Expectation Maximization (MAP-EM) 488
B.3 Estimation of Mixture Models 490
B.3.1 EM for Mixture Models 490
B.3.2 MAP-EM for Mixture Models 492
B.3.3 A Case in Which EM Fails 494
B.4 Model-Selection Criteria 496
B.4.1 Akaike Information Criterion 497
B.4.2 Bayesian Information Criterion 498
B.5 Robust Statistical Methods 498
B.5.1 Influence-Based Outlier Detection 499
B.5.2 Probability-Based Outlier Detection 501
B.5.3 Random-Sampling-Based Outlier Detection 503
B.6 Exercises 506

C Basic Facts from Algebraic Geometry 509
C.1 Abstract Algebra Basics 509
C.1.1 Polynomial Rings 509
C.1.2 Ideals and Algebraic Sets 511
C.1.3 Algebra and Geometry: Hilbert's Nullstellensatz 513
C.1.4 Algebraic Sampling Theory 514
C.1.5 Decomposition of Ideals and Algebraic Sets 516
C.1.6 Hilbert Function, Polynomial, and Series 517
C.2 Ideals of Subspace Arrangements 519
C.3 Subspace Embedding and PL-Generated Ideals 522
C.4 Hilbert Functions of Subspace Arrangements 524
C.4.1 Hilbert Function and Algebraic Subspace Clustering 525
C.4.2 Special Cases of the Hilbert Function 528
C.4.3 Formulas for the Hilbert Function 530
C.5 Bibliographic Notes 534

References 535

Index 553