

1. Record Nr.	UNINA9910677590703321
Autore	Smilde Age K.
Titolo	Multiblock data fusion in statistics and machine learning : applications in the natural and life sciences // Age K. Smilde, Tormod Næs, Kristian Hovde Liland
Pubbl/distr/stampa	Hoboken, New Jersey : , : John Wiley & Sons, Inc., , [2022] ©2022
ISBN	1-119-60097-9 1-119-60098-7
Descrizione fisica	1 online resource (418 pages)
Disciplina	519.52
Soggetti	Science - Statistical methods
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Intro -- Multiblock Data Fusion in Statistics and Machine Learning -- Contents -- Foreword -- Preface -- List of Figures -- List of Tables -- Part I Introductory Concepts and Theory -- chapnumcolor1 Introduction -- 1.1 Scope of the Book -- 1.2 Potential Audience -- 1.3 Types of Data and Analyses -- 1.3.1 Supervised and Unsupervised Analyses -- 1.3.2 High-, Mid- and Low-level Fusion -- 1.3.3 Dimension Reduction -- 1.3.4 Indirect Versus Direct Data -- 1.3.5 Heterogeneous Fusion -- 1.4 Examples -- 1.4.1 Metabolomics -- 1.4.2 Genomics -- 1.4.3 Systems Biology -- 1.4.4 Chemistry -- 1.4.5 Sensory Science -- 1.5 Goals of Analyses -- 1.6 Some History -- 1.7 Fundamental Choices -- 1.8 Common and Distinct Components -- 1.9 Overview and Links -- 1.10 Notation and Terminology -- 1.11 Abbreviations -- chapnumcolor2 Basic Theory and Concepts -- 2.i General Introduction -- 2.1 Component Models -- 2.1.1 General Idea of Component Models -- 2.1.2 Principal Component Analysis -- 2.1.3 Sparse PCA -- 2.1.4 Principal Component Regression -- 2.1.5 Partial Least Squares -- 2.1.6 Sparse PLS -- 2.1.7 Principal Covariates Regression -- 2.1.8 Redundancy Analysis -- 2.1.9 Comparing PLS, PCovR and RDA -- 2.1.10 Generalised Canonical Correlation Analysis -- 2.1.11 Simultaneous Component Analysis -- 2.2 Properties of Data -- 2.2.1 Data Theory -- 2.2.2 Scale-types -- 2.3 Estimation Methods

-- 2.3.1 Least-squares Estimation -- 2.3.2 Maximum-likelihood Estimation -- 2.3.3 Eigenvalue Decomposition-based Methods -- 2.3.4 Covariance or Correlation-based Estimation Methods -- 2.3.5 Sequential Versus Simultaneous Methods -- 2.3.6 Homogeneous Versus Heterogeneous Fusion -- 2.4 Within- and Between-block Variation -- 2.4.1 Definition and Example -- 2.4.2 MAXBET Solution -- 2.4.3 MAXNEAR Solution -- 2.4.4 PLS2 Solution -- 2.4.5 CCA Solution -- 2.4.6 Comparing the Solutions. 2.4.7 PLS, RDA and CCA Revisited -- 2.5 Framework for Common and Distinct Components -- 2.6 Preprocessing -- 2.7 Validation -- 2.7.1 Outliers -- 2.7.1.1 Residuals -- 2.7.1.2 Leverage -- 2.7.2 Model Fit -- 2.7.3 Bias-variance Trade-off -- 2.7.4 Test Set Validation -- 2.7.5 Cross-validation -- 2.7.6 Permutation Testing -- 2.7.7 Jackknife and Bootstrap -- 2.7.8 Hyper-parameters and Penalties -- 2.8 Appendix -- chapnumcolor3 Structure of Multiblock Data -- 3.i General Introduction -- 3.1 Taxonomy -- 3.2 Skeleton of a Multiblock Data Set -- 3.2.1 Shared Sample Mode -- 3.2.2 Shared Variable Mode -- 3.2.3 Shared Variable or Sample Mode -- 3.2.4 Shared Variable and Sample Mode -- 3.3 Topology of a Multiblock Data Set -- 3.3.1 Unsupervised Analysis -- 3.3.2 Supervised Analysis -- 3.4 Linking Structures -- 3.4.1 Linking Structure for Unsupervised Analysis -- 3.4.2 Linking Structures for Supervised Analysis -- 3.5 Summary -- chapnumcolor4 Matrix Correlations -- 4.i General Introduction -- 4.1 Definition -- 4.2 Most Used Matrix Correlations -- 4.2.1 Inner Product Correlation -- 4.2.2 GCD coefficient -- 4.2.3 RV-coefficient -- 4.2.4 SMI-coefficient -- 4.3 Generic Framework of Matrix Correlations -- 4.4 Generalised Matrix Correlations -- 4.4.1 Generalised RV-coefficient -- 4.4.2 Generalised Association Coefficient -- 4.5 Partial Matrix Correlations -- 4.6 Conclusions and Recommendations -- 4.7 Open Issues -- Part II Selected Methods for Unsupervised and Supervised Topologies -- chapnumcolor5 Unsupervised Methods -- 5.i General Introduction -- 5. ii Relations to the General Framework -- 5.1 Shared Variable Mode -- 5.1.1 Only Common Variation -- 5.1.1.1 Simultaneous Component Analysis -- 5.1.1.2 Clustering and SCA -- 5.1.1.3 Multigroup Data Analysis -- 5.1.2 Common, Local, and Distinct Variation -- 5.1.2.1 Distinct and Common Components. 5.1.2.2 Multivariate Curve Resolution -- 5.2 Shared Sample Mode -- 5.2.1 Only Common Variation -- 5.2.1.1 SUM-PCA -- 5.2.1.2 Multiple Factor Analysis and STATIS -- 5.2.1.3 Generalised Canonical Analysis -- 5.2.1.4 Regularised Generalised Canonical Correlation Analysis -- 5.2.1.5 Exponential Family SCA -- 5.2.1.6 Optimal-scaling -- 5.2.2 Common, Local, and Distinct Variation -- 5.2.2.1 Joint and Individual Variation Explained -- 5.2.2.2 Distinct and Common Components -- 5.2.2.3 PCA-GCA -- 5.2.2.4 Advanced Coupled Matrix and Tensor Factorisation -- 5.2.2.5 Penalised-ESCA -- 5.2.2.6 Multivariate Curve Resolution -- 5.3 Generic Framework -- 5.3.1 Framework for Simultaneous Unsupervised Methods -- 5.3.1.1 Description of the Framework -- 5.3.1.2 Framework Applied to Simultaneous Unsupervised Data Analysis Methods -- 5.3.1.3 Framework of Common/Distinct Applied to Simultaneous Unsupervised Multiblock Data Analysis Methods -- 5.4 Conclusions and Recommendations -- 5.5 Open Issues -- chapnumcolor6 ASCA and Extensions -- 6.i General Introduction -- 6.ii Relations to the General Framework -- 6.1 ANOVA-Simultaneous Component Analysis -- 6.1.1 The ASCA Method -- 6.1.2 Validation of ASCA -- 6.1.2.1 Permutation Testing -- 6.1.2.2 Back-projection -- 6.1.2.3 Confidence Ellipsoids -- 6.1.3 The ASCA+ and LiMM-PCA Methods -- 6.2 Multilevel-SCA -- 6.3 Penalised-ASCA -- 6.4 Conclusions and Recommendations -- 6.5 Open Issues --

chapnumcolor7 Supervised Methods -- 7.i General Introduction -- 7.ii Relations to the General Framework -- 7.1 Multiblock Regression: General Perspectives -- 7.1.1 Model and Assumptions -- 7.1.2 Different Challenges and Aims -- 7.2 Multiblock PLS Regression -- 7.2.1 Standard Multiblock PLS Regression -- 7.2.2 MB-PLS Used for Classification -- 7.2.3 Sparse Multiblock PLS Regression (sMB-PLS). 7.3 The Family of SO-PLS Regression Methods (Sequential and Orthogonalised PLS Regression) -- 7.3.1 The SO-PLS Method -- 7.3.2 Order of Blocks -- 7.3.3 Interpretation Tools -- 7.3.4 Restricted PLS Components and their Application in SO-PLS -- 7.3.5 Validation and Component Selection -- 7.3.6 Relations to ANOVA -- 7.3.7 Extensions of SO-PLS to Handle Interactions Between Blocks -- 7.3.8 Further Applications of SO-PLS -- 7.3.9 Relations Between SO-PLS and ASCA -- 7.4 Parallel and Orthogonalised PLS (PO-PLS) Regression -- 7.5 Response Oriented Sequential Alternation -- 7.5.1 The ROSA Method -- 7.5.2 Validation -- 7.5.3 Interpretation -- 7.6 Conclusions and Recommendations -- 7.7 Open Issues -- Part III Methods for Complex Multiblock Structures -- chapnumcolor8 Complex Block Structures -- with Focus on L-Shape Relations -- 8.i General Introduction -- 8.ii Relations to the General Framework -- 8.1 Analysis of L-shape Data: General Perspectives -- 8.2 Sequential Procedures for L-shape Data Based on PLS/PCR and ANOVA -- 8.2.1 Interpretation of X1, Quantitative X2-data, Horizontal Axis First -- 8.2.2 Interpretation of X1, Categorical X2-data, Horizontal Axis First -- 8.2.3 Analysis of Segments/Clusters of X1 Data -- 8.3 The L-PLS Method for Joint Estimation of Blocks in L-shape Data -- 8.3.1 The Original L-PLS Method, Endo-L-PLS -- 8.3.2 Exo- Versus Endo-L-PLS -- 8.4 Modifications of the Original L-PLS Idea -- 8.4.1 Weighting Information from X3 and X1 in L-PLS Using a Parameter " -- 8.4.2 Three-blocks Bifocal PLS -- 8.5 Alternative L-shape Data Analysis Methods -- 8.5.1 Principal Component Analysis with External Information -- 8.5.2 A Simple PCA Based Procedure for Using Unlabelled Data in Calibration -- 8.5.3 Multivariate Curve Resolution for Incomplete Data -- 8.5.4 An Alternative Approach in Consumer Science Based on Correlations Between X3 and X1. 8.6 Domino PLS and More Complex Data Structures -- 8.7 Conclusions and Recommendations -- 8.8 Open Issues -- Part IV Alternative Methods for Unsupervised and Supervised Topologies -- chapnumcolor9 Alternative Unsupervised Methods -- 9.i General Introduction -- 9.ii Relationship to the General Framework -- 9.1 Shared Variable Mode -- 9.2 Shared Sample Mode -- 9.2.1 Only Common Variation -- 9.2.1.1 DIABLO -- 9.2.1.2 Generalised Coupled Tensor Factorisation -- 9.2.1.3 Representation Matrices -- 9.2.1.4 Extended PCA -- 9.2.2 Common, Local, and Distinct Variation -- 9.2.2.1 Generalised SVD -- 9.2.2.2 Structural Learning and Integrative Decomposition -- 9.2.2.3 Bayesian Inter-battery Factor Analysis -- 9.2.2.4 Group Factor Analysis -- 9.2.2.5 OnPLS -- 9.2.2.6 Generalised Association Study -- 9.2.2.7 Multi-Omics Factor Analysis -- 9.3 Two Shared Modes and Only Common Variation -- 9.3.1 Generalised Procrustes Analysis -- 9.3.2 Three-way Methods -- 9.4 Conclusions and Recommendations -- 9.4.1 Open Issues -- chapnumcolor10 Alternative Supervised Methods -- 10.i General Introduction -- 10.ii Relations to the General Framework -- 10.1 Model and Focus -- 10.2 Extension of PCovR -- 10.2.1 Sparse Multiblock Principal Covariates Regression, Sparse PCovR -- 10.2.2 Multiway Multiblock Covariates Regression -- 10.3 Multiblock Redundancy Analysis -- 10.3.1 Standard Multiblock Redundancy Analysis -- 10.3.2 Sparse Multiblock Redundancy Analysis -- 10.4 Miscellaneous Multiblock Regression

Methods -- 10.4.1 Multiblock Variance Partitioning -- 10.4.2 Network Induced Supervised Learning -- 10.4.3 Common Dimensions for Multiblock Regression -- 10.5 Modifications and Extensions of the SO-PLS Method -- 10.5.1 Extensions of SO-PLS to Three-Way Data -- 10.5.2 Variable Selection for SO-PLS -- 10.5.3 More Complicated Error Structure for SO-PLS.
10.5.4 SO-PLS Used for Path Modelling.
