

Preface

As the data size increases and the data structure becomes increasingly complex, considering the nonlinearity, mixed measurement level data, simultaneous analysis and computational efficiency, for example, become increasingly necessary.

Accordingly, we have been investigating data analysis and developing methods related to the above problems: how to jointly analyze data that include not only categorical variables but also numerical variables; matrix operations approaching categorical data analyses; how to select a reasonable subset of variables from mixed measurement level data; analyses based on the sparseness of data and simultaneous estimations, such as a joint method with dimension reduction and clustering, which have recently been investigated; and how to accelerate the computations involving the alternating least squares (ALS) algorithm.

Although we have heretofore been investigating these topics separately, we decided to include the above-mentioned methods/techniques in this series when we started to write this Springer Brief Statistics series.

We present our research interests in two parts as follows:

The first part consists of two chapters that introduce the principles of nonlinear principal component analysis (PCA). After a brief introduction of the ordinary PCA, a PCA for categorical (nominal and ordinal) data is introduced as a nonlinear PCA, in which the optimal scaling technique is used to quantify the categorical variables. The ALS algorithm is the main algorithm used in this method. Next, multiple correspondence analysis (MCA), which is a special case of nonlinear PCA, is introduced, and ALS is also used in the computation. All formulations in these methods are integrated in the same manner as the matrix operations. Since any measurement level data can be treated consistently as numerical data and ALS is a very powerful tool for estimations, these methods can be used in a variety of fields, including biometrics, econometrics, psychometrics, and sociology.

The second part of the book consists of four chapters, which describe applications of the nonlinear PCA: variable selection for mixed measurement level data, sparse MCA, reduced k-means clustering, and an acceleration of the ALS algorithm. The variable selection methods used in PCA, which were originally developed for numerical data, can be applied to any type of measurement level data

using nonlinear PCA. Sparseness and k-means clustering for nonlinear data, which were proposed in recent studies, are extensions obtained using the same matrix operations used in nonlinear PCA and MCA. Finally, an acceleration algorithm is proposed to reduce the computational cost of ALS iteration in nonlinear multivariate methods.

This book demonstrates the usefulness of nonlinear PCA and MCA, which can be applied to different measurement level data in a variety of fields and covers the latest topics, including the extension of traditional statistical methods, newly proposed nonlinear methods, and the computational efficiency of these methods. The ALS algorithm is a key concept in all chapters, and optimal quantifications and matrix operations are also key concepts in this book.

We sincerely hope that the concepts and applications presented herein will be of use.

Okayama, Japan
June 2016

Yuichi Mori
Masahiro Kuroda
Naomichi Makino

Nonlinear Principal Component Analysis and Its
Applications

Mori, Y.; Kuroda, M.; Makino, N.

2016, X, 80 p. 17 illus., 8 illus. in color., Softcover

ISBN: 978-981-10-0157-4