

LIKELIHOOD-BASED DIMENSION FOLDING ON TENSOR DATA

Ning Wang¹, Xin Zhang¹ and Bing Li²

¹*Florida State University and* ²*Pennsylvania State University*

Abstract: Sufficient dimension reduction methods are flexible tools for data visualization and exploratory analysis, typically in a regression of a univariate response on a multivariate predictor. Recently, there has been growing interest in the analysis of matrix-variate and tensor-variate data. For regressions with tensor predictors, a general framework of dimension folding and several moment-based estimation procedures have been proposed in the literature. In this article, we propose two likelihood-based dimension folding methods motivated by quadratic discriminant analysis for tensor data: the maximum likelihood estimators are derived under a general covariance setting and a structured envelope covariance setting. We study the asymptotic properties of both estimators and show using simulation studies and a real-data analysis that they are more accurate than existing moment-based estimators.

Key words and phrases: Dimension folding, quadratic discriminant analysis, sufficient dimension reduction, tensor.

1. Introduction

Tensors, also known as multidimensional arrays, are a direct generalization of vectors and matrices (Hitchcock (1927); Kolda and Bader (2009)). Tensor data are observed in various applied fields. For example, in a study using gene expression time course data (Baranzini et al. (2005)), gene expressions for 53 multiple sclerosis patients were measured over multiple time points. After being given recombinant human interferon beta (rIFN β), which is often used to control the symptoms of multiple sclerosis, patients were classified as good ($Y = 1$) or poor ($Y = 0$) responders to rIFN β based on their clinical characteristics. For each of the 53 subjects, the matrix-variate predictor can be organized as $genes \times times = 76 \times 7$ and is used to predict the binary response Y . Another example is from neuroimaging studies, where we are interested in predicting whether a subject has a neurological disorder based on image scans in the form of three-way or four-way tensors. For such data sets, we may lose important structural

Corresponding author: Xin Zhang, Department of Statistics, Florida State University, Tallahassee, FL, 32306, USA. E-mail: henry@stat.fsu.edu.