Statistica Sinica **35** (2025), 1-22
doi:https://doi.org/10.5705/ss.202022.0007

THE TUCKER LOW-RANK CLASSIFICATION MODEL FOR TENSOR DATA

Junge Li, Qing Mai* and Xin Zhang

Florida State University

Abstract: With the rapid advances of modern technology, tensor data (i.e., multi-way array) have been collected in various scientific research and engineering applications. The classification of tensor data is of great interest, where predictive models and algorithms are proposed for predicting a categorical class label for each tensor-valued sample. Aiming to improve interpretability of tensor classification methods, we consider an intuitive and efficient discriminant analysis approach, referred to as the Tucker Low-rank Classification (TLC) model. The TLC model assumes that the between-class mean differences have a low-rank Tucker decomposition, while the covariance matrix is separable. As such, the TLC model greatly reduces the number of parameters by exploiting the tensor structure. We construct a penalized estimator for the TLC model to achieve a sparse Tucker decomposition on the key discriminant analysis parameters and to further improve the parsimony in the final classifier. We establish estimation, variable selection, and prediction consistency for the penalized estimator to confirm that the proposed estimator achieves efficiency gain compared to standard methods. We demonstrate the superior performance of TLC in extensive simulation studies and real data examples.

Key words and phrases: Classification, dimension reduction, discriminant analysis, Tucker tensor decomposition.

1. Introduction

Tensor data, also known as multi-way arrays, are often collected in modern scientific studies and engineering applications. For example, in gene expression analysis, observations are sometimes in the form of matrices (i.e., two-way tensors) with rows characterizing genes and columns representing experimental conditions, tissues, or time points. Neuroimaging studies work on analyzing electroencephalography (EEG, i.e., two-way tensors), anatomical magnetic resonance imaging (MRI, i.e., three-way tensors), functional magnetic resonance imaging (fMRI, i.e., four-way tensors), and so on.

The increasing popularity of tensor data has posed many challenges to statistical analysis. One such challenge is that tensor data are usually high-dimensional, which results in a large number of parameters and expensive computation. A more distinctive challenge is that multi-way data usually have information embedded in the tensor structure, which is not easy to exploit using

*Corresponding author. E-mail: qmai@fsu.edu