MODE-BASED CLASSIFIER:
A ROBUST AND FLEXIBLE DISCRIMINANT ANALYSIS
FOR HIGH-DIMENSIONAL DATA

Wei Xiong[1], Wolfgang Karl Härdle[2], Jianrong Wang[1],
Keming Yu[3] and Maozai Tian[*4]

[1]*University of International Business and Economics,*
[2]*Humboldt-Universitat zu Berlin,* [3]*Brunel University London*
and [4]*Renmin University of China*

Abstract: High-dimensional classification is both challenging and of interest in numerous applications. Componentwise distance-based classifiers, which utilize partial information with known categories, such as mean, median and quantiles, provide a convenient way. However, when the input features are heavy-tailed or contain outliers, performance of the centroid classifier can be poor. Beyond that, it frequently occurs that a population consists of two or more subpopulations, the mean, median and quantiles in this scenario fail to capture such a structure that can be instead preserved by mode, which is an appealing measure of considerable significance but might be neglected. This paper thus introduces and investigates componentwise mode-based classifiers that can reveal important structures missed by existing distance-based classifiers. We explore several strategies for defining the family of mode-based classifiers, including the unimodal classifiers, the multimodal classifier and the quantile-mode classifier. The unimodal classifiers are proposed based on componentwise unimodal distance and kernel mode estimation, and the multimodal classifier is constructed by identifying all the local modes of a distribution according to a novel introduced algorithm. We establish the asymptotic properties of these methods and demonstrate through simulation studies and three real datasets that the mode-based classifiers compare favorably to the current state-of-art methods.

Key words and phrases: Componentwise modal distance, multimodal classifier, multimodality, quantile-mode, unimodal classifier.

1. Introduction

In this work, we focus on the problem of classification for high-dimensional data, where the task is to assign a new observation to one class out of a finite collection of alternatives. Classification arises frequently from bioinformatics, computer vision, natural language processing, and a broad range of other fields, is an indispensable part of artificial intelligence. Most of the conventional methods depend on sufficient and balanced samples to make classifiers more efficient.

---

*Corresponding author. E-mail: mztian@ruc.edu.cn