Statistica Sinica **35** (2025), 1-22
doi:https://doi.org/10.5705/ss.202022.0380

HIGH-DIMENSIONAL SCALE INVARIANT DISCRIMINANT ANALYSIS

Ming Li[1], Cheng Wang[2], Yanqing Yin[3] and Shurong Zheng*[1]

[1]*Northeast Normal University,* [2]*Shanghai Jiao Tong University and* [3]*Nanjing Audit University*

Abstract: In this paper, we propose a scale invariant linear discriminant analysis classifier for high-dimensional data with dense signals. The method is valid for both cases that the data dimension is smaller or greater than the sample size. Based on recent advances of the sample correlation matrix in random matrix theory, we derive the asymptotic limits of the error rate which characterizes the influences of the data dimension and the tuning parameter. The major advantage of our proposed classifier is scale invariant and it is applicable to any variances of the feature. Several numerical studies are investigated and our proposed classifier performs favorably in comparison to some existing methods.

Key words and phrases: Dimension effect, discriminant analysis, random matrix theory, sample correlation matrix, scale invariant.

1. Introduction

Linear discriminant analysis (LDA), which can be dated back to Fisher (1936), is a fundamental problem in multivariate statistical analysis (Anderson, 2003, Chap. 6). From the perspective of methodology, LDA is closely related to many other important statistical methods such as principal component analysis, analysis of variance and regression analysis etc. In real problems, LDA usually has a reliable performance and can compete with many sophisticated methods such as neural networks and support vector machines (Hand, 2006).

In the era of high-dimensional data, many improved methods have been proposed for linear discriminant analysis. For example, to address the singularity of the sample covariance matrix, Dudoit, Fridlyand and Speed (2002) proposed a diagonal linear discriminant analysis (DLDA) which is valid for the situation that the data dimension p is greater than the sample size n. To reduce the dispersion of the eigenvalues of the sample covariance matrix, Friedman (1989) conducted a regularized linear discriminant analysis (RLDA), and see also Guo, Hastie and Tibshirani (2007) for RLDA. Moreover, various other improved LDA methods were studied under some certain assumptions. One may refer to the papers (Shao et al., 2011; Cai and Liu, 2011; Mai, Zou and Yuan, 2012; Fan, Feng and Tong, 2012; Fan, Jin and Yao, 2013; Hao, Dong and Fan, 2015) and references therein

*Corresponding author. E-mail: zhengsr@nenu.edu.cn