

Distance-based classifier by data transformation for high-dimension, strongly spiked eigenvalue models

Makoto Aoshima¹ · Kazuyoshi Yata¹

Received: 26 December 2016 / Revised: 12 January 2018 / Published online: 10 March 2018
© The Institute of Statistical Mathematics, Tokyo 2018

Abstract We consider classifiers for high-dimensional data under the strongly spiked eigenvalue (SSE) model. We first show that high-dimensional data often have the SSE model. We consider a distance-based classifier using eigenstructures for the SSE model. We apply the noise-reduction methodology to estimation of the eigenvalues and eigenvectors in the SSE model. We create a new distance-based classifier by transforming data from the SSE model to the non-SSE model. We give simulation studies and discuss the performance of the new classifier. Finally, we demonstrate the new classifier by using microarray data sets.

Keywords Asymptotic normality · Data transformation · Discriminant analysis · Large p small n · Noise-reduction methodology · Spiked model

Research of the first author was partially supported by Grants-in-Aid for Scientific Research (A) and Challenging Exploratory Research, Japan Society for the Promotion of Science (JSPS), under Contract Numbers 15H01678 and 26540010. Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), JSPS, under Contract Number 26800078.

✉ Makoto Aoshima
aoshima@math.tsukuba.ac.jp
<http://www.math.tsukuba.ac.jp/~aoshima-lab/>
Kazuyoshi Yata
yata@math.tsukuba.ac.jp

¹ Institute of Mathematics, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan