CrossMark

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

Makoto Aoshima[1] · **Kazuyoshi Yata**[1]

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

✉ 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

Springer