

A model selection criterion for discriminant analysis of several groups when the dimension is larger than the total sample size

Masashi Hyodo¹, Takayuki Yamada², and Takashi Seo¹

¹*Tokyo University of Sciences, Japan*

²*Kitasato University, Tokyo, Japan*

Abstract

The program of the Conference is concerned with the problem of selecting variables in multiple group discriminant analysis for high-dimensional data with fewer observations than the dimension. We consider a selection criterion based on approximately unbiased for AIC type of risk. When the dimension is large compared to the sample size, AIC type of risk cannot be defined. We propose AIC by replacing maximum likelihood estimator with ridge-type estimator. This idea follows Srivastava and Kubokawa (2008). It has been further extended by Yamamura et al. (2010). Simulation revealed that the proposed AIC performs well.

Keywords

Akaike information criterion, Discriminant analysis, Ridge-type estimator, High dimensional data.

References

- Srivastava, M.S. and Kubokawa, T. (2008). Akaike information criterion for selecting components of the mean vector in high dimensional data with fewer observations. *J. Japan Statist. Soc.* 38, 259–283.
- Yamamura, M., Yanagihara, H., and Srivastava, M.S. (2010). Variable selection in multivariate linear regression models with fewer observations than the dimension. *Japanese J. Appl. Statist.* 39(1), 1–19.