Check for
updates

Determining the number of canonical correlation pairs for high-dimensional vectors

Jiasen Zheng[1] · Lixing Zhu[2,3]

© The Institute of Statistical Mathematics, Tokyo 2021

Abstract

For two random vectors whose dimensions are both proportional to the sample size, we in this paper propose two ridge ratio criteria to determine the number of canonical correlation pairs. The criteria are, respectively, based on eigenvalue difference-based and centered eigenvalue-based ridge ratios. Unlike existing methods, the criteria make the ratio at the index we want to identify stick out to show a visualized "valley-cliff" pattern and thus can adequately avoid the local optimal solutions that often occur in the eigenvalues multiplicity cases. The numerical studies also suggest its advantage over existing scree plot-based method that is not a visualization method and more seriously underestimates the number of pairs than the proposed ones and the AIC and C_p criteria that often extremely over-estimate the number, and the BIC criterion that has very serious underestimation problem. A real data set is analyzed for illustration.

✉ Lixing Zhu
lzhu@hkbu.edu.hk

Jiasen Zheng
2017000815@ruc.edu.cn

[1] School of Statistics, Renmin University of China, Beijing 100872, China

[2] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

[3] School of Statistics, Beijing Normal University, Beijing 100875, China