Annals of the Institute of Statistical Mathematics (2022) 74:435–450
https://doi.org/10.1007/s10463-021-00803-5

Robust distributed estimation and variable selection for massive datasets via rank regression

Jiaming Luan[1] · Hongwei Wang[1] · Kangning Wang[1] · Benle Zhang[1]

Received: 23 November 2020 / Revised: 22 March 2021 / Accepted: 6 May 2021 /
Published online: 20 June 2021

Abstract

Rank regression is a robust modeling tool; it is challenging to implement it for the distributed massive data owing to memory constraints. In practice, the massive data may be distributed heterogeneously from machine to machine; how to incorporate the heterogeneity is also an interesting issue. This paper proposes a distributed rank regression (DR^2), which can be implemented in the master machine by solving a weighted least-squares and adaptive when the data are heterogeneous. Theoretically, we prove that the resulting estimator is statistically as efficient as the global rank regression estimator. Furthermore, based on the adaptive LASSO and a newly defined distributed BIC-type tuning parameter selector, we propose a distributed regularized rank regression (DR^3), which can make consistent variable selection and can also be easily implemented by using the LARS algorithm on the master machine. Simulation results and real data analysis are included to validate our method.

Keywords Massive data · Robustness · Communication efficient · Variable selection

K. Wang: The authors are listed in the alphabetical order. The authors would like to thank Dr. Shaomin Li for his valuable suggestions. The authors would like to thank the editor, an associate editor and two anonymous reviewers for their constructive comments that led to a major improvement of this article. The research was supported by NNSF project of China (11901356, 11901149), wealth management project (2019ZBKY047) of Shandong Technology and Business University.

✉ Kangning Wang
wkn1986@126.com

[1] Shandong Technology and Business University, No. 191, Binhai Middle Road, Laishan District, Yantai 264005, China