



Contribution ID : 4

Type : not specified

Subdata selection for big data regression based on leverage scores

mercoledì 28 giugno 2023 15:00 (20)

Data continues to become more abundant, and so the datasets that contain it. Even though big datasets can present insights and opportunities, they can pose significant challenges when it comes to statistical analysis. One of the biggest challenges, required to process and analyze large datasets, is the computational resources. Regression can be problematic in case of big datasets, due to the huge volumes of data. A standard approach is subsampling that aims at obtaining the most informative portion of the big data. We consider an approach based on leverage scores, already existing in the current literature for the selection of subdata for linear model discrimination. However, we highlight its importance on the selection of data points that are the most informative for estimating unknown parameters. We conclude that the approach based on leverage scores improves existing approaches, providing simulation experiments as well as a real data application.

Primary author(s) : CHASIOTIS, Vasilis (Athens University of Economics and Business, Department of Statistics); Prof. KARLIS, Dimitris (Athens University of Economics and Business, Department of Statistics)

Presenter(s) : CHASIOTIS, Vasilis (Athens University of Economics and Business, Department of Statistics)

Session Classification : First Session