



Contribution ID : 5

Type : not specified

A model-robust subsampling approach in presence of outliers

mercoledì 28 giugno 2023 17:40 (20)

In the era of big data, several sampling approaches are proposed to reduce costs (and time) and to help in informed decision making. Some of these proposals (Drovandi et al., 2017; Wang et al., 2019; Deldossi and Tommasi (2022) among others) are inspired to Optimal Experimental Design and require the specification of a model for the big dataset. This model assumption, as well as the possible presence of outliers in the big dataset represent a limitation for the most commonly applied subsampling criterions. Deldossi et al. (2023) introduced non-informative and informative exchange algorithms to select “nearly” D-optimal subsets without outliers in a linear regression model.

In this study, we extend their proposal to account for model uncertainty. More precisely, we propose a model robust approach where a set of candidate models is considered; the optimal subset is obtained by merging the subsamples that would be selected by applying the approach of Deldossi et al. (2023) if each model was considered as the true generating process. The approach is applied in a simulation study and some comparisons with other subsampling procedures are provided.

Key-words: Active learning, D-optimality, Subsampling

References

Deldossi, L., Tommasi C. (2022) Optimal design subsampling from Big Datasets. *Journal of Quality Technology* 54(1): 93–101

Deldossi, L., Pesce, E., Tommasi, C. (2023) Accounting for outliers in optimal subsampling methods, *Statistical Papers*, <https://doi.org/10.1007/s00362-023-01422-3>.

Drovandi CC, Holmes CC, McGree JM, Mengersen K, Richardson S, Ryan EG (2017) Principles of experimental design for big data analysis. *Statistical Sciences* 32(3): 385–404

Wang H, Yang M, Stufken J (2019) Information-based optimal subdata selection for Big Data linear regression. *Journal of American Statistical Association* 114(525): 393–405

Primary author(s) : DELDOSSI, Laura; TOMMASI, Chiara

Presenter(s) : DELDOSSI, Laura

Session Classification : Second Session