Description of the project

Statistical Learning from the Perspective of Random Matrix Theory

BY TARAS BODNAR (TARAS.BODNAR@MATH.SU.SE)
AND JOHANNES HEINY (JOHANNES.HEINY@MATH.SU.SE)

Prediction and classification problems are two main tasks considered in modern statistical learning theory. Different methods, like ridge regression, lasso regression, decision tree, logistic regression, discriminant analysis, etc., have recently been developed in the literature and successfully implemented in practice. Theoretical results, motivating the applicability of the existing approaches, are derived in most cases under the standard asymptotic regime, i.e., when the sample size is considerably larger than the number of predictors in the model. Even though the same procedures are applied in the high-dimensional setting, i.e., when the number of predictors is comparable to the sample size, theoretical justifications are usually not available, and one can face unexpected results.

Within the project, we are going to contribute to statistical learning theory by developing new theoretical findings that explain and justify the applicability of the existing prediction and classification approaches in the high-dimensional setting. The goal of the project will be achieved by developing new theoretical results in random matrix theory and their application to the problems at hand. As a byproduct of the developed theory, it is expected to obtain new methods for both prediction and classification problems, whose finite-sample properties will be assessed via an extensive simulation study. The proposed new approaches will also be compared to the existing ones and implemented for real data.

More information about the supervisors' past and ongoing research can be found at
https://www.su.se/english/profiles/tbodn-1.219689?open-collapse-boxes=body-research
https://www.su.se/english/profiles/johe3032-1.640812.

REFERENCES

Bodnar, T., Dette, H. and Parolya, N. Testing for independence of large dimensional vectors.
Annals of Statistics 47 (2019), no. 5, 2977–3008. [arXiv:1708.03964]

Heiny, J., and Yao, J. Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations.
Annals of Statistics 50 (2022), no. 6, 3249–3280. [arXiv:2003.03857]

Bastian, P., Dette, H. and Heiny, J. Testing for practically significant dependencies in high dimensions via bootstrapping maxima of U-statistics.
Annals of Statistics (2024), to appear. [arXiv:2210.17439]