

Stabilizing variable selection and regression

Niklas Pfister

University of Copenhagen, Denmark, np@math.ku.dk

A common setup in many data-driven scientific fields is to observe data across several different experiments or environments. Often these experiments are explicitly constructed to ensure that parts of the data generating mechanism change – the hope being that correlation structures that are detectable across multiple different settings are more likely to be causal. In this talk, we will introduce a formal framework for this type of analysis by considering a multi-environment regression setting in which a response Y is regressed on a set of predictors X . We will show that we can gain additional insights into the causal structure between Y and X by distinguishing between stable and unstable predictors (i.e., predictors which have a fixed or a changing functional dependence on the response, respectively). We apply these ideas to hypothesis generation in multiomic data.

This talk is based on joint work with Evan G. Williams, Jonas Peters, Ruedi Aebersold and Peter Bühlmann [1].

References

- [1] Pfister, N., Williams, E. G., Peters, J., Aebersold, R. and Bühlmann, P. (2021). Stabilizing variable selection and regression. *Annals of Applied Statistics*, (to appear).