

Computational statistics: Data Science for Social Science

Marina Khismatullina

Institute for Finance and Statistics, University of Bonn

marina.k@uni-bonn.de

The final project should focus on one of the methods that we discussed in class:

1. Linear regression methods:
 - (a) principal component regression;
 - (b) ridge regression;
 - (c) lasso;
2. Tree-based methods:
 - (a) regression trees;
 - (b) random forests;
 - (c) bagging;
 - (d) boosting;
 - (e) stacked methods/ensemble methods;
 - (f) causal forests.

The project should include

1. a description of the method,
2. a simulation study using a realistic set-up from a research paper in economics as a motivation and benchmarking,
3. and (ideally) an empirical application.

The description should also include theoretical properties of the method together with the properties of the data most often encountered in empirical settings. The simulation study should reflect these two components. You may collaborate (i.e. discuss) with other students who are working within the same general method, for example, lasso, but everyone has to hand in an **individual** project.

Some journals that you can browse for inspiration on empirical work:

- American Economic Journal: Applied Economics

- Journal of Labor Economics
- Journal of Health Economics
- Journal of Development Economics
- Journal of Public Economics
- Journal of Applied Econometrics
- ... AER, ReStud, Quantitative Economics

And a nice paper as an inspiration for using data science methods for economics applications: <https://www.aeaweb.org/articles?id=10.1257/aer.p20171040>.