Speaker: Dominik Rothenhaeusler, ETH Zurich

Title: Anchor Regression: Heterogeneous Data Meets Causality

Abstract:

Many traditional statistical prediction methods mainly deal with the problem of overfitting to the given data set. On the other hand, there is a vast literature on the estimation of causal parameters for prediction under interventions. However, both types of estimators can perform poorly when used for prediction on heterogeneous data. We discuss the delicate trade-off between predictive performance on the training data and perturbed data. In particular, under a linear structural equation model with exogenous variables, we show that the change in loss under certain perturbations (interventions) can be written as a convex penalty. This motivates anchor regression, a regularization scheme that encourages the estimator to generalize well to perturbed data. Under instrumental variable (IV) assumptions, the procedure naturally provides an interpolation between the solution to ordinary least squares and the IV estimator. The proposed procedure allows statisticians and practitioners to trade-off predictive performance on the distribution of the training data and on distributions which are perturbed versions of what is seen in the training data.