

**University of California, Irvine
Statistics Seminar**

Statistical Learning with Dependent Data

**Sumanta Basu
Assoc. Professor of Statistics and Data Science
Cornell University**

**4 p.m., Thursday, Oct. 31, 2024
6011 Donald Bren Hall**

With advances in data collection and storage, statistical learning algorithms are becoming increasingly popular for structure learning and prediction with large-scale data sets that exhibit temporal or spatial dependence. Most algorithms in the literature focus on using off-the-shelf machine learning algorithms that ignore the dependent nature of the data. In this talk, we aim to demonstrate the merit of incorporating classical statistical wisdoms for scale and dependence modeling into the statistical learning framework through two algorithms that we developed. The first, called RF-GLS, extends random forests (RF) for dependent error processes in the same way Generalized Least Squares (GLS) fundamentally extends Ordinary Least Squares (OLS) for linear models under dependence. The second algorithm, called AutoTune, offers an automatic tuning parameter selection algorithm for LASSO, by revisiting the well-known problem of scale estimation and adjustment for high-dimensional regression. We illustrate the benefit of these algorithms on simulated data sets, and provide some theoretical analysis to shed insight on their asymptotic properties.