STATISTICS SEMINAR

Weikang Duan

PhD Preliminary Defense

Monday, November 29th, 2021
Dickens 302, 1:00 pm (Central Time)

Bayesian approaches for high-dimensional nonlinear regression challenges

In the past decades, statistical learning has been an increasingly popular topic that has drawn a significant amount of attention from researchers. Nonlinear regression models are powerful tools due to their flexibility to extract information from complex nonlinear datasets. A major challenge with the nonlinear regression modeling in the current big data era is the curse of dimensionality. Although an abundance of variable selection methods have been proposed, the developments in high-dimensional Bayesian nonlinear regression is still in infancy. In this dissertation, we develop fast and powerful variable selection procedures in order to address the problem of high-dimensional nonlinear regression from a Bayesian perspective.

In the first project, we address the challenge in Gaussian process (GP) regression modeling, which is the most popular Bayesian approach to nonlinear regression. Even though approaches to address variable selection for the GP models have been proposed in both frequentist and Bayesian frameworks, the existing methods have many limitations. In this project, we propose a Bayesian model hybrid search algorithm to quickly scan through the model space to search for a set of models having high posterior probabilities. Prediction is then conducted via Bayesian model averaging. In addition to the variable selection problem, another challenge is how to deal with the case of massive sample size under high-dimensional data settings. To address the massive and high-dimensional data problem, we propose an approach which combines quantile subsample hybrid search with nearest neighbor GP. We examine the results of our proposed methods through both simulation studies and real data analysis.

In the second project, we focus on variable selection for reproducing kernel Hilbert space (RKHS) methods under the nonlinear regression framework. For RKHS models, simultaneous variable selection and sparse kernel estimation, called the doubly sparse estimation problem, are needed. To address this issue, some attempts have been made using the penalized likelihood estimation techniques. One of limitations with the penalized likelihood approach is that it does not address the uncertainties associated with the simultaneous variable selection and sparse kernel estimation. In addition, tuning parameter selection tends to greatly increase the computational cost. In this project, we propose a Bayesian doubly sparse RKHS regression method. We demonstrate the advantage of our proposed method through both simulation studies and real data analysis.