



ASA Pittsburgh Chapter 2021 Spring Virtual Banquet

Keynote Speaker: Dr. Bin Yu

Chancellor's Distinguished Professor and Class of 1936 Second Chair in the departments of statistics and EECS at UC Berkeley. Member of the U.S. National Academy of Sciences and of the American Academy of Arts and Sciences.

Title: Veridical Data Science for biomedical discovery: detecting epistatic interactions with epiTree.

Abstract: "A.I. is like nuclear energy -- both promising and dangerous" -- Bill Gates, 2019.

Data Science is a pillar of A.I. and has driven most of recent cutting-edge discoveries in biomedical research. In practice, Data Science has a life cycle (DSLCL) that includes problem formulation, data collection, data cleaning, modeling, result interpretation and the drawing of conclusions. Human judgment calls are ubiquitous at every step of this process, e.g., in choosing data cleaning methods, predictive algorithms and data perturbations. Such judgment calls are often responsible for the "dangers" of A.I. To maximally mitigate these dangers, we developed a framework based on three core principles: Predictability, Computability and Stability (PCS). Through a workflow and documentation (in R Markdown or Jupyter Notebook) that allows one to manage the whole DSLCL, the PCS framework unifies, streamlines and expands on the best



practices of machine learning and statistics – bringing us a step forward towards veridical Data Science.

In this lecture, we will illustrate the PCS framework through the epiTree; a pipeline to discover epistasis interactions from genomics data. epiTree addresses issues of scaling of penetrance through decision trees, significance calling through PCS p-values, and combinatorial search over interactions through iterative random forests (which is a special case of PCS). Using UK Biobank data, we validate the epiTree pipeline through an application to the red-hair phenotype, where several genes are known to display epistatic interactions.

Date: Friday, April 23rd, 2021, 4:30-5:30 PM EST

Location: <https://pitt.zoom.us/j/97891483192>