WNAR Webinar by Dr. Bin Yu
Veridical Data Science for biomedical discovery:
detecting epistatic interactions with epiTree

It is a great pleasure to announce the upcoming WNAR Webinar by our distinguished colleague Dr. Bin Yu from UC Berkeley.

Time: 9:30am - 10:30am PST (12:30 PM - 1:30 PM EST), Friday February 26, 2021

Registration is required:

https://www.eventbrite.com/e/wnar-webinar-veridical-data-science-for-biomedical-discovery-tickets-140230781041

This is a free event. Priority registration is given to WNAR members until Monday, Feb. 15. After that, registration is open to the public and ends on Thursday, Feb. 25. So, hurry up!

Zoom Dial-in information and webinar link will be sent 24 hours before the event.

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 (DSLC) that includes problem formulation, data collection, data cleaning, modeling, result interpretation and the drawing of conclusions. Human judgement 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 DSLC, 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.

Speaker Bio:
Dr. Bin Yu is the Chancellor's Professor in the Departments of Statistics and of Electrical Engineering & Computer Sciences at the University of California, Berkeley. Dr. Yu's work leverages computational developments to solve scientific problems by combining statistical machine learning approaches with the domain expertise of many collaborators, spanning many fields including statistics, machine learning, neuroscience, genomics, and remote sensing. Her recent work has focused on solidifying a vision for data science, including a framework for veridical data science and a framework for interpretable machine learning. Yu has received recent news coverage regarding her veridical data science framework, investigations into the theoretical foundations of deep learning, and work forecasting COVID-19 severity in the US. Dr. Yu is a fellow of the Institute of Mathematical Statistics, the IEEE, the American Statistical Association, the American Association for the Advancement of Science, the American Academy of Arts and Sciences, and the National Academy of Sciences. In 2012, she was the Tukey Lecturer of the Bernoulli Society for Mathematical Statistics and Probability. In 2018, she was awarded the Elizabeth L. Scott Award. She was invited to give the Breiman lecture at NeurIPS 2019 (formally known as NIPS), on the topic of veridical data science.

WNAR is the Western North American Region of The International Biometric Society, an international professional and academic society promoting the development and application of statistical and mathematical theory and methods in the biosciences. Visit www.wnar.org for more details and events or follow us on https://twitter.com/wnar_ibs