

# Genome-Phenome Association Analysis of Complex Diseases a Structured Sparse Regression Approach (Keynote Talk)

Eric Xing

School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA 15213, USA  
epxing@cs.cmu.edu

Genome-wide association (GWA) studies have recently become popular as a tool for identifying genetic variables that are responsible for increased disease susceptibility. A modern statistical method for approaching this problem is through model selection (or structure estimation) of Structured Input-Output Regression Models (SIORM) fitted on genetic and phenotypic variation data across a large number of individuals.

The inputs of such models bear rich structure, because the cause of many complex disease syndromes involves the complex interplay of a large number of genomic variations that perturb disease-related genes in the context of a regulatory network. Likewise, the outputs of such model are also structured, as patient cohorts are routinely surveyed for a large number of traits such as hundreds of clinical phenotypes and genome-wide profiling for thousands of gene expressions that are interrelated. A Structured Input-Output Regression Model nicely captures all these properties, but raises severe computational and theoretical challenge on consistent model identification.

In this talk, I will present models, algorithms, and theories that learn Sparse SIORMs of various kinds in very high dimensional input/output space, with fast and highly scalable optimization procedures, and strong statistical guarantees. I will demonstrate application of our approaches to a number of complex GWA scenarios, including associations to trait networks, to trait clusters, to dynamic traits, under admixed populations, and with epistatic effects.

This is joint work with Seyoung Kim, Mladen Kolar, Seunghak Lee, Xi Chen, and Kriti Puniyani and Judie Howrylak.