Abstract:
Modeling and tractable computation form two fundamental but competing pillars of data science; indeed, fitting good models to data is often computationally challenging in modern applications. Focusing on the canonical tasks of ranking and regression, I introduce problems where this tension is immediately apparent, and present methodological solutions that are both statistically sound and computationally tractable.

I begin by describing a class of “permutation-based” models as a flexible alternative to parametric modeling in a host of inference problems including ranking from ordinal data. I introduce procedures that narrow a conjectured statistical-computational gap, demonstrating that carefully chosen non-parametric structure can significantly improve robustness to mis-specification while maintaining interpretability. Next, I address a complementary question in the context of convex regression, where I show that the curse of dimensionality inherent to non-parametric modeling can be mitigated via parametric approximation. Our provably optimal methodology demonstrates that it is often possible to enhance the interpretability of non-parametric models while maintaining important aspects of their flexibility.