The Small-Data, Large-Scale Optimization Regime

Vishal Gupta
USC Marshall School of Business

Abstract

Despite the promises of Big Data, data in modern operations research applications can be scarce. Such data scarcity can arise because of high-granularity in decision-making, rapidly time-varying environments, prohibitively expensive data collection, large measurement imprecision, or concerns around privacy. Most importantly, this data scarcity is typically unavoidable.

In this talk we introduce the *small-data, large-scale optimization regime*, an asymptotic setting that arguably better describes these types of applications than the more traditional large-sample regime. We highlight unique phenomena that emerge in the small-data, large-scale regime, and show how these phenomena can cause certain traditional data-driven optimization algorithms to perform quite poorly. Finally, we show how to exploit these new phenomena to design novel algorithms with provably good performance in our small-data, large-scale settings and empirically strong performance in those applications plagued by data scarcity.