Learning and Optimization: Separate or Integrate?

Nathan Kallus
Cornell University

Abstract

Predictive side information (aka "context") reduces the uncertainty in optimization under uncertainty, but leveraging this requires we learn a potentially complex predictive relationship. We can potentially use off-the-shelf ML methods to learn it, but the training process ignores the downstream optimization task where we eventually plug in the model. Alternatively, we can train the model in an end-to-end fashion to directly optimize the downstream costs of the decision policy it would induce. In this talk I will tackle the question, which is better? Should we separate or integrate the learning and optimization tasks?

In the first part of this talk I will focus on contextual linear optimization, where the cost function is bilinear in the decision and uncertain variables so, by linearity of expectation, the only relevant aspect of the predictive relationship is the conditional expectation (aka regression function). We show that the naive separated approach actually achieves regret convergence rates that are significantly faster than any end-to-end method that directly optimizes downstream decision performance. Specifically, we leverage the fact that specific problem instances do not have arbitrarily bad near-dual-degeneracy and develop appropriate upper and lower regret bounds in view of this. This is overall positive for practice: predictive models are easy and fast to train using existing ML tools, simple to interpret and reuse as a prediction, and, as shown, lead to decisions that perform very well.

In the second part of this talk I will focus on general (nonlinear) contextual stochastic optimization problems, where we must consider the whole conditional probability model. As this object can be much higher dimensional than any decision policy, it is here better to integrate the tasks and directly learn a policy. We adapt random forests to this task by searching over tree splits to directly optimize downstream decisions, rather than prediction accuracy. This is a seemingly computationally intractable problem that we solve by developing approximate splitting criteria that utilize optimization perturbation analysis to eschew burdensome optimization for each candidate split. We prove that the approximations are consistent and the method is asymptotically optimal, and we empirically validate its superior performance.

This talk is based on the following papers:
Fast Rates for Contextual Linear Optimization
Stochastic Optimization Forests
Bio:

Nathan Kallus is an Assistant Professor in the School of Operations Research and Information Engineering and Cornell Tech at Cornell University. Nathan's research interests include optimization, especially under uncertainty; causal inference; sequential decision making; and algorithmic fairness. He holds a PhD in Operations Research from MIT as well as a BA in Mathematics and a BS in Computer Science from UC Berkeley. Before coming to Cornell, Nathan was a Visiting Scholar at USC's Department of Data Sciences and Operations and a Postdoctoral Associate at MIT's Operations Research and Statistics group.