Bayesian optimization for automating model selection

Roman Garnett, Washington University in St. Louis

Abstract: We discuss a problem of enormous practical importance that is often ignored: the selection of appropriate prior mean/covariance functions for Gaussian process models. Despite the success of kernel-based nonparametric methods, kernel selection still requires considerable expertise, and is often described as a “black art.” We present a sophisticated method for automatically searching for an appropriate kernel from an infinite space of potential choices. Previous efforts in this direction have focused on traversing a kernel grammar, only examining the data via computation of marginal likelihood. Our proposed search method is based on Bayesian optimization in model space, where we reason about model evidence as a function to be maximized. We explicitly reason about the data distribution and how it induces similarity between potential model choices in terms of the explanations they can offer for observed data. In this light, we construct a novel kernel between models to explain a given dataset (a "kernel kernel"). Our method is capable of finding a model that explains a given dataset well without any human assistance, often with fewer computations of model evidence than previous approaches, a claim we demonstrate empirically.