Bayesian Optimization

Abhinav Jain[1] and Yogesh Sharma[2]

[1]Department of Electrical Engineering, I.I.T Kanpur
[2]Department of Aerospace Engineering, I.I.T Kanpur

August 25, 2016

Often, a data scientist faces the problem of tuning up a large set of parameters or hyper parameters manually. And this search space is too vast to navigate. Bayesian optimization is a powerful solution for these problems which provides an excellent tool for finding good Machine Learning hyper parameters.

For instance, Automatic machine learning and hyper parameter tuning pursues the goal to automatically select the best model (e.g., random forests, support vector machines, neural networks, etc.) and its associated hyper parameters for solving a task on a given data set. Bayesian optimization comes handy for this task. Its approach for model selection and tuning can be utilised in tuning CNN, deep belief networks, MCMC etc.

Bayesian optimization is a sequential design strategy for global optimization of black-box functions that doesn't require derivatives. It treats objective function as a random function and places a prior over it. The prior captures our beliefs about the behaviour of the function. After gathering the function evaluations, which are treated as data, the prior is updated to form the posterior distribution over the objective function, f. The posterior distribution, in turn, is used to construct an acquisition function, α (often also referred to as infill sampling criteria) that determines what the next query point should be. Bayesian optimization algorithms select the next query point by maximizing acquisition functions like Thompson sampling, probability of improvement, expected improvement, upper-confidencebounds, and entropy search which trade off exploration and exploitation. Their optima are located where the uncertainty in the surrogate model is large (exploration) and/or where the model prediction is high (exploitation).

By studying the ingredients of Bayesian Optimization in depth, primarily focusing on statistical modelling which leads to general algorithms to solve a broad range tasks, we intend to exploit and review its impact in areas like automatic ML, combinatorial optimisation, adaptive MonteCarlo, reinforcement learning. Several open source packages like Spearmint provide a platform for implementing various forms of Bayesian optimization. As a sanity check for the algorithms that we analyze, we intend to implement the ones that are practically feasible, and test them on publicly available data sets.

References

[1] Ziyu Wang Ryan P. Adams Bobak Shahriari, Kevin Swersky and Nando de Freitas. *Taking the Human Out of the Loop: A Review of Bayesian Optimization.* Proceedings of the IEEE, Volume 104, 2016.

[2] Ryan P. Adams. A tutorial on bayesian optimization for machine learning.