Accelerator Parameter Optimization using high-dimensional Bayesian Optimization

Project Proposal For Bachelor Thesis

Motivation

Tuning hyperparameters is usually considered a computationally intensive and tedious task, be it for hyper-parameters in neural networks, or complex physical instruments, such as free electron lasers. Users for such applications could benefit from a 'one-click-training' feature, which would find optimal parameters given some reward function in as few steps as possible. This project proposal aims to find such an algorithm which is both efficient, and holds certain convergence guarantees. We focus our efforts in Bayesian Optimization techniques that maximize on a Gaussian Process, and revise techniques for high-dimensional BO.

Background

In Bayesian optimization, we want to use a Gaussian Process to find an optimal parameter setting $\mathbf{x}^*$ that maximizes a given utility function f. We assume the response surface to be locally continuous.

We refer to one observation as y, and to multiple observations as $\mathbf{y}$. We refer to a setting of n parameters as $\mathbf{x}$, where each individual scalars represents exactly one parameter. A set of observations then is the matrix $\mathbf{X}$. The relationship between the observations y and individual parameter settings $\mathbf{x}$ is $y = f(\mathbf{x}) + \epsilon$ where $\epsilon \sim \mathcal{N}\left(0, \sigma_n^2\right)$. Any quantity to be predicted has a subscript-star (e.g. y_* is the observation we want to predict).

In it's simplest form, a Gaussian procedure is described by the following equation:

$$\begin{pmatrix} y \\ y_* \end{pmatrix} \sim N\left(\mu(\mathbf{x}), \begin{pmatrix} K & K_*^T \\ K_* & K_{**} \end{pmatrix}\right), \tag{1}$$

Where $K = \text{kernel}(\mathbf{X}, \mathbf{X})$, $K_* = \text{kernel}(\mathbf{x}_*, \mathbf{X})$ and $K_{**} = \text{kernel}(\mathbf{x}_*, \mathbf{x}_*)$. Using this, the prediction for any point y_*, given all previously sampled points y by estimating the probability $p(y_*|y) \sim N(K_* K^{-1} y, K_{**} - K_* K^{-1} K_*')$

This, in turn, can be used to build an acquisition function. This acquisition function describes where to best sample points next. Some popular acquisition functions include GP-UCB, Most probable improvement (MPI) and Expected Improvement (EI). The choice of the acquisition function has great influence on the performance of the optimization procedure.

We will talk about the problems and possible solutions in the next section.

Scope Of The Project

In certain settings, Bayesian optimization suffers from the curse of dimensionality. Thus, the goal of this project is to implement different solutions that resolve the curse of dimensionality with regards

to Bayesian optimization. This project includes, but is not limited to the following methods.

Some proposed solutions are the following:

1. Sampling the active subspace

2. Forming groups in lower dimensions

3.

Assuming enough BO algorithms have been tested out, we are open to switch to other methods, such as other sampling algorithms or even neural networks.