Differentiating BO Trajectories

Darian Nwankwo

Cornell University

don4@cornell.edu

October 23, 2020

Overview

Black-Box Optimization

- Goal: Given domain Ω, find a global minimum $\mathbf{x}^* \in \Omega$:

$$f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \Omega$$

- Assume $f(\mathbf{x})$ is continuous, expensive to evaluate, no gradient information, and possibly noisy.

Bayesian Optimization

Bayesian Optimization in a nutshell:

1. Gather initial samples
2. Initialize our model
3. Get the acquisition function $\alpha(\mathbf{x})$
4. Optimize the acquisition function
5. Sample new data based on results from the optimization of $\alpha(\mathbf{x})$ and update model
6. Repeat until budget is exhausted
7. Make a recommendation

Bayesian Optimization

- Expected improvement (EI) is the standard acquisition function:

$$EI(\mathbf{x}) = \mathbb{E}[max(f(\mathbf{x}) - f(\mathbf{x}^*), 0)],$$

where $f(\mathbf{x})$ is distributed according to a GP posterior.

Bayesian Optimization

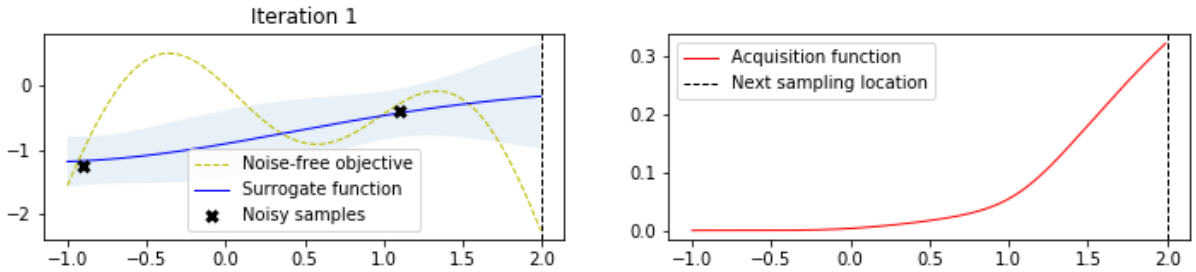

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

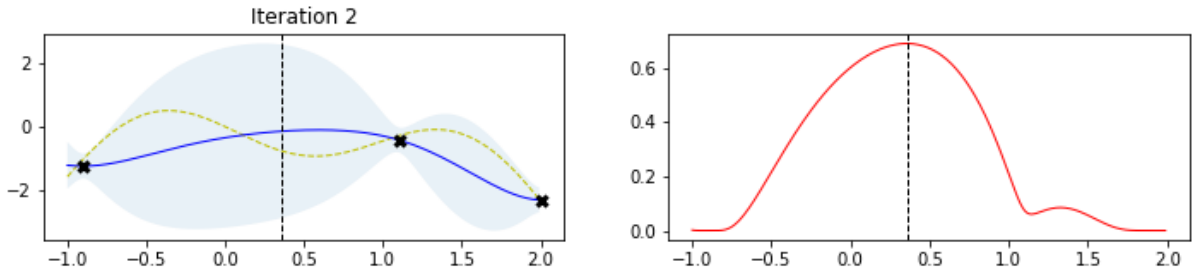

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

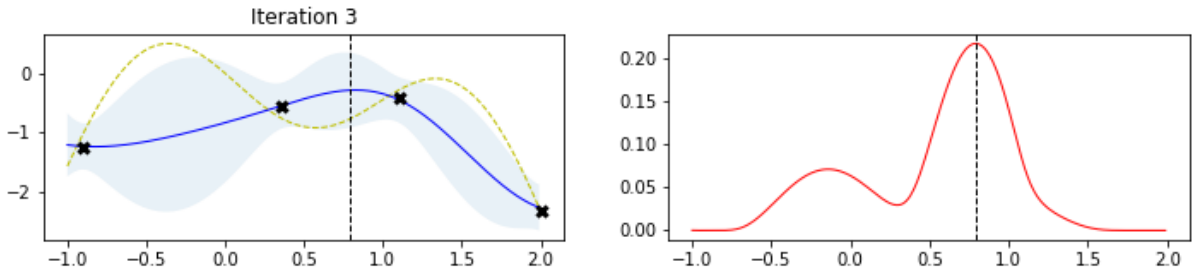

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

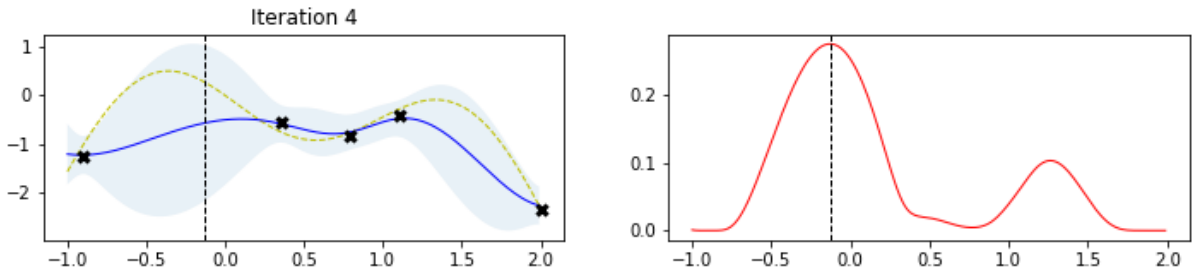

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

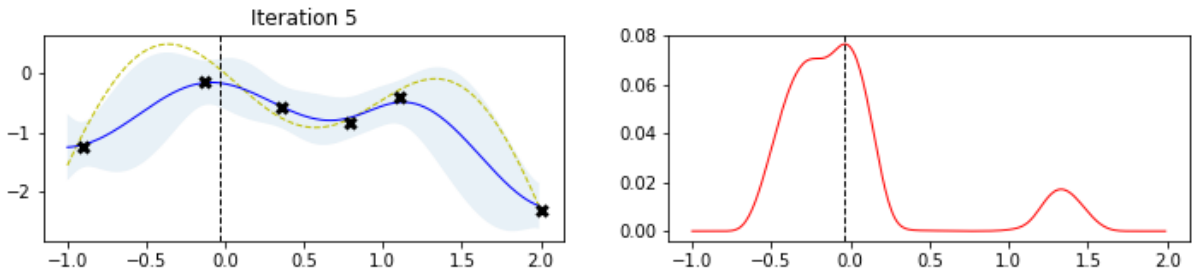

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

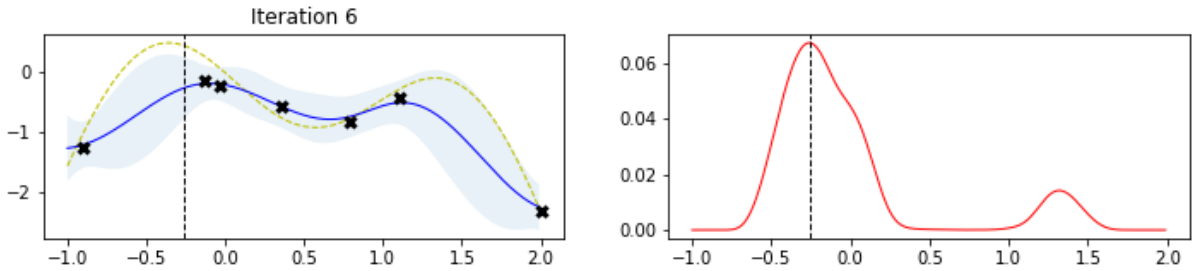

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

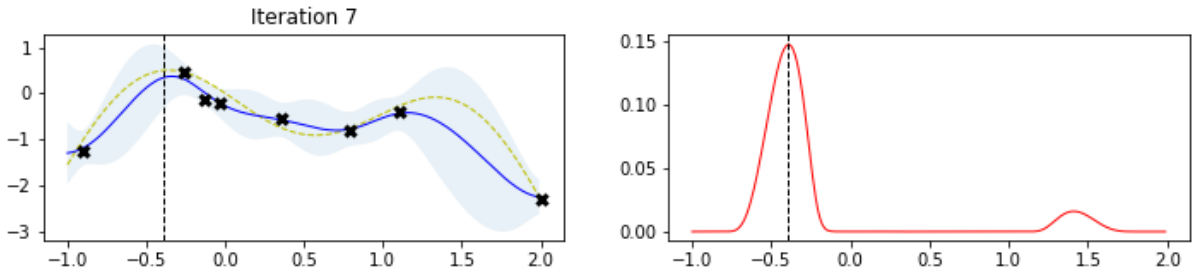

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

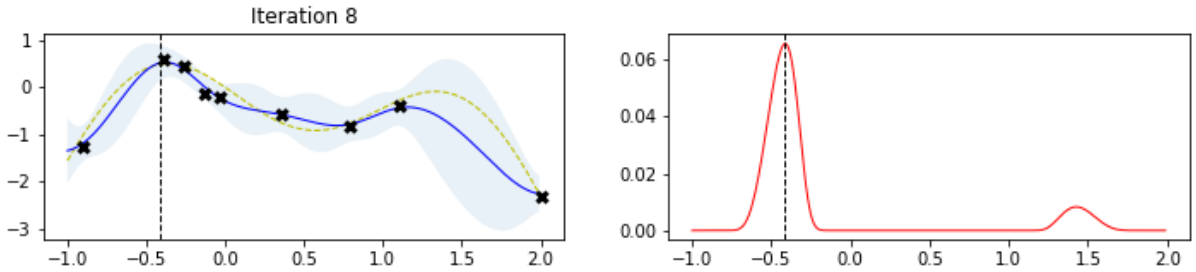

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

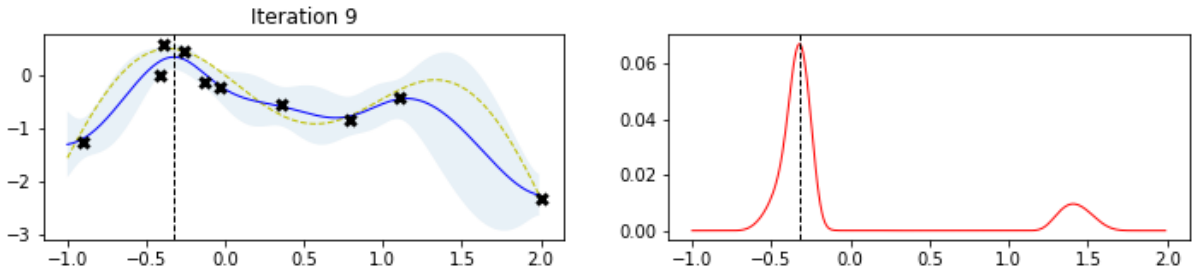

Figure: Left: GP Model. Right: EI Acquisition Function

Bayesian Optimization

Recall step 3 from our previous slide

- Get the acquisition function $\alpha(\mathbf{x})$

There exist two families of acquisitions functions that are of particular interest:

- Myopic
- Non-Myopic

Here, we are particularly interested in non-myopic acquisition functions.

Bayesian Optimization

Note: Non-myopic acquisition functions consider the impact of the next h function evaluations and are typically computed through rollout, in which h steps of BO are simulated.

Differentiating BO Trajectories

An h-step EI *rollout* policy involves choosing x^{n+1} based on the anticipated behavior of the EI algorithm starting from x^{n+1} and proceeding for h steps. That is, we consider the iteration

$$x^{k+1} = \text{argmax}_x \, \alpha(x|X^k, y^k, f^{+k}, \theta^k)$$

where θ^k denotes the kernel hyperparameters chosen at step k.

Differentiating BO Trajectories

Gradient information is particularly useful in this computation, but it requires that we be able to differentiate the whole (anticipated) sequence. That is, we might suppose that $\hat{f}$ is a draw of the GP, conditioned on the data from the first n samples, and then try to compute the derivative with respect to the proposed sample point x^{n+1} of the maximum of $\hat{f}(x^k)$ for k from 1 to $n+s$.

Differentiating BO Trajectories

We are interested in:

$$x^{k+1} = \text{argmax}_x \, \alpha(x|X^k, y^k, f^{+k}, \theta^k)$$
$$\therefore \nabla\alpha(x^{k+1}|X^k, y^k, f^{+k}, \theta^k) = 0$$

It's useful to think of x^{k+1} as an implicit function, thus we differentiate the basic iteration at step k to yield the following:

$$H_\alpha \delta x^{k+1} + \sum_{j=1}^{k} \left(\frac{\partial \nabla \alpha}{\partial x^j} \delta x^j + \frac{\partial \nabla \alpha}{\partial y_j} \delta y_j \right) + \frac{\partial \nabla \alpha}{\partial f^{+k}} \delta f^{+k} + \frac{\partial \nabla \alpha}{\partial \theta^k} \delta \theta^k = 0.$$

Differentiating BO Trajectories

Here we consider the derivatives of the expected improvement function needed for Newton's method and for differentiation of the argmax with respect to data and hyperparameters. We write the expected improvement acquisition function as

$$\begin{aligned}\alpha(x) &= \sigma(x)g(z(x)) \\ g(z) &= z\Phi(z) + \phi(z) \\ z(x) &= \sigma(x)^{-1}\left[\mu(x) - f^{+} - \xi\right]\end{aligned}$$

where Φ and ϕ denote the standard normal CDF and PDF, respectively; f^{+} is the best function value found so far; and ξ is a parameter to encourage additional exploration.

Differentiating BO Trajectories

Given the aforementioned information, we structure the computation from the bottom up. Differentiating the kernel function, then the predictive mean and variance, then z and finally α.

Note: We will use the notation $f_{,i}$ to denote $\partial f / \partial x_i$, and $\dot{f}$ to denote differentiation with respect to data or an arbitrary hyperparameter. Except in this initial paragraph, we will generally suppress the parameter x, leaving it implicit.

Differentiating BO Trajectories

Our goal is two-fold:

1. We want to compute the derivatives necessary for Newton iteration on the problem of maximizing α; that is, we want the gradient components $\alpha_{,i}$ and the Hessian components $\alpha_{,ij}$.
2. Given x^* such that $\alpha_{,i}(x^*) = 0$, we want to view x^* as an implicit function of the data and input hyper-parameters, and compute derivatives of x^* via implicit differentiation:

$$\alpha_{,ij}\dot{x}^*_j + \dot{\alpha}_{,i} = 0.$$

Differentiating BO Trajectories

Sketching out the computation needed for α alone (in the interest of time) we have:

$$
\begin{aligned}
\alpha &= \sigma g(z) \\
\alpha_{,i} &= \sigma_{,i} g(z) + \sigma g'(z) z_{,i} \\
\alpha_{,ij} &= \sigma_{,ij} g(z) + \sigma_{,i} g'(z) z_{,j} + \sigma_{,j} g'(z) z_{,i} + \sigma g'(z) z_{,ij} + \sigma g''(z) z_{,i} z_{,j} \\
&= \sigma_{,ij} g(z) + [\sigma_{,i} z_{,j} + \sigma_{,j} z_{,i} + \sigma z_{,ij}] g'(z) + \sigma g''(z) z_{,i} z_{,j}
\end{aligned}
$$

We also may want the mixed derivative with respect to spatial coordinates and data and hypers:

$$
\dot{\alpha}_{,i} = \dot{\sigma}_{,i} g(z) + \sigma_{,i} g'(z) \dot{z} + \dot{\sigma} g'(z) z_{,i} + \sigma g''(z) \dot{z} z_{,i} + \sigma g'(z) \dot{z}_{,i}
$$

The End