Part 1 - A Useless Tutorial for using KG

Aniket

University of Michigan

22 Aug

Plan

Overview

A presentation in multiple parts that:

1. Adds to our toolbox of acquisition functions: UCB, PI, EI (Expected Improvement) with another acronym
2. (potentially some theory) Sequential sampling policies with KG
3. Applications (covariance emulation)

Some references

1. A Tutorial on Bayesian Optimization (Frazier, 2018)
2. Maximizing Acquisition Functions for BO (Wilson et al., 2018)
3. BO for noisy experiments (Letham et al., 2019)

Bayesian Optimization

1. Derivative-free optimization method for black box functions, almost always selecting one measurement at a time.
   $d^* \in \arg\max_d f(d, \xi)$ or (OUU variant):
   $d^* \in \arg\max_d \mathbb{E}(f(d, \xi))$
2. Two ingredients: Probabilistic surrogate model (99 % of times this is a GP):
3. Maximize an acquisition function quantifying the utility of a given design point:

What does f usually look like?

1. $d < 20$ (this if from a 2016 ;) talk on BO in the Tech Sector - Frazier)
2. Feasible set is simple e.g. box constraints
3. no gradient information from evaluations (note that gradients could be used in optimization of the acquisition, see (Wilson et al., 2018) for an overview)
4. Expensive / Time consuming to evaluate
5. Noisy measurements!

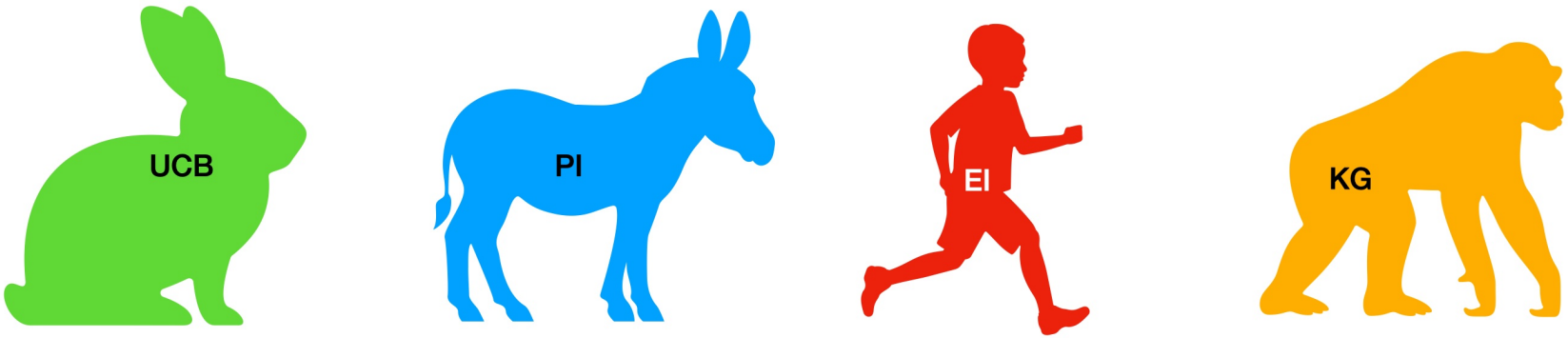
A Vast and Unending Field

Impossible to enumerate everything that goes on here, some examples:

1. Plenty of exotic variants e.g. high-dim BO, composite BO, hedging methods etc. (Eriksson et al., 2020; Miller and Mak, 2024; Astudillo and Frazier, 2019; Brochu et al., 2011) and surprising new findings e.g. (Hvarfner et al., 2024)
2. Plenty of software (some good, some beta) e.g. BoTorch (`https://botorch.org`), vanilla BO (`https://github.com/bayesian-optimization/BayesianOptimization`), CornellMOE (`https://github.com/wujian16/Cornell-MOE`)
3. To achieve fame and success - find your own acquisition function, write software to do domain-agnostic BO, tailor acquisition to very hard problems!

Common Acquisition Functions

Zoo of acquisition functions!



Figure

Digression - regret optimal?

Plan

KG Definition

Why KG right now? We are interested in this acquisition for (potentially) selecting model subsets for pilot sampling / multifidelity estimation and thought this a good opportunity to explore the formulation.

KG Definition

Other KG features?

1. No tuning parameters in the acquisition?

Acquisition and Updates

Acquisition and updates

Comparisons with EI

Takeaways

1. Your general setup for Bayesian Optimization will require you to build a GP, and select a myopic acquisition policy - there are many options for doing so.
2. We looked at two popular policies in noise-free and noisy settings, the empirical observation is that KG outperforms EI in the latter case.
3. We haven't yet made any comments about sub-optimality, and special cases (i.e. fixed number of alternatives, stopping at the k th measurement etc.)

References I

Astudillo, R. and P. Frazier (2019, May). Bayesian Optimization of Composite Functions. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 354–363. PMLR.

Brochu, E., M. W. Hoffman, and N. de Freitas (2011, March). Portfolio Allocation for Bayesian Optimization. arXiv:1009.5419 [cs].

Eriksson, D., M. Pearce, J. R. Gardner, R. Turner, and M. Poloczek (2020, February). Scalable Global Optimization via Local Bayesian Optimization. arXiv:1910.01739 [cs, stat].

Frazier, P. I. (2018, July). A Tutorial on Bayesian Optimization. arXiv:1807.02811 [cs, math, stat].

Hvarfner, C., E. O. Hellsten, and L. Nardi (2024, June). Vanilla Bayesian Optimization Performs Great in High Dimensions. arXiv:2402.02229 [cs, stat].

References II

Letham, B., B. Karrer, G. Ottoni, and E. Bakshy (2019, June). Constrained Bayesian Optimization with Noisy Experiments. *Bayesian Analysis 14*(2), 495–519.

Miller, J. J. and S. Mak (2024, March). Targeted Variance Reduction: Robust Bayesian Optimization of Black-Box Simulators with Noise Parameters. arXiv:2403.03816 [cs, stat].

Wilson, J. T., F. Hutter, and M. P. Deisenroth (2018, December). Maximizing acquisition functions for Bayesian optimization. arXiv:1805.10196 [cs, stat].

Thank you!