Prior-mean-RObust Bayesian Optimization (PROBO)

Julian Rodemann, Thomas Augustin

Department of Statistics, LMU Munich
julian@stat.uni-muenchen.de

Bayesian optimization (BO) with Gaussian processes (GP) as surrogates is used to optimize functions that are expensive to evaluate and lack analytical description. Its applications range from engineering [3] to drug discovery [5] and COVID-19 detection [1]. BO's main popularity, however, stems from machine learning, where it has become one of the predominant hyperparameter optimizers after the seminal work of [8].

In my talk as part of the Young Statisticians Lecture Series, I would like to propose Prior-mean-RObust Bayesian Optimization (PROBO), that outperforms BO on specific problems [7]. In the first part of my talk, the effect of the GP prior specifications on classical BO's convergence is studied. I find the prior's mean parameters to have the highest influence on convergence among all prior components. In response to this result, I introduce PROBO that aims at rendering BO more robust towards GP prior mean parameter misspecification. This is achieved by explicitly accounting for GP imprecision via a prior near-ignorance model [2] [4] from the realm of Imprecise Probabilities (IP). It allows the set-valued specification of the GP prior. At the heart of PROBO is a novel acquisition function, the generalized lower confidence bound (GLCB), which is able to explicitly account for prior-induced model imprecision as opposed to classical data-related uncertainty. In the second part of my talk, I will test my approach against classical BO on a real-world problem from material science and observe PROBO to converge faster. Further experiments on multimodal and wiggly target functions confirm the superiority of PROBO [6].

I will conclude with a brief plea for IP models in heuristic optimization methods based on surrogate models. They not only offer a vivid framework to represent prior ignorance, as I will demonstrate in this very talk, but may also be beneficial in applications where prior knowledge is abundant. In such situations, in the case of data contradicting the prior, precise probabilities often fail to adequately represent uncertainty, whereas IP models can handle these prior-data conflicts, see e.g. [9].

Keywords: Bayesian optimization · Imprecise probabilities · Prior near-ignorance · Model imprecision

Acknowledgments: I warmly thank Prof. Dr. Thomas Augustin for the supervision of my master's thesis, whose results this presentation builds on.

References

1. Awal, M.A., Masud, M., Hossain, M.S., Bulbul, A.A., Mahmud, S.M.H., Bairagi, A.K.: A novel Bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data. IEEE Access **9**, 10263–10281 (2021)
2. Benavoli, A., Zaffalon, M.: Prior near ignorance for inferences in the k-parameter exponential family. Statistics **49**(5), 1104–1140 (2015)
3. Frazier, P.I., Wang, J.: Bayesian optimization for materials design. In: Information Science for Materials Discovery and Design, pp. 45–75. Springer (2016)
4. Mangili, F.: A prior near-ignorance Gaussian process model for nonparametric regression. International Journal of Approximate Reasoning **78**, 153–171 (2016)
5. Pyzer-Knapp, E.O.: Bayesian optimization for accelerated drug discovery. IBM Journal of Research and Development **62**(6), 2–1 (2018)
6. Rodemann, J.: Robust Generalizations of Stochastic Derivative-Free Optimization. Master's thesis, LMU Munich (2021)
7. Rodemann, J., Augustin, T.: Accounting for Gaussian Process Imprecision in Bayesian Optimization. In: Honda, K., Entani, T., Ubukata, S., Huynh, V.N., Inuiguchi, M. (eds.) Integrated Uncertainty in Knowledge Modelling and Decision Making. Springer Lecture Notes in Computer Science (LNCS). pp. 92–104. Springer, Cham (2022)
8. Snoek, J., Larochelle, H., Adams, R.P.: Practical Bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems **25** (2012)
9. Walter, G., Augustin, T.: Imprecision and prior-data conflict in generalized Bayesian inference. Journal of Statistical Theory and Practice **3**(1), 255–271 (2009)