Master thesis project proposal

Rianne de Heide

Over the last ten years it has become clear that Bayesian inference can behave quite badly under model misspecification, i.e., if the true data generating process is not in the model.

This has been shown through simulations in parametric Bayesian (generalized) linear regression in e.g. Van Ommen & Grünwald (2017) and De Heide et. al. (2020). Grünwald and Langford (2007) exhibit a non-parametric classification setting in which the posterior never concentrates on the unique distribution closest to the true data generating process in KL divergence.

In this project, we are going to investigate whether we can recover the same problem in simulations with non-parametric Bayesian regression, e.g. a Dirichlet or Gaussian process. We then are going to investigate whether the problem can be solved by exponentiating the likelihood with a learning rate η, which, if chosen small enough, should fix the problem.

References

[1] Grünwald, Peter, and John Langford. "Suboptimal Behavior of Bayes and MDL in Classification under Misspecification." Machine Learning 66, no. 2 (March 1, 2007): 119–49. https://doi.org/10.1007/s10994-007-0716-7.

[2] Grünwald, Peter, and Thijs van Ommen. "Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It." Bayesian Analysis 12, no. 4 (December 2017): 1069–1103. https://doi.org/10.1214/17-BA1085.

[3] De Heide, Rianne, Alisa Kirichenko, Peter Grunwald, and Nishant Mehta. "Safe-Bayesian Generalized Linear Regression." In International Conference on Artificial Intelligence and Statistics, 2623–33. PMLR, 2020. http://proceedings.mlr.press/v108/heide20a.html.