

6.435 Course Project Proposal: Exploration of Boosting Variational Inference

Geoffrey Liu 03/14/2022

Introduction Recently, there has been some development into boosting variational inference, where the variational family is a mixture of simple distributions, with each component added sequentially. In this project I will explore different types of variational inference (VI) procedures, building up from mean-field variational approximations, automatic and black-box variational inference and ultimately to boosting variational inference (BVI). The goal is to implement boosting variational inference, and apply it to a set of standard distributions, then visualize and compare to mean-field and structured/full-rank VI. The project builds upon variational inference procedures that we have learned in class.

Data The data I will use for the project are purely synthetic. This will include standard 1-D and 2-D distributions, such as a gaussian or cauchy mixtures and the banana distribution. Following the spirit of papers in this area, I will also apply my implementations to a simulated logistic regression as an optional extension.

Literature Review VI is widely used and often preferred to MCMC samplers due to their computational scalability from stochastic optimization and automatic differentiation. One major disadvantage to VI is that the variational family typically does not contain the posterior, thus limiting the best achievable approximation. BVI addresses this limitation by using a mixture of simple distributions as the variational family. Instead of fitting the mixture distribution jointly, BVI iteratively refines the approximation by adding a single mixture component at a time.

There is a small but clear literature developing BVI, including but not limited to Guo et al. (2016), Miller et al. (2016) and Campbell et al. (2019). There are two general formulations of BVI, Miller et al. (2016) propose minimizing the KL divergence over the mixture weights and mixture component jointly while Guo et al. (2016) use a gradient boosting formulation and optimize sequentially. Campbell et al. (2019) proposes using the Hellinger Distance instead of the KL divergence, and show that it prevents degeneracy of gradient-based BVI and avoids the difficult joint optimizations over weight and component. An extension is to explore the possibility of using boosting in Stein Variational Gradient Descent to add particles iteratively.

Project Plan I would like to learn JAX and implement this in Python+JAX.

1. (03/28) Implement Automatic Differentiation Variational Inference (ADVI) and Black Box Variational Inference (BBVI). The rest of the project depends on these implementations, so I will implement these from scratch.
2. (03/28) For both debugging and educational purposes, create a visualization program to show refinement of ADVI and BBVI to the posterior using simple 1-D and 2-D distributions.
3. (04/06) Implement BVI from Miller et al. (2016)
4. (04/13) Implement BVI from Guo et al. (2016) and compare differences and analyze issues with joint weight and mixture optimization from step 3 using visualization program.
5. (04/20) Implement Universal BVI (UBVI) from Campbell et al. (2019) and add to visualization program.
6. (Optional) Implement Stein Variational Gradient Descent (SVGD) and add to visualization program. Use ideas from BVI to formulate a boosting or sequential SVGD where particles are iteratively added.
7. (Optional) Add logistic regression experiments using the different implemented VI algorithms.

Project Risks

- Difficulty in familiarization with authors code and learning JAX. One of the three papers doesn't have code, and code quality appears poor and Guo et al. is in Julia. Check if there is enough tools in JAX for this project.
- Step 3 and 4, where BVI is implemented from scratch may be more difficult than anticipated. De-risk: only implement one of 3 or 4.
- My own implementation of BBVI and ADVI might take longer than expected, putting time-pressure on implementing BVI. De-risk: Front-load and attend office hours.
- Not enough time to complete extension into sequential SVGD, or adding logistic regression experiments.

- References** Campbell, T., & Li, X. (2019). Universal Boosting Variational Inference.
- Guo, F., Wang, X., Fan, K., Broderick, T., & Dunson, D. B. (2016). Boosting Variational Inference.
- Miller, A. C., Foti, N., & Adams, R. P. (2016). Variational Boosting: Iteratively Refining Posterior Approximations.