STCS 6701: Foundations of Graphical Models: Reading 11

Maxime TCHIBOZO (MT3390)

November 2020

1 Automatic Differentiation Variational Inference (2016) - Kucukelbir et al.

Automatic Differentiation Variational Inference (ADVI) allows for scientists to perform inference on data using only a model and a dataset. ADVI performs computations for models usually regardless of conjugacy between model variables. ADVI is fully integrated in Stan.

Previous approaches to automatic inerence mainly focussed on MCMC algorithms, but these are too slow for many real-world applications. Black box variational inference allows for scalable inference on large datasets, and this paper builds on prior work by providing an automated solution to variational inference.

Usually, VI algorithms require extensive mathematical derivations to transform model latent variable definition into an optimization problem. To automate this process, ADVI:

1. transforms the model by replacing the specified latent variables with unconstrained real-valued latent variables. This leads to a reparametrization of $p(x, \theta)$ to $p(x, \xi)$, and a later minimization of $KL(q(\xi)||p(\xi|x))$.
2. reparametrizes the gradient of the ELBO (variational objective function) as an expectation over q, which simplifies computations downstream.
3. reparametrizes the gradient of the ELBO as a standard Gaussian whose parameters are approximated using samples.
4. optimizes noisy gradients once the above steps are completed.

This paper specifies that Stan has its own implementation of automatic differentiation, and offers flexibility in the optimization method used.

ADVI supports models whose latent random variables are continuous and which have a continuous gradient of the log joint $\nabla_\theta \log p(x, \theta)$ on the support domain of the prior $p(\theta)$.

The restriction on continuous latent random variables would at first glance render ADVI unusable for mixture models, hidden Markov models and topic models, but marginalizing out the discrete variables in the likelihoods of these models makes them differentiable.

The VI family q parameters ϕ are updated vi coordinate ascent while respecting constraints on the support of the variational family. Since all transformations and reparametrizations are one-to-one differentiable, applying them simply adds a function into the gradient functions where they appear. All approximations are performed in the continuous real coordinate space, and ADVI is compatible with both the Mean Field approximation and more expressive approximations such as the Full rank Gaussian (similar to the Laplace approximation).

Using all the above points, the ELBO can be rewritten as a function of differentiable parameters - each of which can be estimated. Once the gradient of the ELBO is defined, parameter updates are performed using stochastic optimization steps.

The authors detail a simple algorithm to perform ADVI in all cases which respect the aforementioned constraints.

The efficacy of ADVI is demonstrated in 10 different examples, and the authors conclude by identifying possible improvements of this method. Reparametrizations may affect the accuracy of ADVI when estimating parameters, and the default ADVI setup is not well-suited to discrete latent variable models. The reparametrization of the ELBO as a standard Gaussian is also a costly choice, which could be refined using higher order moments.