STCS 6702: Foundations of Graphical Models: Reading 4

Maxime TCHIBOZO (MT3390)

October 2020

1 Blei et al. (2018) – Variational Inference: A Review Statisticians

MCMCs have been the standard method for approximate inference, notably for computing posterior distributions in Bayesian Statistics. MCMC methods are slow and do not scale well to large datasets or complex data. However, they come with guarantees that the chain will converge to produce exact samples from the target density, whereas variational inference only finds locally optimal distributions. What MCMC simulates, VI approximates.

We use MCMC for smaller but expensive datasets, where we expect precise samples from the target distribution, whereas we use variational inference to quickly explore models which have millions of parameters (notably in computer vision, natural language processing, computational neuroscience). VI's reliance on optimization is what makes it scalable, as we can use stochastic or distributed optimization methods. VI has also proved to be reliable when the geometry of the posterior is complicated (e.g. multimodal).

Variational inference finds the distribution which minimizes the Kullback-Leibler divergence with respect to the posterior over a parametrized family of approximate densities. In practice, the KL cannot be computed, so we optimize the ELBO (evidence lower bound), which does not contain the constant log-evidence term.

In this review, Blei et al. focus on the mean-field variational family, in which latent variables are mutually independent, making the joint of the q conveniently factorizable. The latent factors are then usually taken from a parametric distribution which suits the domain corresponding random variable (i.e. gaussian factors for continuous variables, categorical factors for categorical variables).

This simple principle can be expanded by adding dependencies between the factor variables (structured variational inference).

The Coordinate Ascent (CAVI) algorithm for VI relies on the mean-field family to find a locally optimal distribution with respect to each factor, and is unfortunately sensitive to initialization. Under the mean-field assumption, conditionally conjugate Exponential Family models are typically used to model each complete conditional distribution for each factor with respect to the observations and the other factors. This allows for simple gradient computations when optimizing the ELBO, which is particularly beneficial for stochastic VI.

The VI framework provides flexibility for statisticians, as modelers can adapt a variety of parameters (e.g. step size of the stochastic method) to fit their problem. However, the mean-field family relies on strong independence assumptions, which are not always appropriate. The study of statistical properties of VI approximations is an open area of research, notably regarding their asymptotic behavior. Research is also led towards the expansion of VI methods beyond Exponential Family models, or using other objective functions than the KL divergence.