Foundations of Graphical Models: Project Milestone

Due: F 2018-11-16

Please use the LaTeX template on the website and submit your writeup to Courseworks.

Project Milestone

Please write a page detailing the progress you have made with the final project. This can include experiments and baseline models you've already run, along with a progression of experiments you plan to run. It can also include descriptions of a figure or figures that you hope to get out of your analysis, whether it be certain benchmarks to achieve for an experiment, or how two models compare in capturing specific phenomena in the data.

The milestone is required but it will not be graded.

Optional Problem

This problem is optional and will count as extra credit.

Recall, the goal of variational inference is to maximize the Evidence Lower Bound (ELBO):

$$\mathcal{L}(\nu) = \mathbb{E}_{z \sim q(z;\nu)}[\log p(x, z) - \log q(z; \nu)],$$

where x is the observed data, z denotes the latent variables, and ν denotes the variational parameters.

Score function gradients (Ranganath et al., 2014) and reparameterization gradients (Kingma and Welling, 2013; Rezende et al., 2014) are methods to approximate the gradient of the ELBO using Monte Carlo sampling. The score function gradient uses the following approximation:

$$\nabla_\nu \mathcal{L}(\nu) \approx \frac{1}{S}\sum_{s=1}^{S}[(\log p(x, z^{(s)}) - \log q(z^{(s)}; \nu))(\nabla_\nu \log q(z^{(s)}; \nu))]$$

where $z^{(s)}$ is one of S samples from $q(z; \nu)$. Meanwhile, reparameterization gradients can be used when the variational distribution $q(z; \nu)$ is reparameterizable:

$$\nabla_\nu \mathcal{L}(\nu) \approx \frac{1}{S}\sum_{s=1}^{S} \nabla_\nu[\log p(x, t(\epsilon^{(s)}, \nu)) - \log q(t(\epsilon^{(s)}, \nu); \nu)],$$

where $\epsilon^{(s)}$ is a sample from a distribution that doesn't depend on ν and $t(\cdot)$ is a function such that $t(\epsilon^{(s)}, \nu) \sim q(z; \nu)$.

Low-variance gradients are crucial for the convergence of stochastic gradient-based optimization methods. Using either theoretical or empirical approaches, compare the variance of the gradient approximations provided by the two methods. Feel free to use any model and variational family in your comparison, and to simulate data if you compare the methods empirically.

References

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Ranganath, R., Gerrish, S., and Blei, D. (2014). Black box variational inference. In *Artificial Intelligence and Statistics*, pages 814–822.

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*.