Module 9

Taylor

2022-07-20

Schedule Information

- HW4 Out now
- Keep working on project paper and presentation
- HW5 will be due 8/10 as well
- check your participation tracker data

Big Picture

1. The primary object of interest/inference is the **parameter posterior** $p(\theta \mid y)$
2. We don't have a formula for it, and we can't draw samples from it.
3. We'll review variational inference–finding an approximate formula for it.

Variational Bayes

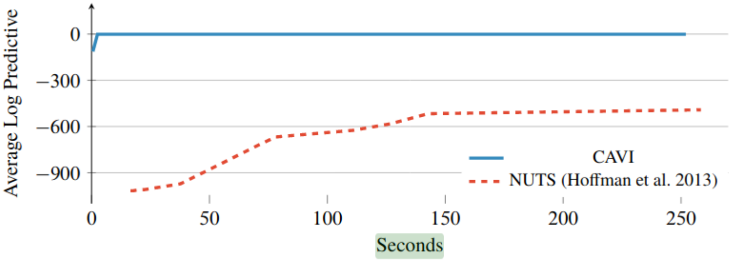

Figure 6: Comparison of CAVI to a Hamiltonian Monte Carlo-based sampling technique. CAVI fits a Gaussian mixture model to ten thousand images in less than a minute.

Figure 1: From https://arxiv.org/pdf/1601.00670.pdf

Today

A complete example of a model with latent variables.

Approximating $p(\theta, x_{1:n} \mid y_{1:n})$.

We will approximate a **joint** posterior, not just $p(\theta \mid y_{1:n})$

Typically the set of "everything that's unknown" is denoted as z.

For us:

$$z = (\mathbf{x}_{1:n}, \theta)$$

Variational Bayes

Say $z = (z^1, \ldots, z^J)$. The approximate posterior is assumed to have the following form:

$$g(z) = g_j(z^j)g_{-j}(z^{-j})$$

Coordinate ascent variational inference will update one component at a time

Variational Bayes

We need something that tells us the "distance" between two **distributions**. We need **Kullback-Leibler divergence**

$$\mathsf{KL}(g||p) = -\mathbb{E}_g\left[\log \frac{p(z \mid y)}{g(z)}\right] \geq 0$$

When the functions in the numerator and denominator are equal at **every** value of θ, then the quantity is equal to 0.

Variational Bayes

Recall

$$\mathsf{KL}(g||p) = -\mathbb{E}_g\left[\log \frac{p(y,z)}{g(z)}\right] + \log p(y)$$

or

$$\mathsf{KL}(g||p) + \mathsf{ELBO}(g) = \log p(y)$$

We deal with the ELBO

Coordinate Ascent Variational Inference

With CAVI (a specific case), we update one component at a time. At the beginning of the update, we have

$$g_j(z^j)g_{-j}(z^{-j})$$

and after updating the jth component, we have

$$\tilde{p}(z_j)g_{-j}(z^{-j})$$

where

$$\tilde{p}(z_j) \propto \exp\left[\int \log p(z, y)g_{-j}(z_{-j})\mathrm{d}z_{-j}\right]$$

This update always decreases KL divergence because it always increases the ELBO.

Coordinate Ascent Variational Inference

Our example in a Jupyter notebook:

- parameters: $\pi, \mu_0, \Lambda_0, \mu_1, \Lambda_1$
- hidden variables: $x_1, \ldots, x_n$
- z is the collection of all of this

```python
# refine g over and over again
for i in range(big_num):
    g = update_x(g, data, prior)
    g = update_pi(g, data, prior)
    g = update_mu_lambda_zero(g, data, prior)
    g = update_mu_lambda_one(g, data, prior)
```

Coordinate Ascent Variational Inference

The bottleneck is usually our ability to derive expectations of

$$\log p(y, z) = \log p(y \mid x, \theta) + \log p(x \mid \theta) + \log p(\theta)$$

Typically this means your models will have to have: - exponential family distributions - lots of independence and conditional independence

Coordinate Ascent Variational Inference

Let's go to `GMM_VI_demo.ipynb`

Variational Bayes

Next class:

- more examples
- a recent “more automatic” version
- using it in `PYMC`