Thermodynamic Variational Objective

1. Overview

Contribution 1: A Theory of Thermodynamic Integration that Generalizes the ELBO

Consider the path between two distributions and use this path to get an expression for the partition function $\log p_\theta(x)$.

First consider the density $\tilde{\pi}_\beta(z)$ along the path $\beta \in [0, 1]$ with corresponding 'potential energy' (by analogy to thermodynamics) $U_\beta(z) = \log \tilde{\pi}_\beta(z)$ and normalizing partition function $Z_\beta := \int \tilde{\pi}_\beta dz$,

$$\begin{aligned}\tilde{\pi}_\beta(z) &:= \tilde{\pi}_1(z)^\beta \tilde{\pi}_0(z)^{1-\beta} \\ &:= p_\theta(x,z)^\beta q_\phi(z|x)^{1-\beta}\end{aligned}$$

where the tildes are the unnormalized distributions. It is intractable for us to compute the partition functions, but physicists have shown us how to find their ratios. And if we conjure up a ratio that reduces to our log likelihood we have a new way to express our log likelihood,

$$\begin{aligned}\log Z_1 - \log Z_0 &= \int_0^1 \mathbb{E}_{\pi_\beta}\Big[U'_\beta(z)\Big] d\beta \\ \log \int p_\theta(x,z)dz - \log \int q_\phi(z|x)dz &= \int_0^1 \mathbb{E}_{\pi_\beta}\Big[U'_\beta(z)\Big] d\beta \\ \log p_\theta(x) &= \int_0^1 \mathbb{E}_{\pi_\beta}\Big[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\Big] d\beta\end{aligned}$$

And this is really nice because now if this expectation is monotonically increasing then a Riemann Sum of it gives us a lower bound on our log likelihood. Proof of monotonicity is in appendix C. Notice that if we take $\beta = 1$ we get back the ELBO:

$$\mathbb{E}_{q_\phi(z|x)}\Big[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\Big]$$

Contribution 2: A Convenient Optimization Method

We play lots of games manipulating equations with grads and expectations to find the gradient of TVO in terms of the expection of a covariance that we can sample. In particular, you can estimate this expectation under $\pi_\beta(z)$ using S-sample importance sampling with $q_\phi(z|x)$ our proposal distribution,

$$\mathbb{E}_{\pi_\beta}[f(z)] \approx \sum_{s=1}^{S} \overline{\omega_s^\beta} f(q_\phi(z|x))$$

where

$$\overline{\omega_s^\beta} := \frac{\omega_s^\beta}{\sum_{s'=1}^{S} \omega_{s'}^\beta} \qquad \omega_s := \frac{p_\theta(x,z_s)}{q_\phi(z_s|x)}$$

and these samples are reusable for each riemann sum step because we just raise to different powers of β. They also don't use reparameterization trick or REINFORCE so we can use them in the general case of non-reparameterizable continuous or discrete latent variables (without any relaxations).

2. Relationship to Variational Inference

TVO lower bounds log evidence by a Riemann sum approximation to the thermodynamic variational identity,

$$\frac{1}{K}\left[ELBO(\theta,\phi,x) + \sum_{k=1}^{K-1} \mathbb{E}_{\pi_{\beta_k}}\Big[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\Big]\right] \le \int_0^1 \mathbb{E}_{\pi_\beta}\Big[\log \frac{p_\theta(x,z)}{q_\phi(z|x)}\Big] d\beta = \log p_\theta(x)$$

where π_β is a geometric combination of $p_\theta(x, z)$ and $q_\phi(z|x)$,

$$\pi_\beta(z) = \frac{p_\theta(x, z)^\beta q_\phi(z|x)^{1-\beta}}{Z_\beta}$$

where Z is the partition function. At $\beta = 0$ we have ELBO, at $\beta = 1$ we have EUBO (upper bound).

But where does this π_β come from?

3. Thermodynamic Integration ("Computing the Ratios of Partition Functions")

Suppose we have densities of the form,

$$\pi_i(z) := \frac{\tilde{\pi}_i(z)}{Z_i}$$

Evaluating the partition functions is usually intractible (you must already know the distribution). But we can compute the log ratio of the normlizing constants by creating a family of unnormalized densities parameterized by $\beta = [0, 1]$ of the form,

$$\tilde{\pi}_\beta(z) := \tilde{\pi}_1(z)^\beta \tilde{\pi}_0(z)^{1-\beta}$$

with partitions,

$$Z_\beta := \int \tilde{\pi}_\beta(z) dz$$

The key observation here is that it is easier to calculate the ratio of normalizing constants than it is to calculate the constants themselves. We do this by deriving an identity. Let $U_\beta(z) := \log \tilde{\pi}_\beta(z)$,

$$\begin{aligned}
\frac{\partial \log Z_\beta}{\partial \beta} &= \frac{1}{Z_\beta} \frac{\partial}{\partial \beta} Z_\beta \\
&= \frac{1}{Z_\beta} \frac{\partial}{\partial \beta} \int \tilde{\pi}_\beta(z) dz \\
&= \int \frac{1}{Z_\beta} \frac{\partial}{\partial \beta} \tilde{\pi}_\beta(z) dz \\
&= \int \frac{\tilde{\pi}_\beta(z)}{Z_\beta} \frac{\partial}{\partial \beta} \log \tilde{\pi}_\beta(z) dz \\
&= \mathbb{E}_{\pi_\beta}\left[U'_\beta(z)\right]
\end{aligned}$$

So we get something we can't compute as the derivative of something we can compute!

We finally integrate out β,

$$\int_0^1 \frac{\partial \log Z_\beta}{\partial \beta} d\beta = \int_0^1 \mathbb{E}_{\pi_\beta}\left[U'_\beta(z)\right] d\beta \implies$$

$$\log(Z_1) - \log(Z_0) = \int_0^1 \mathbb{E}_{\pi_\beta}\left[U'_\beta(z)\right] d\beta$$