Variational Approximations

Setup: We have a posterior π

$$\pi(\theta|x) \propto \frac{P(x,\theta)}{p(x)} = \frac{\overset{\mathcal{L}}{P(x|\theta)}\,\overset{\text{prior}}{p(\theta)}}{\underbrace{p(x)}_{\text{evidence}}}$$

Assume that $\theta \in \mathbb{R}^d$

mcmc to approximate π that is quite slow

Idea: Choose $Q \in \{q,\ q \in Q\}$ $\quad q$ a density on $\mathbb{R}^d$

to approximate π by

$$q_* = \underset{q \in Q}{\arg\min}\ KL(\pi, q) = \underset{q}{\arg\min} \int \log \frac{q(u)}{\pi(u|x)}\, q(u)\, du$$

All inference then done with q_*

Plugging in $\pi(\theta|x)$

$$KL(\pi|q) = \log p(x) + \int \left[\log q(u) - \log p(x,\theta)\right] q(u)\, du$$

Define the evidence lower bound (ELBO)

$$ELBO(q) = \int \left[\log p(x,u) - \log q(u)\right] q(u)\, dx$$

$\Rightarrow \log(p(x)) = KL(\pi | q) + ELBO(q) \geq ELBO(q)$

So then minimizing KL $\iff$ maximizing ELBO

<u>Goal</u>: $\max_{q \in Q} ELBO(q)$

In practice take Q simple as to simplify calculations

<u>Mean field Variation. Approx</u>: $q(\theta_1, \dots, \theta_d) = \prod_{j=1}^{d} q_j(\theta_j)$ $\qquad Q = \{ q = \prod_{j=1}^{d} q_j \ , q_j \in \mathcal{P}_j \}$

We can then solve this problem using <u>coordinate ascent variational inference</u> ((CAVI)

Coordinate Ascent/Descent

- maximize one coordinate at a time

Fix $q_1, \dots, q_d$ and maximize $f \mapsto ELBO(f, q_2, \dots, q_d)$

$$ELBO(f, q_2, \dots, q_d) = \int \log p(x, \theta) \, f(\theta_1) \prod_{i=2}^{d} q_i(\theta_i) \, d\theta_i \, d\theta$$

$$- \int \log f(\theta_1) \, f(\theta_1) \, d\theta_1 \underbrace{- \sum_{i=2}^{d} \int \log q_i(u_i) \, q_i(u_i) \, du_i}_{\text{constant in } f}$$

$$\mathbb{E}(\log p(x, \theta, \theta_{-j})) \stackrel{def}{\equiv} \int \log p(x, \theta) \prod_{k \neq j} q(\theta_k) \, d\theta_k$$

just a function in θ_j. So

$$\text{ELBO}(f, q_{2:d}) = \int \mathbb{E}\left[\log p(x, t, \theta_{-1})\right] f(t) - \int f(t) \log f(t)\, dt$$

$$= - \int \left[\log f(t) - \mathbb{E}\left(\log p(x, t, \theta_{-1})\right)\right] f(t)\, dt$$

Recall $KL(q_1 \| q_2) = \int \log\left(\frac{q_2}{q_1}\right) q_1 \quad = 0 \quad \text{iff} \quad q_1 = q_2$

So we want to find f^* that satisfies

$$f_*(x) \propto \exp\left\{\mathbb{E}\left(\log p(x, t, \theta_{-j})\right)\right\}$$

<u>CAVI</u>

At the k-th iteration $\Rightarrow q_1^{(k)} \dots q_d^{(k)}$

Set $\bar{q} = \prod_{i=1}^{d} q_i^{(k)}$

For $i = 1, 2, \dots, d$

$\bar{q}_j(t) \propto \exp\left[\mathbb{E} \log p(x, t, \theta_{-j})\right]$

$q^{(k+1)} = \bar{q}$

until $\text{ELBO}(q^{(k)})$ convergences

In practice we use parametric families

$$q_\beta(\theta) = \prod_{j=1}^{d} q_{\beta_j}(\theta_j)$$

Ex: $$q_\beta(\theta) = \prod_{j=1}^{d} N(\theta_j, \underbrace{\mu_j, \sigma_j^2}_{\beta_j})$$

Ex: Suppose $$\log p(x, \theta) = \frac{1}{2\sigma^2} \|y - X\theta\|_2^2 - \frac{1}{2c^2} \|\theta\|^2$$

then $$\mathbb{E}[\log p(x, t, \theta_{-j})] = \int \left[-\frac{1}{2c} \|\theta\|_2^2 - \frac{1}{2\sigma^2} \|y - X\theta\|_2^2 \right] \prod_{k \neq j} N(\theta_k, \mu_k, \sigma_k^2)\, d\theta_j$$

All quadratic in θ_j.

then updates based on a closed form solution.