Variational Approximations

Suppose π is a distribution on $\mathbb{R}^d$.

Pick a family $\mathbb{Q}$ of densities on $\mathbb{R}^d$ and solve the problem

$$q_* = \underset{q \in \mathbb{Q}}{\operatorname{argmin}}\ KL(\pi | q)$$

$$KL(\pi | q) \overset{\text{def}}{=} \int \log\left(\frac{q}{\pi}\right) q$$

Simplification: Mean field approximation: set

$$\mathbb{Q} = \left\{ q : q(\theta) = \prod_{i=1}^{d} q_i(\theta_i),\ q_i \in \mathbb{Q}_i \right\}$$

when π is the posterior $\pi(\theta | x) = \dfrac{p(x, \theta)}{p(x)}$

and $KL(\pi | q) = \underbrace{\log p(x)}_{\substack{\text{ind. of} \\ q}} + \int [\log q(\theta) - \log p(x, \theta)]\, q(\theta)$

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

Therefore

$$\min_{q} KL(\pi|q) \Longleftrightarrow \max_{q} ELBO(q)$$

In most applications, we take

$$Q = \left\{ \prod_{i=1}^{d} q_{\varphi_i}(\theta_i),\ (\varphi_1, \dots, \varphi_d) \text{ finite dim parameter} \right\}$$

and then

$$\max_{q} ELBO(q) \Longleftrightarrow \max_{\varphi \in \mathbb{R}^n} ELBO(q)$$

<u>CAVI</u> (Coordinant Ascent Variational Appox.)

Suppose $Q = \left\{ q,\ q = \prod_{i=1}^{d} q_i(\theta_i),\ q_i \in Q_i \right\}$

To maximize $ELBO(q) = ELBO(q_1, \dots, q_d)$

<u>Coord. Ascent:</u>

- fix all coord. but one and maximize
- Repeat on all coordinates
- Repeat 1-2

$$\text{ELBO}(f) \stackrel{\text{def}}{=} \int \left[\log p(x,\theta) - \log f(\theta_1) - \sum_{i=1}^{2} \log q_i(\theta_i) \right] f(\theta_1) \prod_{i=1}^{2} q_i(\theta_i) \, d\theta \, d\theta_1$$

$$= \int \Big[\underbrace{\int \log p(x,\theta) \prod_{i=2}^{d} q_i(\theta_i) \, d\theta_i}_{\text{funct. of } \theta_1} - \log f(\theta_1) \Big] f(\theta_1) \, d\theta_1 \quad + C$$

Consider the density

$$h(\theta_1) \equiv \frac{\exp\left\{ \int \log p(x,\theta) \prod_{i=2}^{d} q_i(\theta_i) \, d\theta_i \right\}}{\int \exp\left\{ \int \log p(x,\theta) \prod_{i=2}^{d} q_i(\theta_i) \, d\theta_i \right\} d\theta_1} = \frac{\#}{c(q_{2:d})}$$

So $\text{ELBO}(f) = \int \left[\log h(\theta_1) - \log f(\theta_1) \right] f(\theta_1) \, d\theta_1 + \underbrace{\log c(q_{2:d}) + C}_{C'}$

$$= -KL(f \| h) + C'$$

So we maximize ELBO by taking $f \equiv h$.

CAVI:

Repeat until Convergence

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

    set $q_i(\theta) \propto \exp\left\{ \int \log p(x,\theta) \prod_{j \neq i} q_j(\theta_j) \, d\theta_j \right\}$

  end for

Example Linear regression $\underbrace{y}_{\mathbb{R}^n} = \underbrace{X}_{\mathbb{R}^{n \times p}} \theta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I)$ known

Prior: $\theta \sim N(0, c^2 I_p)$

$$\pi(\theta|y) \propto e^{\frac{-\|\theta\|_2^2}{2c^2}} e^{-\frac{1}{2\sigma^2}\|X\theta - y\|_2^2}$$

$$\sim N(m, \sigma^2\Sigma), \quad m = \Sigma X^T y$$

$$\Sigma = \left(X^T X + \frac{\sigma^2}{c^2} I\right)^{-1}$$

Here $\log p(y,\theta) = \frac{-\|\theta\|_2^2}{2c^2} - \frac{1}{2\sigma^2}\|y - X\theta\|_2^2 + C$

Let's approximate with $Q = \left\{ q, q(\theta) = \prod_{i=1}^{d} N(\theta_i, \mu_i, \sigma_i^2) \right\}$

Fix j: write $\|y - X\theta\|_2^2 = \|y - X_{-j}\theta_{-j} - \theta_j X_j\|_2^2$

$= \|y - X_{-j}\theta_{-j}\|_2^2 + \theta_j^2\|X_j\|_2^2 - 2\theta_i \langle X_j, y - X_{-j}\theta_{-j}\rangle$

$\Rightarrow \int \log p(y,\theta) \prod_{i \neq j} N(\theta_i, \mu_i, \sigma_i^2)\, d\theta_i$

$= \frac{-\theta_j^2}{2c^2} - \frac{1}{2c^2}\sum_{i \neq j}(\mu_i^2 + \sigma_i^2) - \frac{1}{2\sigma^2}\theta_j^2\|X_j\|_2^2 + \frac{1}{\sigma^2}\langle X_j, y - X\mu_{-j}\rangle + C$

$\Rightarrow h(q) \propto \exp\left\{\int \log p(y,\theta) \prod_{j \neq i} q_i(\theta_j)\, d\theta_j\right\} \sim N(\bar{\mu}_j, \bar{\sigma}_j^2)$

when

$$\bar{\sigma}_j^2 = \frac{1}{\frac{1}{c^2} + \frac{\|X_j\|_2^2}{\sigma^2}} \qquad \bar{\mu}_j = \frac{\langle X_j, y - X_{-j}\mu_{-j}\rangle}{\|X_j\|_2^2 + \frac{\sigma^2}{c^2}}$$

<u>Alg:</u> Choose initial $\mu^{(0)}$, $\sigma_j^2 = \dfrac{\sigma^2}{\|x_j\|_2^2 + \frac{\sigma}{c^2}}$

Do until convergence

$\bar{\mu}^{(k)} = \mu^{(k)}$

For $j = 1, \dots, p$

$$\bar{\mu}_j^{(k)} = \frac{\langle x_j, y - X_{-j}\bar{\mu}_{-j}^{(k)} \rangle}{\|x_j\|_2^2 + \sigma^2/c^2}$$

end for

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