03/18/25 Intro to VI + CAVI

Alternative to sampling from posterior
Use optimization to approximate the posterior.

$$p_\theta(z|x) = \frac{p_\theta(x, z)}{p_\theta(x)}$$

observe x
latent z
parameters θ

$$p_\theta(x) = \int p_\theta(x, z)\, dz$$

Find q to minimize KL-divergence

$$D_{KL}(q(z) \,||\, p_\theta(z|x))$$

$$= \int q(z) \log \frac{q(z)}{p_\theta(z|x)} dz$$

$$= \int q(z) \left[\log q(z) - \log p_\theta(z|x)\right] dz$$

$$= \underbrace{\int q(z) \log q(z)\, dz}_{\substack{H(q) \\ \text{Entropy}}} - \int q(z) \log p(z|x)\, dz$$

$$\int q(z) \log p_\theta(z|x)\, dz =$$

$$\underbrace{\int q(z) \log p_\theta(x,z)\, dz}_{-\text{"energy"}} - \underbrace{\int q(z) \log p_\theta(x)\, dz}_{= \log p_\theta(x)}$$

$D_{KL}(q(z) \| p_\theta(z|x))$

$$= \underbrace{H(q) - \int q(z) \log p_\theta(x,z)\, dz}_{\text{Minimize.}} + \underbrace{\log p_\theta(x)}_{\text{Free of } q.}$$

Re-arrange:

$$\log p_\theta(x) - D_{KL}(q(z) \| p_\theta(z|x))$$

$$= \underbrace{E_{q(z)}[\log p_\theta(x,z)] - H(q)}_{\text{Evidence lower bound (ELBO)}}$$

Since $D_{KL}(q(z) \| p_\theta(z|x)) \geq 0$

$\text{ELBO} \leq \log p_\theta(x) \leftarrow$ "evidence"

$\therefore$ maximize ELBO is equivalent to minimizing KL-divergence.

$$\mathbb{E}_{q(z)}[\log p_\theta(x,z)] - H(q)$$

$$= \mathbb{E}_{q(z)}\left[\log p_\theta(x|z) + \log p_\theta(z) - \log q(z)\right]$$

$$= \underbrace{\mathbb{E}_{q(z)}[\underbrace{\log p_\theta(x|z)}_{\text{log likelihood}}]}_{\substack{\text{expected log likelihood} \\ \text{under } q(z).}} - \underbrace{\mathbb{E}_{q(z)}\left[\log \frac{q(z)}{p_\theta(z)}\right]}_{D_{KL}(q(z)\|p_\theta(z))}$$

- The KL term acts as a regularizer preventing the approximate posterior from diverging "too much" from the prior

We need to assume some form for q.

$\rightarrow q_\phi(z)$: a parametric family

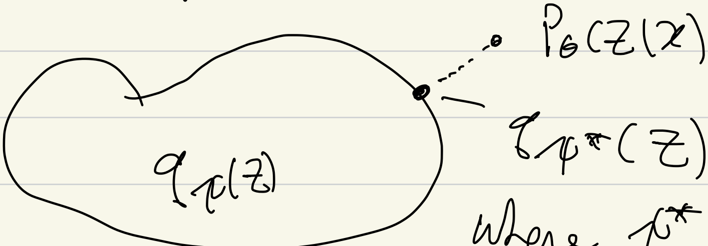
ϕ: variational parameters

In essence,

$$\min_q D_{KL}[q(z) \,||\, p_\theta(z|x)]$$

$$\equiv \min_\phi D_{KL}(q_\phi(z) \,||\, p_\theta(z|x)) \quad (1)$$

Once a parametric family is chosen, it determines the space of distributions



where ϕ^* minimizes (1)

Toy Example

Derivation + Notebook → HW

- 2D NIG target $p(\mu_1, \mu_2, \sigma_1, \sigma_2 | y_{1:n})$
- $q_{\phi_j} = \text{Normal}(m_j, s_j^2)$ so $\phi_j = (m_j, s_j^2)$

Gradient based optimization

Mean-field VI

Let $z = (z_1, \dots, z_J)$ J: # of latent variables.

Choose $q_\phi(z) = \prod_{j=1}^{J} q_{\phi_j}(z_j)$

$$\mathbb{E}_{q(z)}[\log p_\theta(x, z)] - H(q)$$

$$H(q) = \int q(z) \log q(z)\, dz$$

$$= \int \prod_{i=1}^{J} q_i(z_i) \sum_{i=1}^{J} \log q_i(z_i)\, dz_{1:J}$$

$$= \sum_{i=1}^{J} H(q_i)$$

Update coordinate i: [maximize ELBO wrt $q_i(z_i)$]

$$\mathbb{E}_{q(z)}[\log p_\theta(x, z)]$$

$$= \mathbb{E}_{z_i} \mathbb{E}_{z_{-i}} [\log p_\theta(x, z)]$$

$$\mathbb{E}_{z_i}\Big[\underbrace{\mathbb{E}_{z_{-i}}[\log p_\theta(x, z)]}_{\text{function of } z_i \text{, call it } g_i(z_i)} - \log q_i(z_i)\Big]$$

Define a probability distribution

$$p(z_i) \propto \exp(g_i(z_i))$$

$$\mathbb{E}_{z_i}[g_i(z_i) - \log q(z_i)] + c$$

$$\propto \mathbb{E}_{z_i}[\log p_i(z_i) - \log q(z_i)]$$

$$= \mathbb{E}_{z_i}\left[\log \frac{p_i(z_i)}{q_i(z_i)}\right]$$

$$= \underbrace{-D_{KL}(q_i(z_i) \| p_i(z_i))}_{=0 \text{ iff } q_i(z_i) = p_i(z_i)}$$

2) So $q_i^*(z_i) = p_i(z_i) \propto \exp(g_i(z_i))$

In general

$$q_i^*(z_i) \propto e^{\mathbb{E}_{z_{mb(i)}}[\log p(x, z)]}$$

Coordinate ascent, repeat until convergence.

GMM → example

GMM $\mu_k \sim N(0, \sigma^2)$

$$c_i \sim \mathrm{Cat}(1/K, \dots, 1/K)$$

$$x_i \mid c_i, \mu \sim N(c_i^T \mu, 1)$$

$$p(\mu, c, x) = p(\mu) \prod_{i=1}^{N} p(c_i)\, p(x_i \mid c_i, \mu)$$

$$p(x) = \sum_{c} p(c) \int p(\mu) \prod p(x_i \mid c_i, \mu)\, d\mu$$

K^n possible configurations

Latent variables $\mu_{1:K}$, $c_{1:N}$.

Mean field VI

$$q(\mu_{1:K}, c_{1:N}) = \prod_{k=1}^{K} q_{\phi_k}(\mu_k) \prod_{i=1}^{N} q_{\psi_{c_i}}(c_i)$$

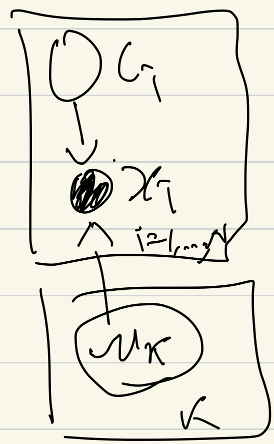
$$\phi_k = (m_k, s_k^2) \quad \psi_{c_i} = (\psi_{c_{i,k}})$$

$$\mathrm{ELBO}(m, s^2, \psi) = \mathbb{E}[\log p(x, c, \mu)] - H(q)$$

$$\boxed{q^*(z_i) \propto \exp\left(\mathbb{E}_{-z_i}\left[\log p(x, z)\right]\right)}$$

$$p(\mu, c, x) = p(\mu) \prod_{i=1}^{N} p(c_i)\, p(x_i \mid c_i, \mu)$$

$$\log p(\mu, c, x) = \sum_k \log p(\mu_k) + \sum_{i=1}^{N} \left[\log p(c_i) + \log p(x_i \mid c_i, \mu)\right]$$



Update for $q(c_i)$

$$\mathbb{E}_{-c_i}\left[\log p(\mu, c, x)\right] = \log p(c_i) + \mathbb{E}_{q(\mu)}\left[\log p(x_i \mid c_i, \mu)\right]$$

$$p(x_i \mid c_i, \mu) = \prod_k p(x_i \mid c_i, \mu_k)^{c_{ik}}$$

$$\mathbb{E} \log p(x_i \mid c_i, \mu) = \sum_k c_{ik}\, \mathbb{E}_{q(\mu_k)}\Big[\underbrace{\log p(x_i \mid c_i, \mu_k)}_{-\frac{1}{2}(x_i - c_i^T \mu)^2}\Big]$$

$$\sum_k c_{ik}\, \mathbb{E}_{q(\mu_k)}\left[-\frac{1}{2}(x_i^2 + \mu_k^2 - 2x_i\mu_k)\right]$$

$$\propto \sum_k c_{ik}\left[\mathbb{E}_{q(\mu_k)}[\mu_k]x_i - \frac{1}{2}\mathbb{E}_{q(\mu_k)}[\mu_k^2]\right]$$

x_i^2 ?

$\sum_k c_{ik}\, \mathbb{E}_{q(\mu_k)}[x_i^2]$

$= x_i^2$ indep of c_{ik}

$$\log p(c_i) = -\sum_k c_{ik} \log K$$

So

$$q^*_{c_i}(c_i) \propto \exp\left(-\sum_k c_{ik}\left(\log K + \mathbb{E}_{q(\mu_k)}[\mu_k]x_i - \frac{1}{2}\mathbb{E}_{q(\mu_k)}[\mu_k^2]\right)\right)$$

So

$$\psi_{ik} = \exp\left(\mathbb{E}_{q(\mu_k)}[\mu_k]x_i - \frac{1}{2}\mathbb{E}_{q(\mu_k)}[\mu_k^2]\right)$$