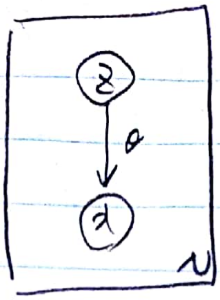
MLE for Generative Models



- Objective : $\underset{\theta}{\arg\max}\ l(\theta)$

  , where $l(\theta)$ is log-likelihood :

$$l(\theta) = \mathbb{E}_{x\sim p(x)} \log P_\theta(x)$$

- ELBO : Evidence Lower Bound

$$\Rightarrow \ l(\theta) = \mathbb{E}_{x\sim p(x)} \log P_\theta(x)$$

$$= \mathbb{E}_{x\sim p(x)} \log \int P_\theta(x,z)\, dz$$

$$= \mathbb{E}_{x\sim p(x)} \log \int q_\phi(z) \frac{P_\theta(x,z)}{q_\phi(z)}\, dz$$

$$\geq \mathbb{E}_{x\sim p(x)} \int q_\phi(z) \log \frac{P_\theta(x,z)}{q_\phi(z)}\, dz \quad (\because \text{Jensen's inequality})$$

$$= \mathbb{E}_{x\sim p(x)}\Big[\underbrace{\mathbb{E}_{z\sim q_\phi(z)}\big[\log P_\theta(x,z) - \log q_\phi(z)\big]}_{\text{ELBO}(\phi)}\Big]$$

(θ is fixed w.r.t. ϕ)

- KL Divergence $\left(D_{KL}(q \| p) = \mathbb{E}_q [\log q - \log p] \right)$

- $\log p_\theta(x) = \mathrm{ELBO}(\phi) + D_{KL}(q_\phi(z) \| p_\theta(z|x))$

pf)

$$\mathrm{ELBO}(\phi) + D_{KL}(q_\phi(z) \| p_\theta(z|x))$$

$$= \mathbb{E}_{z \sim q_\phi(z)} [\log p_\theta(x,z) - \log q_\phi(z)]$$

$$+ \mathbb{E}_{z \sim q_\phi(z)} \left[\log q_\phi(z) - \log \frac{p_\theta(x,z)}{p_\theta(x)} \right]$$

$$= \mathbb{E}_{z \sim q_\phi(z)} \log p_\theta(x)$$

$$= \log p_\theta(x)$$

■

Scanned by CamScanner

- Rethinking the objective.

  - Assumption : we already know $\begin{cases} q_\phi(z|x) \\ p_\theta(x|z) \\ p(z) \end{cases}$

$$l(\theta) = \mathbb{E}_{x\sim p(x)}\, p_\theta(x)$$

$$= \mathbb{E}_{x\sim p(x)}\left[ELBO(\phi ; \theta) + D_{KL}(q_\phi(z) \| p_\theta(z|x)) \right]$$

$$\geq \mathbb{E}_{x\sim p(x)}\, ELBO(\phi ; \theta)$$

$$= \mathbb{E}_{x\sim p(x)}\left[\mathbb{E}_{z\sim q_\phi(z)}\left[\log p_\theta(x,z) - \log q_\phi(z) \right] \right]$$

$$= \mathbb{E}_{x\sim p(x)}\left[\underline{\mathbb{E}_{z\sim q_\phi(z)}\left[\log p_\theta(x|z) \right] - D_{KL}(q_\phi(z) \| p(z))} \right]$$

It is tractable if we use Monte-Carlo method.

⇒ Instead of finding actual objective ($l(\theta)$), continually increasing lower bound of $l(\theta)$ using ELBO and gradient descent.

· Optimization of Variational Auto-Encoder (VAE)

◦ Replace $q_\phi(z) \rightarrow q_\phi(z|x)$

$\Rightarrow \mathbb{E}_{x\sim p(x)} \, ELBO(\phi;\theta)$

$$= \mathbb{E}_{x\sim p(x)}\left[\mathbb{E}_{z\sim q_\phi(z|x)}\left[\log p_\theta(x|z)\right] - D_{KL}\left(q_\phi(z|x)\,\|\,p(z)\right)\right]$$

Monte-Carlo

$$\approx \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{M}\sum_{j=1}^{M}\left(\log p_\theta(x_i|z_j) - \left(\log q_\phi(z_j|x_i) - \log p(z_j)\right)\right)\right]$$

, where $\{x_i\}$ are given data (ex: images)

and $\{z_j\}$ are sampled data s.t. $z_j \sim N(\mu_i, \Sigma_i)$

$$\Rightarrow \begin{cases} \dfrac{\partial}{\partial z_j} p_\theta(x_i|z_j) = \dfrac{\partial p_\theta(x_i|z_j)}{\partial\theta}\dfrac{\partial\theta}{\partial z_j} \\[2ex] \dfrac{\partial}{\partial x_i} q_\phi(z_j|x_i) = \dfrac{\partial q_\phi(z_j|x_i)}{\partial\phi}\cdot\dfrac{\partial\phi}{\partial x_i} \end{cases}$$

⇒ differentiable!

⇒ To differentiate it, we should use "reparametrization trick"

「reparametrization trick」

$$z_j|x_i = g_\phi(x_i, \varepsilon) = \underbrace{\mu_i + \Sigma_i}_{\text{output of NN.}}\,\varepsilon$$

where $\varepsilon \sim N(0, I^2)$

Appendix.

- Reparametrization Trick

$$\mathbb{E}_{z \sim q_\phi(z)}\left[f(z)\right] = \mathbb{E}_{\varepsilon \sim q(\varepsilon)}\left[f(g(\phi, \varepsilon))\right]$$

$$\overset{\text{Monte Carlo}}{\approx} \frac{1}{S} \sum_{i}^{S} q(\varepsilon_i) f(g(\phi, \varepsilon))$$

where $z = g(\phi, \varepsilon)$ and

ϕ is deterministic and

ε is a r.v. .