[Paper review 22]

Semi-Implicit Variational Inference

(MYin, 2018)

[Contents]

- 1. Review (VI with Implicit Distributions)
- 2. Semi-Implicit Distributions
- 3. SIVI (Semi-Implicit Variational Inference)

1. Review (VI with Implicit Distributions

$$\mathsf{ELBO}: \mathcal{L}(heta) = \mathbb{E}_{q_{ heta}(z)}[\underbrace{\log p(x,z)}_{\mathrm{model}} - \underbrace{\log q_{ heta}(z)}_{\mathrm{entropy}}]$$

 $\text{Gradient of ELBO}: \nabla_{\theta} \mathcal{L}(\theta) = \mathbb{E}_{q(\varepsilon)} \left[\nabla_{\theta} \left(\log p \left(x, f_{\theta}(\varepsilon) \right) - \log q_{\theta} \left(f_{\theta}(\varepsilon) \right) \right) \right]$

• (1) model term : (with MC approximation)

$$\mathbb{E}_{q(arepsilon)}\left[
abla_{ heta}\log p\left(x,f_{ heta}(arepsilon)
ight)
ight]pproxrac{1}{S}\sum_{s=1}^{S}
abla_{ heta}\log p\left(x,f_{ heta}\left(arepsilon^{(s)}
ight)
ight),\quad arepsilon^{(s)}\sim q(arepsilon)$$

• (2) entropy term:

$$abla_{ heta} \log q_{ heta}\left(f_{ heta}(arepsilon)
ight) =
abla_{z} \log q_{ heta}(z) imes
abla_{ heta} f_{ heta}(arepsilon) + \underbrace{
abla_{ heta} \log q_{ heta}(z)|_{z=f_{ heta}(arepsilon)}}_{=0(ext{ in expectation })}$$

but $\nabla_z \log q_\theta(z)$ is not available!

2. Semi-Implicit Distributions

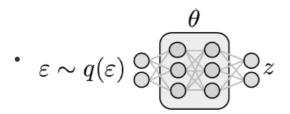
Goal: instead of "density ratio estimation"...

- method 1) lower bound of ELBO (SIVI)
- method 2) estimate gradients with sampling (UIVI)

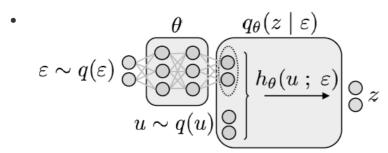
going to talk about "SIVI"

Implicit vs Semi-Implicit

Implicit



Semi-Implicit



- $q(\epsilon)$ is still implicit
- ex) $q_{\theta}(z \mid \epsilon) = N(z \mid \mu_{\theta}(\epsilon), \Sigma_{\theta}(\epsilon))$ (output of NN with input ϵ is used as "mean" & "variance")

 $q_{ heta}(z)$ is implicit

- 1) easy to sample $\begin{aligned} & \text{sample } \varepsilon \sim q(\varepsilon) \\ & \text{obtain } \mu_{\theta}(\varepsilon) \text{ and } \Sigma_{\theta}(\varepsilon) \\ & \text{sample } z \sim \mathcal{N}\left(z \mid \mu_{\theta}(\varepsilon), \Sigma_{\theta}(\varepsilon)\right) \end{aligned}$
- 2) but intractable $q_{\theta}(z) = \int q(\varepsilon) q_{\theta}(z \mid \varepsilon) d\varepsilon$

Assumptions on Conditional $q_{ heta}(z \mid \epsilon)$

- assumption 1) reparameterizable
- assumption 2) tractable gradient (= $\nabla_z \log q_\theta(z \mid \varepsilon)$ ($\nabla_z \log q_\theta(z)$ is intractable)

Gaussian

meets those two assumptions!

- assumption 1) reparameterizable $u\sim \mathcal{N}(u\mid 0,I),\quad z=h_{\theta}(u;\varepsilon)=\mu_{\theta}(\varepsilon)+\Sigma_{\theta}(\varepsilon)^{1/2}u$
- assumption 2) tractable gradient (= $\nabla_z \log q_{\theta}(z \mid \varepsilon)$ $\nabla_z \log q_{\theta}(z \mid \varepsilon) = -\Sigma_{\theta}(\varepsilon)^{-1} \left(z - \mu_{\theta}(\varepsilon)\right)$

3. SIVI (Semi-Implicit Variational Inference)

lower bound of ELBO

$$egin{aligned} \mathcal{L}(heta) &\geq \overline{\mathcal{L}}(heta), \quad ext{where} \ &\overline{\mathcal{L}}(heta) &= \mathbb{E}_{arepsilon \sim q(arepsilon)} \left[\mathbb{E}_{z \sim q_{ heta}(z|arepsilon)} \left[\mathbb{E}_{arepsilon^{(1)}, \ldots, arepsilon^{(L)} \sim q(arepsilon)} [\log p(x, z) - \log \left(rac{1}{L+1} \left(q_{ heta}(z \mid arepsilon) + \sum_{\ell=1}^{L} q_{ heta} \left(z \mid arepsilon^{(\ell)}
ight)
ight)
ight]
ight] \end{aligned}$$

- $\overline{\mathcal{L}}(\theta)$: SIVI bound
- optimize ELBO (X) ${\rm optimize\ lower\ bound\ of\ ELBO\ (O)}$ (since, lower bound does not depend on $q_{\theta}(z)$, which is intractable)
- as $L o \infty$, $\mathcal{L}(heta) o \overline{\mathcal{L}}(heta)$ (L controls the tightness of the bound) (computational complexity increases with L)

SIVI allows for semi-implicit constribution of prior in VAEs