[Paper review 22]

Semi-Implicit Variational Inference

(M Yin, 2018)

[Contents]

1. Review (VI with Implicit Distributions)

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

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)

  $\mathbb{E}_{q(\varepsilon)}\left[\nabla_\theta \log p\left(x, f_\theta(\varepsilon)\right)\right] \approx \frac{1}{S} \sum_{s=1}^S \nabla_\theta \log p\left(x, f_\theta\left(\varepsilon^{(s)}\right)\right), \quad \varepsilon^{(s)} \sim q(\varepsilon)$
- (2) entropy term :

  $\nabla_\theta \log q_\theta\left(f_\theta(\varepsilon)\right)=\nabla_z \log q_\theta(z) \times \nabla_\theta f_\theta(\varepsilon)+\underbrace{\left.\nabla_\theta \log q_\theta(z)\right|_{z=f_\theta(\varepsilon)}}_{=0(\text { in expectation })}$

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

2. Semi-Implicit Distributions

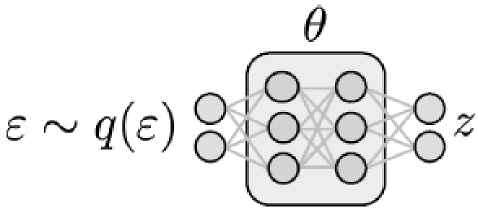
Goal : instead of "density ratio estimation"...

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

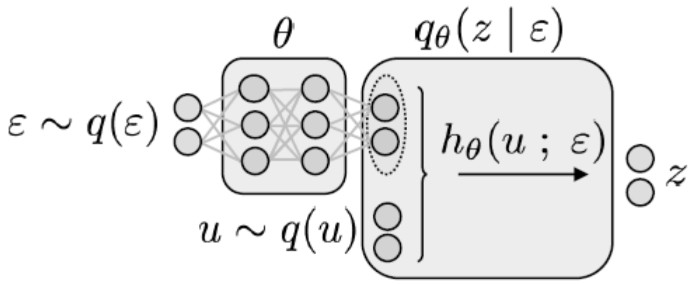
going to talk about "SIVI"

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")

$q_\theta(z)$ is implicit

- 1) easy to sample

  $\text{sample } \varepsilon \sim q(\varepsilon)$
  $\text{obtain } \mu_\theta(\varepsilon) \text{ and } \Sigma_\theta(\varepsilon)$
  $\text{sample } z \sim \mathcal{N}(z \mid \mu_\theta(\varepsilon), \Sigma_\theta(\varepsilon))$
- 2) but intractable

  $q_\theta(z) = \int q(\varepsilon) q_\theta(z \mid \varepsilon) d\varepsilon$

Assumptions on Conditional $q_\theta(z \mid \epsilon)$

- 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 (= $\nabla_z \log q_\theta(z \mid \varepsilon)$

  $\nabla_z \log q_\theta(z \mid \varepsilon) = -\Sigma_\theta(\varepsilon)^{-1}(z - \mu_\theta(\varepsilon))$

3. SIVI (Semi-Implicit Variational Inference)

lower bound of ELBO

$\mathcal{L}(\theta) \geq \overline{\mathcal{L}}(\theta), \quad \text{where}$

$\overline{\mathcal{L}}(\theta) = \mathbb{E}_{\varepsilon \sim q(\varepsilon)} \left[\mathbb{E}_{z \sim q_\theta(z|\varepsilon)} \left[\mathbb{E}_{\varepsilon^{(1)}, \ldots, \varepsilon^{(L)} \sim q(\varepsilon)} \left[\log p(x, z) - \log \left(\frac{1}{L+1} \left(q_\theta(z \mid \varepsilon) + \sum_{\ell=1}^{L} q_\theta \left(z \mid \varepsilon^{(\ell)} \right) \right) \right) \right] \right] \right]$

- $\overline{\mathcal{L}}(\theta)$: SIVI bound
- optimize ELBO (X)

  optimize lower bound of ELBO (O)

  (since, lower bound does not depend on $q_\theta(z)$, which is intractable)
- as $L \rightarrow \infty$, $\mathcal{L}(\theta) \rightarrow \overline{\mathcal{L}}(\theta)$

  (L controls the tightness of the bound)

  (computational complexity increases with L)

SIVI allows for semi-implicit constribution of prior in VAEs