[Paper review 21]

Variational Inference using Implicit Distributions

(Ferenc Huszar , 2017)

[Contents]

1. Review
2. Introduction
3. Implicit Distributions and Adversarial Training
   1. Expressive Implicit Distribution
   2. Implicit Distribution is HARD in VI
   3. Density Ratio Estimation
4. Limitations

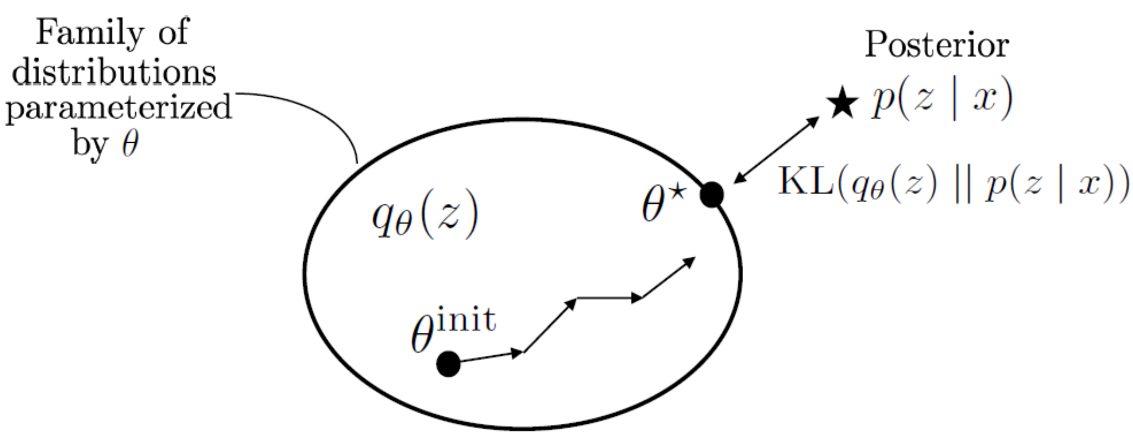
1. Review

hard to calculate posterior .. use Bayesian Inference

Variational Inference

fitting θ by ..

- Minimize KL : $\theta^{\star}=\underset{\theta}{\arg \min } \mathrm{KL}\left(q_{\theta}(z) \| p(z \mid x)\right)$
- Maximize ELBO : $\mathcal{L}(\theta)=\mathbb{E}_{q_{\theta}(z)}\left[\log p(x, z)-\log q_{\theta}(z)\right]$



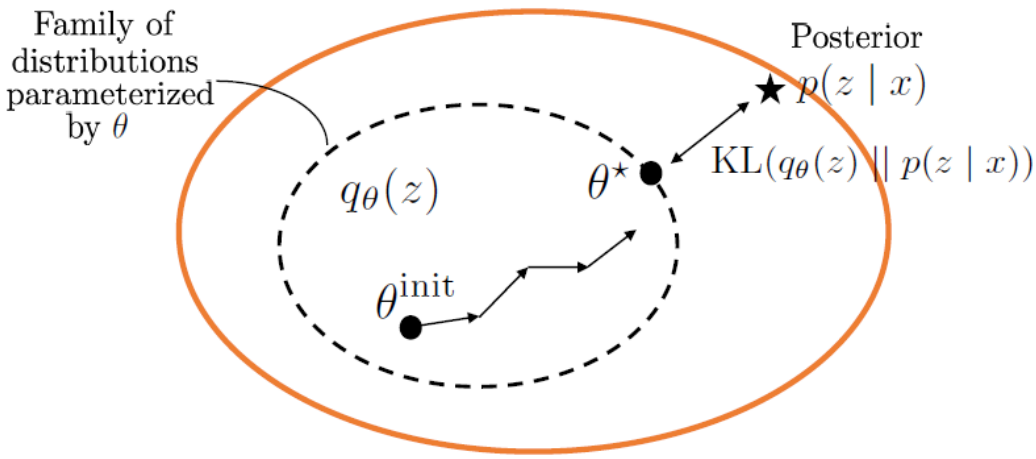
Mean Field Variational Inference (MFVI)

$q_\theta(z) = \prod_n q_{\theta_n}(z_n)$

- simple, fast but may be inaccurate

Point

how to expand the variational family?



2. Introduction

1) expand the variational family $q_\theta(z)$

2) use IMPLICIT distribution

- easy to sample : $z \sim q_\theta(z)$
- but intractable $q_\theta(z)$

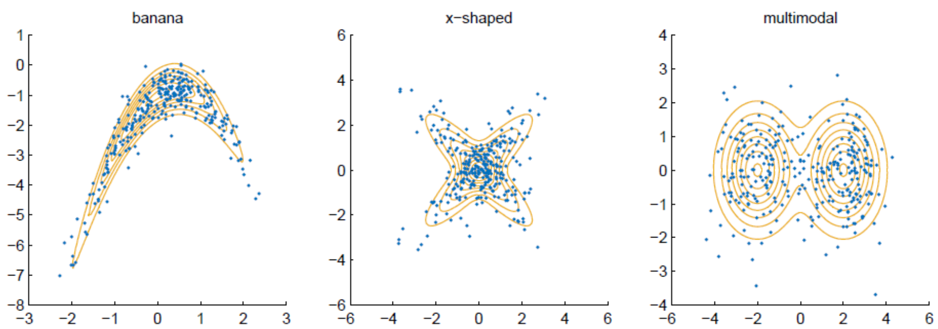
3) Challenge :

- "Solve the optimization problem with intractable $q_\theta(z)$"

  (= maximize $\mathcal{L}(\theta) = \mathbb{E}_{q_\theta(z)}\left[\log p(x,z) - \log q_\theta(z)\right]$)

4) Goal:

- "More Expressive variational distributions"

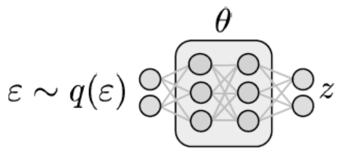


Blue dots: samples from $q_\theta(z)$

3. Implicit Distributions and Adversarial Training

3.1 Expressive Implicit Distribution

how to form EXPRESSIVE implicit distribution?



step 1) generate random noise

- $\epsilon \sim q(\epsilon)$

step 2) pass ϵ through NN with param θ

- $z = f_\theta(\epsilon)$

$q_\theta(z)$

- "Implicit" Distribution : easy to sample from $q(\epsilon)$, but can not evaluate density :
- "Flexible" Distribution : by using NN
- GOAL : tune θ so that $q_\theta(z) \approx p(z \mid x)$

3.2 Implicit Distribution is HARD in VI

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 :

$\nabla_\theta \mathcal{L}(\theta) = \mathbb{E}_{q(\varepsilon)}\left[\nabla_\theta\left(\log p\left(x, f_\theta(\varepsilon)\right) - \log q_\theta\left(f_\theta(\varepsilon)\right)\right)\right]$ (by using Reparameteriaztion)

- (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 :

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

  but $\nabla_z \log q_\theta(z)$ is not available : REWRITE ELBO as below

Another Expression of ELBO:

$$\begin{aligned}\mathcal{L}(\theta) &= \mathbb{E}_{q_\theta(z)}[\underbrace{\log p(x,z)}_{\text{model}} - \underbrace{\log q_\theta(z)}_{\text{entropy}}] \\ &= \mathbb{E}_{q_\theta(z)}[\log p(x \mid z)] - \mathrm{KL}\left(q_\theta(z) \| p(z)\right) \\ &= \mathbb{E}_{q_\theta(z)}[\log p(x \mid z)] - \mathbb{E}_{q_\theta(z)}\left[\log \frac{q_\theta(z)}{p(z)}\right]\end{aligned}$$

3.3 Density Ratio Estimation

Approximate Density ratio :

- ELBO : $\mathcal{L}(\theta) = \mathbb{E}_{q_\theta(z)}[\log p(x \mid z)] - \mathbb{E}_{q_\theta(z)}\left[\log \frac{q_\theta(z)}{p(z)}\right]$
- density ratio : $\log \frac{q_\theta(z)}{p(z)}$

Classifier $D(z)$

- Class y = 1: The sample z comes from $q_\theta(z)$

  Class y = 0: The sample z comes from $p(z)$
- Optimal Classifier : $D^\star(z) = \frac{q_\theta(z)}{q_\theta(z)+p(z)}$

How to train $D(z)$

- re-express density ratio & ELBO

  $\log \frac{q_\theta(z)}{p(z)} = \log D^\star(z) - \log(1 - D^\star(z))$, so

  $\mathcal{L}(\theta) = \mathbb{E}_{q_\theta(z)}[\log p(x \mid z)] - \mathbb{E}_{q_\theta(z)}[\log D(z) - \log(1 - D(z))]$
- $D^\star(z) = \max_D \mathcal{L}(\theta) = \max_D \mathbb{E}_{q_\theta(z)}[D(z)] + \mathbb{E}_{p(z)}[\log(1 - D(z))]$

Algorithm Summary

- ELBO objective : $\mathbb{E}_{q_\theta(z)}[\log p(x \mid z)] - \mathbb{E}_{q_\theta(z)}[\log D(z) - \log(1 - D(z))]$
- step 1) follow gradient estimate of the ELBO w.r.t θ (use reparameteriation trick!)

  step 2) for each θ, fit $D(z)$ so that $D(z) \approx D^*(z)$

4. Limitations

- $D(z)$ needs to be trained to optimum after EACH UPDATE of θ

  (in practice, optimization is truncated to a few iterations)
- Unstable training when discriminator does not catch up!
- Overfits in high dimension