Variational Inference with Normalizing Flows

Mou Minghao

SSE, CUHK(SZ)

July 31, 2022

Outline

Normalizing Flows[1]

one limitation of the variational methodology due to the available choices of approximating families, is that even in an asymptotic regime we can not obtain the true posterior. Thus, an ideal family of variational distributions $q_\phi(z|x)$ is one that is highly flexible, preferably flexible enough to contain the true posterior as one solution. One path towards this ideal is based on the principle of *normalizing flows*

A normalizing flow describes the transformation of a probability density through a sequence of invertible mappings

[1][Rezende and Mohamed, 2015]

Finite Flows

consider transforming a random variable z by $z' := f(z)$, then z' has a distribution

$$q(z') = q(z) \left| det \frac{\partial f^{-1}}{\partial z'} \right| = q(z) \left| det \frac{\partial f}{\partial z} \right|^{-1} \quad (1)$$

We can construct arbitrarily complex densities by composing several simple maps and successively applying (1)

Applying a chain of K transformations f_k on a RV $z_0 \sim q_0(z)$

$$z_K := f_K \circ \cdots f_2 \circ f_1(z_0)$$

$$\ln q_K(z_K) := \ln q_0(z_0) - \sum_{k=1}^{K} \ln det \left| \frac{\partial f_k}{\partial z_k} \right| \quad (2)$$

The path traversed by the random variables is called the *flow* and the path formed by the successive distributions q_k is called a *normalizing flow*

Finite Flows

A property of such transformations, often referred to as the law of the unconscious statistician (LOTUS), is that expectations w.r.t. the transformed density q_K can be computed without explicitly knowing q_K.

$$\mathbb{E}_{q_K}[h(z)] = \mathbb{E}_{q_0}[h(f_K \circ \cdots \circ f_1(z_0))] \tag{3}$$

We can understand the effect of invertible flows as a sequence of expansions or contractions on the initial density.

The formalism of normalizing flows now gives us a systematic way of specifying the approximate posterior distributions $q(z|x)$ required for variational inference. With an appropriate choice of transformations f_K, we can initially use simple factorized distributions such as an independent Gaussian, and apply normalizing flows of different lengths to obtain increasingly complex and multi-modal distributions

Infinitesimal Flows

It is natural to consider the case in which the length of the normalizing flow tends to infinity. In this case, we obtain an *infinitesimal flow*

a partial differential equation describing how the initial density $q_0(z)$ evolves over 'time': $\frac{\partial}{\partial t} q_t(z) = \mathcal{T}_t[q_t(z)]$, where $\mathcal{T}$ describes the continuous-time dynamics

Example (Hamiltonian Flow)

Hamiltonian Monte Carlo can also be described in terms of a normalizing flow on an augmented space $\tilde{z} = (z, \omega)$ with dynamics resulting from the Hamiltonian $\mathcal{H}(z, \omega) = -\mathcal{L}(z) - \frac{1}{2}\omega^T H \omega$

References I

Rezende, D. and Mohamed, S. (2015).

Variational inference with normalizing flows.

In *International conference on machine learning,* pages 1530–1538. PMLR.