EBM and MCMC: Moreau Yosida

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

To be completed...

1 Introduction

Algorithm 1 MY for Energy-Based model

Input: Total number of iterations T, number of MCMC transitions K and of samples M, sequence of global learning rate $\{\eta_t\}_{t>0}$, sequence of MCMC stepsizes $\gamma_{k\,k>0}$, initial value θ_0, MCMC initialization $\{z_0^m\}_{m=1}^M$ and observations $\{x_i\}_{i=1}^n$. Moreau Yosida decreasing sequence of parameters $\{\lambda_t\}$

for $t = 1$ to T **do**

  Draw M samples $\{z_t^m\}_{m=1}^M$ from the objective potential via Langevin diffusion:

  for $k = 1$ to K **do**

    Construct the Markov Chain as follows:

$$z_k^m = z_{k-1}^m + \gamma_k/2\left[\nabla f_{\theta_t}(z_{k-1}^m) + \lambda_t^{-1}\left(z_{k-1}^m - \mathrm{prox}_g^{\lambda}(z_{k-1}^m)\right)\right] + \sqrt{\gamma_k}\mathsf{B}_k \ , \tag{1}$$

    where B_t denotes the Brownian motion (Gaussian noise).

  end for

  Assign $\{z_t^m\}_{m=1}^M \leftarrow \{z_K^m\}_{m=1}^M$.

  Sample m positive observations $\{x_i\}_{i=1}^m$ from the empirical data distribution.

  Compute the gradient of the empirical log-EBM:

$$\begin{aligned}\nabla \log p(\theta_t) =& \mathbb{E}_{p_{\text{data}}}\left[\nabla_\theta f_{\theta_t}(x)\right] - \mathbb{E}_{p_\theta}\left[\nabla_{\theta_t} f_\theta(z_t)\right] \\ \approx& \frac{1}{n}\sum_{i=1}^n \nabla_\theta f_{\theta_t}(x_i) - \frac{1}{m}\sum_{i=1}^m \nabla_\theta f_{\theta_t}(z_t^m) \ .\end{aligned}$$

  Update the vector of global parameters of the EBM:

$$\theta_{t+1} = \theta_t + \eta_t \nabla \log p(\theta_t) \ .$$

end for

Output: Vector of fitted parameters θ_{T+1}.

2 Conclusion

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

A Appendix