Report_AlgriPart

4/25/2021

Algorithm

This part contains two parts: Hamiltonian Monte Carlo and Stochastic Gradient Hamiltonian Monte Carlo.

Hamiltonian Monte Carlo

HMC provides momentum variables (r) to establish a new joint distribution (θ, r) so that it can draw samples from the posterior distribution.

Let U be the potential energy function described in the article, we have

$$\pi(\theta, r) \propto \exp(-U(\theta) - \frac{1}{2} r^T M^{-1} r)$$

$$U = \sum_{x \in D} ln[p(x|D)] - ln[p(\theta)]$$

Since Hamilton function can be described as $H(\theta, r) = U(\theta) + \frac{1}{2} r^T M^{-1} r$ in the physics field, we have $d\theta = M^{-1} r dt$ and $dr = -\nabla U(\theta) dt$

Stochastic Gradient Hamiltonian Monte Carlo

Instead of computing the ∇U in $dr = -\nabla U(\theta) dt$, SGHMC uniformly chooses $\tilde{D}$ (minibatch) from the dataset and compute $\tilde{U}$. The equattion can be shown as the following:

$$\nabla \tilde{U}(\theta) = -\frac{|D|}{|\tilde{D}|} \sum_{x \in \tilde{D}} \nabla logp(x|\theta) - \nabla logp(\theta)$$

From the article, we can approximate the above equation by $\nabla U(\theta) + N(0, V(\theta))$. Since the equation can be regarded as a discreted system, we may want to introduce Metroplis-Hasting sampling. Let ϵ be the step term in in Metroplis-Hasting sampling, we can rewrite $dr = -\nabla U(\theta) dt$ as $dr = -\nabla U(\theta) + N(0, 2B(\theta) dt)$, where $B(\theta) = \frac{1}{2} \epsilon V(\theta)$ (positive semi-definite). In addition, the article shows we need to add a friction term because $\pi(\theta, r)$ is not invariant in this case. Finally the dynamic equation can be show as

. Therefore, the dynamic equations become the following:

$$d\theta = M^{-1} r dt \quad and \quad dr = -\nabla U(\theta) dt - BM^{-1} r dt + N(0, 2B(\theta) dt)$$

$$dr = -\nabla U(\theta) dt - BM^{-1} r dt + N(0, 2B(\theta) dt) \quad (10)$$