Weekly Report KARIMI-2021-11-05

My work this week has mainly been towards

1. Distributed and Private EBM
2. AAAI22 rebuttal (STANLEY paper)

1 Distributed and Private EBM

Focus on running experiments for this project. Talked to Jianwen several times to narrow down the project.

Algorithm 1 Distributed and private EBM

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 $\{z_0^m\}_{m=1}^M$. Set of selected devices $\mathcal{D}^t$.

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

Data: $\{x_i^p\}_{i=1}^{n_p}$, n_p number of observations on device p. $n = \sum_{p=1}^P n_p$ total.

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

  /* Happening on distributed devices */

  for *For device* $p \in \mathcal{D}^t$ **do**

    Draw M negative samples $\{z_K^{p,m}\}_{m=1}^M$ // local langevin diffusion

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

$$z_k^{p,m} = z_{k-1}^{p,m} + \gamma_k/2 \nabla_z f_{\theta_t}(z_{k-1}^{p,m})^{p,m} + \sqrt{\gamma_k}\mathsf{B}_k^p \,,$$

      where B_k^p denotes the Brownian motion (Gaussian noise).

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

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

    Compute the gradient of the empirical log-EBM // local - and + gradients

$$\delta^p = \frac{1}{M}\sum_{i=1}^M \nabla_\theta f_{\theta_t}(x_i^p) - \frac{1}{M}\sum_{m=1}^M \nabla_\theta f_{\theta_t}(z_K^{p,m})$$

    Use black box compression operators

$$\Delta^p = \mathcal{C}(\delta^p)$$

    Devices broadcast Δ^p to Server

  /* Happening on the central server */

  Aggregation of devices gradients: $\nabla \log p(\theta_t) \approx \frac{1}{|\mathcal{D}^t|}\sum_{p=1}^{|\mathcal{D}^t|}\Delta^p$.

  Update the vector of global parameters of the EBM: $\theta_{t+1} = \theta_t + \eta_t \nabla \log p(\theta_t)$

2 AAAI22 rebuttal (STANLEY paper)

Wrote the rebuttal and added some complexity comparison for our method.

*** Complexity Analysis:** We provide the running times of our method and the baselines in Table 1 on CIFAR-10 and Celeb-A datasets with a batchsize of 100. We would like to stress on the similar computational

complexity between the vanilla Langevin and our method STANLEY since our newly introduced stepsize uses the already computed gradient vector. On the contrary, the HMC method has recourse to both the gradient and the Hessian of the target distribution, resulting in longer computation time as reported on Table 1.

Table 1: Runtime (in s) for training our EBM during 1 epoch.

	Vanilla Langevin	HMC	GD	**STANLEY**
CIFAR-10 Dataset	232.5	698.4	211.3	265.2
Celeb-A dataset	376.3	640.1	345.2	414.8