Project Proposal - Prithvi

October 2023

1 Overview

A recent paper presenting a Latent Space Energy-Based Prior Model [4], leverages the concept of latent space modelling, wherein the focus lies on developing a low-dimensional representation of the data that encapsulates its fundamental characteristics. A key aspect of the proposed model is its integration of an energy-based framework, to represent and learn the latent space distribution. This learned distribution subsequently functions as a prior for a top-down generator, facilitating the mapping of acquired latent representations back to the original observable data. Additionally, the paper highlights the use of short-run Markov Chain Monte Carlo (MCMC) sampling from both the prior and posterior distributions. It argues that this simplistic Energy-Based Model (EBM) in the latent space effectively captures patterns within the data, while also leveraging the efficiency and strong mixing capabilities of MCMC sampling in the latent space.

However, the paper adopts a statistically flexible approach in its implementation. Notably, it operates **without error correction** in the Langevin sampling loop, which is used to estimate samples from complex and intractable probability distributions.

Despite these relaxed choices, the research demonstrates the model's capability to generate highly defined images when trained on the SVHN, CIFAR-10, and CelebA datasets. This leads to our proposed inquiry: would enhancing the model's ability to produce samples with lower variance in their evaluated likelihoods result in improved performance? Taking a step further, we propose an investigation into the **gradients of the log-likelihoods**, building upon the original paper's utilisation of the gradient of the log-marginal likelihood as the learning gradient, represented by $\nabla_\theta \log(p(x|\theta)) = \mathbb{E}p(z|x,\theta)\,[\nabla_\theta \log(p(x,z))]$ [4]. This would bypass the need for additional computations, a crucial advantage considering that the computational expenses associated with marginalising across the latent space in an MCMC estimation of the marginal likelihood would be very costly.

Consequently, the proposed project aims to address this question by including **thermodynamic integration** in the evaluation of the marginal likelihood. Drawing from insights provided in the previous work of [1], it is suggested that the variance in the MCMC estimate can be altered through the use of power posteriors [3]. This offers a promising avenue to gain a more nuanced understanding of the impact of likelihood variance on the model's performance and its implications for the latent energy-based model under consideration.

With these questions in mind, the proposed approach is to follow an experimental procedure similar to previous work investigating gradient variation, [2]. The proposed experiment procedure is outlined as follows:

1. Train both models, (the vanilla model from the original paper, and an altered model incorporating thermodynamic integration), on training datasets for the same number of epochs.

   - In the altered model, we sample z from the distribution $p(z|x,\theta,t) = \frac{p(x|z,\theta)^t p(z|\alpha)}{Z_t}$, where $p(z|\alpha) = \frac{1}{Z_\alpha} \exp(f(z|\alpha))\pi_0(z)$ represents the exponentially tilted prior distribution. We then optimise for $\theta^* = \arg\max_\theta \log(p(x|\theta)) \propto \int_0^1 \mathbb{E}_{p(z|x,t,\theta)} [\log(p(x|z,\theta))]\, dt$. Here, t denotes a temperature within the temperature schedule, x is the observed data, θ represents the generator parameters, and α signifies the EBM parameters. Z_t and Z_α are the respective normalising constants.
   - In the unaltered model, the posterior sampling is instead conducted on $p(z|x,\theta) = \frac{p(z|\alpha)p(x|z,\beta)}{p(x|\theta)}$, where $p(z|\alpha)$ is the same exponentially-tilted prior mentioned above, and $p(x|z,\beta)$ is obtained from the generator model [4]. Optimisation is then conducted on $\nabla_\theta \log(p(x|\theta)) = \mathbb{E}_{p(z|x,\theta)} [\nabla_\theta \log(p(x,z))]$

2. Similar to the work presented in [2], during the training procedure periodically extract the grad-marginal log-likelihoods, $\nabla_\theta \log(p(x|\theta))$ of a particular batch, and calculate the average variances. These gradients should be readily available from the optimisation procedure mentioned above. During this periodic break, also sample an image and collect FID scores to serve as a quantitative measure of the models' generative capabilities.

3. Once training is complete, additionally evaluate the models on test datasets and collect the same metrics mentioned in the previous stage.

4. Repeat the entire procedure until a comprehensive range of variances and FID scores are obtained. If a wider range of variances is required, consider adopting different temperature schedules, or network architectures.

5. Once the data is sufficient, plot variance against training iteration, FID score against training iteration, and FID score against variance for each model. Compare and contrast.

2 Milestones

Please note that other coursework deadlines have been included in blue. Project work may be relaxed in the lead up to these.

03/11/23 - **4F13 Deadline #1** - Implement the latent energy-based prior model as a Juypter notebook. Train to generate MNIST images. In a different Juypter notebook, adapt the model to use power posteriors.

10/11/23 - Continue working on the notebooks. Once complete, transfer and modularise them into a platform for experimentation in file system. Now that the notebook implementations have been completed, refer back to the literature to consolidate your understanding.

17/11/23 - **4F13 Deadline #2** - Continue adapting the modularised code into the actual experiment listed above. Investigate methods to store your models and data in a readily accessible way, (save to file).

24/11/23 - Try investigating the models with different temperature schedules and networks. The aim is to build some intuition on directions to take if a wider variance range is required. During the training, keep evaluating the model's grad-likelihood estimates to verify the range of variances that you are gathering.

01/12/23 - **4F13 Deadline #3** - Continue training and saving models. Begin evaluating on test datasets and collecting variances.

07/12/23 - **4M17 Deadline** - Present your findings to Justin and Mark. Discuss next steps.

References

[1] Ben Calderhead and Mark Girolami. Estimating bayes factors via thermodynamic integration and population mcmc. *Computational Statistics Data Analysis*, 53(12):4028–4045, 2009.

[2] Fartash Faghri, David Duvenaud, David J Fleet, and Jimmy Ba. A study of gradient variance in deep learning. July 2020.

[3] N. Friel and A. N. Pettitt. Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(3):589–607, 2008.

[4] Bo Pang, Tian Han, Erik Nijkamp, Song-Chun Zhu, and Ying Nian Wu. Learning latent space energy-based prior model. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 21994–22008. Curran Associates, Inc., 2020.