Adaptive Multi-stage Density Ratio Estimation for Learning Latent Space Energy-based Model

Zhisheng Xiao, Tian Han

NEURAL INFORMATION PROCESSING SYSTEMS

1 Motivations and Backgrounds

Backgrounds:

- Given two distributions with density $p(x)$ and $q(x)$, the density ratio $r(x)= (p(x))/(q(x))$ can be estimated by **training a classifier to distinguish samples** from p and q
- Such a technique can be useful for training Energy-based models (EBMs)
- However, the density ratio estimation is severely inaccurate when the **gap between p and q is large**
- In practice, it is difficult to apply this technique to train EBMs, since we typically do not have a parametrized base distribution that is close to the data distribution

Motivations for adaptive multi-stage density ratio estimation:

- The biggest issue for density ratio estimation is **the discrimination being too easy**, so the discriminator does not learn meaningful information
- We design **multiple stages** of density ratio estimation, where in each stage we **update the base distribution and target distribution adaptively**, so that the discrimination task is increasing harder

2 Adaptive Multi-stage Density Ratio Estimation

- We train our latent space EBM **adaptively** and **in multiple stages**
- Our latent EBM prior has the form below, where the prior model $p_\varphi(z)$ can be sequentially and adaptively learned to bridge the gap between **prior** and **posterior** densities in the previous stages.

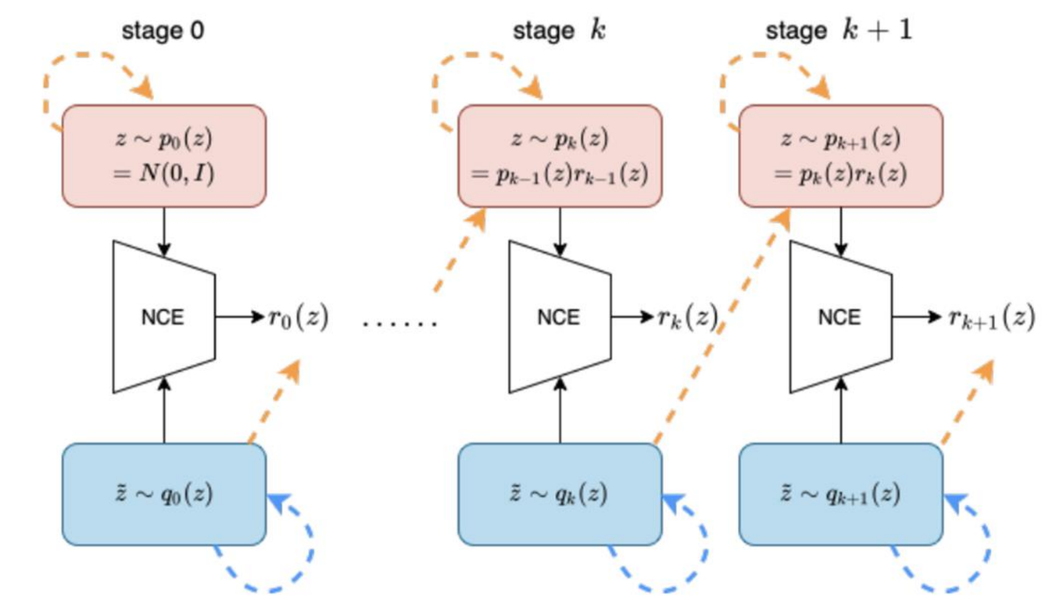
$$p_\phi(\mathbf{z}) = \prod_{k=0}^{m-1} r_{\phi_k}(\mathbf{z})\, p_0(\mathbf{z})$$

- At the $k+1^{th}$ stage, a density ratio estimator $r_{\varphi_k}(z)$ is introduced to estimates the ratio between the **aggregated posterior** and new **prior** at current stage
- The new prior is the previous prior corrected by $r_{\varphi_k}(z)$:

$$p_{\phi_{k+1}}(\mathbf{z}) = r_{\phi_k}(\mathbf{z}) p_{\phi_k}(\mathbf{z})$$

- An illustration of training process is presented in the next section

3 Method



- We estimate the density ratio $r_k(z)$ in each stage using contrastive estimation which trains a classifier to distinguish samples from the prior $p_\varphi(k)$
- Posterior samples are obtained by short-run LD (blue dashed curve), prior samples can be obtained either by short-run LD (orange dashed curve) or using persistent chain (orange dashed line)
- The ratio estimated in stage k can be integrated to form a new prior in stage $k+1$

4 Results on Sample Quality

	SVHN		CelebA		CIFAR-10	
	MSE	FID	MSE	FID	MSE	FID
VAE [23]	0.019	46.78	0.021	65.75	0.057	106.37
ABP [13]	-	49.71	-	51.50	-	-
SRI [33]	0.018	44.86	0.020	61.03	-	-
SRI (L=5) [33]	0.011	35.32	0.015	47.95	-	-
2s-VAE [4]	0.019	42.81	0.021	44.40	0.056	72.90
RAE [9]	0.014	40.02	0.018	40.95	0.027	74.16
NCP-VAE [1]	0.020	33.23	0.021	42.07	0.054	78.06
LEBM [36]	0.008	29.44	0.013	37.87	0.020	70.15
Adaptive CE (ours)	**0.004**	**26.19**	**0.009**	**35.38**	**0.008**	**65.01**

MSE(↓) and FID(↓) obtained from models trained on different datasets.

5 Qualitative Samples

Qualitative samples on SVHN, CelebA-64, CIFAR-10 and CelebAHQ-256

6 Summary

- In this paper, we propose adaptive multi-stage density ratio estimation, which is an effective method for learning an EBM prior for a generator model
- Our method learns the latent EBMs by introducing multiple density ratio estimators that learn the density ratio between prior and posterior sequentially and adaptively
- Empirical results show the advantage of our method on generation, reconstruction and anomaly detection tasks.