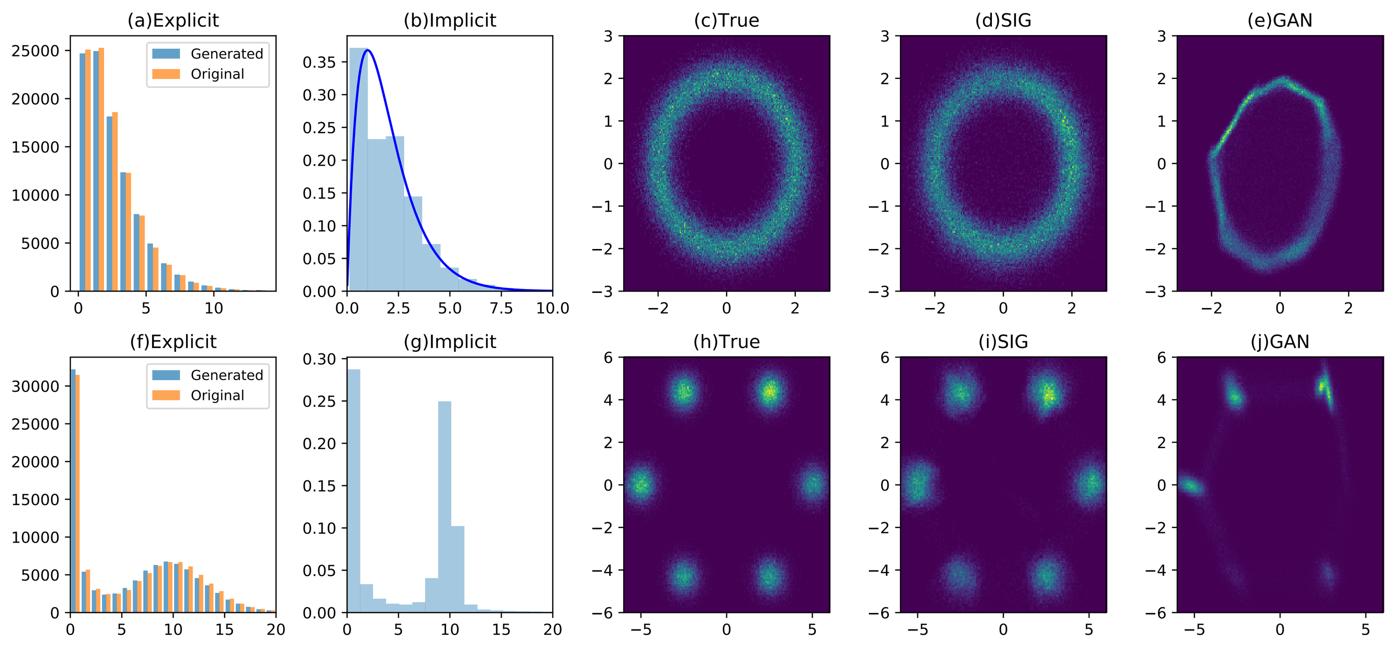
Semi-implicit generative model

Mingzhang Yin, Mingyuan Zhou

Semi-implicit generator



Types of generative models

- Explicit generative models: has tractable probability density; can be trained with MLE; for example: VAE, PixelRNN, RealNVP, SBN
- Implicit generative models: has no point-wise evaluable PDF; can be trained adversarially but not with MLE; for example: GAN

Semi-implicit Generator

- Implicit-step: $\boldsymbol{\theta}_i = g_{\boldsymbol{\phi}}(\boldsymbol{z}_i)$, $\boldsymbol{z}_i \sim p(\boldsymbol{z})$; Explicit-step: $\boldsymbol{x}_i \sim p(\boldsymbol{x} \mid \boldsymbol{\theta}_i)$.
- Can be considered as an infinite mixture of analytic densities with an implicit mixing distribution $\boldsymbol{x}_i \sim \int p(\boldsymbol{x}|\boldsymbol{\theta}) p_{\boldsymbol{\phi}}(\boldsymbol{\theta}) d\boldsymbol{\theta}$
- With finite mixture approximation, SIG can be trained in the MLE framework

$$\min_{\boldsymbol{\phi}} \mathbb{H}_M = -\mathbb{E}_{p_{\text{data}}(\boldsymbol{x})} \mathbb{E}_{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_M \sim p_{\boldsymbol{\phi}}(\boldsymbol{\theta})} \log \frac{1}{M} \sum_{j=1}^{M} p(\boldsymbol{x} \mid \boldsymbol{\theta}_j) \quad (1)$$

SIG Generation in Multi-modal Space

Definition 1: Discrete multi-modal space

Suppose $(\mathcal{X}, \nu)$ is a metric space with metric $\nu : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$, $\mathcal{X} = \bigcup_{i=1}^{K} U_i$, where $U_i \cap U_j = \emptyset$ for $i \neq j$. Let the distance between two sets be $D(U_i, U_j) = \inf\{\nu(x, y); x \in U_i, y \in U_j\}$ and let the diameter of a set be $d(U) = sup\{\nu(x, y); x, y \in U\}$. Suppose there exists $c_0 > \epsilon_0 > 0$ such that $\min_{i,j} D(U_i, U_j) > c_0$, $\max_i d(U_i) < \epsilon_0$. Then $\mathcal{X} = \bigcup_{k=1}^{K} U_k$ is a discrete multi-modal space under mesure ν.

- We can study a simplified optimal assignment problem: assuming that N data points have been sampled from the true data distribution, how to assign M generated data to the neighborhood of the true data such that $\mathbb{H}_M$ defined in (1) is minimized under expectation?

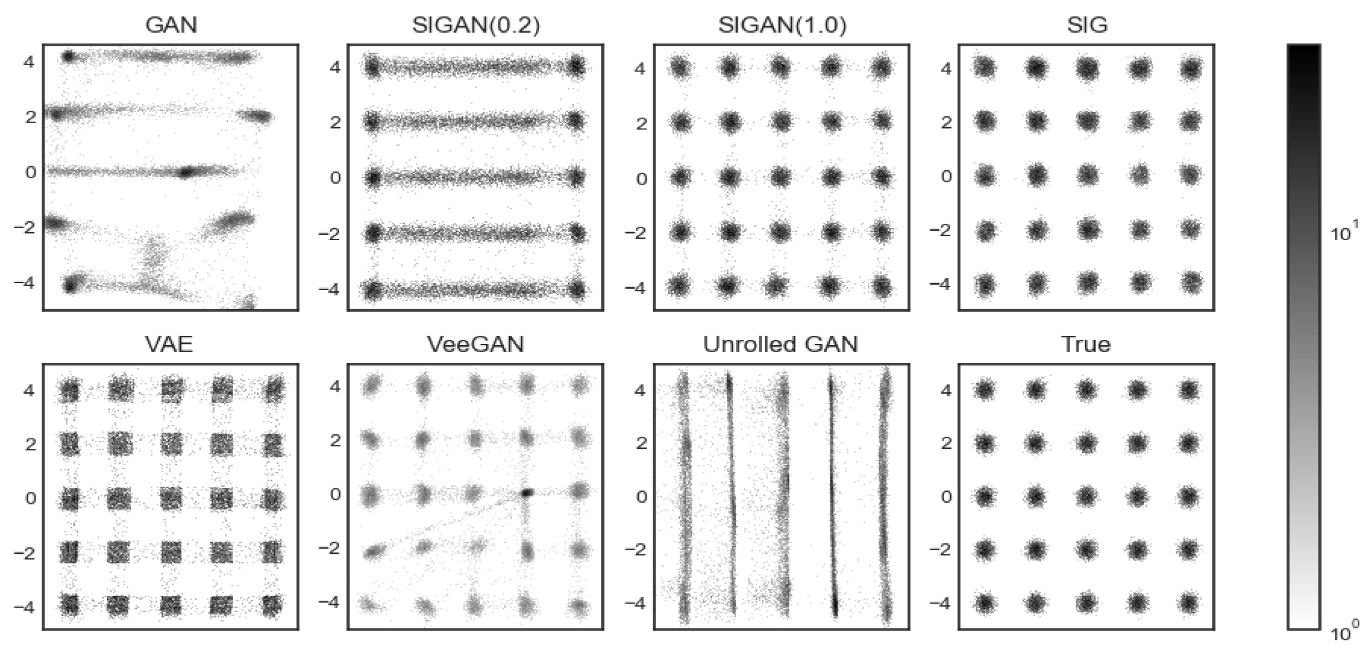
$$\min_{\{m_1, \cdots, m_K\}} -\frac{1}{N} \sum_{i=1}^{N} \log \frac{1}{M} \sum_{j=1}^{M} \mathbb{E}_{\boldsymbol{x}_i \in U_{t_i}, \boldsymbol{\theta}_j \in U_{z_j}} [p(\boldsymbol{x}_i \mid \boldsymbol{\theta}_j)], \quad (2)$$

- The discrete multi-modal space $\mathcal{X} = \bigcup_{i=1}^{K} U_i$, $\boldsymbol{x}_i \in U_{t_i}$, $\boldsymbol{\theta}_j \in U_{z_j}$, $t_i, z_j \in \{1, \ldots, K\}$ and $\{m_k\}_{k=1}^{K}$ are the number of $\boldsymbol{\theta}$'s that are assigned to be in U_k.

Theorem 1: SIG in multi-modal space

Suppose P_{data} is defined on a discrete multi-modal space $\mathcal{X} = \bigcup_{i=1}^{K} U_i$ with l_2-norm. Suppose there are N data points $\boldsymbol{x}_i \sim P_{data}, i = 1, \cdots, N$, among which n_k points belong to U_k. Suppose we need to sample $\boldsymbol{\theta}_j \sim p_{\boldsymbol{\phi}}(\boldsymbol{\theta}), j = 1, \cdots, M$, and m_k denotes the number of $\boldsymbol{\theta}$'s in U_k. Denoting r as a radial basis function (RBF), we let $u = \mathbb{E}[r(\boldsymbol{x}, \boldsymbol{\theta})]$ if $\boldsymbol{x}, \boldsymbol{\theta} \in U_i$, and $v = \mathbb{E}[r(\boldsymbol{x}, \boldsymbol{\theta})]$ if $\boldsymbol{x} \in U_i$, $\boldsymbol{\theta} \in U_j, i \neq j$. Then the objective in (2) is convex and the optimum m_k to maximize (2) satisfies $\frac{m_k^*}{M} = \frac{n_k}{N} + (\frac{n_k}{N} - \frac{1}{K})\frac{Kv}{(u-v)}$. In particular, $m_k^* \neq 0$ if $n_k > \frac{N}{K} \frac{1}{1+\frac{u-v}{Kv}}$.

- We compare different generative models on a 5×5 Gaussian mixture model by sampling 50,000 points from trained generator.



GAN with Semi-implicit Regularizer

- Generative adversarial network (GAN) solves a minimax problem

$$\min_G \max_D V(D, G) = \mathbf{E}_{\boldsymbol{x} \sim p_{data}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbf{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z}))]$$

- Changing generator loss from $\mathbf{E}_{\boldsymbol{z}}[\log(1 - D(G(\boldsymbol{z}))]$ to $\frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \exp(\sigma^{-1}(D(G(\boldsymbol{z}))))$, and setting discriminator as ideally optimal one, the generator loss of GAN is identical to the SIG loss.
- Combining SIG and GAN objective as GAN-SI can interpolate between the adversarial training (weak fitting) and MLE training (strong fitting).
- For GAN-SI, the discriminative loss is

$$\min_{\gamma} -\mathbb{E}_{\boldsymbol{x} \sim P_d} \log D_{\gamma}(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})} \log(1 - D_{\gamma}(T_{\boldsymbol{\phi}}(\boldsymbol{z}))) \quad (3)$$

- the generator loss is a linear combination of the original GAN loss and SIG loss

$$\min_{\boldsymbol{\phi}} -\mathbb{E}_{\boldsymbol{z} \sim g(\boldsymbol{z})}[\log D_{\gamma}(T_{\boldsymbol{\phi}}(\boldsymbol{z})) - \lambda \mathbb{E}_{\boldsymbol{x} \sim P_d} \log \int p(\boldsymbol{x} \mid \boldsymbol{\theta}) p_{\boldsymbol{\phi}}(\boldsymbol{\theta}) d\boldsymbol{\theta}], \quad (4)$$

- γ are the discriminator network parameters, $\boldsymbol{\theta} = T_{\boldsymbol{\phi}}(\boldsymbol{z})$ is the deterministic transform in the implicit distribution and $\lambda \geq 0$ is a hyperparameter to balance the strength between the GAN and SIG objectives.

GAN-SI Experiments

Stacked MNIST

- To measure the performance on discrete multimode data, we stack 3 randomly chosen MNIST images on the RGB color channels to form a $28 \times 28 \times 3$ image (MNIST-3).
- MNIST-3 contains 1000 modes corresponding to 3-digit between 0 and 999.
-

	IS	High quality	$\exp(H(y\|x))$	$\exp(H(y))$	Mode	KL$(P_g)\|\|P_d$
DCGAN(S)	2.9±0.52	**0.63±0.14**	**1.96±0.32**	5.1±1.19	21.0±8.12	4.99±0.24
DCGAN-SI(S)	**4.33±0.59**	0.6±0.07	2.05±0.2	**8.78±0.41**	**279.2±296.52**	**2.63±1.0**
DCGAN(M)	5.59±0.36	0.7±0.03	1.71±0.09	9.51±0.31	811.8±116.24	0.75±0.35
DCGAN-SI(M)	**5.93±0.47**	**0.72±0.04**	**1.65±0.11**	**9.75±0.11**	**969.0±29.19**	**0.3±0.13**
DCGAN(L)	4.71±1.12	0.67±0.08	1.78±0.17	8.25±1.32	389.8±477.24	2.95±2.33
DCGAN-SI(L)	**6.05±0.68**	**0.73±0.06**	**1.62±0.17**	**9.75±0.12**	**957.0±32.74**	**0.36±0.12**

- High quality image and entropy $\exp(H(y|x))$ reflect sample quality while $\exp(H(y))$, Mode and KL reflect sample diversity. For Inception score, High quality image, $\exp(H(y))$, higher is better; for $\exp(H(y|x))$ and KL, lower is better.

CIFAR10

- We test the semi-implicit regularizer on the CIFAR-10 dataset.
- We combine semi-implicit regularizer with DCGAN and WGAN-GP to balance the quality and diversity of generated samples.
-

Real data	Unsupervised, standard CNN			
	DCGAN	DCGAN-SI	WGAN-GP	WGAN-GP-SI
11.24 ± .12	6.16 ± .14	6.85± .06	6.43 ± .07	6.67 ± .11

(a)DCGAN

(b)DCGAN-SI

References

[1] Martin Arjovsky and Léon Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.

[2] Tong Che, Yanran Li, Athul Paul Jacob, Yoshua Bengio, and Wenjie Li. Mode regularized generative adversarial networks. *arXiv preprint arXiv:1612.02136*, 2016.

[3] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled generative adversarial networks. *arXiv preprint arXiv:1611.02163*, 2016.

Extended version at `https://github.com/mingzhang-yin`