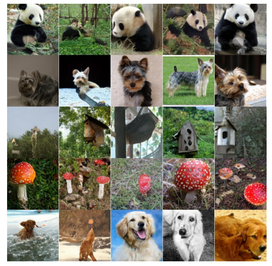
An Information-Theoretic Evaluation of Generative Models in Learning Multi-modal Distribution

Mohammad Jalali, Cheuk Ting Li, Farzan Farnia *mjalali@ec.iut.ac.ir, ctli@ie.cuhk.edu.hk, farnia@cse.cuhk.edu.hk*

Diversity evaluation of generative models

Which model can generate more modes?

StyleGAN-XL



BigGAN

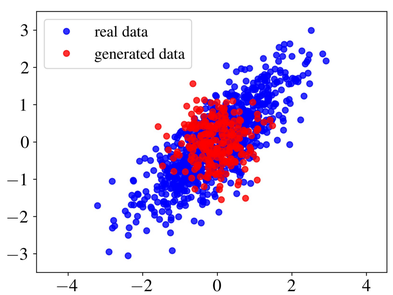
LDM

Existing Metrics:

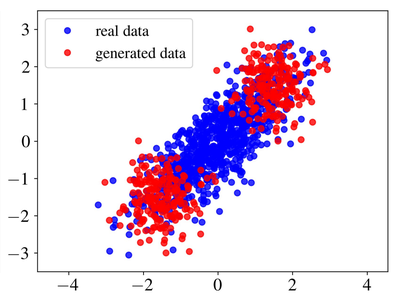
- **Recall (Kynkäänniemi 2019)**
- **Coverage (Naeem 2020)**

Evaluation of generative models under multi-modal distributions

What if the we know that the underlying distribution have multiple modes?



Recall: 0.77, Coverage: 0.42



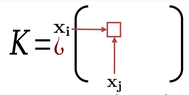
Recall: 0.74, Coverage: 0.42

Recall and Coverage can not count the number of modes.

Can we count the number of modes?

Estimating the number of modes in multi-modal distributions

An entropy based approach:



$$k_\sigma(\mathbf{x_i}, \mathbf{x_j}) := \exp\left(-\frac{\|\mathbf{x_i} - \mathbf{x_j}\|_2^2}{2\sigma^2}\right)$$

$$\mathrm{RE}_\alpha(K) := \frac{1}{1-\alpha}\log\left(\mathrm{Tr}(K^\alpha)\right) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{d}\lambda_i^\alpha\right)$$

Close form expression of order-2 Renyi Kernel Entropy (RKE) for Gaussian Mixture model if the kernel bandwidth dominates the spectral norm (maximum eigenvalue) of the component-wise covariance matrices:

$$\widehat{\mathrm{RKE}}_2(\mathbf{X}) = -\log\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}k^2(\mathbf{x}_i, \mathbf{x}_j)\right) = -\log\left(\|K_{XX}\|_F^2\right)$$

Can we measure number of common modes between dataset and generated data?

RRKE as the Number of Common Modes:

$$\mathrm{RRE}_{1/2}(X, Y) := -2\log\left(\mathrm{Tr}\left(\sqrt{Y^{1/2}XY^{1/2}}\right)\right)$$

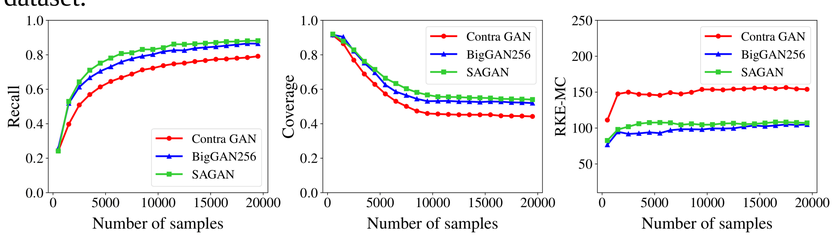
Close form expression of order ½ for Gaussian Mixture model:

$$\widehat{\mathrm{RRKE}}_{\frac{1}{2}}(\mathbf{X}, \mathbf{Y}) = -\log\left(\|K_{XY}\|_*^2\right)$$

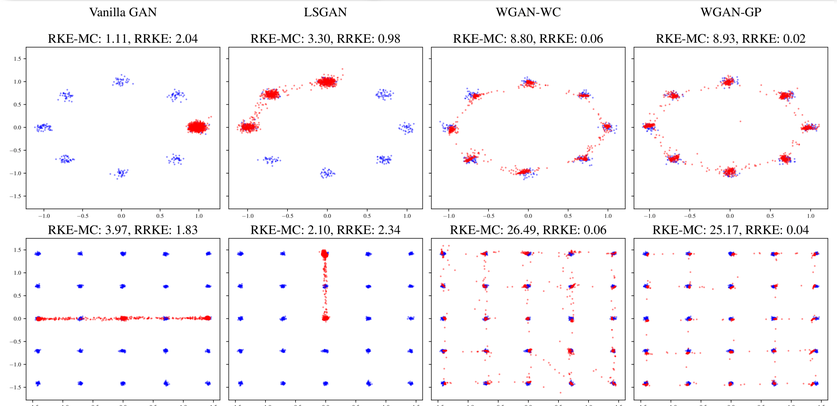
Where K_{XY} denotes the normalized cross kernel matrix and * denotes the nuclear norm, i.e. the sum of matrix's singular values.

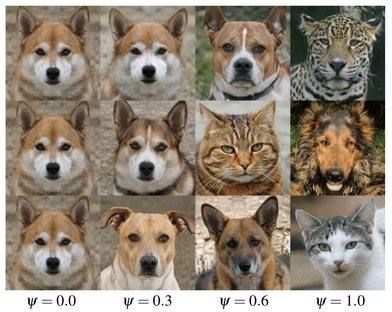
Numerical Result: Comparing sample complexity

Comparing convergence of Recall, Coverage, and RKE score on ImageNet dataset.

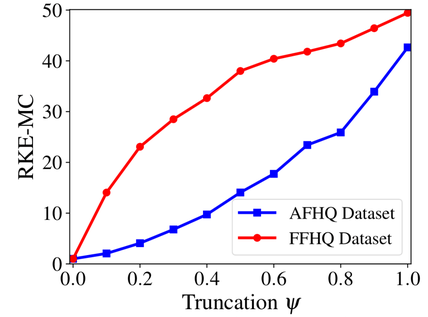


Numerical Result: Synthetic and Real datasets





AFHQ generated samples with different truncations



RKE Mode Count for StyleGAN samples on AFHQ and FFHQ varying with truncation factor

FFHQ generated samples with different truncations

References

Tuomas Kynkäänniemi et al. (2019). "Improved Precision and Recall Metric for Assessing Generative Models". In: Advances in Neural Information Processing Systems 32.

Muhammad Ferjad Naeem et al. (2020). "Reliable Fidelity and Diversity Metrics for Generative Models". In: International Conference on Machine Learning 37.