Evaluating Generative Models using Nonparametric Estimation of Rényi Divergence with Sample-level Auditing

Student: Alexis Fox
Mentors: Dr. Samarth Swarup and Dr. Abhijin Adiga

Background

Generative Models

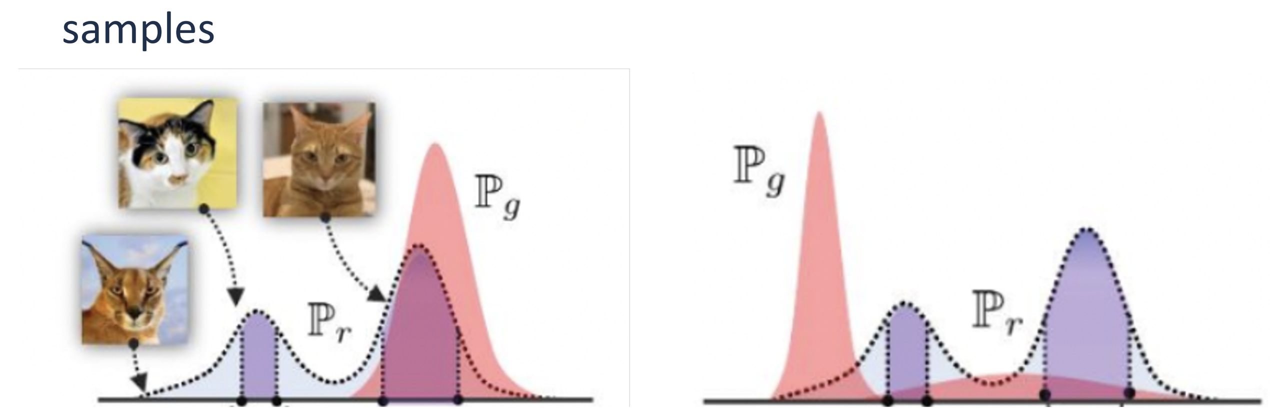
- **Are a class of machine learning models designed to generate new data samples similar to existing datasets**
  - They learn underlying patterns & distribution of given dataset
  - *Examples*: Text generation (ChatGPT) , image generation (DALLE)

Need for Metric

- **No ground truth labels for generated samples**
  - "Good" vs. "bad" samples is unclear
  - Compared to discriminative models, which have clear false positives, true negatives, etc.
- **Need a way to quantify quality of model based on difference between real and generated probability distributions**

Current Metrics

- **Popular ones include Fréchet Inception Distance (finds distance between feature vectors) & Parzen Window (density estimator, computes likelihood)**
  - ***Issues***: 1D scores can't distinguish failure cases, domain-specific
- **Recent proposed metrics are 2D, measure precision & recall tradeoff**
  - ***Precision***: Measures fidelity, ability of model to produce realistic samples
  - ***Recall***: Measures diversity, ability of model to produce wide range of distinct samples



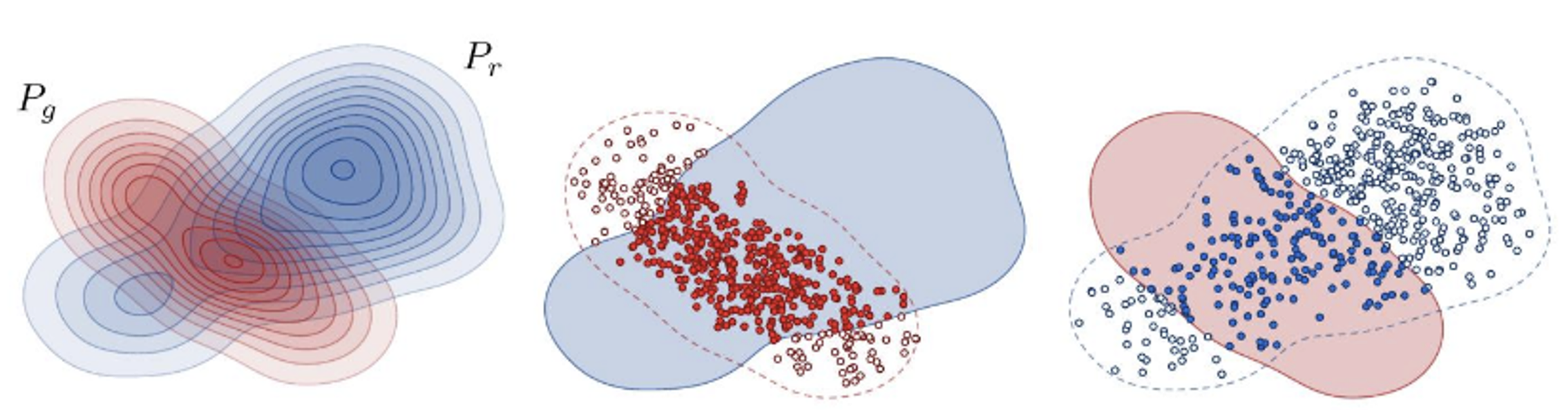
High Precision, Low Recall *High Recall, Low Precision*

Rényi Divergence

- **Family of divergence measures that quantify difference between 2 probability distributions (P & Q), can be used as metric for generative models**

$$R_\alpha(P\|Q) = \frac{1}{\alpha-1}\ln \sum_{i=1}^{n} \frac{p_i^\alpha}{q_i^{\alpha-1}}$$

- **Let P = Real distribution, Q = Generated distribution**
  - **R_α(Q||P)** can measure precision (how much of Q is in P)
  - **R_α(P||Q)** can measure recall (how much of P is in Q)



(a) Example distributions (b) Precision (c) Recall

- **Parametrized by α: $\alpha > 1$ weights popular events more, $\alpha < 1$ weights rare events, α->1 converges to KL divergence**

Estimating Divergence

- **Often lack full distribution of P or Q, or the integrals are intractable, so Renyi divergence can't be directly calculated**
  - Therefore, need to estimate divergence
- **Parametric/Nonparametric Estimators**
  - ***Parametric***: Strong assumptions about distribution form (ex: fitting Gaussian), fast but less accurate
  - ***Nonparametric***: Minimal assumptions about distribution (ex: Kernel Density Estimator), slow but more accurate
- **Plug-in/Direct Estimators**
  - ***Plug-in***: Estimate density, plug-into divergence equation
  - ***Direct***: Divergence directly estimated from data without explicit density estimations

Purpose

1. **Find/develop an asymptotically-unbiased, direct nonparametric estimator of Rényi Divergence**
2. **Use second-order Rényi divergence to quantify precision (fidelity) and recall (diversity) of generative model & create frontier to visualize trade-off**
3. **Analyze sample contribution to divergence metric to audit unrealistic outliers**

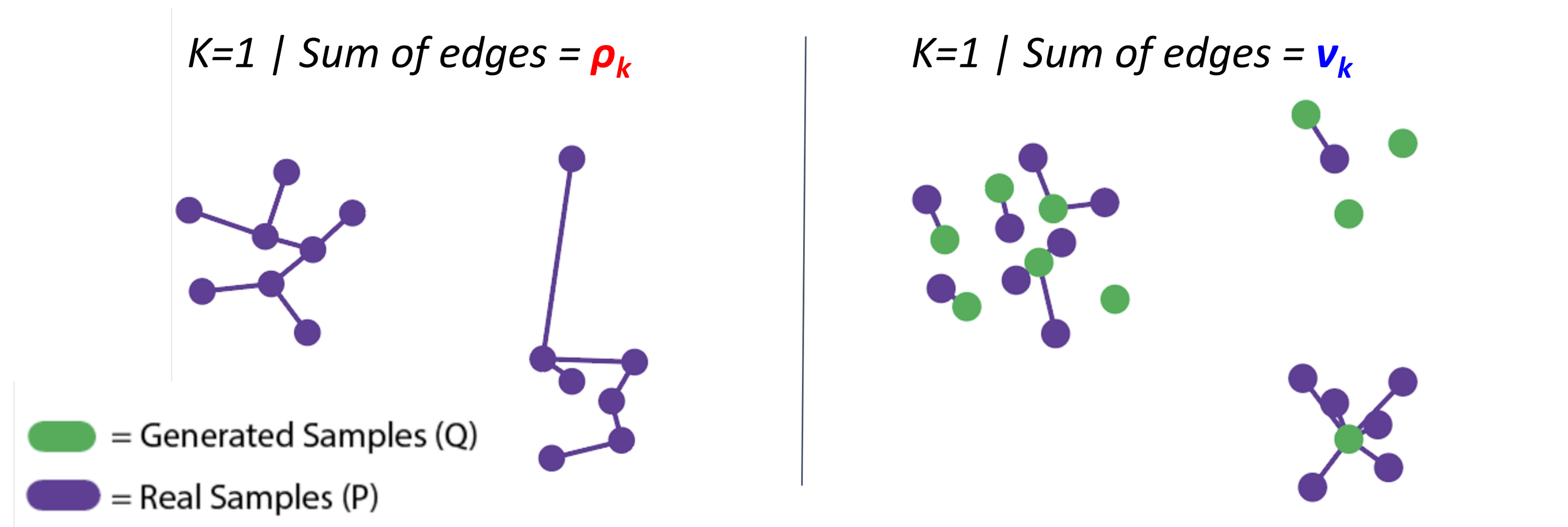
Methodology

Use a Fixed-K Bias-Corrected Estimator of Divergence

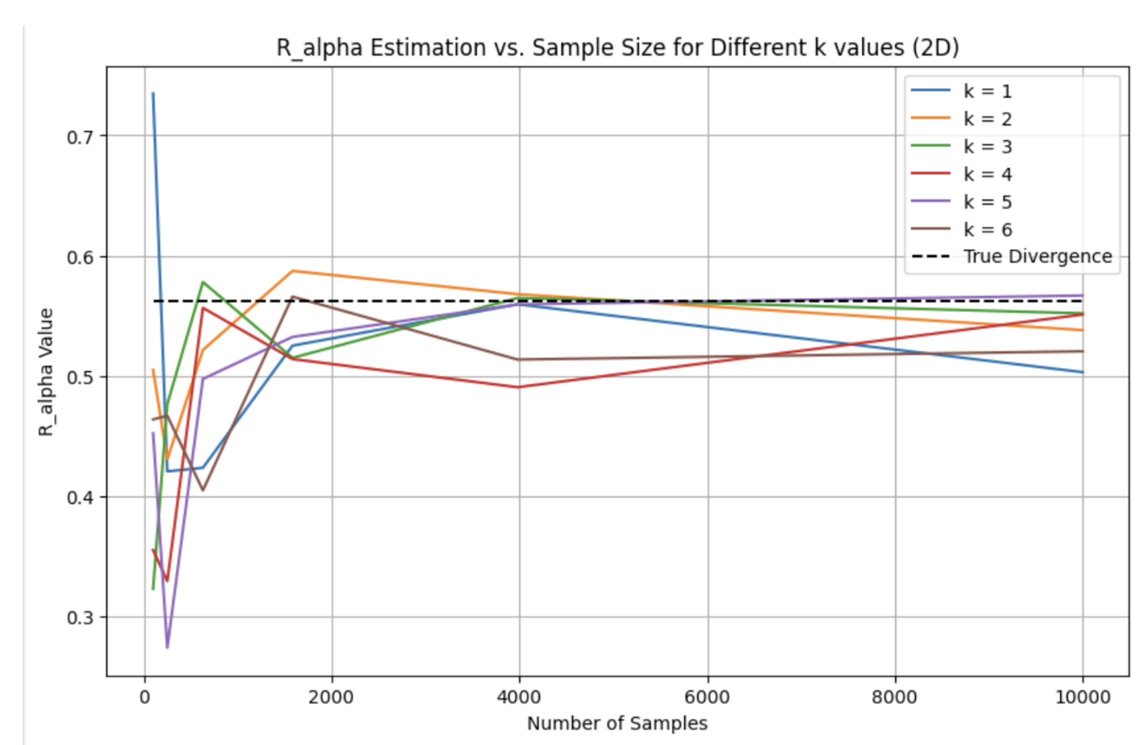
- Edges of K-nearest-neighbor graphs can be used in divergence estimation
  - Points nearer each other contribute less to divergence, farther apart bigger contribution
- A direct, nonparametric estimator developed by Póczos & Schneider:

$$\widehat{R}_\alpha\left(P_{1:n}\|Q_{1:m}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{(n-1)\rho_k^d(i)}{m\nu_k^d(i)}\right)^{1-\alpha} B_{k,\alpha}$$

  - **ρ_k(i)** := the Euclidean distance of the kth nearest neighbor of **P_i** in the sample **$P_{1:n}$**
  - **ν_k(i)** := the distance of the kth nearest neighbor of **P_i** in the sample **$Q_{1:n}$**
  - **$B_{k,a}$** := Multiplicative bias-correcting term that depends on *k* and α

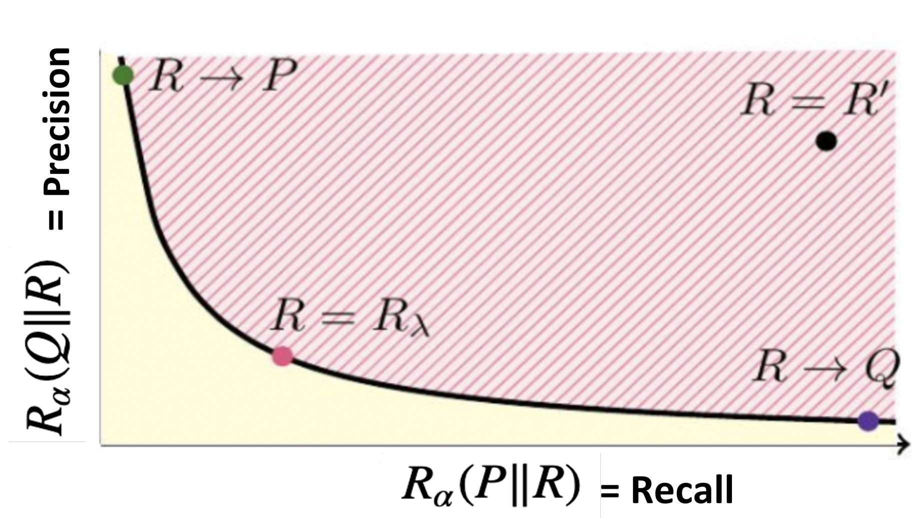


  - **Properties**
    - Asymptotically-unbiased (converges to true divergence value)
    - Curse of dimensionality, convergence rate: $O\left(\frac{1}{\sqrt[d]{n}}\right)$
    - Choice of k doesn't matter; fixed-k



Use estimation of Renyi Divergence (a=2) to calculate Precision and Recall

- Similar to precision-recall curve for classification, can create curve that represents tradeoff
  - Generated by confidence-thresholding **Q** (using intermediate distribution R) than calculating corresponding **R_α(Q||P)** and **R_α(P||Q)**



Divergence Frontier: Precision-Recall Curve

- Will analyze sensitivity to outliers

Sample-level Auditing

- Will examine each generated samples' contribution to divergence metric, larger **R_α(Q||P)** means less realistic sample
  - Can create a realism score
  - Will analyze thresholding outlier rate

References

Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y., & Yoo, J. (2020, November). Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning* (pp. 7176-7185). PMLR.
Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems, 32*.
Póczos, B., & Schneider, J. (2011, June). On the Estimation of a-Divergences. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (pp. 609-617). JMLR Workshop and Conference Proceedings.
Noshad, M., Moon, K. R., Sekeh, S. Y., & Hero, A. O. (2017, June). Direct estimation of information divergence using nearest neighbor ratios. In *2017 IEEE International Symposium on Information Theory (ISIT)* (pp. 903-907). IEEE.
Alaa, A., Van Breugel, B., Saveliev, E. S., & van der Schaar, M. (2022, June). How faithful is your synthetic data? sample-level metrics for evaluating and auditing generative models. In *International Conference on Machine Learning* (pp. 290-306). PMLR.