Deep Generative Models

Lecture 9

Roman Isachenko

AI Masters

2024, Spring

Recap of previous lecture

WGAN objective

$$\min_{\boldsymbol{\theta}} W(\pi || p) = \min_{\boldsymbol{\theta}} \max_{\phi \in \boldsymbol{\Phi}} \left[\mathbb{E}_{\pi(\mathbf{x})} f_\phi(\mathbf{x}) - \mathbb{E}_{p(\mathbf{z})} f_\phi(G_{\boldsymbol{\theta}}(\mathbf{z})) \right].$$

- Function f in WGAN is usually called *critic*.
- If parameters ϕ lie in a compact set $\boldsymbol{\Phi} \in [-c, c]^d$ then $f(\mathbf{x}, \phi)$ will be K-Lipschitz continuous function.

$$K \cdot W(\pi || p) = \max_{\|f\|_L \leq K} \left[\mathbb{E}_{\pi(\mathbf{x})} f(\mathbf{x}) - \mathbb{E}_{p(\mathbf{x})} f(\mathbf{x}) \right] \geq$$
$$\geq \max_{\phi \in \boldsymbol{\Phi}} \left[\mathbb{E}_{\pi(\mathbf{x})} f_\phi(\mathbf{x}) - \mathbb{E}_{p(\mathbf{x})} f_\phi(\mathbf{x}) \right]$$

"Weight clipping is a clearly terrible way to enforce a Lipschitz constraint"

Arjovsky M., Chintala S., Bottou L. Wasserstein GAN, 2017

Recap of previous lecture

Theorem

Let $\pi(\mathbf{x})$ and $p(\mathbf{x})$ be two distribution in $\mathcal{X}$, a compact metric space. Let γ be the optimal transportation plan between $\pi(\mathbf{x})$ and $p(\mathbf{x})$. Then

$$\mathbb{P}_{(\mathbf{y},\mathbf{z})\sim\gamma}\left[\nabla f^*(\hat{\mathbf{x}}_t) = \frac{\mathbf{z} - \hat{\mathbf{x}}_t}{\|\mathbf{z} - \hat{\mathbf{x}}_t\|}\right] = 1.$$

Gradient penalty

$$W(\pi||p) = \underbrace{\mathbb{E}_{\pi(\mathbf{x})} f(\mathbf{x}) - \mathbb{E}_{p(\mathbf{x})} f(\mathbf{x})}_{\text{original critic loss}} + \lambda \underbrace{\mathbb{E}_{U[0,1]}\left[\left(\|\nabla f(\hat{\mathbf{x}})\|_2 - 1\right)^2\right]}_{\text{gradient penalty}}.$$

Samples $\hat{\mathbf{x}}_t = t\mathbf{y} + (1-t)\mathbf{z}$ with $t \in [0, 1]$ are uniformly sampled along straight lines between pairs of points: $\mathbf{y}$ from the data distribution $\pi(\mathbf{x})$ and $\mathbf{z}$ from the generator distribution $p(\mathbf{x}|\boldsymbol{\theta})$.

Arjovsky M., Chintala S., Bottou L. Wasserstein GAN, 2017

Gulrajani I. et al. Improved Training of Wasserstein GANs, 2017

Recap of previous lecture

f-divergence minimization

$$D_f(\pi||p) = \mathbb{E}_{p(\mathbf{x})} f\left(\frac{\pi(\mathbf{x})}{p(\mathbf{x})}\right) \rightarrow \min_p .$$

Here $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a convex, lower semicontinuous function satisfying $f(1) = 0$.

Variational divergence estimation

$$D_f(\pi||p) \geq \sup_{T \in \mathcal{T}} \left[\mathbb{E}_\pi T(\mathbf{x}) - \mathbb{E}_p f^*(T(\mathbf{x}))\right],$$

Fenchel conjugate

$$f^*(t) = \sup_{u \in \mathrm{dom}_f} \left(ut - f(u)\right), \quad f(u) = \sup_{t \in \mathrm{dom}_{f*}} \left(ut - f^*(t)\right)$$

Note: To evaluate lower bound we only need samples from $\pi(\mathbf{x})$ and $p(\mathbf{x})$. Hence, we could fit implicit generative model.

Nowozin S., Cseke B., Tomioka R. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization, 2016

Recap of previous lecture

How to evaluate likelihood-free models?

$p(y|\mathbf{x})$ – pretrained image classification model (e.g. ImageNet classifier).

What do we want from samples?

- **Sharpness**

Low sharpness

High sharpness

$p(y|\mathbf{x})$ has low entropy (each image $\mathbf{x}$ should have distinctly recognizable object).

- **Diversity**

Low diversity

High diversity

$p(y) = \int p(y|\mathbf{x})p(\mathbf{x})d\mathbf{x}$ has high entropy (there should be as many classes generated as possible).

---

image credit: https://deepgenerativemodels.github.io

Recap of previous lecture

Let take some pretrained image classification model to get the conditional label distribution $p(y|\mathbf{x})$ (e.g. ImageNet classifier).

Evaluation of likelihood-free models

- Sharpness $\Rightarrow$ low $H(y|\mathbf{x}) = -\sum_y \int_{\mathbf{x}} p(y, \mathbf{x}) \log p(y|\mathbf{x}) d\mathbf{x}$.
- Diversity $\Rightarrow$ high $H(y) = -\sum_y p(y) \log p(y)$.

Inception Score

$$IS = \exp(H(y) - H(y|\mathbf{x})) = \exp\left(\mathbb{E}_{\mathbf{x}} KL(p(y|\mathbf{x}) || p(y))\right)$$

Frechet Inception Distance

$$D^2(\pi, p) = \|\mathbf{m}_\pi - \mathbf{m}_p\|_2^2 + \text{Tr}\left(\boldsymbol{\Sigma}_\pi + \boldsymbol{\Sigma}_p - 2\sqrt{\boldsymbol{\Sigma}_\pi \boldsymbol{\Sigma}_p}\right).$$

FID is related to moment matching.

Salimans T. et al. Improved Techniques for Training GANs, 2016
Heusel M. et al. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, 2017

Recap of previous lecture

- $\mathcal{S}_\pi = \{\mathbf{x}_i\}_{i=1}^n \sim \pi(\mathbf{x})$ – real samples;
- $\mathcal{S}_p = \{\mathbf{x}_i\}_{i=1}^n \sim p(\mathbf{x}|\boldsymbol{\theta})$ – generated samples.

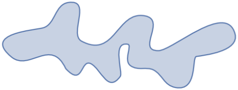
Embed samples using pretrained classifier network (as previously):

$$\mathcal{G}_\pi = \{\mathbf{g}_i\}_{i=1}^n, \quad \mathcal{G}_p = \{\mathbf{g}_i\}_{i=1}^n.$$

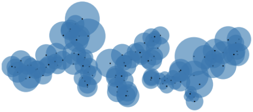
Define binary function:

$$f(\mathbf{g}, \mathcal{G}) = \begin{cases} 1, \text{if exists } \mathbf{g}' \in \mathcal{G} : \|\mathbf{g} - \mathbf{g}'\|_2 \leq \|\mathbf{g}' - \text{NN}_k(\mathbf{g}', \mathcal{G})\|_2; \\ 0, \text{otherwise.} \end{cases}$$

$$\text{Precision}(\mathcal{G}_\pi, \mathcal{G}_p) = \frac{1}{n} \sum_{\mathbf{g} \in \mathcal{G}_p} f(\mathbf{g}, \mathcal{G}_\pi); \quad \text{Recall}(\mathcal{G}_\pi, \mathcal{G}_p) = \frac{1}{n} \sum_{\mathbf{g} \in \mathcal{G}_\pi} f(\mathbf{g}, \mathcal{G}_p).$$



(a) True manifold



(b) Approx. manifold

Kynkäänniemi T. et al. Improved precision and recall metric for assessing generative models, 2019

Outline

1. Evaluation of likelihood-free models
   Maximum Mean Discrepancy (MMD)
   Precision-Recall

Outline

1. Evaluation of likelihood-free models
   - Maximum Mean Discrepancy (MMD)
   - Precision-Recall

Outline

1. Evaluation of likelihood-free models
   - Maximum Mean Discrepancy (MMD)
   - Precision-Recall

Maximum Mean Discrepancy (MMD)

Theorem

$\pi(\mathbf{x}) = p(\mathbf{y})$ if and only if $\mathbb{E}_{\pi(\mathbf{x})} f(\mathbf{x}) = \mathbb{E}_{p(\mathbf{y})} f(\mathbf{y})$ for any bounded and continuous f.

$$\mathrm{MMD}(\pi, p) = \sup_f \left[\mathbb{E}_{\pi(\mathbf{x})} f(\mathbf{x}) - \mathbb{E}_{p(\mathbf{x})} f(\mathbf{x})\right]$$

Theorem (Reproducing Kernel Hilbert Space)

$$\mathrm{MMD}^2(\pi, p) = \mathbb{E}_{(\mathbf{x},\mathbf{x}')} k(\mathbf{x}, \mathbf{x}') + \mathbb{E}_{(\mathbf{y},\mathbf{y}')} k(\mathbf{y}, \mathbf{y}') - 2\mathbb{E}_{(\mathbf{x},\mathbf{y})} k(\mathbf{x}, \mathbf{y}),$$

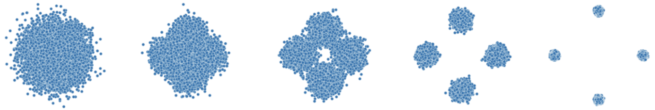
Here

- $k(\mathbf{x}, \mathbf{y})$ is a positive definite, symmetric kernel function (for example $k(\mathbf{x}, \mathbf{y}) = \frac{\exp(\|\mathbf{x}-\mathbf{y}\|^2)}{\sigma^2}$);
- $f(\mathbf{x}) = \langle f, \phi(\mathbf{x}) \rangle_{\mathcal{H}}$;
- $\phi(\mathbf{x}) = k(\cdot, \mathbf{x})$.

Jayasumana S. et al. Rethinking FID: Towards a Better Evaluation Metric for Image Generation, 2024

Maximum Mean Discrepancy (MMD)

$$\mathrm{MMD}^2(\pi, p) = \mathbb{E}_{(\mathbf{x},\mathbf{x}')} k(\mathbf{x}, \mathbf{x}') + \mathbb{E}_{(\mathbf{y},\mathbf{y}')} k(\mathbf{y}, \mathbf{y}') - 2\mathbb{E}_{(\mathbf{x},\mathbf{y})} k(\mathbf{x}, \mathbf{y}),$$



FD	0.0	0.0	0.0	0.0	0.0
MMD	0.0	5.794	17.21	78.88	202.8

- Needs less sample size for evaluation.
- High dependence on the pretrained classification model.
- Works with any distribution!

Jayasumana S. et al. Rethinking FID: Towards a Better Evaluation Metric for Image Generation, 2024

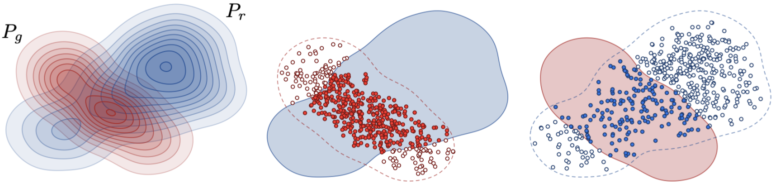
Outline

1. Evaluation of likelihood-free models
    Maximum Mean Discrepancy (MMD)
    Precision-Recall

Precision-Recall

What do we want from samples

- **Sharpness**: generated samples should be of high quality.
- **Diversity**: their variation should match that observed in the training set.



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

- **Precision** denotes the fraction of generated images that are realistic.
- **Recall** measures the fraction of the training data manifold covered by the generator.

Kynkäänniemi T. et al. Improved precision and recall metric for assessing generative models, 2019

Precision-Recall

- $\mathcal{S}_\pi = \{\mathbf{x}_i\}_{i=1}^n \sim \pi(\mathbf{x})$ – real samples;
- $\mathcal{S}_p = \{\mathbf{x}_i\}_{i=1}^n \sim p(\mathbf{x}|\boldsymbol{\theta})$ – generated samples.

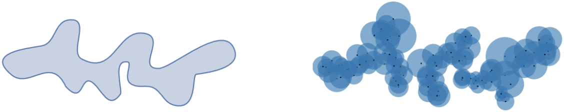
Embed samples using pretrained classifier network (as previously):

$$\mathcal{G}_\pi = \{\mathbf{g}_i\}_{i=1}^n, \quad \mathcal{G}_p = \{\mathbf{g}_i\}_{i=1}^n.$$

Define binary function:

$$f(\mathbf{g}, \mathcal{G}) = \begin{cases} 1, & \text{if exists } \mathbf{g}' \in \mathcal{G} : \|\mathbf{g} - \mathbf{g}'\|_2 \leq \|\mathbf{g}' - \text{NN}_k(\mathbf{g}', \mathcal{G})\|_2; \\ 0, & \text{otherwise.} \end{cases}$$

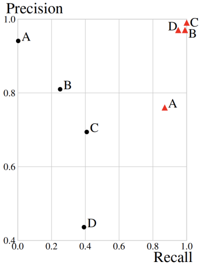
$$\text{Precision}(\mathcal{G}_\pi, \mathcal{G}_p) = \frac{1}{n} \sum_{\mathbf{g} \in \mathcal{G}_p} f(\mathbf{g}, \mathcal{G}_\pi); \;\; \text{Recall}(\mathcal{G}_\pi, \mathcal{G}_p) = \frac{1}{n} \sum_{\mathbf{g} \in \mathcal{G}_\pi} f(\mathbf{g}, \mathcal{G}_p).$$



(a) True manifold (b) Approx. manifold

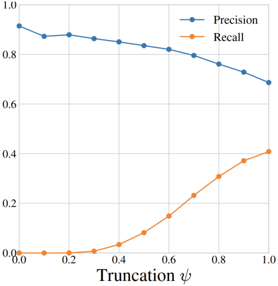
Kynkäänniemi T. et al. Improved precision and recall metric for assessing generative models, 2019

Precision-Recall



A (FID = 91.7) **B** (FID = 16.9) **C** (FID = 4.5) **D** (FID = 16.7)

$\psi = 0.0$ $\psi = 0.3$ $\psi = 0.7$ $\psi = 1.0$



Kynkäänniemi T. et al. Improved precision and recall metric for assessing generative models, 2019

Truncation trick

BigGAN: truncated normal sampling

$$p(\mathbf{z}|\psi) = \mathcal{N}(\mathbf{z}|0, 1) / \int_{-\infty}^{\psi} \mathcal{N}(\mathbf{z}|0, 1) d\mathbf{z}$$

Components of $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ which fall outside a predefined range are resampled.

StyleGAN

$$\mathbf{z}' = \hat{\mathbf{z}} + \psi \cdot (\mathbf{z} - \hat{\mathbf{z}}), \quad \hat{\mathbf{z}} = \mathbb{E}_{\mathbf{z}} \mathbf{z}$$

- Constant ψ is a tradeoff between diversity and fidelity.
- $\psi = 0.7$ is used for most of the results.

Brock A., Donahue J., Simonyan K. Large Scale GAN Training for High Fidelity Natural Image Synthesis, 2018
Karras T., Laine S., Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks, 2018

Summary

- Maximum Mean Discrepancy tries to fix some of the FID drawbacks.

- Precision-recall allow to select model that compromises the sample quality and the sample diversity.

- Truncation tricks help to select model with compromised samples: diverse and sharp.